FORECASTING TIME-VARYING VOLATILITY IN GLOBAL STOCK MARKETS

Lappeenranta–Lahti University of Technology LUT

School of Business and Management

Strategic Finance and Business Analytics

Master's thesis

2022

Ville Tillgren

Examiners: Associate professor Jan Stoklasa

Associate professor Sheraz Ahmed

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppatieteet

Ville Tillgren

Ajan mukaan muuttuvan volatiliteetin ennustaminen globaaleilla osakemarkkinoilla

Kauppatieteiden pro gradu -tutkielma

84 sivua, 6 kuvaa ja 15 taulukkoa

Tarkastajat: Apulaisprofessori Jan Stoklasa ja Apulaisprofessori Sheraz Ahmed

Avainsanat: Volatiliteetin ennustaminen, GARCH, EGARCH, GJR-GARCH, EWMA, liukuva keskiarvo, asymmetrinen volatiliteetti

Tämä pro gradu -tutkielma tarkastelee volatiliteetin ennustamista kuudessa eri osakeindeksissä. Indeksit ovat DAX30 (Saksa), FTSE100 (Iso-Britannia), Shanghai SE Composite (Kiina), NIKKEI225 (Japani), S&P500 (Yhdysvallat) ja Dow Jones Industrial Average (Yhdysvallat). Erilaisten GARCH-mallien ja liukuvan keskiarvon päivittäiset ja kuukausitason volatiliteettiennusteet aikajaksolla 1.1.2016-31.12.2020 vertaillaan ja laitetaan järjestykseen MSE ja MAE tappiofunktioiden mukaisesti. Yang-Zhang volatiliteettiestimaattoria käytetään kuvaamaan todellista volatiliteettia. Mallien volatiliteettiennusteiden tilastollista merkitsevyyttä mitataan Diebold-Mariano testillä. Tutkielman keskeinen tulos on se, että mikään yksittäinen malli ei ole toistuvasti paras tutkimuksessa mukana olevissa indekseissä. Päivätason ennusteissa vähintään yksi EGARCH-mallin ennuste on kuitenkin jokaisessa indeksissä kolmen parhaan mallin joukossa. Ainoana poikkeuksena tähän on DAX30, jossa GJR-GARCH ennustemallit ovat molempien tappiofunktioiden mukaan parhaimmat. Liukuva keskiarvo sekä eksponentiaalinen painotettu liukuva keskiarvo ovat parhaimmat päivätason volatiliteetin ennustemallit Shanghai SE Composite indeksissä. Asymmetriset EGARCH -ja GJR-GARCH-mallit suoriutuvat volatiliteettiennusteissa lähes toistuvasti tavallista GARCH-mallia paremmin. Kuukausitason volatiliteettiennusteissa DAX30 ja NIKKEI225 ovat ainoat indeksit, joissa EGARCH ei ole kolmen parhaan ennustemallin joukossa. Huipukkuutta paremmin mallintavat Studentin t-jakauman volatiliteettiennusteet eivät toistuvasti suoriudu paremmin verrattuna normaalijakauman volatiliteettiennusteisiin. Diebold-Mariano testin tulokset osoittavat, että suurimmassa osassa tapauksista kolmen parhaan mallin volatiliteettiennusteissa ei ole tilastollisesti merkitsevää eroa ja testi vahvistaa viiden päivän liukuvan keskiarvon hyvänä volatiliteetin ennustemallina, sillä mallin volatiliteettiennusteet ovat lähes kaikissa tapauksissa tilastollisesti yhtä tarkkoja parhaan mallin ennusteen kanssa.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

School of Business and Management

Business Administration

Ville Tillgren

Forecasting time-varying volatility in global stock markets

Master's thesis

2022

84 pages, 6 figures and 15 tables

Examiners: Associate professor Jan Stoklasa and Associate professor Sheraz Ahmed

Keywords: Volatility forecasting, GARCH, EGARCH, GJR-GARCH, EWMA, moving average, asymmetric volatility

This master's thesis examines volatility forecasting in six global equity indices, namely DAX30 (Germany), FTSE100 (UK), Shanghai SE Composite (China), NIKKEI225 (Japan), S&P500 (US) and Dow Jones Industrial Average (US). The forecasting window is from 1.1.2016 to 31.12.2020. The daily and monthly volatility forecasts of different GARCH-type and moving average models are compared and ranked based on mean squared error and mean absolute error loss functions. The Yang-Zhang volatility estimator is used as a proxy for the actual volatility. The Diebold-Mariano test is used to test statistical difference of the models' forecasts. The overall finding is that there's no single model outperforming across indices, although at least one of the EGARCH model was constantly ranked in top three forecasting model under both loss functions. The exception to this was the mean squared error ranking for German DAX30 where the GJR-GARCH and the five-day moving average ranked best daily volatility forecasting models. The moving average and exponentially weighted moving average are best daily volatility forecasting models in Chinese Shanghai SE Composite index. With few exceptions the EGARCH and GJR-GARCH models constantly outperformed the standard GARCH across the indices. With monthly forecasts, the DAX30 and NIKKEI225 were the only indices where the EGARCH was not ranked in top three forecasting model under both loss functions. Although the data show signs of excess kurtosis and fat tails, models with t-distributions are not constantly outperforming their normal distributed counterparts. The Diebold-Mariano test results indicate that in most cases, there's no statistical difference between the first ranked model forecast and following second or third ranked model forecasts. The Diebold-Mariano test confirm the good performance of the five-day moving average, as the model's forecasts are almost always statistically equally accurate with the best ranked models' forecasts.

ACKNOWLEDGEMENTS

As I’m finishing this master’s thesis it is time to say few words. First, I’m grateful and happy that I decided to chase my dreams back in 2014. It was after hard work and determination that I got the news of getting in to study in university. It really was a dream come true. I moved to Lappeenranta and got to know many great people and the spirit of Skinnarila. Working full time these past years, it has not always been easy to schedule studies along other things in life. But now, thanks to my determination it is finally happening and I’m graduating as a Master of Science in Economics and Business Administration. I’m really thrilled and extremely proud of myself – I did it!

There are great people who have helped me along my journey, and I would like to express my gratitude. I want to say thank you to my family – mom Raija, dad Asko, sister Erja and brother Tomi and his family. You are always there for me and believe in me, I love you. My gratitude goes also to my girlfriend and friends. I love you all. At LUT I want to say thank you to my fellow students and all the personnel. I’ve learned so much and appreciate the work that professors and every single person does at LUT. I want to say many thanks to my supervisors, professor Jan Stoklasa and professor Sheraz Ahmed for their great guidance and help. I also want to thank you Jan and Sheraz for your excellent teaching. The lectures were all great and inspiring. Now my journey continues, and I know it will be an awesome ride.

In Helsinki 22.03.2022

Ville Tillgren

Abbreviations

ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
ARCH	Autoregressive Conditional Heteroscedasticity
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information criterion
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedasticity
EWMA	Exponentially Weighted Moving Average
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GJR	Glosten, Jagannathan and Runkle
KPSS	Kwiatkowski, Phillips, Schmidt and Shin
MA	Moving Average
MAE	Mean Absolute Error
MSE	Mean Squared Error

Table of contents

Figures

Tables

1 Introduction

Volatility and its forecasts are perhaps one of the most important subjects in finance and widely used e.g., in risk management, derivative pricing and hedging and portfolio management. Different volatility-based trading strategies are also actively used by investors. Volatility, often described as variance or standard deviation, is a statistical measure of variation and counts the dispersion of the data around its mean. In finance, one of the most common representations of the volatility is the standard deviation which denotes the security's return variation from the mean over a certain period of time. Volatility is also a measure of risk and a security with a higher standard deviation is often noted riskier because of the bigger fluctuation between the returns and the mean. This thesis concentrates on volatility forecasting in global stock markets and next, before the research questions, few conceptual issues regarding volatility are presented.

Volatility and financial time series have some well-known stylized facts such as fat tails, volatility clustering, leverage effects and long memory in volatility. Fat tails means deviation from normal distribution and exhibition of excess kurtosis. Volatility clustering can be understood as periods of high volatility following further high volatility and similar pattern with lower volatility following low volatility. Leverage effects means that there is negative correlation in prices with volatility. Leverage effects can be understood also as an asymmetry in volatility, which means that there is higher volatility after negative shock compared to similar size positive shock. Long memory indicates that volatility is persistent and correlated. (Knight and Satchell 2007, 3)

Having these certain characteristics volatility has raised a lot of attention among research and many different volatility models have been developed. Early econometric models however were not created to capture all the characteristics of volatility. In fact, volatility was considered as a constant and time-independent parameter in the early models although volatility clustering was first mentioned by Mandelbrot (1963) already in 1960s. These early models assumed that returns are independent of each other and that the return-generating process is linear with parameters that are independent of past realizations (MSCI 2021). It was the ARCH (Autoregressive Conditional Heteroscedasticity) by Engle (1982) that modelled non-linearity and replaced the assumption of constant volatility with conditional volatility. The ARCH was not only intended to estimate but also forecast volatility and its groundbreaking concept of conditional heteroscedasticity refers to the future volatility being conditional (time-dependent) on current and past volatility. Modelling volatility clustering the ARCH however has some limitations, such as that the model usually requires several parameters to capture the dynamics of volatility and therefore quite soon extended to GARCH (Generalized Autoregressive Conditional Heteroscedasticity) by Bollerslev (1986). The GARCH is widely recognized and used in financial applications. Perhaps the model's biggest contribution is however volatility forecasting. Several other forecasting models have also been developed, such as the EGARCH (Exponential Generalized Autoregressive Conditional Heteroscedasticity) by Nelson (1991) and the GJR-GARCH by Glosten, Jagannathan and Runkle (1993). One of the key differences between these GARCH-type models is that both EGARCH and GJR-GARCH were created to cover also the leverage effects whereas the standard GARCH does not account the asymmetric volatility (Brooks 2019, 405-406).

Beside of the sophisticated GARCH-type models, some more simpler models are also still widely used to forecast volatility. Perhaps one of the well-known are the moving average (MA) and exponentially weighted moving average (EWMA). Moving average volatility forecasting simply means that the model uses some fixed length of arithmetic average of standard deviations which are rolling forward keeping the averaging length same. Exponentially weighted moving average does the same, with the exception that there's a weighting factor, usually denoted with lambda. The value of lambda is between one and zero and higher values indicate more weight to most recent observations compared to the simple moving average with equal weights. Compared to asymmetric GARCH models, the moving average models are not accounting the leverage effects. (Brooks 2019, 390-391)

Previous research related to volatility forecasting is broad and wide and there's almost countless papers considering different forecasting models including sophisticated GARCH-type and more simpler moving average models. The results from previous studies are however mixed and there's no solid and consistent view of superior volatility forecasting model. Some of the research gives support to GARCH (see e.g., Franses and Van Dijk 1996; Bera and Higgins 1997; Andersen and Bollerslev 1998; Hansen and Lunde 2005; Sharma and Sharma 2015), while others show the outperformance of asymmetric models (for EGARCH see e.g., Pagan and Schwert 1990; Cumby, Figlewski and Hasbrouck 1993; Awartani and

Corradi 2005 and for GJR-GARCH see e.g., Engle and Ng 1993; Brailsford and Faff 1996). For research supporting moving average and EWMA see e.g., Kuen and Tung (1992); Walsh and Tsou (1998); McMillan, Speight, and Gwilym (2000). Poon and Granger (2003) review 93 research papers of volatility forecasting. They conclude that historical volatility methods, such as random walk, moving average, exponential weights and autoregressive models perform equally well compared to other sophisticated models including the GARCH-type models. The evaluation of the forecasting performance is usually done with some loss functions, such as mean absolute error, mean squared error, root mean squared error and mean absolute percentage error. There's however no unanimous conclusion that which of the loss functions one should always use and as Poon and Granger (2003) outline, with different data, loss functions or volatility proxies, the forecasting results might be different. Poon and Granger (2003) also mention that it is rarely discussed if one forecasting method is significantly better than another and although some particular method of forecasting volatility can be suggested being the best, there's no discussion about the cost-benefit from using it. Based on these aspects and results from previous research, interesting topic regarding the comparison of the GARCH-type and moving average models is that which of the models should we choose when considering forecasting volatility.

The main purpose of this thesis is to compare the volatility forecasts of different GARCH-type and moving average models in six global equity indices that are DAX30 (Germany), FTSE100 (UK), Shanghai SE Composite (China), NIKKEI225 (Japan), S&P500 (US) and Dow Jones Industrial Average (US). Daily and monthly volatility forecasts are produced for the out-of-sample period from 1.1.2016 to 31.12.2020. The volatility forecasts of different models are evaluated and ranked based on mean squared error and mean absolute error loss functions. Although fat tails and excess kurtosis are a well-known phenomenon in financial data, it is not however clear whether volatility forecasting models with normal distributions or distributions allowing more kurtoses are better. Some of the previous research has showed the superiority of more leptokurtic distributions (see e.g., Wilhelmsson 2006; Liu & Morley 2009). Therefore, to compare the volatility forecasting results the normally distributed and t-distributed models are included. The Diebold-Mariano test is used to test statistical difference of the models' forecasts. To study these issues in global stock markets, four research questions are formed:

Which of the volatility forecasting models perform best according to loss functions MSE and MAE for the one-day and one-month forecasting horizon?

How does the GARCH, EGARCH and GJR-GARCH volatility forecasting models perform compared to moving average and EWMA models for the one-day ahead forecasting horizon?

Is there a difference between normally distributed and t-distributed models based on volatility forecasting results?

Is there a statistical difference in volatility forecasting accuracy between models?

The purpose of the results from this thesis is two-sided. First, the findings from this thesis brings new information and contributes to the existing volatility forecasting literature about the performance of the models across six large equity indices. Second, the practical side of the results can be considered for traders that are interested in short-term volatility predictions. These can be for instance day traders that are using volatility-based trading strategies and models.

Limitations of the study

Next, some of the limitations regarding the research behind this thesis. First, the models used on forecasting in this thesis rely on GARCH, EGARCH, GJR-GARCH, EWMA and MA. The choice of these models was based on their broad use in previous literature. These models are also easy to adapt and calibrate to equity-based volatility forecasting. However, some other models such as AGARCH, IGARCH, TGARCH and QGARCH to name a few, could also provide interesting volatility forecasting results (see e.g., Awartani and Corradi 2005; Ederington and Guan 2005; Franses and Van Dijk 1996; McMillan, Speight, and Gwilym 2000).

This thesis examines univariate GARCH-type models and multivariate models, such as the BEKK and DCC-GARCH are left out of this study. The volatility predictions of the GARCH-type models are based on historical volatility. Another way of volatility forecasting is to use options and extract volatility using some option pricing model such as the Black-Scholes-Merton model (Black and Scholes 1973; Merton 1973). This implied volatility is often considered as a market's expectation of future volatility. In research, both the historical estimation of volatility and the GARCH-type models and implied volatility are widely examined in the area of volatility forecasting. In this thesis the focus is on GARCH-type models and therefore the implied volatility was left out of scope.

For some readers, the use of the results from this thesis can be more theoretical because of the short forecasting horizon. Longer prediction horizon, such as six months or one year, can be more useful in practice regarding for instance strategic portfolio asset management. However, day traders and other short-term investors may use shorter period volatility forecasts. Also, as the market movements can be large on a daily and even on an intraday level, the volatility forecasts of next day's movements can be useful for investors with speculative trading.

Another limitation involved is the use of daily data. Again, the background of using daily stock market data comes from the literature as it has claimed that the GARCH-models work well on daily data (see e.g., Figlewski 1997, 34-35). Another common assumption is that the high-frequency data could also be fruitful basis for volatility forecasting. As it was not possible to acquire the high-frequency intraday data the natural choice was the readily available daily stock market data, and this thesis contributes to the existing volatility forecasting literature by using six global equity indices with a long twenty-year period from 2000 to 2020.

One limitation is also the choice of volatility proxy. Volatility proxy means the actual or true volatility. Historical close-to-close volatility calculated from today's and yesterday's closing prices has been claimed to be inaccurate and noisy estimator (Andersen and Bollerslev 1998). Same historical estimator using high-frequency intraday data could provide more realistic picture of the true realized volatility. As the high-frequency data was not possible to get, the focus was directed to range-based volatility estimators. Range-based volatility estimators use open, close, high, and low prices over certain fixed sampling interval. Alizadeh, Brandt and Diebold (2002) and Shu and Zhang (2006) claim that logarithmic range-based estimators are powerful to obtain market microstructure effects compared to intraday high-frequency data estimators. From the four different estimators, Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991) and Yang and Zhang (2000), the choice was the Yang-Zhang estimator as it has claimed to account for overnight price jumps and offer precise volatility estimate. Previous literature has favored Garman-Klass estimator as well as Parkinson's model (Diebold and Yilmaz 2012). Using different volatility proxy may naturally offer different aspects to volatility forecasting.

Together with volatility proxy, the loss functions or error functions, are important aspects when comparing the performance of volatility forecasting models as the ranking of the models is based on the loss functions. Some common loss functions are root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and mean mixed error (MME). MSE has stated to be a robust loss function (Hansen and Lunde 2006) but one of the shortcomings of MSE is that it is sensitive to outliers and penalizes high volatility (Brooks 2019, 285). Because of this the MAE loss function was also chosen to be included in this thesis. Using several different loss functions might give more profound results of the forecasting performance of the models. However, in the context of this thesis, the MSE and MAE were considered reasonable selection.

After the introduction the thesis is organized as follows. Theoretical framework and literature review are discussed in chapter 2. Chapter 3 presents the data and methodology followed by results in chapter 4. Discussion and summary are presented in chapter 5.

2 Theoretical framework and literature review

In this chapter the theoretical framework behind the volatility is discussed. Volatility forecasting models, statistical tests and literature review are also presented.

2.1 Volatility and its proxies

Volatility, often described as variance or standard deviation, is a measure of variation. Being more precise, volatility counts the dispersion around the mean of the data. High volatility can be understood as a large deviation from the mean, whereas low volatility fluctuates closer from the mean. In finance, common representation of the volatility is the standard deviation which denotes the asset's price variation from the mean over a certain period of time. Volatility is also a measure of risk and a security with a higher standard deviation is often noted riskier because of the bigger fluctuation between the returns and the mean. A well-known indicator of the market risk is the VIX, also known as the Fear Index, a volatility index by the Chicago Board Options Exchange. Derived from the S&P500 index options, the VIX is a representation of the market's sentiment and expectations for the future volatility. Under turbulent periods, the VIX tends to rise and reflects uncertainty and the probability of price fluctuations in the market. The volatility of the VIX is called implied volatility as it is extracted from the current option prices.

Broadly speaking, the estimation and forecast of volatility is usually done either from some historical dataset or derived from option markets using for instance the Black-Scholes-Merton option pricing model. Implied volatility is more of market expectations of future volatility because it is derived from the current market price of the option whereas historical volatility is computed from previous historical asset price changes. The option pricing models often assume constant volatility. Other distinctive feature of the implied volatility is that options with same time to maturity but with different strike price produce different volatility for the same underlying asset. (Poon and Granger 2003) The forecasts of implied and historical volatility models are widely discussed in research and the results are not unanimous whether implied or historical volatility models are superior. The volatility predictions of the GARCH-type models are based on historical volatility and because this thesis examines these models the implied volatility is not discussed further. Common presentation of the historical volatility is as follows:

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \tag{1}$$

where square root of the variance denotes the standard deviation of the returns. N is the sample size, x_i is the daily logarithmic return computed from the closing prices and $\bar{x}$ is the mean return. The annualized volatility is obtained by multiplying the standard deviation by the square root of the number of trading days per year which in this thesis was decided to be 252. For monthly returns, the annualized volatility is obtained by multiplying the standard deviation of monthly logarithmic returns by the square root of 12.

Next, the choice of volatility proxy is discussed. The volatility proxy is important because it is a measure of true realized volatility, and the volatility forecasts of different models are then compared against the volatility proxy. With daily data, the daily close-to-close volatility can be calculated from todays and yesterday's closing prices. Close-to-close volatility is easy to adapt and might be useful on some occasions. Close-to-close volatility estimation however doesn't include any information about the possible intraday volatility. Asset price can fluctuate intensively during trading days but end up near the previous day's closing price. Another drawback of using close-to-close estimator is that the squared returns are claimed to be inaccurate and noisy estimators of the realized volatility (Andersen and Bollerslev 1998). Squared returns might offer a proxy of the realized volatility but on a given day the estimate can be very different compared to the actual (unobserved) volatility. Using high frequency intraday data might give more realistic picture of the true realized volatility (Andersen, Bollerslev, Diebold and Labys 2003).

Alizadeh, Brandt and Diebold (2002) describe realized volatility as the sum of squared high-frequency returns over a given sampling period. Daily realized variance series can be computed for instance by summing over each day a sequence of squared intraday returns. Sampling period for intraday returns can be e.g., five-minute returns. Market microstructure biases can however skew high-frequency prices and returns. Authors explain that in the presence of a bid-ask spread, the observed price is a noisy version of the true price because it effectively equals the true price plus or minus half the spread, depending on whether a trade is buyer or seller initiated. Buying and selling transactions affect to bid-ask spread and therefore increases the measured volatility of high-frequency returns. Authors continue that bid-ask bounce affects so that by summing the squared high-frequency returns, each of which is biased upward, the realized volatility contains a cumulate and therefore potentially large bias, which becomes more severe if returns are sampled more frequently.

Alizadeh, Brandt and Diebold (2002) and Shu and Zhang (2006) claim that logarithmic range-based volatility estimators are powerful to obtain market microstructure effects compared to intraday high-frequency data estimators. The idea of range-based volatility estimators is to use open, close, high, and low prices over certain fixed sampling interval. The range-based estimation is less likely to be contaminated by bid-ask bounce. Alizadeh, Brandt and Diebold (2002) explain also that the observed maximum daily price is likely to be at the ask and hence too high by half the spread, whereas the observed minimum is likely to be at the bid and too low by half the spread. This means that on average the range is inflated only by the average spread, which is small in liquid markets. The range-based estimator might be a less efficient volatility proxy than realized volatility under ideal conditions. However, the range-based might still be superior in real-world in which market microstructure biases distort high-frequency prices and returns. (Alizadeh, Brandt and Diebold 2002)

For this thesis, it was not possible to acquire high frequency intraday data and therefore we'll focus on range-based volatility estimators. Next, four of these estimators are described: Parkinson (1980), Garman and Klass (1980), Rogers and Satchell (1991) and Yang and Zhang (2000). These range-based volatility proxies are commonly discussed among literature and therefore were chosen to be included in this thesis.

Parkinson's (1980) volatility estimator uses high and low intraday trading prices. The assumption is that the logarithmic prices follow Brownian motion with zero mean. Parkinson's estimator does not account possible overnight volatility. Parkinson's estimator can be derived as follows:

$$\sigma_{Parkinson} = \sqrt{\frac{N}{4Zln(2)}\sum_{i=1}^{Z} ln\left(\frac{H_i}{L_i}\right)^2} \tag{2}$$

where N is the number of periods per year. Z is the number of periods used in volatility estimation. H_i is the highest intraday price and L_i is the lowest intraday price.

Another range-based volatility estimator is the Garman-Klass (1980) model. A distinctive difference to Parkinson's estimator is that open and closing prices are added in Garman-Klass model, which also assumes Brownian motion with zero drift. Overnight volatility is not captured by the Garman-Klass model either, which might result to underestimating volatility. The Garman-Klass estimator is given by:

$$\sigma_{Garman-Klass} = \sqrt{\frac{N}{Z}\sum_{i=1}^{Z}\left[\frac{1}{2}ln\left(\frac{H_i}{L_i}\right)^2 - (2ln(2)-1)\left(ln\left(\frac{C_i}{O_i}\right)\right)^2\right]} \tag{3}$$

where N is the number of periods per year, Z is the number of periods used in volatility estimation, H_i is the highest intraday price, L_i is the lowest intraday price, C_i is the closing price and O_i is the opening price.

Rogers and Satchell (1991) demonstrated a model with a non-zero drift (allowing non-zero mean of returns). As with other previous models, their model does not account for overnight price variation. The Rogers and Satchell estimator is as follows:

$$\sigma_{Rogers-Satchell} = \sqrt{\frac{N}{Z}\sum_{i=1}^{Z}\left[ln\left(\frac{H_i}{C_i}\right)ln\left(\frac{H_i}{O_i}\right) + ln\left(\frac{L_i}{C_i}\right)ln\left(\frac{L_i}{O_i}\right)\right]} \tag{4}$$

where, N is the number of periods per year, Z is the number of periods used in volatility estimation, H_i is the highest intraday price, L_i is the lowest intraday price, C_i is the closing price and O_i is the opening price.

Yang and Zhang (2000) volatility estimator can account for non-zero drift and also overnight price jumps. The model is a sum of overnight volatility and weighted open-to-close volatility and Rogers and Satchell volatility. The Yang-Zhang estimator can be derived as:

$$\sigma_{YZ} = \sqrt{N}\sqrt{\sigma^2_{overnight\ volatility} + k\sigma^2_{open-to-close\ volatility} + (1-k)\sigma_{Rogers-Satchell}} \tag{5}$$

$$\sigma^2_{overnight\ volatility} = \frac{1}{Z-1}\sum_{i=1}^{Z}\left(ln\left(\frac{O_i}{C_{i-1}}\right) - \overline{ln\left(\frac{O_i}{C_{i-1}}\right)}\right)^2$$

$$\sigma^2_{open-to-close\ volatility} = \frac{1}{Z-1}\sum_{i=1}^{Z}\left(ln\left(\frac{C_i}{O_i}\right) - \overline{ln\left(\frac{C_i}{O_i}\right)}\right)^2$$

$$k = \frac{0.34}{1.34 + \frac{Z+1}{Z+1}}$$

where, N is the number of periods per year, Z is the number of days used in volatility estimate, C_i is the closing price and O_i is the opening price on trading day i and $C_{i-1}$is the closing price of the previous day. The Rogers-Satchell volatility estimator is the same as provided in formula 4.

After presenting these different volatility proxies it is now time to choose one to be used in the empirical part of this thesis. It is rather clear that any of these proxies offer a bit better estimate from true volatility than the close-to-close estimator. Allowing drift, the Rogers and Satchell model would provide good volatility proxy. Previous literature has favored Garman and Klass estimator as well as Parkinson's model (Diebold and Yilmaz 2012). Shu and Zhang (2006) stated that Yang-Zhang model is precise and offers accurate volatility estimates during large overnight price jumps. Therefore, accounting for both non-zero drifts and overnight price jumps, the Yang and Zhang estimator was chosen as a volatility proxy in this thesis.

2.2 Volatility forecasting models

One of the well-known forecasting model is the Autoregressive Moving Average (ARMA) model by Box and Jenkins (1976). The ARMA model however has some shortcomings, as the model assumes homoscedasticity and it doesn't account the stylized facts of financial time series, such as volatility clustering. The ARMA model or Autoregressive Integrated Moving Average (ARIMA) model is however included in this thesis to model the conditional mean of the equity indices. Therefore, the combinations of ARIMA and GARCH-type models are used to model both conditional mean and conditional volatility of the indices. The "I" in ARIMA process refers to integrated autoregressive process in which the characteristic equation has a unit root on the unit circle. When the process is non-stationary, then by differentiating the variables by d times, the model is stationary. Hence, the ARMA(p, q) differenced d times, is equivalent to ARIMA(p,d,q). (Brooks 2019, 272)

The autoregressive part models the dependence of the current value of y on the past values of the same variable y in addition with a white noise error term. Moving average part models the linear combination of white noise processes, so that the dependent variable y depends on the current and past values of the white noise error terms. (Brooks 2019, 251, 254) Combining the autoregressive and moving average parts, the ARMA(p, q) model is described as follows:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{6}$$

where y_t denotes the form of time series data, which in thesis is the logarithmic returns data from the equity indices, ϕ is the coefficient of the autoregressive part of lag p, θ is the coefficient of the moving average process of lag q. ε_t is a white noise process with zero mean and constant variance. (Brooks 2019, 264)

After ARMA model, Engle (1982) developed the ARCH model which quite soon extended to GARCH model by Bollerslev (1986). The ARCH(q) model can be derived as follows:

$$\varepsilon_t = v_t \sigma_t , v_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{7}$$

where the conditional variance depends on previous q lags of squared errors. v_t is normally distributed with zero mean and unit variance, so that ε_t will also be normally distributed with zero mean and variance σ_t^2. For an ARCH(q) model, all coefficients α_0 and α_i are required to be non-negative to ensure the positive conditional variance estimates. (Brooks 2019, 394-395) ARCH is reported having some limitations, such as that it is problematic to decide the number of lags of the squared residuals and that the number of lags to capturing all the dependence in the conditional variance, can be very large. This can result to large conditional variance model not being parsimonious. Other problematic issue is that non-negative coefficients estimates might occur if there are large amounts of parameters. (Brooks 2019, 396)

Next, the GARCH, EGARCH and GJR-GARCH models are presented. The GARCH(p,q) process is given by:

$$\varepsilon_t = v_t \sigma_t , v_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{8}$$

where the conditional variance σ_t^2 depends upon q lags of the squared errors and p lags of the conditional variance itself. Constraints for non-negativity are: $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, p, \beta_j \geq 0, j = 1, \dots, q.$ The α_i coefficient is an ARCH term and explain volatility clustering and the β_j is a GARCH term representing the persistence of volatility. The unconditional variance of ε_t is constant and given by $var(\varepsilon_t) = \alpha_0/1 - \alpha - \beta$ so long as $\alpha_i + \beta_j < 1$. For $\alpha_i + \beta_j > 1$, the unconditional variance is not defined and $\alpha_i + \beta_j = 1$ is termed integrated GARCH or IGARCH. For stationary GARCH models ($\alpha_i + \beta_j < 1$), the conditional variance forecasts converge to the long-term mean variance. (Brooks 2019, 398-399)

Compared to ARCH, the GARCH model is less likely to breach non-negativity constraints and the GARCH is more parsimonious and avoids overfitting. Overcoming the limitations of ARCH model, the basic GARCH still has its own shortcomings, such as that it assumes positive and negative shocks affecting similarly to volatility. For financial time series there’s however evidence that volatility tends to rise more in response to bad news compared to a positive shock of the same magnitude. This asymmetric volatility is also called the leverage effects. Other limitation of the GARCH is that the non-negative constraints might be violated by the model. This could be avoided only if there are artificial constraints on the model coefficients. (Brooks 2019, 396, 404-405) According to Nelson (1991) it is also difficult to evaluate whether shocks to variance are persistent or not. If volatility shocks persist forever, they may move the whole term structure of risk premia, and therefore likely to have a significant impact on investment in long term capital goods (Poterba and Summers 1986).

Asymmetric GARCH models named the GJR-GARCH model by Glosten, Jagannathan and Runkle (1993) and the EGARCH model by Nelson (1991) were developed to model the asymmetric behavior of volatility and to overcome the shortcomings of the basic GARCH model. (Brooks 2019, 396, 404-405)

The EGARCH(p,q) model can be formulated by:

$$\varepsilon_t = v_t \sigma_t , v_t \sim N(0,1)$$

$$log(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i \left[\frac{|\varepsilon_{t-i}|}{\sqrt{\sigma_{t-i}^2}} - \sqrt{\frac{2}{\pi}} \right] + \sum_{i=1}^{q} \gamma_i \frac{\varepsilon_{t-i}}{\sqrt{\sigma_{t-i}^2}} + \sum_{j=1}^{p} \beta_j \, log\left(\sigma_{t-j}^2\right) \tag{9}$$

where the ARCH term, α_i indicates the magnitudes of past standardized innovations and the GARCH term β_j indicates past logarithmic conditional variances and the leverage term γ_i capture asymmetry in volatility clustering. Compared to the GARCH model, the EGARCH models the conditional variance with logarithm, which means that the variance will be positive even if the parameters are negative. If large and sudden negative shocks cause volatility to rise more than a positive shock of the same magnitude, then the leverage term γ_i is expected to be negative. (Brooks 2019, 405-406)

The GJR-GARCH(p,q) model is given by:

$$\varepsilon_t = v_t \sigma_t , v_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \gamma_i \, \varepsilon_{t-i}^2 I_{t-i} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{10}$$

where similarly with EGARCH, the additional term γ_i accounts possible asymmetries in volatility. For these leverage effects (negative shocks have larger impact on conditional variance than positive shocks) it is assumed that $\gamma_i > 0$. Even if $\gamma_i < 0$ the model can be used as long as $\alpha_i + \gamma_i \geq 0$. The dummy variable I_{t-i} gets value 1 if $\varepsilon_{t-i} < 0$, and zero otherwise. Constraints for non-negativity are: $\alpha_0 > 0$, $\alpha_i > 0$, $\beta_j \geq 0$, $\alpha_i + \gamma_i \geq 0$. (Brooks 2019, 405)

Next, the moving average models are presented. The simple moving average model computes an arithmetic average from certain length of previous observations. It is then rolling one-step forward keeping the averaging length same. We use the five-day moving average (MA5), one month moving average (MA21) and three month (MA63) moving average models. The length of these moving average models is calculated based on the approximation of 252 trading days per year. There is no theoretical background of choosing the right length for the model. It is although rather obvious that shorter moving average models are following and adjusting to current trends in data faster than longer moving average models. One distinctive feature of the moving average model is that it is not incorporating the abovementioned stylized facts of financial time-series data.

The moving average model is given by:

$$\sigma_t = \frac{1}{T}\sum_{j=1}^{T} \sigma_{t-j} \tag{11}$$

where T is the averaging length and σ_{t-j} represents historical volatility from daily closing prices.

Exponentially weighted moving average (EWMA) is another moving average model and can be expressed as follows:

$$\sigma_t^2 = (1-\lambda)\sum_{j=1}^{\infty} \lambda^{j-1}(r_{t-j} - \bar{r})^2 \tag{12}$$

where σ_t^2 is the variance estimate for period t, and this also becomes the forecast of future volatility for next periods, $\bar{r}$ is the average return and λ is a weighting factor. The value of the lambda is always between one and zero. Putting more weight to most recent observations, the EWMA is more responsive and reacts more rapidly to new information compared to the simple moving average model with equal weights. A lower value of the lambda indicates less impact for the recent observations. RiskMetrics introduced by the financial services firm

J.P. Morgan use 0.94 lambda value for daily returns and this decay factor is also used in this thesis (MSCI 2021). 0.94 lambda means that the most recent observation is weighted by 6%. The next observation is weighted by 5.64% (*(1-0.94)*(0.94)^1*) and so on. As with simple moving average, the EWMA does not model leverage effects. The EWMA is not accounting for mean-reverting volatility either. (Brooks 2019, 390-391)

2.3 Statistical tests

In this chapter statistical tests used in this thesis are described. We begin with stationarity tests. Stationarity is a preferable condition in financial time series and a stationary series can be described having constant mean, constant variance and constant autocovariance for each given lag. A stationary variable is one that does not contain a unit root. (Brooks 2019, 334)

There are various tests available for testing stationarity such as the augmented Dickey-Fuller (ADF) test and KPSS test by Kwiatkowski, Phillips, Schmidt and Shin (1992). Other widely used stationarity test is Phillips-Perron test but in this thesis we consider the ADF and KPSS tests. For the augmented Dickey-Fuller test the null hypothesis is that the time series contains a unit root and alternative hypothesis is that the series is stationary, and no unit root exists.

The Dickey-Fuller model with null hypothesis that the series contains a unit root ($\phi = 1$) is described as:

$$y_t = \phi y_{t-1} + u_t \tag{13}$$

Where y_t is data process, u_t is a mean zero innovation process. Another way of expressing the model is by:

$$\Delta y_t = \psi y_{t-1} + u_t \tag{14}$$

Where a test of $\phi = 1$ is equivalent to a test of $\psi = 0$ (because $\phi - 1 = \psi$). From this the augmented Dickey-Fuller test for a unit root can be modelled as:

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t \tag{15}$$

Where p is the number of lags of the dependent variable, u_t is a mean zero innovation process, Δ is the differencing operator and as Brooks (2019, 345) explains, the lags of Δy_t absorb any dynamic structure present in the dependent variable, to ensure that u_t is not autocorrelated. Brooks (2019, 344) say that u_t is assumed not to be autocorrelated but would be so if there was autocorrelation in the dependent variable of the regression Δy_t which has not been modelled. The Dickey-Fuller test would then be oversized, meaning that the true size of the test (the proportion of times a correct null hypothesis is incorrectly rejected) would be higher than the nominal size used (e.g. 5%). Answer to this is the augmented test using p lags of the dependent variable. The optimal number of lags is however user specific and quite arbitrary. For monthly data the use of twelve lags might be appropriate. For daily data used in this thesis there's no clear guideline for the number of lags. Information criterion can be used to decide the number of lags. In this thesis the number of lags were chosen without the use of information criterion and decided to be twenty lags.

Augmented Dickey-Fuller test for a unit root can be modelled as:

$$\Delta y_t = c + \delta t + \phi y_{t-1} + \sum_{i=1}^{p} \beta_p \Delta y_{t-p} + u_t \tag{16}$$

where c is a constant, p is the number of lags of the differences of the dependent variable, Δ is the differencing operator and u_t is a mean zero innovation process. The ADF test can be done with model variants with drift and trend stationary. The model with $\delta = 0$ has no trend and the model with $c = 0$ and $\delta = 0$ has no drift or trend.

Augmented Dickey-Fuller test statistic is:

$$ADF = \frac{(\hat{\phi} - 1)}{SE(\hat{\phi})} \tag{17}$$

Another way to assess stationarity is to use KPSS test by Kwiatkowski, Phillips, Schmidt, Shin (1992), in which the null is that the data is stationary. The structural model behind the test is as follows:

$$y_t = c_t + \phi_t + u_{1t}, \qquad c_t = c_{t-1} + u_{2t} \tag{18}$$

where ϕ_t is the trend coefficient, u_{1t} is a stationary process, u_{2t} is an independent and identically distributed process with mean zero and variance σ^2. The null hypothesis is then that $\sigma^2 = 0$, which considers that the random walk term c_t is constant and interprets as the model intercept.

KPSS test statistic is given by:

$$KPSS = \frac{\sum_{t=1}^{N} S_t^2}{N^2 s^2} \tag{19}$$

where N is the sample size, S_t^2 is the squared cumulative residuals and s^2 is the Newey-West estimate of the long-run variance. Authors of the test used Monte Carlo simulations and tabulated critical values from those simulations. (Mathworks 2022)

The Ljung-Box Q-test (1978) is a test for possible autocorrelation in returns time series. It can be also used with squared residuals for testing conditional heteroscedasticity. In this thesis the Ljung-Box is used to test autocorrelation in returns. The Ljung-Box test statistic can be derived from:

$$Q = N(N+2)\sum_{k=1}^{m} \frac{\hat{\tau}_k^2}{N-k} \sim \chi_m^2 \tag{20}$$

Where N is the sample size, $\hat{\tau}_k^2$ is autocorrelation coefficient and m is maximum lag length. In this thesis the maximum lag length is 20. Under the null hypothesis of no autocorrelation (that all m autocorrelations are jointly zero), the test statistic follows chi-squared distribution. (Brooks 2019, 249)

Engle's (1982) ARCH test is used to test for conditional heteroscedasticity on the residuals. Engle's ARCH test is also used to justify the use of GARCH type models. The idea of Engle's ARCH test is a joint null hypothesis that all q lags of the squared residuals have coefficient values that are not significantly different from zero. If squared residuals indicate autocorrelation, then it means that variance of returns is significantly autocorrelated and thus returns are conditionally heteroscedastic. In practice, the test is a test for autocorrelation in the squared residuals.

Engle's ARCH test can be derived as follows:

$$u_t = v_t \sigma_t \qquad v_t \sim N(0,1)$$

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_j x_{jt} + u_t \tag{21}$$

Test begins with linear regression of the form given in equation X and saving the residuals $\hat{u}_t$. v_t is normally distributed with zero mean and unit variance. Test continues by squaring the residuals and regressing them on q own lags:

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2 + \cdots + \gamma_q \hat{u}_{t-q}^2 + v_t \tag{22}$$

Obtaining the R^2 from the regression and multiplying it by the number of observations T we get the test statistic TR^2. Under the null hypothesis of no ARCH effects the test statistic follows the chi-squared distribution with q degrees of freedom. (Brooks 2019, 395)

The Jarque-Bera tests the null hypothesis that the data is normally distributed with an unknown mean and variance. The test statistic is:

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right), \tag{23}$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis. For large sample sizes, the test statistic has a chi-square distribution with two-degrees of freedom. For normally distributed data the skewness is zero and kurtosis gets value of three. Skewness can be defined as the shape of the distribution and measures the extent to which it is not symmetric about its mean value. Kurtosis indicates the fatness of the tails of the distribution and peakedness at the mean. Coefficient of excess kurtosis equal to the coefficient of kurtosis minus three. A normal distribution has a coefficient of excess kurtosis of zero. (Brooks 2019, 55)

Selecting the right models for data can be a difficult task. One can use graphical evaluation with autocorrelation and partial autocorrelation functions. However, interpreting the plots can be too subjective. Another evaluation method, called the information criteria, can help to evaluate different models' fit to data. Two well-known information criterions are used in this thesis and those are the Akaike's (1974) information criterion (AIC) and the Bayesian information criterion (BIC) by Schwarz (1978).

AIC and BIC consider the model's complexity and the maximum likelihood by considering a penalty term for including additional parameters. AIC incorporate a weak penalty term whereas the BIC uses more strict penalty term. The goal is to choose the number of parameters which minimizes the value of the information criteria. (Brooks 2019, 271)

Akaike's and Schwarz's information criteria are described as follows:

$$AIC = \ln(\hat{\sigma}^2) + \frac{2k}{T} \tag{24}$$

$$BIC = \ln(\hat{\sigma}^2) + \frac{k}{T}\, ln\, T$$

where $\hat{\sigma}^2$ is the residual variance, T is the sample size and k is the total number of parameters estimated. Brooks (2019, 272) asks important question, which criterion should be preferred if they suggest different model orders? It is not straightforward to say that the other should be preferred over the other. Brooks (2019, 272) answers that BIC is consistent but inefficient whereas AIC is not consistent but is generally more efficient. Author continues that BIC will deliver the correct model order, while AIC gives on average too large model. He also notes that, the average variation in selected model orders from different samples within a given population will be greater in the context of BIC than AIC. Conclusion is that neither criterion is superior to other. In this thesis, both criterions will be considered, and ARIMA-GARCH type models are chosen according to both AIC and BIC. Model offering smallest AIC and BIC values are chosen to volatility forecasting.

2.4 Literature review of volatility forecasting

In this chapter the previous research of volatility forecasting is presented. Research papers are presented in a chronological order starting from the early papers of volatility forecasting. This literature review is not a complete list of research related to volatility forecasting and the main purpose is to offer a compact view of previous research.

Akgiray (1989) forecast volatility using daily returns from value-weighted and equal-weighted stock indices from 1963 to 1986. The 24-year period is divided into four different periods of six years each and each period is analyzed separately. Forecasts are generated for the following month (20 days) and for each subsequent forecast, the estimation sample is shifted forward by one month by dropping the initial 20 observations and adding in the new 20 observations. Volatility forecasts are created using the simple historical variance estimate, exponentially weighted moving average, ARCH and GARCH models. The forecasts are compared under mean error, root mean square error, mean absolute error, and mean absolute percent error statistics. Results show that all error statistics indicate the superiority of the GARCH(1,1) and the model is capable of providing the most accurate forecasts through all periods. Based on model fitting and volatility forecasting, the GARCH(1,1) indeed outperforms other models but the research highlight that the results only hold for daily data. The return series for weekly and monthly series are not as leptokurtic as those of daily returns and there is no statistically significant autocorrelation in weekly and monthly returns. For squared and absolute series, the weekly data exhibit some autocorrelation up to four lags, but monthly series have no significant autocorrelation and hence the monthly returns concluded as a strict white noise.

Pagan and Schwert (1990) use monthly US stock return data from July 1835 to December 1925. Forecast models are estimated using data from 1835-1899 and forecasts are made for 1900-1925. Also, estimates from 1835-1925 are used to forecast for 1926-1937. Models are two-step conditional variance estimator, GARCH(1,2), EGARCH(1,2), Hamilton's switching-regime Markov model, nonparametric kernel model with one lag and nonparametric Fourier model with one and two lags. For the sample period 1835-1925, the results imply that the nonparametric models have the best explanatory power. However, for out-of-sample 1900-1925, the authors claim that the nonparametric models work poorly because of the few large returns during the sample period. Authors explain that the nonparametric models obtain their explanatory power from a few extreme returns. For the 1926-1937 predictions, the two-step, GARCH(1,2) and EGARCH(1,2) are the best performing models. Authors explain that these models can capture the volatility persistence rising from volatile period including the Great Depression (1929-1939). Authors claim that the large negative returns caused by the banking crisis of 1837, the banking panic of 1857, the start of the Civil War 1860 and the banking crisis 1907, have the biggest impact why the nonparametric models provide poor predictive results. Large negative returns are also the reason why the EGARCH(1,2) is the best performing model. Authors explain that the kernel and Fourier estimates react rapidly and adapt changes in volatility fastly, while the parametric GARCH, EGARCH and Hamilton estimates do it more gradually with slow adjustment to large volatility shocks and with volatility persistence. Although EGARCH(1,2) is the best model, Pagan and Schwert (1990) discuss that combining parametric models with terms recommended by non-parametric models, the explanatory power would be increasing.

Cao and Tsay (1992) compare volatility forecasting of threshold autoregressive (TAR) models with ARMA, GARCH and EGARCH models using data from US stock indices from January 1928 to December 1989. Using loss functions mean squared error and average absolute deviation, study claim that the EGARCH model is best for longer period volatility forecasting for small stock returns. Results also show that the TAR models outperform GARCH, EGARCH and ARMA models in volatility forecasting of large stock returns.

With daily Australian stock market data 1974-1993 Brailsford and Faff (1996) reported GJR-GARCH(1,1) as a best volatility forecasting model. However, comparing forecast results with different error statistics (the mean error, the mean absolute error, the root mean squared error and the mean absolute percentage error) indicate that there is no single model outperforming. Models included were random walk, historical mean, moving average (5-year) and moving average (12-year), an exponential smoothing model, EWMA, a regression model, GARCH(1,1), GARCH(3,1), GJR-GARCH(1,1) and GJR-GARCH(3,1). Model parameters were estimated using the period 1974-1985. To obtain monthly volatility forecasts, the daily rolling forecasts for each month was done and repeated until June 1993. The mean absolute error imply that the GJR-GARCH(1,1) is the most accurate forecasting model and the GJR-GARCH(3,1) is the second most accurate. The root mean squared error puts the historical mean and a regression model to the first place followed by the moving average (12-year), GJR-GARCH(1,1) and the EWMA. The mean absolute percentage error indicate that the GJR-GARCH(1,1) is the best model followed by the other GARCH models. Authors report also over- and under-predictions with mean mixed error MME(O) which penalize over-predictions and mean mixed error MME(U) which penalize under-predictions. They state that only the random walk and the GJR-GARCH(3,1) models provide equal number of over- and under-predictions. Other models, except the GJR-GARCH(1,1), over-predict volatility. According to MME(U) the GARCH models and the GJR-GARCH(3,1) are the best forecasting models whereas the GJR-GARCH(1,1) and random walk models ranks the bottom line. Another distinctive result from Brailsford and Faff (1996) is that the exponential smoothing model was ranked last, seventh and tenth by the MAE, RMSE and MAPE but offers the second-best forecasting model according to the MME(U). The MME(O) ranks the GJR-GARCH(1,1) as a best forecasting model. Authors highlight that the final purpose of the forecasting is important while choosing the error measurement. They explain that a buyer of a call option being more interested with over-predictions, would prefer the MME(O) statistic and would favor the GJR-GARCH(1,1) model. Brailsford and Faff (1996) conclude that they choose the GJR-GARCH(1,1) as a best forecasting model but also state the importance of choosing the appropriate error statistic and that the obtained rankings should be carefully assessed.

Franses and Van Dijk (1996) compare the volatility forecasting performance of asymmetric QGARCH(1,1) and GJR-GARCH(1,1) models relative to GARCH(1,1) model and also the performance of all three models relative to the simple random walk forecasting. Stock index data 1986-1994 from Germany, The Netherlands, Spain, Italy, and Sweden is used and weekly forecasts for the years 1990-1994 is evaluated using the median of squared error. Results show that the QGARCH and the random walk are the best models, and they do the forecasting almost equally well. GARCH(1,1) is ranked as a best forecasting model for two times, whereas GJR-GARCH(1,1) ranks the worst and authors recommend not to use the model. Authors note that the random walk model performs well with extreme values in the data (such as the 1987 stock market crash).

McMillan, Speight, and Gwilym (2000) study volatility forecasting using the historical mean, moving average, random walk, exponential smoothing, exponentially weighted moving average, simple regression, GARCH, TGARCH, EGARCH and component-GARCH models. Data from UK FTA All Share and FTSE100 stock index is analyzed with monthly, weekly, and daily frequencies and both with and without adjustment for the 1987 stock market crash. For the FTSE100 index, the data covers January 1984 to July 1996 (in-sample estimation 1984-1994 and out-of-sample forecasting 1995-1996). For the FTA All Share index, the data covers January 1969 to July 1996 (in-sample estimation 1969-1994 and out-of-sample forecasting 1995-1996). Forecast evaluation is made using the mean error (ME), root mean squared error (RMSE) and mean absolute error (MAE). Asymmetric loss functions, mean mixed error (MME(U) and MME(O)) statistics are also reported. Results for monthly frequency indicate that the random walk model offers the best forecasting with smallest ME, RMSE and MAE for both FTA All Share index and FTSE100 index. The simple regression and historical mean offer the poorest forecasting accuracy. The accuracy of the GARCH models is rather weak and specially when the 1987 stock market crash is included in FTSE100 index. Results for weekly frequency imply that the random walk model offers the best forecasting accuracy based on the MAE followed by the exponential smoothing model for the FTA All Share index. The moving average and recursive EWMA models are the next best forecasting models for both indices followed by the GARCH. The moving average, recursive exponential smoothing and EWMA models offer the best forecasting under the RMSE error statistic for the FTA All Share index. For FTSE data, the moving average models and the recursive EWMA provide superior forecasts. Under the ME statistic, the recursive exponential smoothing model provides the best forecast except in the crash-adjusted FTA, where the 3-month moving average model is preferred. For the FTA series and crash-unadjusted (1987 market crash excluded) FTSE data, the exponential smoothing and 3-month moving average models provide the best forecasts on the MAE statistic. With 1987 market crash included, the 3-month moving average is marginally superior to the random walk and other smoothing models for the FTSE data. The three-month moving average is also marginally superior for FTA and FTSE crash-unadjusted data under the RMSE statistic. The GARCH model is favored for the crash-adjusted FTSE data and exponential smoothing for the adjusted FTA data. Lastly, with asymmetric loss functions authors summarize that if overpredictions are penalized more heavily than underpredictions, then the random walk model outperforms. If underpredictions are more heavily penalized, then the historical mean is favored for the forecasting of daily FTA and FTSE volatility, while the historical mean and simple regression are jointly favored for weekly FTA volatility, and exponential smoothing is the best one for forecasting weekly FTSE volatility.

McMillan, Speight, and Gwilym (2000) sum that the random walk model provides the most accurate forecasts at the monthly and weekly frequencies, but model performance is rather poor for the daily returns. The exponential smoothing and moving average models provide more accurate weekly and daily forecasts. The historical mean and simple regression rank poorly compared to the other models. Overall, the moving average models provide good relative forecasts. The GARCH models also provide consistently relative fair performance in which GARCH and EGARCH outperforming the TGARCH and CGARCH models. Authors conclude that the moving average and GARCH models provide the most consistent forecasting performance if one considers forecasting method for all frequencies and symmetric loss functions.

Loudon, Watt, and Yadav (2000) study different ARCH models with daily UK stock index data from January 1971 to October 1997. The whole sample period is divided to three sub periods of January 1971 to December 1980, January 1981 to December 1990 and January 1991 to October 1997. Results show that the model estimates are all statistically significant meaning that the conditional variance is related to its previous level and to past innovations in returns. Volatility asymmetry is also revealed as the asymmetry parameters are highly significant although for the EGARCH, GJR-GARCH and TGARCH, the asymmetry was stronger during 1981 to 1990. For the NGARCH and VGARCH, the asymmetry coefficient was highest during 1991 to 1997. Model evaluation include comparing the predictability of the models with West-Cho (1995) test and Diebold-Mariano (1995) test. West-Cho statistic examine statistically significant differences in the conditional variance forecast errors across the models. Out-of-sample tests show highly significant differences. In-sample results show significance both for the first period and the last period. Pairwise tests (of the equality of the conditional variance forecast errors for the benchmark LGARCH and alternative models) show no in-sample or out-of-sample significance except in the third period out-of-sample performance where conditional variance forecast errors are significantly different between the LGARCH and the MGARCH, GJR-GARCH and NGARCH. For the Diebold-Mariano tests, none of the mean differences is significantly non-zero on an in-sample. For LGARCH and MGARCH and LGARCH and VGARCH the Diebold-Mariano statistic show significant difference for the second (1981-1990) out-of-sample period. Skewness and kurtosis tests in the standardized residuals show that the models can only partly capture the observed skewness and kurtosis. Also, volatility persistence is only partially captured by the models. None of the models outperform consistently across different sub-periods and authors conclude that the optimal choice of a model is period specific.

Comparing 330 different ARCH models using daily exchange rate data of Deutsche Mark-USD and IBM daily stock returns, Hansen and Lunde (2005) state that GARCH(1,1) works well with exchange rate data but is outperformed in stock returns data. Models covering leverage effects can produce better forecasting results with stock returns data.

Comparing the volatility forecasting ability of the historical standard deviation, an exponentially weighted moving average, GARCH(1,1), AGARCH, EGARCH and two regression models Ederington and Guan (2005) reveal that the GARCH(1,1) model overweight the most recent observations and put too little weight on older observations. This however has little impact on out-of-sample forecast accuracy, which is evaluated with root mean squared forecast error. Authors create a new model, the least squares regression model in which the forecast volatility for the future period is a weighted average of recent absolute return deviations with exponentially declining weights. The model outperforms both GARCH and EGARCH models. Authors use daily data, and the forecast horizons are 10, 20, 40, 80 and 120 trading days. Authors note that the GARCH model perform relatively better at short forecast horizons (10-40 days). Using data from S&P500 index, the Japanese yen/dollar exchange rate, the three-month Eurodollar rate, the 10-year treasury bond rate and five equities (Boeing, GM, International Paper, McDonald’s, and Merck), the result for out-of-sample forecasting is that the regression model provide best forecasts in T-bonds, Eurodollars, and yen, while GARCH outperforms in the S&P500. With equities both GARCH and regression model indicate the best forecasts.

Awartani and Corradi (2005) study predictive ability of GARCH, IGARCH, ABGARCH, EGARCH, TGARCH, GJR-GARCH, AGARCH and QGARCH models. Exponential smoothing model by RiskMetrics is also included. Models covering asymmetry are EGARCH, TGARCH, GJR-GARCH, AGARCH and QGARCH. Authors use dividend adjusted daily S&P-500 price index data from January 1990 to September 2001. Study reveals that asymmetric GARCH models, specially EGARCH outperforms the GARCH(1,1) for one-step ahead and longer forecast horizons using mean squared errors (MSE). In the multiple comparison, the asymmetric GARCH models again outperforms the GARCH(1,1). However, GARCH models that do not cover asymmetry, are not able to beat the GARCH(1,1). According to results, exponential smoothing model by RiskMetrics has poorest predictive ability.

Wilhelmsson (2006) with S&P500 futures data from 1996 to 2002 report that the GARCH(1,1) with t-distribution outperformed all the other volatility forecast models, including the moving average models. One of the more recent research by Sharma and Sharma (2015) compare the daily volatility forecasts of standard GARCH, and the more advanced EGARCH, GJR-GARCH, TGARCH, AVGARCH, APARCH and NGARCH models using 21 global stock indices from 2000 to 2013. The results show that the standard GARCH is the best volatility forecasting model. The other models do not offer any additional value and do not create better volatility forecasts. Sharma and Sharma (2015) however note that the GARCH models' volatility forecasting performance is sensitive to the choice of data set and there is no single GARCH model that provides the best forecast for all the 21 stock indices.

These results from previous research show that the forecasting results and performance of the GARCH-type models is not unanimous and there's no consistent view of superior volatility forecasting model. In their famous article, Poon and Granger (2003) review 93 research papers of volatility forecasting. They conclude that historical volatility methods, such as moving average, exponential weights and autoregressive models perform equally well compared to other sophisticated models including the GARCH-type models. However, Poon and Granger (2003) outline that with different data, loss functions or volatility proxies, the forecasting results might be different. Also, there's no unanimous conclusion that which of the loss functions one should always use and Poon and Granger (2003) mention that it is rarely discussed if one forecasting method is significantly better than another and although some particular method of forecasting volatility can be suggested being the best, there's no discussion about the cost-benefit from using it. Based on these aspects and results from previous research, interesting topic concerning the comparison of the GARCH-type and moving average models is that which of the models should we choose when considering forecasting volatility. Making empirical research with stock market data this thesis seeks to find more profound answer to model selection.

3 Data and methodology

The data in this thesis consists of six large stock market indices DAX30, FTSE100, NIKKEI225, Shanghai SE Composite, S&P500 and Dow Jones Industrial Average. These indices were chosen because they cover large amount of equity markets, and to compare how different GARCH-type and moving average models perform across global indices. Daily time series data includes the period from January 1, 2000 through December 31, 2020 and the data was obtained from Thomson Reuters Datastream. All the indices are price indices, except the DAX30 which is a total return index. The difference between price indices and total return indices are that dividends and other cash distributions are included in total return index. Price index tracks only the price movements of the component. For Shanghai SE Composite index it was not possible to get total return index observations for the whole sample period. The S&P500 is commonly displayed as a price index, whereas the DAX30 as a total return index. In addition of closing prices, the daily opening and high and low prices were obtained in purpose of the volatility proxy used in this thesis. National holidays were excluded from the data. No other preparations were done for the data. Full sample was divided to in-sample period of January 1, 2000 to December 31, 2015 and to out-of-sample forecasting period January 1, 2016 to December 31, 2020. Following approach used in literature, the returns in this thesis are expressed in logarithmic returns. Next, the indices are briefly presented.

The DAX30 is a German stock market index. It includes the 30 largest and most liquid companies on the Frankfurt Stock Exchange. The DAX30 includes global giants, such as BMW, Daimler, Siemens, and Volkswagen Group, and offers information about the market sentiment in Europe. (Deutsche Börse 2021) The FTSE100 is another European stock market index covering companies from wide range of industries. It is a capitalization-weighted share index of the 100 largest companies listed on the London Stock Exchange. (FTSE Russell 2021) The NIKKEI225 is a stock market index for the Tokyo Stock Exchange. Index consists of 225 large companies from different fields of businesses. (Nikkei 2021) The Shanghai Stock Exchange Composite index is a stock market index of all shares listed on the Shanghai Stock Exchange (China Securities Index 2021). One of the most well-known index, the S&P500 includes 500 largest companies traded on the New York Stock Exchange and

Nasdaq. The S&P500 covers large amount of US equities and is often used as an indicator of the US economy. The S&P500 is a free-float capitalization-weighted index and its current market cap is approximately 33 trillion USD. (S&P Dow Jones Indices 2021a) Another iconic index, the Dow Jones Industrial Average consists of the thirty largest US companies (S&P Dow Jones Indices 2021b).

Figure 1 displays the development of the indices from 2000 to 2020. The twenty-year period is next shortly discussed. Rather clear upward trend after 2010 is visible in all indices except in Shanghai SE Composite. Another distinctive feature is the development of the indices during the early years of 2000s. Downward trend turned upwards, especially in China due some speculative matters and rocketing markets, until the 2008 financial crisis which led to market crash. The economic situation was globally severe, but especially the drop in China was large and although Chinese markets have emerged, the Shanghai SE Composite has not reached its pre-2008 level. The aftermath of the 2008 crisis led to the European sovereign debt crisis which caused different problems and uncertainty in the eurozone. Some of this uncertainty emerged in 2012 as the concerns expanded from Greece to Spain and Italy which reported high budget deficits and public debt ratios. Together with disorder from many European banks, the crisis led to massive bailouts. Support from the European Union, European Central Bank and International Monetary Fund returned trust to markets. Rather steep slope of rising indices from 2012 is also visible in DAX30 and FTSE100. Another volatile period occurred in 2015-2016 which led to market drop in 2016. Reasons behind the drop are many, as it has been explained that the fall in oil prices, uncertainty in US and Chinese stock market (massive selloffs and speculation with Chinese stock market bubble) and United Kingdom decision to European Union membership referendum (Brexit) might cause the market downturn. During the most recent years, the indices dropped in 2018 mainly because of worries caused by the US and China trade war and a global economic slowdown and especially in US where the Federal Reserve informed about the rising interest rates. In 2020 the global stock markets saw an extraordinary dive due to COVID-19 pandemic. In March 2020 the markets crashed heavily as the pandemic caused fear and massive selloffs. The markets however recovered quite soon as the announcements from the governments and central banks financial stimulus convinced investors and markets turned upwards.

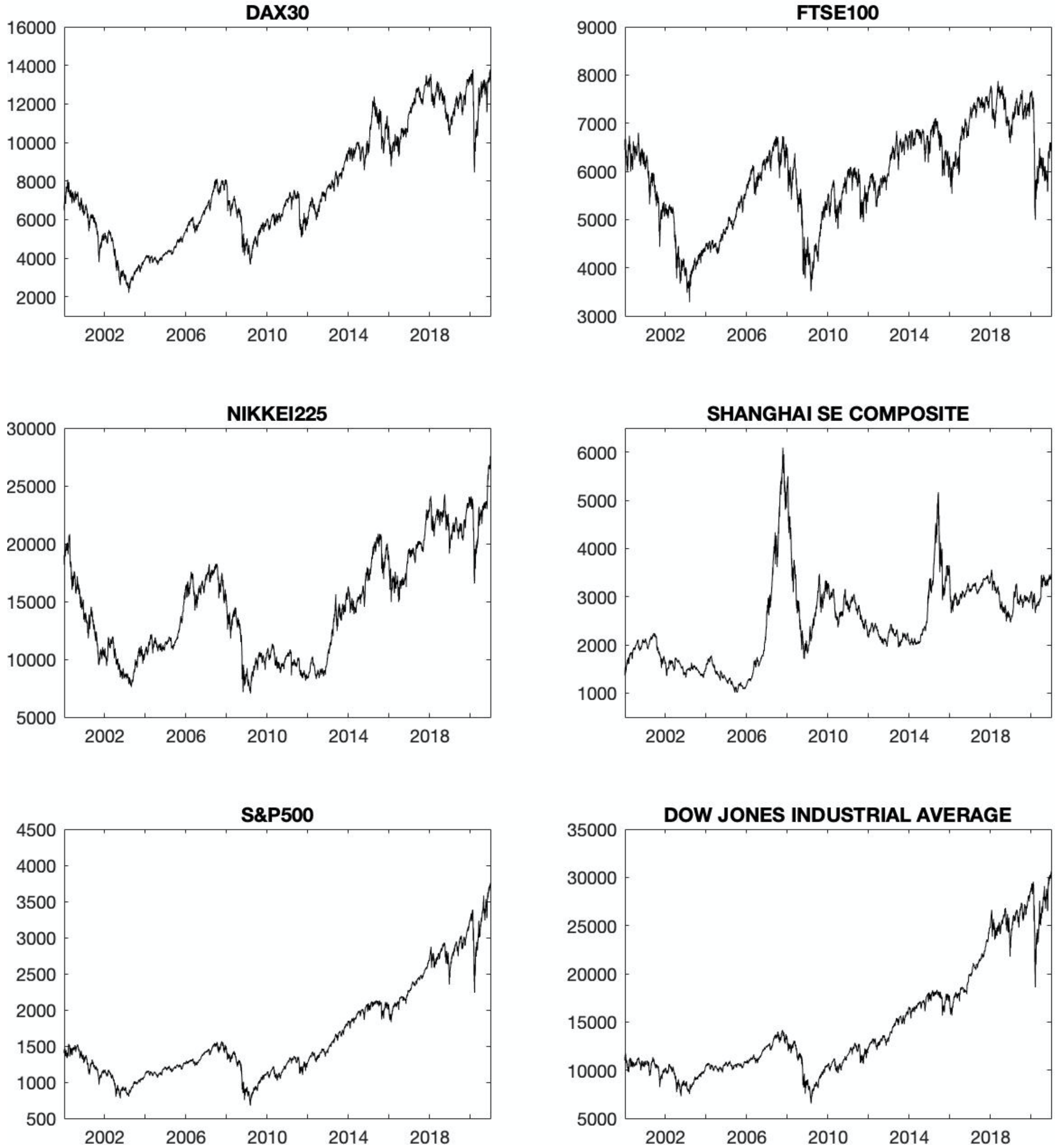

Figure 1. Price development of the equity indices from 2000 to 2020

3.1 Descriptive statistics

Logarithmic returns derived from the price indices are presented in Figure 2. Volatility clustering is clearly visible in all indices as high volatility following further high volatility and similar pattern with lower volatility following low volatility. Previously described market conditions are also visible and as the markets face turbulent periods it is evident that the returns fluctuate widely. Based on pure graphical evaluation, it seems that volatility bursts and persistent volatility is more present in Shanghai SE Composite index returns. The minimum daily returns seemed to happen during the 2020 COVID-19 pandemic in DAX30, FTSE100, S&P500 and Dow Jones indices. In Shanghai SE Composite, the minimum returns seemed to occur right before the 2008 financial crisis and during 2015-2016 market disorder. The biggest daily returns seemed to happen during the 2008 crisis and the 2020 pandemic.

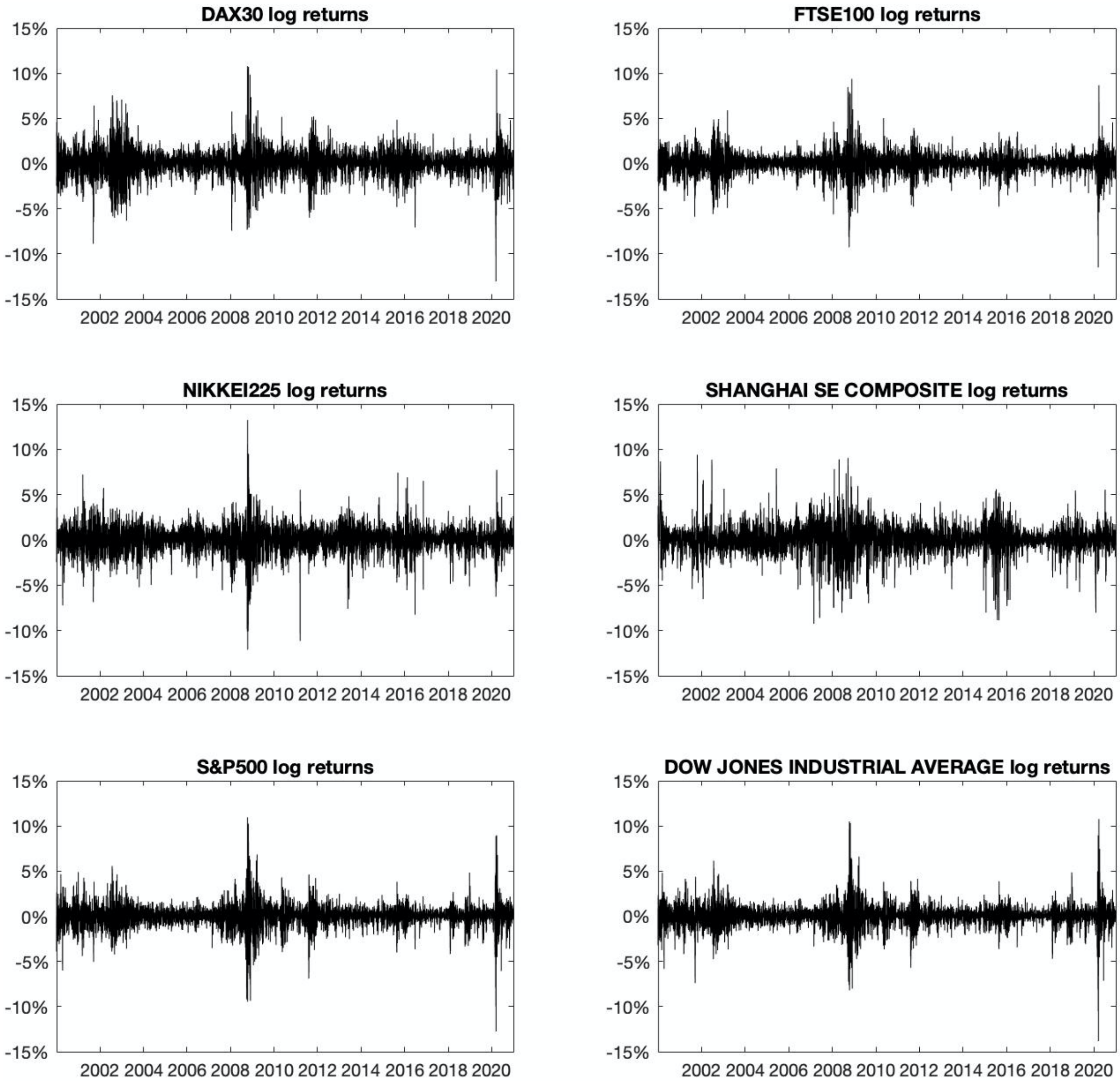

Figure 2. Logarithmic returns of equity indices from 2000 to 2020

Next, the in-sample returns are presented in QQ-plots in Figure 3. X-axis describes the standard normally distributed quantiles. The QQ-plots the data (log returns) against the normal distribution. The information from the QQ-plots indicate very strong nonnormality and leptokurtosis of the data. The returns have heavier tails as the distribution of sample data has observations in the upper right-hand side and in the lower left-hand side. Sample data observations are not along the red line, and this is a notion of non-normal distribution.

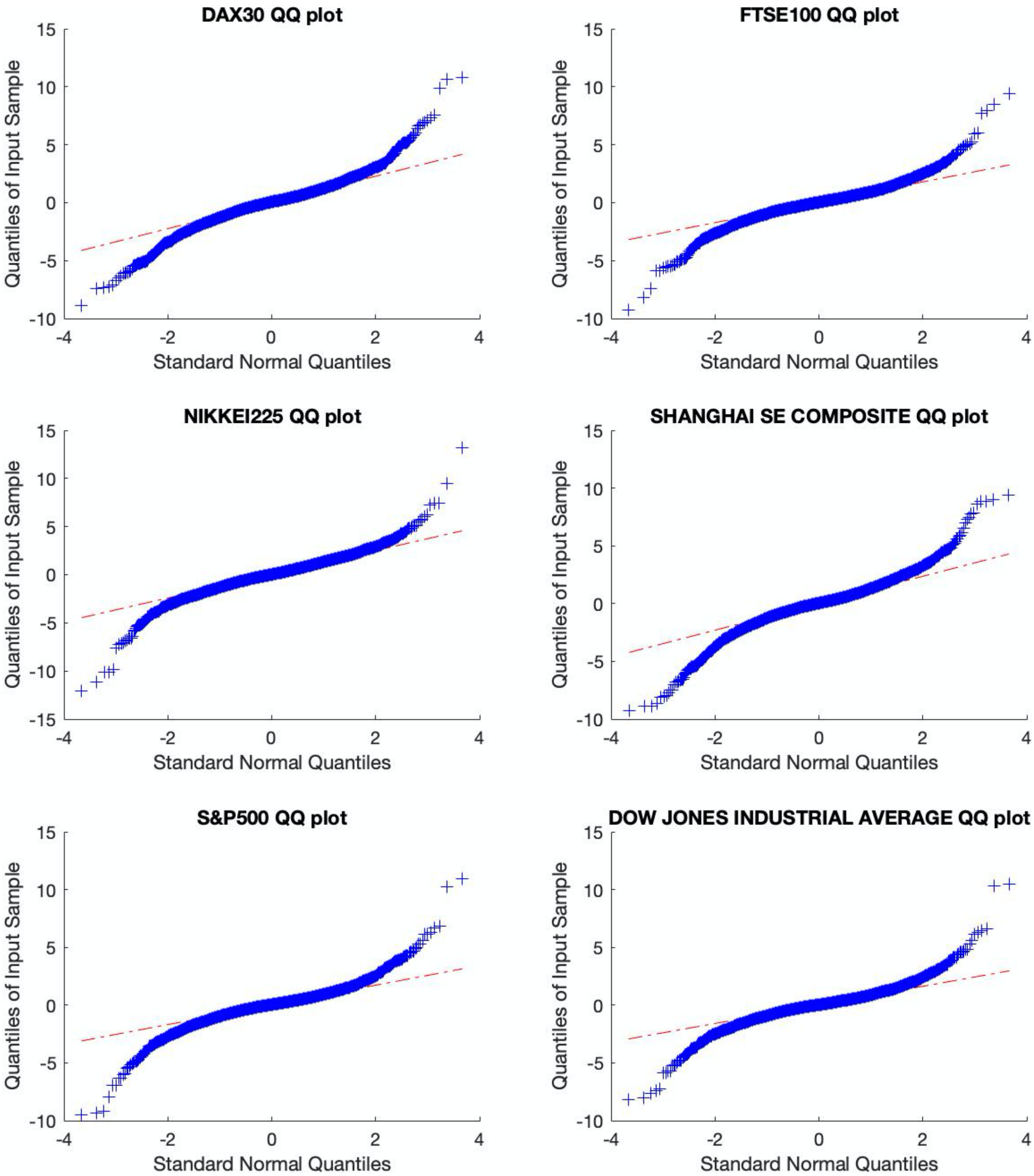

Figure 3. QQ-plots of the in-sample returns

Histograms of the in-sample returns are presented in Figure 4. Not so much can be said from the histograms. Visually evaluated noteworthy is the peakedness of the histograms as the highest peaks are around zero. Histograms seems to be quite symmetric and as noticed also from the Figure 2, few extreme observations can be picked from the tails. Non-normal distribution of the returns is verified by the histograms.

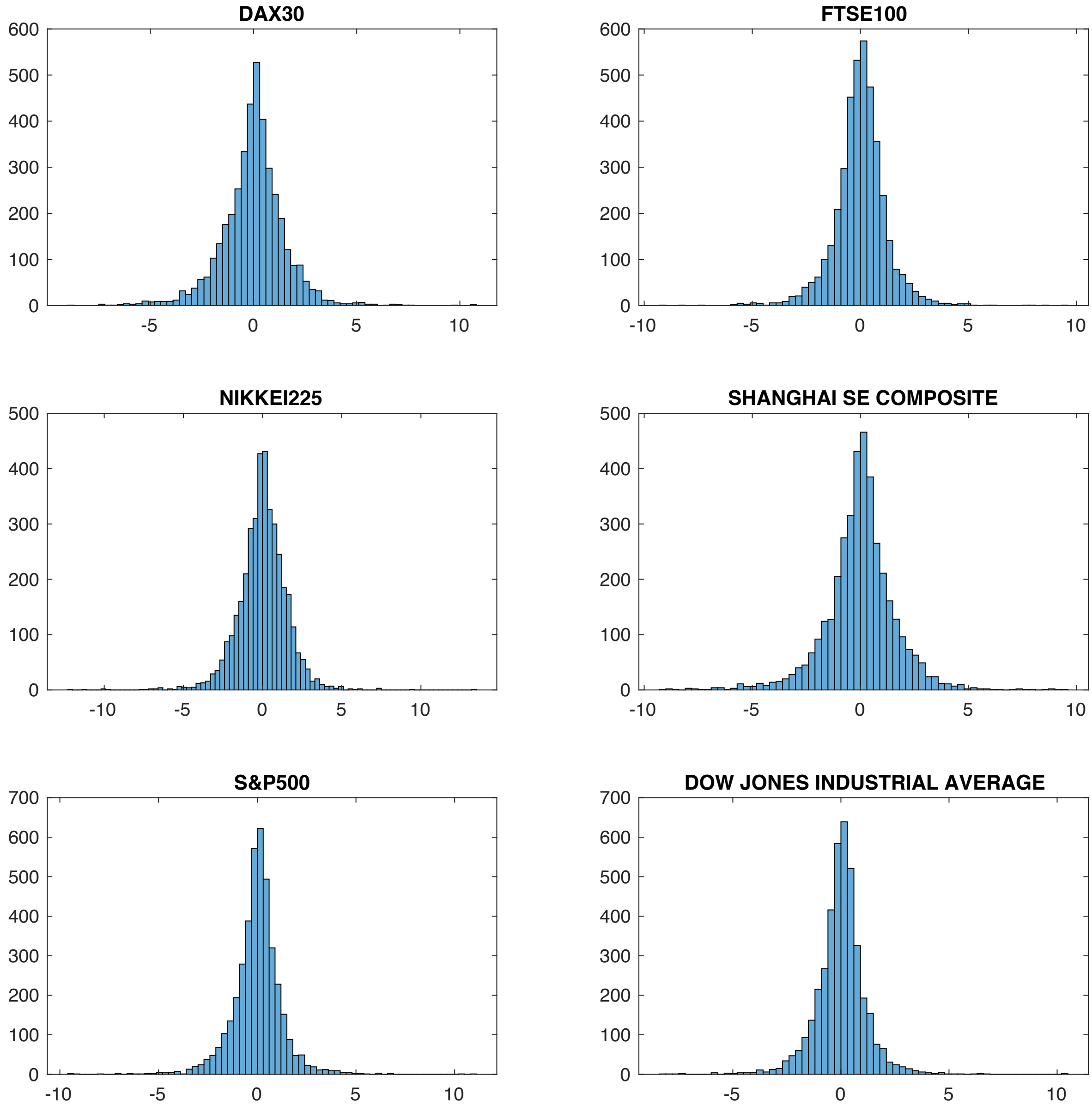

Figure 4. Histograms of the in-sample returns

Last visual diagnostics of the data are sample autocorrelation (ACF) and partial autocorrelation (PACF) functions. First, the ACF and PACF for the in-sample returns are presented in Figure 5.

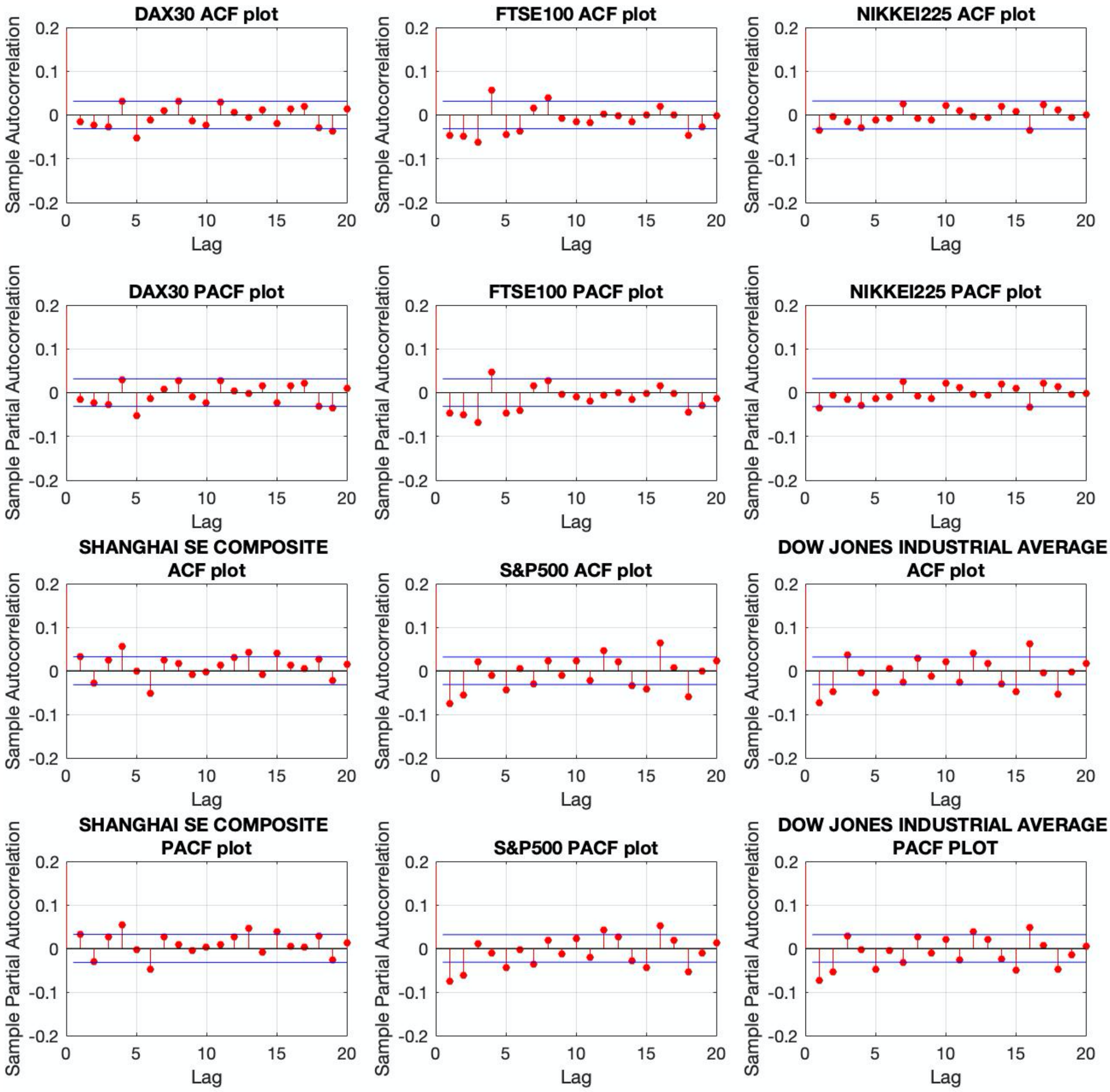

Figure 5. ACF and PACF of the in-sample returns

With five percent significance level there seems to be no autocorrelation in DAX30 returns, except in lag 5. Autocorrelation and partial autocorrelation functions for FTSE100 in-sample returns indicate significant autocorrelation of both signs in lags 1 to 6. Autocorrelation and partial autocorrelation functions for NIKKEI225 returns indicate no autocorrelation. Autocorrelation and partial autocorrelation functions for Shanghai SE Composite imply some autocorrelation at least in lags 4 and 6. Autocorrelation and partial autocorrelation functions for S&P500 and Dow Jones indicate autocorrelation at least in lags 2 and in lags 16 and 18. Overall, little can be said from the autocorrelation and partial autocorrelation functions as it seems that there is no or rather little autocorrelation.

Next, the ACF and PACF for the squared returns are presented in Figure 6. The ACF and PACF indicate persistent and significant autocorrelation. Visually this implies conditional heteroscedasticity of the returns.

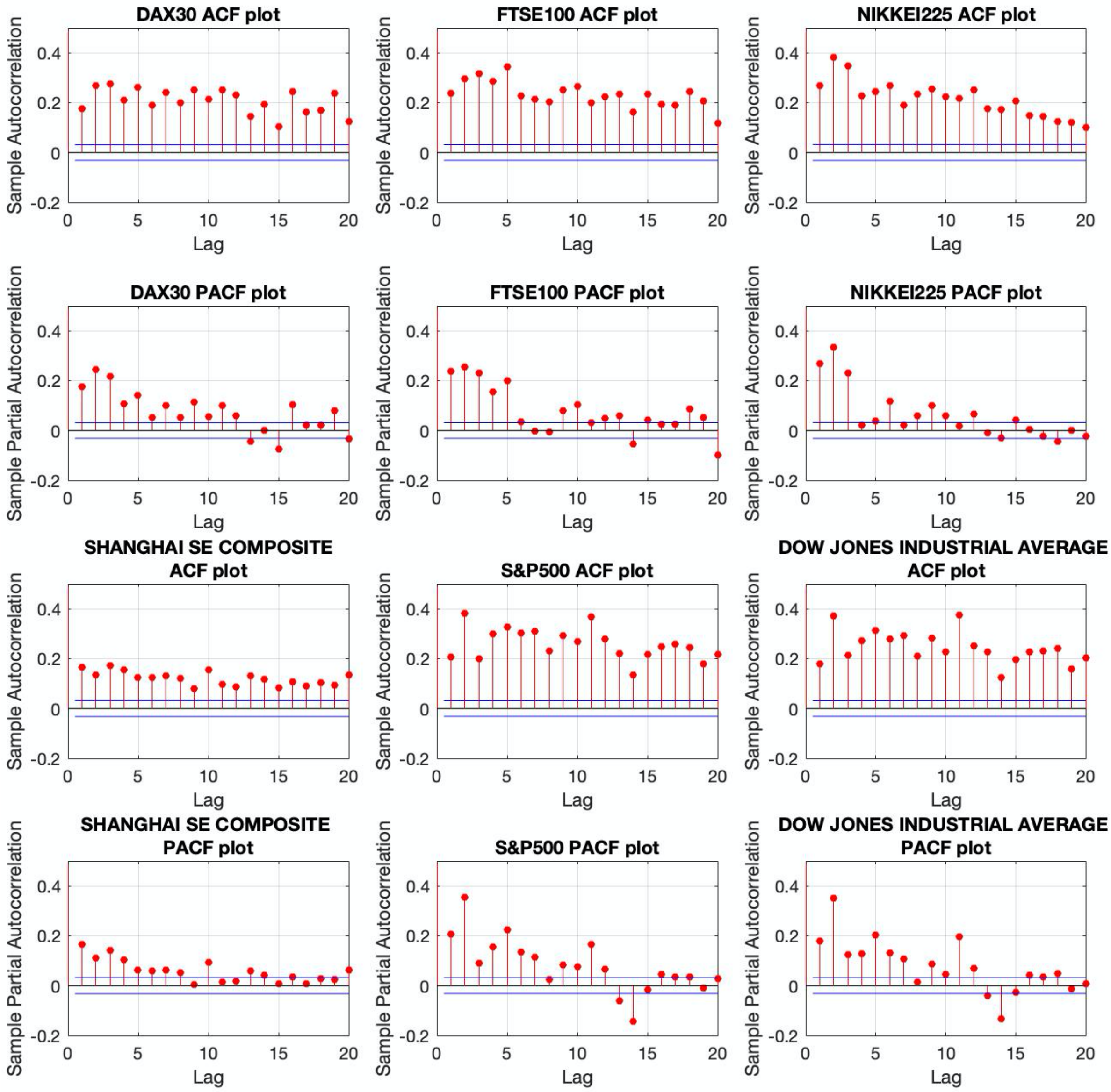

Figure 6. ACF and PACF of the in-sample squared returns

After visual representations of the data, the descriptive statistics for both full sample and in-sample are presented in Table 1. Similarity can be found through the indices as the sample mean, median and variance are all small. As the visual information from the QQ-plots imply, the high kurtosis verifies fat tails of the returns. Small negative skewness indicate that the returns are slightly left-tailed. Kurtosis and skewness both indicate that the returns are not normally distributed. Non-normality is verified with Jarque-Bera test as the null hypothesis of normal distribution is strongly rejected. ADF test rejects the null hypothesis of a unit root and show signs of stationary data. ADF test was done in three different ways; with autoregressive, autoregressive with drift and trend stationary models up to twenty lags. KPSS test verify the stationarity as the null hypothesis of time series being stationary is not rejected. Ljung-Box Q test is showing significant autocorrelation up to twenty lags as the null of no autocorrelation is rejected with 1% significance level. However, with NIKKEI225 data, the Ljung-Box Q test indicates no autocorrelation. Engle's ARCH test confirms visual evaluation of the squared returns as test imply conditional heteroscedasticity in returns by clearly rejecting the null hypothesis of no ARCH effects. Both visual inspection and tests confirm the use of GARCH-type models. Conditional mean model is justified by the autocorrelation in returns.

Table 1. Descriptive statistics for full sample and in-sample returns. Statistical significance at the 5% and 1% levels is denoted by ** and ***.

	DAX30	**FTSE100**	**NIKKEI225**	**Shanghai SE**	**S&P500**	**Dow Jones**
Panel A: In-sample 2000-2015						
N	4067	4040	3929	3872	4024	4024
Mean	0.00011	-0.00002	4.12964E-07	0.00025	0.00008	0.00011
Median	0.00079	0.00037	0.00033	0.00068	0.00053	0.00044
Standard Deviation	0.01549	0.01229	0.01545	0.01646	0.01267	0.01186
Sample Variance	0.00024	0.00015	0.00024	0.00027	0.00016	0.00014
Kurtosis	4.18406	6.08763	6.16320	4.34607	8.01334	7.91706
Skewness	-0.01878	-0.15312	-0.39776	-0.28014	-0.18535	-0.06631
Minimum	-0.08875	-0.09266	-0.12111	-0.09256	-0.09470	-0.08201
Maximum	0.10797	0.09384	0.13235	0.09401	0.10957	0.10508
ADF	-14.3548***	-15.4930***	-14.0431***	-12.3881***	-14.6808***	-15.0166***
KPSS	0.0629	0.0633	0.0819	0.0822	0.0517	0.0376
Ljung-Box	47.5233***	83.5720***	24.8912	59.5365***	111.4298***	107.7610***
Engle's ARCH test	879.4908***	1032.5000***	992.1632***	396.1117***	1244.6000***	1182.9000***
Jarque-Bera	2968.3645***	6235.6187***	6301.4952***	3087.0901***	10756.0847***	10477.0040***
Panel B: Full sample 2000-2020						
N	5330	5305	5149	5090	5282	5282
Mean	0.00013	-5.89805E-06	0.00007	0.00018	0.00018	0.00019
Median	0.00076	0.00041	0.00041	0.00065	0.00059	0.00051
Standard Deviation	0.01491	0.01199	0.01490	0.01549	0.01255	0.01208
Sample Variance	0.00022	0.00014	0.00022	0.00024	0.00016	0.00015
Kurtosis	5.74919	7.83999	6.34990	5.00961	10.94638	13.02634
Skewness	-0.16468	-0.32603	-0.37359	-0.37041	-0.39325	-0.37702
Minimum	-0.13055	-0.11512	-0.12111	-0.09256	-0.12765	-0.13842
Maximum	0.10797	0.09384	0.13235	0.09401	0.10957	0.10764
ADF	-16.5012***	-17.2605***	-16.1274***	-14.3887***	-16.5782***	-17.0102***
KPSS	0.0596	0.0581	0.0557	0.0547	0.0344	0.0237
Ljung-Box	35.0808**	82.7594***	22.9644	66.7635***	173.7501***	190.9219***
Engle's ARCH test	1026.3000***	1164.2000***	1185.8000***	527.8228***	1657.0000***	1687.0000***
Jarque-Bera	7346.6308***	13650.8543***	8748.6224***	5424.7910***	26446.7912***	37378.6864***

3.2 Loss functions MSE & MAE and Diebold-Mariano test

Next, a short discussion about the loss functions and the Diebold-Mariano test used in this thesis. Together with volatility proxy, the loss functions or error functions, are important aspects when comparing the performance of volatility forecasting models. Previous literature has used different methods to evaluate and compare the forecasting performance of the models. Some common methods are root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and mean mixed error (MME). Mean squared error measures the average squared difference between the estimated (forecasted) values and the actual values, which is the Yang-Zhang volatility proxy in this thesis. Mean absolute error measures the average absolute difference between the forecasted and the actual values. Hansen and Lunde (2006) present six different loss functions. Authors state that MSE is a robust loss function. MSE however penalize large errors more than MAE and Brooks (2019, 285) claim that the usefulness of MSE depends on whether large forecast errors are disproportionately more serious than smaller errors.

MSE is computed as a sum of squared differences between the forecasted values and the actual values divided by the sample size. MSE is given by:

$$MSE = \frac{1}{T-(T_1-1)}\sum_{t=T_1}^{T}(\sigma_t - \sigma_{f,t})^2 \tag{25}$$

where T is the number of observations and T_1 is the first out-of-sample forecast observation. σ_t and $\sigma_{f,i}$ are the volatility proxy and forecasted volatility at time t.

MAE is computed as a sum of absolute differences between the forecasted values and the actual values divided by the sample size. MAE is given by:

$$MAE = \frac{1}{T-(T_1-1)}\sum_{t=T_1}^{T}|\sigma_t - \sigma_{f,t}| \tag{26}$$

where T is the number of observations and T_1 is the first out-of-sample forecast observation σ_t and $\sigma_{f,i}$ are the volatility proxy and forecasted volatility at time t.

In order to assess more deeply the performance of the volatility forecasting models some other loss functions could also be used. However, in the context of this thesis, the MSE and

MAE are good candidates because they emphasize different things and there's conceptual difference between these two loss functions. Also, their wide use in previous research is another reason why they are chosen in this thesis to rank the volatility forecasting models.

The loss functions MAE and MSE show the ranking of the volatility forecasting models. To assess whether these rankings hold statistically significant difference, the Diebold-Mariano test (1995) is used. White's (2000) Reality Check and Hansen's (2005) Superior Predictive Ability tests are other ways to compare the forecasts. In this thesis we'll focus on the Diebold-Mariano test and the test is done so that the best ranked model under MAE and MSE from each index and its volatility forecast σ_{fi} is tested against all the other model forecasts σ_{fj} separately. The two forecast errors are then computed as $e_i = \sigma_{YZ} - \sigma_{fi}$ and $e_j = \sigma_{YZ} - \sigma_{fj}$, where σ_{YZ} is the Yang-Zhang volatility proxy. The loss differentials are then defined as $d_t = e_{it}^2 - e_{jt}^2$ for the squared forecast errors and as $d_t = |e_{it}| - |e_{jt}|$ for the absolute forecast errors. The test is based on the loss differentials of the forecast errors and the null hypothesis of equal forecast accuracy means that the population mean of the loss differential series is zero. The alternative hypothesis is that the two forecasts have different levels of accuracy. (Diebold and Mariano 1995; Diebold 2015)

Test statistic of the Diebold and Mariano is:

$$DM = \frac{\overline{d}}{\sqrt{\left(\left(\frac{1}{T}\right) * \widehat{Var}\left(\overline{d}\right)\right)}} \tag{27}$$

where $\overline{d}$ is the sample mean loss differential and $\widehat{Var}\left(\overline{d}\right)$ is a consistent estimate of the variance of loss differential. Under the null hypothesis of equal forecast accuracy, the Diebold-Mariano test statistic follows a standard normal distribution $N(0,1)$. (Diebold and Mariano 1995)

3.3 Making of volatility forecasts

This chapter explains how volatility forecasting was done. First, the full sample data was divided to in-sample period of January 1, 2000 to December 31, 2015 and to out-of-sample forecasting period of January 1, 2016 to December 31, 2020. After the data sampling the second decision was to choose the specific GARCH-type models. The basic GARCH(1,1), EGARCH(1,1) and GJR(1,1) models were selected first as a benchmarking models for volatility forecasting. Previous studies, e.g., Engle and Ng (1993), Brailsford and Faff (1996), have shown that these basic models offer good volatility forecasting results. Other GARCH-type models were then chosen based on purely AIC and BIC information criterion. Because these information criteria can give different results it was chosen that if the results vary (i.e., the AIC and BIC suggest different models), then both of these were considered when choosing the models. As an example with the DAX30 data, the AIC suggested the ARIMA(1,0,1)-EGARCH(2,3) whereas, the BIC proposed the ARIMA(0,0,0)-EGARCH(1,3), so therefore both of these models were chosen. Similar model selection was done with all of the six indices and after the selection the models were fitted with the in-sample data.

Before forecasting the standardized residuals of the fitted models were tested with Ljunq-Box test, Engle's ARCH test and Jarque-Bera test. Log-likelihood values were also considered to check the adequacy of the fitted models. Making the goodness-of-fit tests it was clear that the selected models based on AIC and BIC were not the best ones. There was e.g., some ARCH effects still left, and the Ljunq-Box test indicated that there was also some significant autocorrelation left in the standardized residuals. With NIKKEI225 the models based on AIC and BIC seemed to be fitted quite well since all the initial models did not reject Ljunq-Box or ARCH test. The forecasting was done in Matlab and the process was to run daily and monthly volatility forecasts with one-step ahead rolling window all the way through the out-of-sample period 2016-2020. In this way, the oldest observation dropped out and models were re-estimated until to the end of the out-of-sample period. However, the computational problems appeared when trying to run other, some better fitting models. With some bigger models, e.g. ARIMA(2,0,2)-GARCH(3,3) Matlab was not able to fully run the forecasts and there was missing forecast values. For these reasons it was decided to make the daily volatility forecasts based on the AIC and BIC model selection and the author of this thesis recognizes that these models may not be the best ones and some alternative models could offer better volatility forecasting results. Also, the monthly volatility forecasts were chosen to be done only with the GARCH(1,1), EGARCH(1,1) and GJR(1,1) models. With monthly forecasts there was also some computational problems as there was missing forecast values for EGARCH(1,1) with NIKKEI225 data. Again, author of this thesis recognize that some other models could be better. It was also decided not to include the moving average models for monthly forecasting because of time and length limitation of making this thesis.

Selected daily volatility forecasting models for DAX30 were ARIMA(1,0,0)-GARCH(1,2), ARIMA(0,0,0)-GARCH(1,1), ARIMA(1,0,1)-EGARCH(2,3), ARIMA(0,0,0)-EGARCH(1,3), ARIMA(1,0,1)-GJR-GARCH(1,1), ARIMA(0,0,0)-GJR-GARCH(1,1). Forecasting models for FTSE100 were ARIMA(1,0,1)-GARCH(1,2), ARIMA(1,0,0)-GARCH(1,1), ARIMA(1,0,1)-EGARCH(1,2), ARIMA(1,0,1)-EGARCH(1,1), ARIMA(1,0,1)-GJR-GARCH(1,1), ARIMA(0,0,0)-GJR-GARCH(1,1). For NIKKEI225 the forecasting models were ARIMA(2,0,2)-GARCH(1,2), ARIMA(0,0,0)-GARCH(1,1), ARIMA(1,0,1)-EGARCH(1,2), ARIMA(0,0,0)-EGARCH(2,2), ARIMA(1,0,1)-GJR-GARCH(1,1), ARIMA(0,0,0)-GJR-GARCH(1,1). Forecasting models for Shanghai SE Composite were ARIMA(1,0,1)-GARCH(1,2), ARIMA(0,0,0)-GARCH(1,1), ARIMA(1,0,1)-EGARCH(2,2), ARIMA(1,0,1)-EGARCH(1,1), ARIMA(1,0,1)-GJR-GARCH(1,2), ARIMA(0,0,0)-GJR-GARCH(1,1). For S&P500 the volatility forecasting models were ARIMA(1,0,1)-GARCH(2,2), ARIMA(1,0,1)-GARCH(1,2), ARIMA(1,0,1)-EGARCH(1,2), ARIMA(1,0,1)-EGARCH(1,1), ARIMA(1,0,1)-GJR-GARCH(1,1). And lastly, the forecasting models for DOW Jones were ARIMA(1,0,1)-GARCH(2,2), ARIMA(0,0,0)-GARCH(2,2), ARIMA(0,0,0)-EGARCH(2,2), ARIMA(1,0,1)-EGARCH(1,1), ARIMA(1,0,1)-GJR(2,2), ARIMA(1,0,1)-GJR(1,1). Volatility forecasts were also made with the basic GARCH(1,1), EGARCH(1,1) and GJR(1,1) models in all of the indices. In addition of normal distribution, the t-distributed forecasting models were also included to compare their volatility forecasting performance. For the moving average models, the five-day moving average (MA5), one month moving average (MA21), three month moving average (MA63) and the exponentially weighted moving average (EWMA) forecasting models were included for all of the indices. After obtaining the daily and monthly volatility forecasts, each forecasting window was computed with the MSE and MAE loss functions from the equations 25 and 26. Tables from 2 to 8 in the next chapter are presenting the average of these errors and the models are ranked based on these values. Lastly, the Diebold-Mariano test was used to test statistical difference of the models' forecasts.

4 Results

In this chapter the daily and monthly volatility forecasting results for each index are presented. The results from the Diebold-Mariano test of equal predictive accuracy are also included.

4.1 Volatility forecasting results

DAX30

The daily volatility forecasting results for DAX30 are presented in Table 2. ARIMA(1,0,1)-GJR-GARCH(1,1) with t-distribution is the best performing volatility forecasting model under both loss functions. The MA5 and ARIMA(1,0,1)-GJR-GARCH with normal distribution are on a second and third place based on MSE. Normally distributed ARIMA(1,0,1)-EGARCH(2,3) and ARIMA(1,0,1)-GJR-GARCH(1,1) holds second and third place based on MAE loss function. Being the second best model under MAE, the normally distributed ARIMA(1,0,1)-EGARCH(2,3) ranks among the middle class based on the MSE. Basic GJR-GARCH(1,1) models are performing well under MSE whereas same models are on a middle cast based on MAE. According to MSE, the basic EGARCH(1,1) models are right after the basic GJR(1,1) models (although the numerical difference between the models is rather large). However, under MAE, the basic GJR-GARCH(1,1) models are outperformed by the basic EGARCH(1,1) models. The bottom three ranking with MSE loss function is MA21, t-distributed ARIMA(1,0,0)-GARCH(1,2) and MA63. Under MAE, the bottom ranked models are GARCH(1,1) and ARIMA(1,0,0)-GARCH(1,2) both t-distributed, and MA63. GARCH models are among the worst performing models with both MSE and MAE.

Table 2. Volatility forecasting performance and ranking of the models for DAX30

DAX30

Model	Average MSE	Rank		Average MAE	Rank
ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.686116	1	ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.051951	1
MA5	0.688841	2	ARIMA(1,0,1)-EGARCH(2,3)	0.052063	2
ARIMA(1,0,1)-GJR-GARCH(1,1)	0.695545	3	ARIMA(1,0,1)-GJR-GARCH(1,1)	0.052656	3
ARIMA(0,0,0)-GJR-GARCH(1,1)	0.695892	4	ARIMA(0,0,0)-GJR-GARCH(1,1)	0.052679	4
ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.695892	5	ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.052679	5
GJR-GARCH(1,1)	0.707818	6	EGARCH(1,1)	0.053046	6
GJR-GARCH(1,1) t	0.708878	7	EGARCH(1,1) t	0.053118	7
EGARCH(1,1) t	0.737638	8	ARIMA(0,0,0)-EGARCH(1,3)	0.053193	8
EGARCH(1,1)	0.766232	9	ARIMA(1,0,1)-EGARCH(2,3) t	0.053224	9
ARIMA(0,0,0)-EGARCH(1,3) t	0.837488	10	ARIMA(0,0,0)-EGARCH(1,3) t	0.053226	10
ARIMA(0,0,0)-EGARCH(1,3)	0.839728	11	GJR-GARCH(1,1)	0.054015	11
ARIMA(0,0,0)-GARCH(1,1)	0.863346	12	GJR-GARCH(1,1) t	0.054197	12
ARIMA(1,0,1)-EGARCH(2,3)	0.872928	13	MA5	0.055224	13
GARCH(1,1)	0.877539	14	EWMA	0.056255	14
ARIMA(1,0,0)-GARCH(1,2)	0.885136	15	MA21	0.056995	15
ARIMA(0,0,0)-GARCH(1,1) t	0.890407	16	ARIMA(0,0,0)-GARCH(1,1)	0.057456	16
GARCH(1,1) t	0.909510	17	GARCH(1,1)	0.057738	17
ARIMA(1,0,1)-EGARCH(2,3) t	0.909779	18	ARIMA(1,0,0)-GARCH(1,2)	0.058147	18
EWMA	0.927982	19	ARIMA(0,0,0)-GARCH(1,1) t	0.058216	19
MA21	0.968938	20	GARCH(1,1) t	0.058438	20
ARIMA(1,0,0)-GARCH(1,2) t	0.980727	21	ARIMA(1,0,0)-GARCH(1,2) t	0.060740	21
MA63	1.443383	22	MA63	0.068780	22

It is also noteworthy from the German market results that the EWMA and MA21 are ranked higher under MAE close to the middle ranking whereas under MSE both models are ranked to the bottom. The EWMA is ranked just below the middle class under MAE and the model is outperforming all the GARCH models. Under MSE, EWMA's performance is among the bottom range models. The best performing model under both loss functions is t-distributed. This is an expected result as the data showed signs of excess kurtosis. The t-distributed models are not however constantly outperforming models with normal distributions. In fact, the results indicate that under both loss functions, the relative performance, and the ranking of the models with normal distributions is better compared to models with t-distributions.

FTSE100

Table 3 presents the daily volatility forecasting results for FTSE100. Under MSE loss function, the top three ranked models are ARIMA(1,0,1)-EGARCH(1,1) with t-distribution, basic EGARCH(1,1) and ARIMA(1,0,1)-EGARCH(1,1) with normal distributions. Under MAE loss function, the ARIMA(1,0,1)-EGARCH(1,1) with t-distribution ranks first, the ARIMA(1,0,1)-EGARCH(1,2) with t-distribution ranks second and basic EGARCH(1,1) holds third position. Under MSE, the EWMA is among bottom ranked models followed by MA21 and MA63. EWMA, ARIMA(1,0,1)-GARCH(1,2) with t-distribution and moving average model MA63 are ranked as bottom three models under MAE loss function. Being the second best model under MAE, the ARIMA(1,0,1)-EGARCH(1,2) with t-distribution ranks sixth under MSE. Otherwise model rankings are rather similar with both loss functions as the EGARCH and GJR type models holds the upper and middle class rankings. Basic GARCH(1,1) is outperformed by all the EGARCH and GJR models. Under MAE, the MA5 is close to basic GJR-GARCH(1,1) model and the model is also outperforming all the GARCH models.

Table 3. Volatility forecasting performance and ranking of the models for FTSE100

FTSE100

Model	Average MSE	Rank		Average MAE	Rank
ARIMA(1,0,1)-EGARCH(1,1) t	0.376954	1	ARIMA(1,0,1)-EGARCH(1,1) t	0.042879	1
EGARCH(1,1)	0.382116	2	ARIMA(1,0,1)-EGARCH(1,2) t	0.043708	2
ARIMA(1,0,1)-EGARCH(1,1)	0.382485	3	EGARCH(1,1)	0.043879	3
EGARCH(1,1) t	0.390204	4	ARIMA(1,0,1)-EGARCH(1,1)	0.043967	4
ARIMA(1,0,1)-EGARCH(1,2)	0.396519	5	EGARCH(1,1) t	0.044322	5
ARIMA(1,0,1)-EGARCH(1,2) t	0.400224	6	ARIMA(1,0,1)-EGARCH(1,2)	0.044420	6
ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.448037	7	ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.045160	7
ARIMA(1,0,1)-GJR-GARCH(1,1)	0.450159	8	ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.045540	8
ARIMA(0,0,0)-GJR-GARCH(1,1)	0.454161	9	ARIMA(1,0,1)-GJR-GARCH(1,1)	0.045755	9
ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.454963	10	ARIMA(0,0,0)-GJR-GARCH(1,1)	0.046024	10
GJR-GARCH(1,1)	0.457192	11	GJR-GARCH(1,1)	0.046169	11
ARIMA(1,0,0)-GARCH(1,1)	0.469458	12	MA5	0.046548	12
ARIMA(1,0,1)-GARCH(1,2)	0.472088	13	GJR-GARCH(1,1) t	0.046656	13
GJR-GARCH(1,1) t	0.472404	14	ARIMA(1,0,0)-GARCH(1,1)	0.047825	14
GARCH(1,1)	0.473553	15	GARCH(1,1)	0.047931	15
MA5	0.482676	16	ARIMA(1,0,1)-GARCH(1,2)	0.047980	16
ARIMA(1,0,0)-GARCH(1,1) t	0.484868	17	MA21	0.048165	17
ARIMA(1,0,1)-GARCH(1,2) t	0.487855	18	ARIMA(1,0,0)-GARCH(1,1) t	0.048320	18
GARCH(1,1) t	0.488531	19	GARCH(1,1) t	0.048406	19
EWMA	0.570884	20	EWMA	0.048482	20
MA21	0.606920	21	ARIMA(1,0,1)-GARCH(1,2) t	0.048521	21
MA63	0.835566	22	MA63	0.056862	22

A distinctive feature from the FTSE100 results is that all asymmetric EGARCH models are performing well and all GJR-GARCH models are ranked among middle class according to both loss functions. Model rankings based on normal distributions and t-distributions vary. Both loss functions show that the best model is t-distributed and with MAE also the second best model holds t-distribution. With basic GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) the normally distributed models outperform their t-distribution counterparts. Based on the data with excess kurtosis and non-normality, it is surprising result that the t-distributed models are not constantly outperforming models with normal distributions.

NIKKEI225

The daily volatility forecasting results for NIKKEI225 are presented in Table 4. According to both loss functions, the ARIMA(0,0,0)-EGARCH(2,2) with normal distribution is the best performing model. Based on MAE loss function, the model is followed by MA5 and ARIMA(2,0,2)-EGARCH(2,2) with t-distribution. Basic EGARCH(1,1) with normal and t-distribution ranks second and third under MSE. Same models are doing pretty well also under MAE. Top performer among MAE, the MA5 ranks around middle position under MSE. According to both loss functions, the GJR-GARCH models ranks after the EGARCH type models. Being the second worst forecasting model under MSE, the MA21 ranks a bit higher under MAE and outperforms both EWMA and all GARCH models. As was the case with German DAX30 and UK FTSE100, the performance of the GARCH type models is again rather poor under both loss functions.

Table 4. Volatility forecasting performance and ranking of the models for NIKKEI225

NIKKEI225

Model	Average MSE	Rank		Average MAE	Rank
ARIMA(0,0,0)-EGARCH(2,2)	0.513319	1	ARIMA(0,0,0)-EGARCH(2,2)	0.056959	1
EGARCH(1,1)	0.521137	2	MA5	0.057159	2
EGARCH(1,1) t	0.524877	3	ARIMA(0,0,0)-EGARCH(2,2) t	0.057637	3
ARIMA(0,0,0)-EGARCH(2,2) t	0.530426	4	EGARCH(1,1)	0.058659	4
ARIMA(1,0,1)-EGARCH(1,2)	0.532504	5	ARIMA(1,0,1)-EGARCH(1,2)	0.058952	5
ARIMA(1,0,1)-EGARCH(1,2) t	0.548547	6	EGARCH(1,1) t	0.058966	6
ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.580790	7	ARIMA(1,0,1)-EGARCH(1,2) t	0.059609	7
ARIMA(1,0,1)-GJR-GARCH(1,1)	0.593452	8	ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.060684	8
ARIMA(0,0,0)-GJR-GARCH(1,1)	0.593531	9	ARIMA(1,0,1)-GJR-GARCH(1,1)	0.061531	9
ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.593531	10	ARIMA(0,0,0)-GJR-GARCH(1,1)	0.061540	10
GJR-GARCH(1,1)	0.605000	11	ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.061540	11
GJR-GARCH(1,1) t	0.607939	12	GJR-GARCH(1,1)	0.062313	12
MA5	0.674646	13	GJR-GARCH(1,1) t	0.062452	13
ARIMA(0,0,0)-GARCH(1,1)	0.691834	14	MA21	0.063843	14
GARCH(1,1)	0.692283	15	EWMA	0.064012	15
EWMA	0.696202	16	ARIMA(0,0,0)-GARCH(1,1)	0.066121	16
ARIMA(2,0,2)-GARCH(1,2)	0.704979	17	GARCH(1,1)	0.066216	17
ARIMA(0,0,0)-GARCH(1,1) t	0.710006	18	ARIMA(2,0,2)-GARCH(1,2)	0.066424	18
GARCH(1,1) t	0.711554	19	ARIMA(0,0,0)-GARCH(1,1) t	0.066868	19
ARIMA(2,0,2)-GARCH(1,2) t	0.741796	20	GARCH(1,1) t	0.067021	20
MA21	0.771787	21	ARIMA(2,0,2)-GARCH(1,2) t	0.067962	21
MA63	0.972174	22	MA63	0.075250	22

Nikkei index data shows signs of asymmetry and it is a surprising result that simple moving average beats more complex GARCH type models which are created to capture well-known characteristics of stock market data. Moving average is purely an average of a certain time period and the only dynamic feature is adding new variables and dropping the old ones as time period goes further. EWMA on the other hand puts more weight on more recent data and it is memorizing past events based on weighting factor lambda. This could be understood as an asset in changing market conditions and hence ability to produce more accurate forecasts. The forecasting performance of the EWMA is far from the EGARCH and GJR models but still the model beats all the GARCH models under MAE and gets rather close to GARCH(1,1) with MSE. According to results from the table 4, the top performing model is normally distributed and all the basic GARCH(1,1), EGARCH(1,1) and GJR(1,1) models with normal distributions are performing slightly better than their t-distributed counterparts.

Shanghai SE Composite

Table 5 provides the daily volatility forecasting results for Shanghai SE Composite index. According to MSE loss function, the top three volatility forecasting models are EWMA, MA5 and MA21. Same models are in top three also with MAE loss function as the MA5 is ranked first followed by MA21 and on a third position EWMA. ARIMA(1,0,1)-GJR-GARCH(1,2) with t-distribution is the poorest forecasting model under MAE. The same model ranks the second worst forecasting model under MSE while MA63 holds the bottom position. In general, the EGARCH type models are performing well as they are ranked after the MA5, MA21 and EWMA. Normally distributed basic GARCH(1,1) outperforms all GJR-GARCH models and is on a middle position based on the rankings of MSE and MAE. This is again surprising result as it would’ve been expected that the GJR-GARCH models follow the similar kind asymmetric EGARCH models more closely. Ranked on a bottom under MSE, the MA63 rise few steps higher under MAE. Otherwise, the model rankings are rather similar on both loss functions. Moving average models MA5, MA21 and EWMA are showing good volatility forecasting performance and the differences are quite big compared to more complex GARCH, EGARCH and GJR-GARCH models. This is rather surprising result, and one possible explanation can be that moving average models are able to response and react more quickly to changing market conditions in a high volatile environment. Based on the results, the GARCH-type models with normal distributions are persistently outperforming the t-distributed models.

Table 5. Volatility forecasting performance and ranking of the models for Shanghai SE Composite

Shanghai SE Composite

Model	Average MSE	Rank		Average MAE	Rank
EWMA	0.540193	1	MA5	0.052253	1
MA5	0.556823	2	MA21	0.053951	2
MA21	0.598278	3	EWMA	0.054216	3
ARIMA(1,0,1)-EGARCH(2,2)	0.604012	4	ARIMA(1,0,1)-EGARCH(1,1)	0.059726	4
ARIMA(1,0,1)-EGARCH(1,1)	0.612422	5	EGARCH(1,1)	0.059805	5
EGARCH(1,1)	0.613252	6	ARIMA(1,0,1)-EGARCH(2,2)	0.060107	6
ARIMA(1,0,1)-EGARCH(2,2) t	0.625067	7	ARIMA(1,0,1)-EGARCH(1,1) t	0.061355	7
ARIMA(1,0,1)-EGARCH(1,1) t	0.636503	8	ARIMA(1,0,1)-EGARCH(2,2) t	0.061441	8
EGARCH(1,1) t	0.641823	9	EGARCH(1,1) t	0.061527	9
ARIMA(0,0,0)-GARCH(1,1)	0.646364	10	ARIMA(0,0,0)-GARCH(1,1)	0.062297	10
GARCH(1,1)	0.646405	11	GARCH(1,1)	0.062309	11
ARIMA(0,0,0)-GJR-GARCH(1,1)	0.657852	12	ARIMA(1,0,1)-GARCH(1,2)	0.062734	12
GJR-GARCH(1,1)	0.658809	13	ARIMA(0,0,0)-GJR-GARCH(1,1)	0.062784	13
ARIMA(1,0,1)-GARCH(1,2)	0.658951	14	GJR-GARCH(1,1)	0.062838	14
ARIMA(1,0,1)-GJR-GARCH(1,2)	0.661192	15	ARIMA(1,0,1)-GJR-GARCH(1,2)	0.062861	15
GARCH(1,1) t	0.674023	16	GARCH(1,1) t	0.064388	16
ARIMA(0,0,0)-GARCH(1,1) t	0.674065	17	ARIMA(0,0,0)-GARCH(1,1) t	0.064402	17
ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.686723	18	MA63	0.064516	18
GJR-GARCH(1,1) t	0.690028	19	ARIMA(0,0,0)-GJR-GARCH(1,1) t	0.064978	19
ARIMA(1,0,1)-GARCH(1,2) t	0.717509	20	GJR-GARCH(1,1) t	0.065143	20
ARIMA(1,0,1)-GJR(1,2) t	0.732184	21	ARIMA(1,0,1)-GARCH(1,2) t	0.065646	21
MA63	0.772009	22	ARIMA(1,0,1)-GJR-GARCH(1,2) t	0.066443	22

S&P500

Table 6 presents the daily volatility forecasting results for S&P500. T-distributed and normally distributed ARIMA(1,0,1)-EGARCH(1,1) followed by MA5 are top three forecasting models according to MAE loss function. Top three models under MSE loss functions are normally distributed ARIMA(1,0,1)-EGARCH(1,1), EGARCH(1,1) and ARIMA(1,0,1)-EGARCH(1,1) with t-distribution. EWMA, MA21 and MA63 are the bottom ranked models under MSE. ARIMA(1,0,1)-GARCH(1,2) with t-distribution, EWMA and MA63 are the last three models under MAE loss function. Being the top performer under MSE, the normally distributed ARIMA(1,0,1)-EGARCH(1,1) ranks second under MAE. Same model t-distributed counterpart holds first place under MAE, while its ranking is third under MSE. Another difference between the rankings of the models is that the basic EGARCH(1,1) with normal distribution ranks fourth under MAE, while the model holds second place under MSE. MA5 is ranked third under MAE and sixth under MSE loss function and there's quite big numerical difference to the winner ARIMA(1,0,1)-EGARCH(1,1) with normal distribution. MA21 is among worst performing models under MSE whereas ranking changes few positions higher under MAE.

Table 6. Volatility forecasting performance and ranking of the models for S&P500

S&P500

Model	Average MSE	Rank		Average MAE	Rank
ARIMA(1,0,1)-EGARCH(1,1)	0.335936	1	ARIMA(1,0,1)-EGARCH(1,1) t	0.040685	1
EGARCH(1,1)	0.346272	2	ARIMA(1,0,1)-EGARCH(1,1)	0.040798	2
ARIMA(1,0,1)-EGARCH(1,1) t	0.361836	3	MA5	0.041613	3
EGARCH(1,1) t	0.374369	4	EGARCH(1,1)	0.042301	4
ARIMA(1,0,1)-EGARCH(1,2)	0.392890	5	ARIMA(1,0,1)-EGARCH(1,2)	0.043211	5
MA5	0.413721	6	EGARCH(1,1) t	0.043288	6
ARIMA(1,0,1)-GJR-GARCH(1,1)	0.456754	7	ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.045581	7
ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.483563	8	ARIMA(1,0,1)-GJR-GARCH(1,1)	0.046155	8
GJR-GARCH(1,1)	0.485374	9	GJR-GARCH(1,1)	0.047510	9
GARCH(1,1)	0.523860	10	GJR-GARCH(1,1) t	0.048355	10
ARIMA(1,0,1)-GARCH(1,2)	0.528586	11	GARCH(1,1)	0.050915	11
ARIMA(1,0,1)-GARCH(2,2)	0.528806	12	ARIMA(1,0,1)-GARCH(1,2)	0.051156	12
GJR-GARCH(1,1) t	0.544563	13	ARIMA(1,0,1)-GARCH(2,2)	0.051227	13
GARCH(1,1) t	0.606601	14	GARCH(1,1) t	0.052723	14
ARIMA(1,0,1)-GARCH(2,2) t	0.628026	15	MA21	0.053226	15
ARIMA(1,0,1)-GARCH(1,2) t	0.629620	16	ARIMA(1,0,1)-GARCH(2,2) t	0.053632	16
EWMA	0.747593	17	ARIMA(1,0,1)-GARCH(1,2) t	0.053654	17
MA21	0.813991	18	EWMA	0.055225	18
MA63	1.298019	19	MA63	0.069468	19

The results from the S&P500 show that the benchmark models GARCH(1,1) and GJR-GARCH(1,1) are all performing moderately. EGARCH(1,1) is particularly top performer under both loss functions. Asymmetric EGARCH and GJR-GARCH models outperform GARCH models under MAE loss function. Results are similar under MSE, although t-distributed GJR-GARCH(1,1) drops a bit lower and is being outperformed by the ARIMA(1,0,1)-GARCH(1,2) and ARIMA(1,0,1)-GARCH(2,2) models. The results also reveal that the normally distributed ARIMA(1,0,1)-EGARCH(1,1) and EGARCH(1,1) followed by ARIMA(1,0,1)-EGARCH(1,1) with t-distribution are the best models according MSE. T-distributed ARIMA(1,0,1)-EGARCH(1,1) takes the lead under MAE followed by normally distributed models. T-distributed GARCH models are among the worst volatility forecasting models under MAE. With MSE some of the t-distributed models are also close to bottom rankings. Also, normally distributed GARCH(1,1), EGARCH(1,1) and GJR(1,1) are all outperforming their t-distributed counterparts. It's not straightforward to say which distribution is the best one as the rankings vary depending on the loss function. However, normally distributed models seem to have an upper hand when comparing models separately under the loss functions.

Dow Jones Industrial Average

The daily volatility forecasting results for Dow Jones index are provided in Table 7. According to results, the ARIMA(1,0,1)-EGARCH(1,1) and ARIMA(0,0,0)-EGARCH(2,2) both with normal distributions, followed by t-distributed ARIMA(1,0,1)-EGARCH(1,1), are the top three volatility forecasting models under MAE loss function. According to MSE loss function, the top three models are normally distributed ARIMA(1,0,1)-EGARCH(1,1), EGARCH(1,1) and ARIMA(0,0,0)-EGARCH(2,2). MA21, EWMA and MA63 are the bottom three models under MAE. Worst three volatility forecasting models under MSE are EWMA, MA21 and MA63. EGARCH-type models are all ranked as a best performing models followed by MA5 under MAE loss function. Results are similar with MSE as the EGARCH models perform well, although the MA5 is outperforming the t-distributed ARIMA(0,0,0)-EGARCH(2,2) model. Compared to more complex models, the simple moving average model MA5 is also performing well under both loss functions. A distinctive result between the loss functions is the performance of MA5 and MA21 as the MA5 is superior and MA21 is ranked among the worst performing models.

Table 7. Volatility forecasting performance and ranking of the models for Dow Jones

Dow Jones

Model	Average MSE	Rank	Model	Average MAE	Rank
ARIMA(1,0,1)-EGARCH(1,1)	0.404091	1	ARIMA(1,0,1)-EGARCH(1,1)	0.040584	1
EGARCH(1,1)	0.413114	2	ARIMA(0,0,0)-EGARCH(2,2)	0.040783	2
ARIMA(0,0,0)-EGARCH(2,2)	0.416344	3	ARIMA(1,0,1)-EGARCH(1,1) t	0.040883	3
ARIMA(1,0,1)-EGARCH(1,1) t	0.425303	4	EGARCH(1,1)	0.041880	4
EGARCH(1,1) t	0.433174	5	EGARCH(1,1) t	0.042880	5
MA5	0.469753	6	ARIMA(0,0,0)-EGARCH(2,2) t	0.043432	6
ARIMA(0,0,0)-EGARCH(2,2) t	0.470570	7	MA5	0.044311	7
ARIMA(1,0,1)-GJR-GARCH(1,1)	0.552562	8	ARIMA(1,0,1)-GJR-GARCH(2,2) t	0.046662	8
ARIMA(1,0,1)-GJR-GARCH2,2)	0.571168	9	ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.046672	9
GJR-GARCH(1,1)	0.583272	10	ARIMA(1,0,1)-GJR-GARCH(1,1)	0.046705	10
ARIMA(1,0,1)-GJR-GARCH(2,2) t	0.586441	11	ARIMA(1,0,1)-GJR-GARCH(2,2)	0.047299	11
ARIMA(1,0,1)-GJR-GARCH(1,1) t	0.587503	12	GJR-GARCH(1,1)	0.048222	12
GARCH(1,1)	0.633181	13	GJR-GARCH(1,1) t	0.049068	13
GJR-GARCH(1,1) t	0.638675	14	GARCH(1,1)	0.051801	14
ARIMA(0,0,0)-GARCH(2,2)	0.640260	15	ARIMA(0,0,0)-GARCH(2,2)	0.052292	15
ARIMA(1,0,1)-GARCH(2,2)	0.641253	16	ARIMA(1,0,1)-GARCH(2,2)	0.052317	16
GARCH(1,1) t	0.713474	17	GARCH(1,1) t	0.053465	17
ARIMA(0,0,0)-GARCH(2,2) t	0.727646	18	ARIMA(0,0,0)-GARCH(2,2) t	0.054232	18
ARIMA(1,0,1)-GARCH(2,2) t	0.733538	19	ARIMA(1,0,1)-GARCH(2,2) t	0.054434	19
EWMA	0.940432	20	MA21	0.055535	20
MA21	1.008261	21	EWMA	0.057705	21
MA63	1.592943	22	MA63	0.072078	22

As was the case with S&P500 results, the EWMA is ranked among bottom class under both loss functions. As a benchmark model, the basic EGARCH(1,1) is also rather good based on both loss functions. Asymmetric EGARCH and GJR-GARCH models volatility forecasting performance is again relatively good and the models are outperforming the GARCH models. For some reason the difference between the EGARCH models and GJR-GARCH models is however rather big and again the EGARCH beats the GJR-GARCH. Comparing the performance of normally distributed and t-distributed models, we can see that the first ranked model is normally distributed based on both loss functions. As was the case with S&P500, it seems also that with the Dow Jones, the normally distributed models are outperforming their t-distributed counterparts when comparing models separately under the loss functions. The results of the ARIMA(1,0,1)-GJR-GARCH(1,1) and ARIMA(1,0,1)-GJR-GARCH(2,2) are interesting as under MSE, the normally distributed models ranks higher whereas under MAE the t-distributed models outperform, although the relative numerical difference is rather small.

MONTHLY VOLATILITY FORECASTS

Lastly, the results from the monthly volatility forecasts are presented in Table 8. For the DAX, the GJR-GARCH(1,1) with normal distribution ranks first under both MAE and MSE loss functions. GARCH(1,1) with t-distribution ranks bottom based on MSE and EGARCH(1,1) with t-distribution ranks last under MAE. For the FTSE, the EGARCH(1,1) with t-distribution holds first place based on both loss functions. Normally distributed GJR-GARCH(1,1) ranks last under both loss functions. For the Nikkei, the EGARCH(1,1) and the GJR-GARCH(1,1) both with normal distributions ranks first according to MSE and MAE loss functions. T-distributed GARCH(1,1) ranks last under MSE and EGARCH(1,1) with normal distribution is bottom ranked under MAE loss function. For the Shanghai, the EGARCH(1,1) with normal distribution ranks first based on MSE whereas the GARCH(1,1) also with normal distribution holds first place under MAE. T-distributed GJR-GARCH(1,1) ranks last under both loss functions. For the S&P500, the normally distributed EGARCH(1,1) ranks first based on both loss functions. GARCH(1,1) with t-distribution is bottom ranked. Lastly, for the Dow Jones, the GJR-GARCH(1,1) with normal distribution ranks first under MSE and the EGARCH(1,1) with normal distribution takes the first place under MAE. As was the case with the S&P500, the GARCH(1,1) with t-distribution is bottom ranked also with Dow Jones index.

Table 8. Monthly volatility forecasting performance and ranking of the models

DAX30						**FTSE100**					
	Average MSE	Rank		Average MAE	Rank		Average MSE	Rank		Average MAE	Rank
GJR-GARCH(1,1)	1.038112	1	GJR-GARCH(1,1)	0.085131	1	EGARCH(1,1) t	0.520153	1	EGARCH(1,1) t	0.059988	1
GJR-GARCH(1,1) t	1.067352	2	GJR-GARCH(1,1) t	0.086057	2	EGARCH(1,1)	0.520848	2	EGARCH(1,1)	0.060044	2
EGARCH(1,1)	1.347843	3	GARCH(1,1)	0.093033	3	GARCH(1,1)	0.527053	3	GARCH(1,1)	0.060056	3
GARCH(1,1)	1.445399	4	EGARCH(1,1)	0.094602	4	GARCH(1,1) t	0.534777	4	GARCH(1,1) t	0.060347	4
EGARCH(1,1) t	1.580262	5	GARCH(1,1) t	0.096440	5	GJR-GARCH(1,1) t	0.610669	5	GJR-GARCH(1,1) t	0.060782	5
GARCH(1,1) t	1.627665	6	EGARCH(1,1) t	0.103237	6	GJR-GARCH(1,1)	0.611860	6	GJR-GARCH(1,1)	0.060902	6

Nikkei225						**Shanghai SE Composite**					
	Average MSE	Rank		Average MAE	Rank		Average MSE	Rank		Average MAE	Rank
EGARCH(1,1)	1.310782	1	GJR-GARCH(1,1)	0.104735	1	EGARCH(1,1)	2.048330	1	GARCH(1,1)	0.115751	1
GJR-GARCH(1,1)	1.36129	2	GARCH(1,1)	0.106331	2	GARCH(1,1)	2.051822	2	GARCH(1,1) t	0.116991	2
GARCH(1,1)	1.432529	3	GJR-GARCH(1,1) t	0.107769	3	EGARCH(1,1) t	2.053889	3	EGARCH(1,1)	0.119691	3
EGARCH(1,1) t	1.451752	4	GARCH(1,1) t	0.109959	4	GARCH(1,1) t	2.078758	4	EGARCH(1,1) t	0.119976	4
GJR-GARCH(1,1) t	1.551990	5	EGARCH(1,1) t	0.114264	5	GJR-GARCH(1,1)	2.126378	5	GJR-GARCH(1,1)	0.122096	5
GARCH(1,1) t	1.586940	6	EGARCH(1,1)	0.115243	6	GJR-GARCH(1,1) t	2.132338	6	GJR-GARCH(1,1) t	0.122419	6

S&P500						**Dow Jones**					
	Average MSE	Rank		Average MAE	Rank		Average MSE	Rank		Average MAE	Rank
EGARCH(1,1)	0.733942	1	EGARCH(1,1)	0.072242	1	GJR-GARCH(1,1)	0.701103	1	EGARCH(1,1)	0.071893	1
GJR-GARCH(1,1)	0.740654	2	EGARCH(1,1) t	0.074052	2	GJR-GARCH(1,1) t	0.719029	2	GJR-GARCH(1,1)	0.072663	2
EGARCH(1,1) t	0.756713	3	GJR-GARCH(1,1)	0.074229	3	EGARCH(1,1)	0.731785	3	EGARCH(1,1) t	0.072728	3
GJR-GARCH(1,1) t	0.761025	4	GJR-GARCH(1,1) t	0.074742	4	EGARCH(1,1) t	0.752979	4	GJR-GARCH(1,1) t	0.073125	4
GARCH(1,1)	1.206325	5	GARCH(1,1)	0.087995	5	GARCH(1,1)	1.046227	5	GARCH(1,1)	0.083679	5
GARCH(1,1) t	1.300972	6	GARCH(1,1) t	0.089548	6	GARCH(1,1) t	1.205571	6	GARCH(1,1) t	0.086795	6

As the results of daily and monthly volatility forecasts indicate, the asymmetric EGARCH and GJR-GARCH performed well, and these models almost constantly outperformed the standard GARCH model across the indices. The results from the daily volatility forecasts also show the good performance of the five-day moving average MA5, as the model outperformed many of the more complex GARCH-type models. The moving average models performed well also with Shanghai data as the EWMA was the best performing model followed by the MA5 and MA21 under MSE loss function. Based on MAE, the MA5 was ranked first followed by the MA21 and EWMA. According to results from the monthly forecasts, the exception to top performance of the asymmetric models are the results from Shanghai as under MAE, the GARCH(1,1) with normal and t-distribution ranks first and second place. Under MSE, the EGARCH(1,1) holds first place followed by the GARCH(1,1). Next, the Diebold-Mariano test results for daily and monthly volatility forecasts are presented.

4.2 Diebold-Mariano test results

DAX30

Table 9 provides the Diebold-Mariano test results for DAX30 index. According to the Diebold-Mariano test under squared loss differentials, there's no statistically significant difference between the forecasts of t-distributed ARIMA(1,0,1)-GJR(1,1) and MA5, normally distributed ARIMA(1,0,1)-GJR(1,1) and ARIMA(0,0,0)-GJR(1,1) and t-distributed ARIMA(0,0,0)-GJR(1,1). Similar result with equal forecast accuracy holds for all EGARCH-type model forecasts with 1% significance level. The only exception is the t-distributed ARIMA(1,0,1)-EGARCH(2,3) indicating significant difference between its forecast and the t-distributed ARIMA(1,0,1)-GJR(1,1). With 5% significance level results vary a bit as the ARIMA(0,0,0)-EGARCH(1,3) and ARIMA(1,0,1)-EGARCH(2,3) indicate rejecting the null hypothesis of equal predictive accuracy. For all the other model forecasts, the null hypothesis of equal predictive accuracy is rejected. Under absolute loss differentials the Diebold-Mariano test shows equal predictive accuracy between the forecasts of t-distributed ARIMA(1,0,1)-GJR(1,1) and all EGARCH-type model forecasts. Compared to squared loss functions, there's clear significant difference between the forecasts of t-distributed ARIMA(1,0,1)-GJR(1,1) and all other GJR-type models.

Table 9. Results of the Diebold-Mariano tests for DAX30. Statistical significance at the 5% and 1% levels is denoted by ** and ***

DAX30

Model ranking based on the average MSE	Diebold-Mariano test statistic	p value	Model ranking based on the average MAE	Diebold-Mariano test statistic	p value
ARIMA(1,0,1)-GJR-GARCH(1,1) t			ARIMA(1,0,1)-GJR-GARCH(1,1) t		
MA5	-0.03962	0.9684	ARIMA(1,0,1)-EGARCH(2,3)	-0.10433	0.9169
ARIMA(1,0,1)-GJR-GARCH(1,1)	-1.51378	0.1301	ARIMA(1,0,1)-GJR-GARCH(1,1)	-4.90804***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1)	-1.57125	0.1161	ARIMA(0,0,0)-GJR-GARCH(1,1)	-5.08472***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1) t	-1.57125	0.1161	ARIMA(0,0,0)-GJR-GARCH(1,1) t	-5.08472***	0.0000
GJR-GARCH(1,1)	-5.21325***	0.0000	EGARCH(1,1)	-1.29651	0.1948
GJR-GARCH(1,1) t	-5.61228***	0.0000	EGARCH(1,1) t	-1.51839	0.1289
EGARCH(1,1) t	-1.28542	0.1986	ARIMA(0,0,0)-EGARCH(1,3)	-1.27764	0.2014
EGARCH(1,1)	-1.55598	0.1197	ARIMA(1,0,1)-EGARCH(2,3) t	-1.11509	0.2648
ARIMA(0,0,0)-EGARCH(1,3) t	-2.55912**	0.0105	ARIMA(0,0,0)-EGARCH(1,3) t	-1.32415	0.1855
ARIMA(0,0,0)-EGARCH(1,3)	-2.31365**	0.0207	GJR-GARCH(1,1)	-13.42225***	0.0000
ARIMA(0,0,0)-GARCH(1,1)	-3.36206***	0.0008	GJR-GARCH(1,1) t	-15.97602***	0.0000
ARIMA(1,0,1)-EGARCH(2,3)	-2.40455**	0.0162	MA5	-2.38750**	0.0170
GARCH(1,1)	-3.42521***	0.0006	EWMA	-3.73110***	0.0002
ARIMA(1,0,0)-GARCH(1,2)	-3.63059***	0.0003	MA21	-4.30312***	0.0000
ARIMA(0,0,0)-GARCH(1,1) t dist	-3.86656***	0.0001	ARIMA(0,0,0)-GARCH(1,1)	-6.43379***	0.0000
GARCH(1,1) t	-3.81001***	0.0001	GARCH(1,1)	-6.57157***	0.0000
ARIMA(1,0,1)-EGARCH(2,3) t	-2.87150***	0.0041	ARIMA(1,0,0)-GARCH(1,2)	-7.13829***	0.0000
EWMA	-2.96734***	0.0030	ARIMA(0,0,0)-GARCH(1,1) t	-7.14021***	0.0000
MA21	-3.53357***	0.0004	GARCH(1,1) t	-7.05015***	0.0000
ARIMA(1,0,0)-GARCH(1,2) t	-3.83261***	0.0001	ARIMA(1,0,0)-GARCH(1,2) t	-8.89246***	0.0000
MA63	-4.71824***	0.0000	MA63	-8.94356***	0.0000

FTSE100

The Diebold-Mariano test results for FTSE100 index are presented in Table 10. The results under squared loss imply statistically equal predictive accuracy with 1% significance level, for the first ranked t-distributed ARIMA(1,0,1)-EGARCH and the second ranked EGARCH(1,1) with normal distribution and third ranked ARIMA(1,0,1)-EGARCH(1,1) with normal distribution. Similar result of equal predictive accuracy holds also for the MA5, which is ranked below the middle class under MSE. With 5 % significance level the null hypothesis is rejected with these same models indicating the t-distributed ARIMA(1,0,1)-EGARCH(1,1) offering statistically more accurate forecast. For all the other models, the null is clearly rejected. Same result holds for the absolute loss differentials as the null hypothesis is clearly rejected with all of the models.

Table 10. Results of the Diebold-Mariano tests for FTSE100. Statistical significance at the 5% and 1% levels is denoted by ** and ***

FTSE100

Model ranking based on the average MSE	Diebold-Mariano test statistic	p value	Model ranking based on the average MAE	Diebold-Mariano test statistic	p value
ARIMA(1,0,1)-EGARCH(1,1) t			ARIMA(1,0,1)-EGARCH(1,1) t		
EGARCH(1,1)	-2.54849**	0.0108	ARIMA(1,0,1)-EGARCH(1,2) t	-2.76737***	0.0057
ARIMA(1,0,1)-EGARCH(1,1)	-2.50467**	0.0123	EGARCH(1,1)	-11.08488***	0.0000
EGARCH(1,1) t	-15.59526***	0.0000	ARIMA(1,0,1)-EGARCH(1,1)	-11.02160***	0.0000
ARIMA(1,0,1)-EGARCH(1,2)	-2.72854***	0.0064	EGARCH(1,1) t	-23.33731***	0.0000
ARIMA(1,0,1)-EGARCH(1,2) t	-3.01652***	0.0026	ARIMA(1,0,1)-EGARCH(1,2)	-5.83900***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,1) t	-3.61624***	0.0003	ARIMA(1,0,1)-GJR-GARCH(1,1) t	-3.96910***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,1)	-3.84172***	0.0001	ARIMA(0,0,0)-GJR-GARCH(1,1) t	-4.55443***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1)	-4.00014***	0.0000	ARIMA(1,0,1)-GJR-GARCH(1,1)	-5.14122***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1) t	-3.87899***	0.0001	ARIMA(0,0,0)-GJR-GARCH(1,1)	-5.57694***	0.0000
GJR-GARCH(1,1)	-4.09372***	0.0000	GJR-GARCH(1,1)	-5.76506***	0.0000
ARIMA(1,0,0)-GARCH(1,1)	-5.48305***	0.0000	MA5	-2.93982***	0.0033
ARIMA(1,0,1)-GARCH(1,2)	-5.61408***	0.0000	GJR-GARCH(1,1) t	-6.21184***	0.0000
GJR-GARCH(1,1) t	-4.44669***	0.0000	ARIMA(1,0,0)-GARCH(1,1)	-6.97453***	0.0000
GARCH(1,1)	-5.71082***	0.0000	GARCH(1,1)	-7.08941***	0.0000
MA5	-2.54443**	0.0000	ARIMA(1,0,1)-GARCH(1,2)	-7.15131***	0.0000
ARIMA(1,0,0)-GARCH(1,1) t	-6.07910***	0.0000	MA21	-5.02017***	0.0000
ARIMA(1,0,1)-GARCH(1,2) t	-6.21269***	0.0000	ARIMA(1,0,0)-GARCH(1,1) t	-7.53076***	0.0000
GARCH(1,1) t	-6.26295***	0.0000	GARCH(1,1) t	-7.60259***	0.0000
EWMA	-6.61250***	0.0000	EWMA	-5.69692***	0.0000
MA21	-5.69444***	0.0000	ARIMA(1,0,1)-GARCH(1,2) t	-7.75036***	0.0000
MA63	-7.17499***	0.0000	MA63	-8.89700***	0.0000

NIKKEI225

The Diebold-Mariano test results for NIKKEI225 index are provided in Table 11. Being the best performing model under the MSE, the normally distributed ARIMA(0,0,0)-EGARCH(2,2) and following EGARCH(1,1) models show statistically equal predictive accuracy. With 1% significance level the same result of equal accuracy holds also for the ARIMA(1,0,1)-EGARCH(1,2) with normal distribution whereas with 5% level the null is slightly rejected. For all the other models, the null is clearly rejected implying that the forecast of the best ranked ARIMA(0,0,0)-EGARCH(2,2) with normal distribution, is also statistically more accurate. For the MAE rankings, the Diebold-Mariano test indicate that the null hypothesis of equal predictive accuracy is rejected for each of the models. The only exception is the second ranked MA5 as there's no statistical difference between it and the ARIMA(0,0,0)-EGARCH(2,2) forecasts.

Table 11. Results of the Diebold-Mariano tests for NIKKEI225. Statistical significance at the 5% and 1% levels is denoted by ** and ***

NIKKEI225

Model ranking based on the average MSE	Diebold-Mariano test statistic	p value	Model ranking based on the average MAE	Diebold-Mariano test statistic	p value
ARIMA(0,0,0)-EGARCH(2,2)			ARIMA(0,0,0)-EGARCH(2,2)		
EGARCH(1,1)	-0.64615	0.5182	MA5	-0.12040	0.9042
EGARCH(1,1) t	-0.93902	0.3477	ARIMA(0,0,0)-EGARCH(2,2) t	-2.98602***	0.0028
ARIMA(0,0,0)-EGARCH(2,2) t	-3.08054***	0.0021	EGARCH(1,1)	-3.05564***	0.0023
ARIMA(1,0,1)-EGARCH(1,2)	-1.98773**	0.0468	ARIMA(1,0,1)-EGARCH(1,2)	-4.03923***	0.0000
ARIMA(1,0,1)-EGARCH(1,2) t	-3.31936***	0.0009	EGARCH(1,1) t	-3.54559***	0.0004
ARIMA(1,0,1)-GJR-GARCH(1,1) t	-3.72636***	0.0002	ARIMA(1,0,1)-EGARCH(1,2) t	-4.81633***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,1)	-4.29216***	0.0000	ARIMA(1,0,1)-GJR-GARCH(1,1) t	-5.13137***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1)	-4.29063***	0.0000	ARIMA(1,0,1)-GJR-GARCH(1,1)	-6.17653***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1) t	-4.29063***	0.0000	ARIMA(0,0,0)-GJR-GARCH(1,1)	-6.18817***	0.0000
GJR-GARCH(1,1)	-4.82014***	0.0000	ARIMA(0,0,0)-GJR-GARCH(1,1) t	-6.18817***	0.0000
GJR-GARCH(1,1) t	-4.97760***	0.0000	GJR-GARCH(1,1)	-7.13589***	0.0000
MA5	-2.96934***	0.0030	GJR-GARCH(1,1) t	-7.29487***	0.0000
ARIMA(0,0,0)-GARCH(1,1)	-7.76964***	0.0000	MA21	-5.33031***	0.0000
GARCH(1,1)	-7.85447***	0.0000	EWMA	-6.62948***	0.0000
EWMA	-6.68427***	0.0000	ARIMA(0,0,0)-GARCH(1,1)	-10.38245***	0.0000
ARIMA(2,0,2)-GARCH(1,2)	-7.96957***	0.0000	GARCH(1,1)	-10.51088***	0.0000
ARIMA(0,0,0)-GARCH(1,1) t	-8.64463***	0.0000	ARIMA(2,0,2)-GARCH(1,2)	-10.55660***	0.0000
GARCH(1,1) t	-8.77482***	0.0000	ARIMA(0,0,0)-GARCH(1,1) t	-11.05315***	0.0000
ARIMA(2,0,2)-GARCH(1,2) t	-9.27097***	0.0000	GARCH(1,1) t	-11.20630***	0.0000
MA21	-6.87580***	0.0000	ARIMA(2,0,2)-GARCH(1,2) t	-11.87693***	0.0000
MA63	-10.31441***	0.0000	MA63	-12.21235***	0.0000

SHANGHAI SE Composite

The Diebold-Mariano test results for Shanghai SE Composite index are presented in Table 12. The Diebold-Mariano test under MSE ranking imply that the forecast of the EWMA, MA5 and normally distributed ARIMA(1,0,1)-EGARCH(2,2) are statistically equally accurate. However, with 5% significance level the null is rejected for the ARIMA(1,0,1)-EGARCH(2,2). For all the other models, the null is clearly rejected implying that the forecast of the best ranked EWMA, is also statistically more accurate. Under MAE ranking the Diebold-Mariano test indicate statistically similar forecasting accuracy between the top three ranked models MA5, MA21 and EWMA. For all the other models, the null is rejected.

Table 12. Results of the Diebold-Mariano tests for Shanghai SE Composite. Statistical significance at the 5% and 1% levels is denoted by ** and ***

Shanghai SE Composite

Model ranking based on the average MSE	Diebold-Mariano test statistic	p value	Model ranking based on the average MAE	Diebold-Mariano test statistic	p value
EWMA			MA5		
MA5	-0.34531	0.7299	MA21	-1.00759	0.3137
MA21	-4.26145***	0.0000	EWMA	-1.25388	0.2099
ARIMA(1,0,1)-EGARCH(2,2)	-2.51715**	0.0118	ARIMA(1,0,1)-EGARCH(1,1)	-4.36973***	0.0000
ARIMA(1,0,1)-EGARCH(1,1)	-2.94717***	0.0032	EGARCH(1,1)	-4.41572***	0.0000
EGARCH(1,1)	-2.99376***	0.0028	ARIMA(1,0,1)-EGARCH(2,2)	-4.65222***	0.0000
ARIMA(1,0,1)-EGARCH(2,2) t	-3.59315***	0.0003	ARIMA(1,0,1)-EGARCH(1,1) t	-5.23993***	0.0000
ARIMA(1,0,1)-EGARCH(1,1) t	-3.91990***	0.0000	ARIMA(1,0,1)-EGARCH(2,2) t	-5.40209***	0.0000
EGARCH(1,1) t	-4.07555***	0.0000	EGARCH(1,1) t	-5.31074***	0.0000
ARIMA(0,0,0)-GARCH(1,1)	-4.63848***	0.0000	ARIMA(0,0,0)-GARCH(1,1)	-5.93696***	0.0000
GARCH(1,1)	-4.63682***	0.0000	GARCH(1,1)	-5.94050***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1)	-5.00539***	0.0000	ARIMA(1,0,1)-GARCH(1,2)	-6.14941***	0.0000
GJR-GARCH(1,1)	-5.04139***	0.0000	ARIMA(0,0,0)-GJR-GARCH(1,1)	-6.27538***	0.0000
ARIMA(1,0,1)-GARCH(1,2)	-4.82905***	0.0000	GJR-GARCH(1,1)	-6.30547***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,2)	-4.97705***	0.0000	ARIMA(1,0,1)-GJR-GARCH(1,2)	-6.30143***	0.0000
GARCH(1,1) t	-5.87536***	0.0000	GARCH(1,1) t	-7.00162***	0.0000
ARIMA(0,0,0)-GARCH(1,1) t	-5.88443***	0.0000	ARIMA(0,0,0)-GARCH(1,1) t	-7.01963***	0.0000
ARIMA(0,0,0)-GJR-GARCH(1,1) t	-6.25127***	0.0000	MA63	-5.98337***	0.0000
GJR-GARCH(1,1) t	-6.38049***	0.0000	ARIMA(0,0,0)-GJR-GARCH(1,1) t	-7.43117***	0.0000
ARIMA(1,0,1)-GARCH(1,2) t	-6.48960***	0.0000	GJR-GARCH(1,1) t	-7.51840***	0.0000
ARIMA(1,0,1)-GJR(1,2) t	-6.78686***	0.0000	ARIMA(1,0,1)-GARCH(1,2) t	-7.57576***	0.0000
MA63	-7.55097***	0.0000	ARIMA(1,0,1)-GJR-GARCH(1,2) t	-8.14056***	0.0000

S&P500

According to the results from Table 13, the outperformance of the normally distributed ARIMA(1,0,1)-EGARCH(1,1) under MSE is also statistically significant as its forecast is more accurate than the other models. The only exception to this is the MA5 as the null hypothesis of equal predictive accuracy between the model and the best performing ARIMA(1,0,1)- EGARCH(1,1) is not rejected. Under MAE ranking, the Diebold-Mariano test result indicate that the forecast of best performing model ARIMA(1,0,1)-EGARCH(1,1) with t-distribution is not statistically more accurate than the forecasts of second and third ranked normally distributed ARIMA(1,0,1)-EGARCH(1,1) or MA5. For all other models under MAE, the null hypothesis of equal predictive accuracy is rejected.

Table 13. Results of the Diebold-Mariano tests for S&P500. Statistical significance at the 5% and 1% levels is denoted by ** and ***

S&P500

Model ranking based on the average MSE	Diebold-Mariano test statistic	p value	Model ranking based on the average MAE	Diebold-Mariano test statistic	p value
ARIMA(1,0,1)-EGARCH(1,1)			ARIMA(1,0,1)-EGARCH(1,1) t		
EGARCH(1,1)	-6.04090***	0.0000	ARIMA(1,0,1)-EGARCH(1,1)	-0.37860	0.7050
ARIMA(1,0,1)-EGARCH(1,1) t	-2.75769***	0.0058	MA5	-0.73455	0.4626
EGARCH(1,1) t	-4.13165***	0.0000	EGARCH(1,1)	-4.80071***	0.0000
ARIMA(1,0,1)-EGARCH(1,2)	-4.05211***	0.0000	ARIMA(1,0,1)-EGARCH(1,2)	-5.90147***	0.0000
MA5	-1.34243	0.1795	EGARCH(1,1) t	-10.85862***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,1)	-3.55708***	0.0004	ARIMA(1,0,1)-GJR-GARCH(1,1) t	-7.34919***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,1) t	-3.64551***	0.0003	ARIMA(1,0,1)-GJR-GARCH(1,1)	-8.93693***	0.0000
GJR-GARCH(1,1)	-3.81456***	0.0001	GJR-GARCH(1,1)	-10.03098***	0.0000
GARCH(1,1)	-5.64639***	0.0000	GJR-GARCH(1,1) t	-9.33059***	0.0000
ARIMA(1,0,1)-GARCH(1,2)	-5.64697***	0.0000	GARCH(1,1)	-13.10768***	0.0000
ARIMA(1,0,1)-GARCH(2,2)	-5.50327***	0.0000	ARIMA(1,0,1)-GARCH(1,2)	-13.29954***	0.0000
GJR-GARCH(1,1) t	-4.08361***	0.0000	ARIMA(1,0,1)-GARCH(2,2)	-13.14224***	0.0000
GARCH(1,1) t	-5.93066***	0.0000	GARCH(1,1) t	-13.23679***	0.0000
ARIMA(1,0,1)-GARCH(2,2) t	-6.02355***	0.0000	MA21	-9.03932***	0.0000
ARIMA(1,0,1)-GARCH(1,2) t	-6.05653***	0.0000	ARIMA(1,0,1)-GARCH(2,2) t	-13.82448***	0.0000
EWMA	-7.43999***	0.0000	ARIMA(1,0,1)-GARCH(1,2) t	-13.92687***	0.0000
MA21	-5.88927***	0.0000	EWMA	-11.18563***	0.0000
MA63	-8.33392***	0.0000	MA63	-13.05475***	0.0000

Dow Jones Industrial Average

Based on Diebold-Mariano test results in Table 14, the forecast of the best performing model under MSE, the ARIMA(1,0,1)-EGARCH((1,1) with normal distribution, is statistically equally accurate with the forecast of the third ARIMA(0,0,0)-EGARCH(2,2), fourth ARIMA(1,0,1)- EGARCH(1,1) with t-distribution and sixth MA5 model. However, for the

ARIMA(1,0,1)-EGARCH(1,1) with t-distribution, the result hold with 1% significance, whereas the null hypothesis is rejected with 5% significance level. For all other model forecasts the null hypothesis is rejected, including also the second ranked model, the EGARCH(1,1) with normal distribution. Based on MAE ranking, the Diebold-Mariano test result indicate that the forecast of best performing model ARIMA(1,0,1)-EGARCH(1,1) with normal distribution is not statistically more accurate than the forecasts of the second, ARIMA(0,0,0)-EGARCH(2,2) with normal distribution and third, the ARIMA(1,0,1)-EGARCH(1,1) with t-distribution models. For all other models under MAE, the null hypothesis of equal predictive accuracy is rejected.

Table 14. Results of the Diebold-Mariano tests for Dow Jones. Statistical significance at the 5% and 1% levels is denoted by ** and ***

Dow Jones

Model ranking based on the average MSE	Diebold-Mariano test statistic	p value	Model ranking based on the average MAE	Diebold-Mariano test statistic	p value
ARIMA(1,0,1)-EGARCH(1,1)			ARIMA(1,0,1)-EGARCH(1,1)		
EGARCH(1,1)	-6.97574***	0.0000	ARIMA(0,0,0)-EGARCH(2,2)	-0.54809	0.5836
ARIMA(0,0,0)-EGARCH(2,2)	-1.07283	0.2834	ARIMA(1,0,1)-EGARCH(1,1) t	-1.12855	0.2591
ARIMA(1,0,1)-EGARCH(1,1) t	-2.01178**	0.0442	EGARCH(1,1)	-14.68249***	0.0000
EGARCH(1,1) t	-2.92371***	0.0035	EGARCH(1,1) t	-9.77063***	0.0000
MA5	-0.76377	0.4450	ARIMA(0,0,0)-EGARCH(2,2) t	-5.11418***	0.0000
ARIMA(0,0,0)-EGARCH(2,2) t	-4.22876***	0.0000	MA5	-2.56239**	0.0104
ARIMA(1,0,1)-GJR-GARCH(1,1)	-2.90128***	0.0037	ARIMA(1,0,1)-GJR-GARCH(2,2) t	-5.96209***	0.0000
ARIMA(1,0,1)-GJR-GARCH2,2)	-2.84764***	0.0044	ARIMA(1,0,1)-GJR-GARCH(1,1) t	-5.97731***	0.0000
GJR-GARCH(1,1)	-3.17797***	0.0015	ARIMA(1,0,1)-GJR-GARCH(1,1)	-6.73790***	0.0000
ARIMA(1,0,1)-GJR-GARCH(2,2) t	-3.11526***	0.0018	ARIMA(1,0,1)-GJR-GARCH(2,2)	-7.16408***	0.0000
ARIMA(1,0,1)-GJR-GARCH(1,1) t	-3.13387***	0.0017	GJR-GARCH(1,1)	-7.77987***	0.0000
GARCH(1,1)	-4.93672***	0.0000	GJR-GARCH(1,1) t	-7.43052***	0.0000
GJR-GARCH(1,1) t	-3.46440***	0.0005	GARCH(1,1)	-10.87458***	0.0000
ARIMA(0,0,0)-GARCH(2,2)	-4.74823***	0.0000	ARIMA(0,0,0)-GARCH(2,2)	-11.24089***	0.0000
ARIMA(1,0,1)-GARCH(2,2)	-4.75786***	0.0000	ARIMA(1,0,1)-GARCH(2,2)	-11.24006***	0.0000
GARCH(1,1) t	-5.23180***	0.0000	GARCH(1,1) t	-10.73875***	0.0000
ARIMA(0,0,0)-GARCH(2,2) t	-5.19419***	0.0000	ARIMA(0,0,0)-GARCH(2,2) t	-11.38009***	0.0000
ARIMA(1,0,1)-GARCH(2,2) t	-5.26767***	0.0000	ARIMA(1,0,1)-GARCH(2,2) t	-11.50236***	0.0000
EWMA	-7.25247***	0.0000	MA21	-8.67828***	0.0000
MA21	-5.78639***	0.0000	EWMA	-10.71311***	0.0000
MA63	-7.91488***	0.0000	MA63	-12.98017***	0.0000

Lastly, the results of Diebold-Mariano test for monthly forecasts are presented in Table 15. According to results for DAX30, the null hypothesis of equal predictive accuracy is rejected under both loss functions for the first ranked GJR-GARCH(1,1) with normal distribution and the second ranked GJR-GARCH(1,1) with t-distribution. Under squared loss differentials, the null hypothesis is also rejected for the EGARCH(1,1) with t-distribution. Under absolute loss differentials, third ranked GARCH(1,1) with normal distribution and its forecast is statistically equally accurate with the GJR-GARCH(1,1), as the null hypothesis is not rejected. For FTSE100, the null hypothesis of equal predictive accuracy is rejected under both loss functions for the first ranked EGARCH(1,1) with t-distribution and the second ranked EGARCH(1,1) with normal distribution. For other models, the null hypothesis is not rejected implying statistically equally accurate forecasts. For Nikkei225, the results from Diebold-Mariano test show that the null hypothesis of equal predictive accuracy is not rejected for any of the models. The only exception to this is the result from squared loss differentials as the null hypothesis is rejected with 10% significance level for the first ranked EGARCH(1,1) and bottom ranked GARCH(1,1) with t-distribution. For Shanghai, the null hypothesis of equal predictive accuracy is rejected only for EGARCH(1,1) models under squared loss differentials. Based on absolute loss differentials, the null hypothesis is rejected only for the second ranked GARCH(1,1) with t-distribution. For other models the result indicate statistically similar forecasting accuracy with the first ranked EGARCH(1,1) and GARCH(1,1). For S&P500, the first ranked model under both MSE and MAE, the normally distributed EGARCH(1,1) and its forecast is statistically equally accurate with the GJR-GARCH(1,1) and t-distributed EGARCH(1,1) models forecasts, as the null hypothesis of equal predictive accuracy is not rejected. For bottom ranked GARCH(1,1) models the null hypothesis is rejected with 10% and 5% levels. For Dow Jones, the results are similar indicating statistically equally accurate forecasts for EGARCH(1,1) and GJR-GARCH(1,1) models.

Table 15. Results of the Diebold-Mariano tests for monthly volatility forecasts. Statistical significance at the 10%, 5% and 1% levels is denoted by *, ** and ***

DAX30						**FTSE100**					
Model ranking based on the average MSE	DM test statistic	p value	Model ranking based on the average MAE	DM test statistic	p value	Model ranking based on the average MSE	DM test statistic	p value	Model ranking based on the average MAE	DM test statistic	p value
GJR-GARCH(1,1)			GJR-GARCH(1,1)			EGARCH(1,1) t			EGARCH(1,1) t		
GJR-GARCH(1,1) t	-1.82918	0.0674*	GJR-GARCH(1,1) t	-2.11451	0.0345**	EGARCH(1,1)	-2.11269	0.0346**	EGARCH(1,1)	-1.94483	0.0518*
EGARCH(1,1)	-1.55203	0.1207	GARCH(1,1)	-1.50802	0.1315	GARCH(1,1)	-0.13559	0.8921	GARCH(1,1)	-0.01909	0.9848
GARCH(1,1)	-1.31483	0.1886	EGARCH(1,1)	-1.72650	0.0843*	GARCH(1,1) t	-0.27907	0.7802	GARCH(1,1) t	-0.09965	0.9206
EGARCH(1,1) t	-2.28606	0.0223**	GARCH(1,1) t	-1.80495	0.0711*	GJR-GARCH(1,1) t	-0.74998	0.4533	GJR-GARCH(1,1) t	-0.26321	0.7924
GARCH(1,1) t	-1.43599	0.1510	EGARCH(1,1) t	-2.64856	0.0081***	GJR-GARCH(1,1)	-0.75495	0.4503	GJR-GARCH(1,1)	-0.29981	0.7643

Nikkei225						**Shanghai SE Composite**					
Model ranking based on the average MSE	DM test statistic	p value	Model ranking based on the average MAE	DM test statistic	p value	Model ranking based on the average MSE	DM test statistic	p value	Model ranking based on the average MAE	DM test statistic	p value
EGARCH(1,1)			GJR-GARCH(1,1)			EGARCH(1,1)			GARCH(1,1)		
GJR-GARCH(1,1)	0.17646	0.8599	GARCH(1,1)	-0.51997	0.6031	GARCH(1,1)	-0.02215	0.9823	GARCH(1,1) t	-3.93289	0.0000***
GARCH(1,1)	-0.64567	0.5185	GJR-GARCH(1,1) t	-0.35801	0.7203	EGARCH(1,1) t	-2.68942	0.0072***	EGARCH(1,1)	-0.87624	0.3809
EGARCH(1,1) t	-1.00959	0.3127	GARCH(1,1) t	-1.15360	0.2487	GARCH(1,1) t	-0.19885	0.8423	EGARCH(1,1) t	-0.93864	0.3479
GJR-GARCH(1,1) t	-1.48800	0.1368	EGARCH(1,1) t	0.30720	0.7587	GJR-GARCH(1,1)	-1.52303	0.1278	GJR-GARCH(1,1)	-1.51643	0.1294
GARCH(1,1) t	-1.75305	0.0796*	EGARCH(1,1)	0.22423	0.8226	GJR-GARCH(1,1) t	-1.63736	0.1016	GJR-GARCH(1,1) t	-1.59352	0.1110

S&P500						**Dow Jones**					
Model ranking based on the average MSE	DM test statistic	p value	Model ranking based on the average MAE	DM test statistic	p value	Model ranking based on the average MSE	DM test statistic	p value	Model ranking based on the average MAE	DM test statistic	p value
EGARCH(1,1)			EGARCH(1,1)			GJR-GARCH(1,1)			EGARCH(1,1)		
GJR-GARCH(1,1)	-0.09426	0.9249	EGARCH(1,1) t	-1.14355	0.2528	GJR-GARCH(1,1) t	-1.20089	0.2298	GJR-GARCH(1,1)	-0.22034	0.8256
EGARCH(1,1) t	-0.57360	0.5662	GJR-GARCH(1,1)	-0.69856	0.4848	EGARCH(1,1)	-0.35980	0.7190	EGARCH(1,1) t	-0.83275	0.4050
GJR-GARCH(1,1) t	-0.29167	0.7705	GJR-GARCH(1,1) t	-0.75404	0.4508	EGARCH(1,1) t	-0.61611	0.5378	GJR-GARCH(1,1) t	-0.33195	0.7399
GARCH(1,1)	-1.74716	0.0806*	GARCH(1,1)	-2.14962	0.0316**	GARCH(1,1)	-2.01965	0.0434**	GARCH(1,1)	-2.18639	0.0288**
GARCH(1,1) t	-1.69613	0.0899*	GARCH(1,1) t	-2.10460	0.0353**	GARCH(1,1) t	-1.78611	0.0741*	GARCH(1,1) t	-2.19216	0.0284**

5 Discussion and summary

The main purpose of this thesis was to compare the volatility forecasts of different GARCH-type and moving average models in six global equity indices, namely DAX30 (Germany), FTSE100 (UK), Shanghai SE Composite (China), NIKKEI225 (Japan), S&P500 (US) and Dow Jones Industrial Average (US). Daily time series data of these indices covered the period from January 1, 2000 through December 31, 2020. Daily and monthly volatility forecasts were produced for the out-of-sample period from January 1, 2016 to December 31, 2020. To work properly, the conditional volatility model usually requires also some conditional mean model. In this thesis, the conditional mean was computed with ARIMA-type models. Hence, the forecasting models used in this thesis were combinations of ARIMA-GARCH, ARIMA-EGARCH and ARIMA-GJR-GARCH models. Dropping the conditional mean, the basic GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models were included as a benchmark forecasting models. Previous studies, e.g., Engle and Ng (1993), Brailsford and Faff (1996), have shown that these basic models offer good volatility forecasting results. Monthly volatility forecasts were done only with the GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. According to previous research and empirical evidence financial time series data is often described having excess kurtosis and fatter tails compared to normal distributions. However, many of the models still make use of the normal distributions. The data in this thesis indicated non-normal distributions and therefore all the volatility forecasts with GARCH-type models were done with both normal and Student's t-distributions to compare the forecasting results of these different distributions. The Diebold-Mariano test was used to test statistical difference of the models' forecasts. To study these issues in global stock markets, four research questions were formed. Next, the results regarding the research questions are discussed.

Which of the volatility forecasting models perform best according to loss functions MSE and MAE for the one-day and one-month forecasting horizon?

The results from the daily volatility forecasts for the DAX30 indicate that the best performing model under both MAE and MSE is the ARIMA(1,0,1)-GJR(1,1) with t-distribution. For the FTSE100 the best performing model according to MAE and MSE is the ARIMA(1,0,1)-

EGARCH(1,1) with t-distribution. For the NIKKEI225 the best performing model under MAE and MSE is the ARIMA(0,0,0)-EGARCH(2,2) with normal distribution. For the Shanghai SE Composite the best performing model under MAE is the five-day moving average (MA5) and under MSE, the EWMA ranks first. For the S&P500 the best performing model under MAE is the t-distributed ARIMA(1,0,1)-EGARCH(1,1). According MSE, the ARIMA(1,0,1)-EGARCH(1,1) with normal distribution ranks first. For the Dow Jones the best performing model under MAE and MSE is the ARIMA(1,0,1)-EGARCH(1,1) with normal distribution. The results from the monthly volatility forecasts imply that for the DAX30, the best performing model under both MAE and MSE loss functions is the GJR-GARCH(1,1) with normal distribution. For the FTSE100 the best performing model according to MAE and MSE is the EGARCH(1,1) with t-distribution. For the NIKKEI225 the best performing model under MAE is the GJR-GARCH(1,1) with normal distribution and under MSE, the normally distributed EGARCH(1,1) ranks first. For the Shanghai SE Composite the best performing model under MAE is the GARCH(1,1) with normal distribution whereas the EGARCH(1,1) also with normal distribution ranks first based on MSE. For the S&P500 the best performing model under both loss functions is the EGARCH(1,1) with normal distribution. Lastly, for the Dow Jones the best performing model under MSE is the GJR-GARCH(1,1) with normal distribution and under MAE the EGARCH(1,1) with normal distribution takes the first place.

As these results outline, based on MAE and MSE loss functions there's no single model outperforming across the indices. However, the performance of the EGARCH-type models was convincing with both daily and monthly forecasts and across the indices as at least one of the EGARCH model was constantly ranked in top three under both loss functions. The exception to this was the MSE ranking with daily forecasts for the German DAX30 index where the ARIMA(1,0,1)-GJR-GARCH(1,1) with t-distribution ranked first followed by the five-day moving average MA5 and normal distributed ARIMA(1,0,1)-GJR-GARCH(1,1). Overall, the performance of the GJR-GARCH models was good with German data as the models outperformed most of the EGARCH models. Another exception to the top performance of the EGARCH models were the daily volatility forecasting results from the Shanghai data as the moving average models were the top three forecasting models. With monthly forecasts, the DAX30 and NIKKEI225 were the only indices where the EGARCH was not ranked in top three under both loss functions. In these indices the GJR-GARCH(1,1) and GARCH(1,1) ranked top three under MAE.

Next, the performance of the basic GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) are shortly discussed. When comparing these models relative performance and based on MSE and MAE loss functions the EGARCH(1,1) and GJR-GARCH(1,1) constantly outperformed the standard GARCH(1,1) across the indices and with both daily and monthly forecasts. With daily forecasts, the combinations of different conditional mean and conditional volatility models often outperformed these basic models. However, the basic EGARCH(1,1) was frequently highly ranked so depending on the objective of the forecast and evaluation criteria, the model can be useful in some circumstances.

The good performance of the EGARCH and GJR-GARCH models is in line with previous results (see e.g., Engle and Ng 1993; Brailsford and Faff 1996; Awartani and Corradi 2005). However, the results from this thesis cannot be generalized covering all possible situations. As Poon and Granger (2003) have also outlined, with different data, loss functions or volatility proxies the results might be different. Poon and Granger (2003) also mention that it is rarely discussed if one forecasting method is significantly better than another and although some particular method of forecasting volatility can be said being the best, there's no discussion about the cost-benefit from using it. Clearly, these aspects are important when evaluating different volatility forecasting models and their use in practice. GARCH, EGARCH and GJR-GARCH require statistical software whereas moving average and exponentially weighted moving average can be applied with a single spreadsheet. Another aspect regarding the GARCH-type models is that they usually require large datasets and fitting the models, which in some cases may be problematic. On the other hand, if the objective is to make forecasts and evaluate the out-of-sample performance of the models, then the in-sample fit of the models is not so relevant. A practical point of view is also the accuracy of the volatility forecasts. If there's marginal difference between the GARCH-type and simpler models, perhaps the biggest benefit is gained when there's possibility to combine and choose different models that perform well in different markets and market conditions and based on the objective of the forecast. Some private investors doing day trading would perhaps try the combinations of GARCH-type models and some of the simpler models. Risk managers may assess the benefits from a different point of view and compare some models and their performance more profoundly. Lastly, the evaluation of the forecasting performance is usually done with some loss functions. As the results from this thesis show, the ranking of the models varies based on MAE and MSE loss functions. Choosing the loss function always includes at least some user specific matters and there's no unanimous conclusion that which of the loss functions one should always use. Comparing the forecasting models with several different loss functions may give better understanding of the performance of the models. MAE and MSE were chosen to be used in this thesis because of their wide use in previous research. However, some alternative loss functions, such as root mean squared error and mean absolute percentage error could give interesting results. Asymmetric loss functions that penalize over-predictions and under-predictions of volatility could also provide different ranking of the models.

How does the GARCH-type volatility forecasting models perform compared to moving average and EWMA models for the one-day ahead forecasting horizon?

In general, the asymmetric EGARCH and GJR-GARCH models performance was superior and the models outperformed most of the moving average models. However, under MAE loss function, the MA5 was ranked second with NIKKEI225 and third with S&P500. Under MSE, the MA5 was ranked second with DAX30 and sixth with S&P500 and Dow Jones. As these results indicate, the performance of the MA5 was also relatively good and the model outperformed many of the more complex GARCH-type models. The moving average models performed well also with Shanghai data as the EWMA was the best performing model followed by the MA5 and MA21 under MSE loss function. Based on MAE, the MA5 was ranked first followed by the MA21 and EWMA. Simple models outperforming more sophisticated GARCH-type models can be understood as a rather surprising result. The data indicated clear signs of heteroscedasticity, fat tails, volatility clustering and leverage effects. The GARCH-type models were created to capture these effects so therefore it would've been expected that these models dominate the volatility forecasts. Different EGARCH-type models were however also performing well with Shanghai data as they were ranked just after the top three moving average models although the margin was rather clear favoring the simpler models. China's stock market can be considered rather volatile as the statistics from the Shanghai SE Composite index also show. Can the high volatile market explain why the moving average models perform so well? If the models are able to adapt to changing market conditions more quickly than the GARCH-type models, than there's evidence to also consider these models when choosing the appropriate forecasting models. Another distinctive feature about the Shanghai results was that some of the GARCH-models were ranked after the EGARCH-models. This is surprising since they are not asymmetric models whereas the GJR-GARCH models are, and it would've been expected that the GJR-GARCH follow the ranking of the similar kind asymmetric EGARCH models. The three-month moving average MA63 was bottom ranked model in all indices. The only exception to this was again the result from Shanghai data with MAE loss function in which the ARIMA(1,0,1)-GJR-GARCH(1,2) with t-distribution was ranked worst and MA63 was few positions higher.

Another not so expected result was the performance of the GARCH compared to EWMA models. EWMA and GARCH are closely related, and both models can capture volatility clustering. It would've been expected that the models' forecasting performance is similar and therefore it is surprising result that there's also relatively large numerical difference according to loss functions of the models. The EWMA is performing very well with the Shanghai data as the model is ranked first under MSE and third under MAE whereas the GARCH is ranked on a mediocre level. With other indices the results change as almost all the GARCH models outperform and EWMA ranks lower. Under MAE loss function, the EWMA ranked higher only in DAX30 and NIKKEI225. Previous research (e.g., Ederington and Guan 2005; Awartani and Corradi 2005) show GARCH models outperforming EWMA. On the other hand, Walsh and Tsou (1998) and McMillan and Kambouroudis (2009) give some support of EWMA outperforming.

Is there a difference between normally distributed and t-distributed models based on volatility forecasting results?

According to previous research and empirical evidence financial time series data is often described having excess kurtosis and fatter tails compared to normal distributions. The data used in this thesis showed similar characteristics indicating non-normal distributions and therefore Student's t-distributions were included along normal distributions. If the data is non-normal, then it would be expected that the t-distribution models would outperform their normal distributed counterparts. The results show that this is not the case and models with t-distributions are not constantly outperforming their normal distributed counterparts. Results from the DAX30 and FTSE100 show that the best daily volatility forecasting model under both loss functions is t-distributed model. Results from the other indices however show the normally distributed models holding the first place. Even the results from the Shanghai show that the normal distributed models are all outperforming their t-distributed counterparts. The result from the US data is interesting as with S&P500 the ARIMA(1,0,1)-EGARCH(1,1) with t-distribution is ranked first based on MAE, whereas the same model is ranked third based on MSE. With Dow Jones, the normal distribution model ranks first under both loss functions and the ARIMA(1,0,1)-EGARCH(1,1) with t-distribution holds third position under MAE. Also, based on MAE, the ARIMA(1,0,1)-GJR-GARCH(2,2) and ARIMA(1,0,1)-GJR-GARCH(1,1) both with t-distributions, slightly outperform their normal distributed models. Under MSE these same models rankings change and t-distribution models are outperformed by the normal distribution models. For monthly forecasts, the FTSE100 is the only index where the t-distributed EGARCH(1,1) ranks first and with all the other indices the normally distributed models hold the first place.

Is there a statistical difference in volatility forecasting accuracy between models?

The loss functions MAE and MSE showed the ranking of the volatility forecasting models. To assess whether these rankings hold statistically significant difference, the Diebold-Mariano test was used. The test was done so that the best ranked model under MAE and MSE from each index and its volatility forecast was tested against all the other model forecasts separately. The null hypothesis of the Diebold-Mariano test was equal forecast accuracy between the two forecasts.

For the DAX30 the Diebold-Mariano test show that there’s no statistically significant difference between the forecasts of the best ranked ARIMA(1,0,1)-GJR-GARCH(1,1) with t-distribution and four following model forecasts, MA5, normally distributed ARIMA(1,0,1)-GJR-GARCH(1,1), ARIMA(0,0,0)-GJR-GARCH(1,1) and t-distributed ARIMA(0,0,0)-GJR-GARCH(1,1). Similar result with equal forecast accuracy holds for almost all EGARCH-type model forecasts. For all the other model forecasts under squared loss differentials, the null hypothesis of equal predictive accuracy was rejected indicating that the forecast of the first ranked t-distributed ARIMA(1,0,1)-GJR-GARCH(1,1) would also be statistically more accurate. Under absolute loss differentials the Diebold-Mariano test shows equal predictive accuracy between the forecasts of the best ranked t-distributed ARIMA(1,0,1)-GJR-GARCH(1,1) and all EGARCH-type model forecasts, including the second best ranked ARIMA(1,0,1)-EGARCH(2,3) with normal distribution. Compared to squared loss functions, there’s clear significant difference between the forecasts of t-distributed ARIMA(1,0,1)-GJR-GARCH(1,1) and all other GJR-GARCH-type models again indicating that the forecast of the first ranked t-distributed ARIMA(1,0,1)-GJR-GARCH(1,1) would also be statistically more accurate. The performance of the t-distributed ARIMA(1,0,1)-EGARCH(1,1) and its forecast was convincing in FTSE100 data as the model was ranked first under both loss functions and there’s also clear significance between the model's forecast and for all other models forecasts. Under squared loss differentials the only exception to this were the forecasts of EGARCH(1,1) and ARIMA(1,0,1)-EGARCH(1,1) with normal distributions and MA5 as they showed equal predictive accuracy with 1% significance level.

For the NIKKEI225, the result from the Diebold-Mariano test under absolute loss differentials implies that the forecasts of the best performing ARIMA(0,0,0)-EGARCH(2,2) and the second ranked MA5 are statistically equally accurate. For all the other model forecasts, there's statistical difference indicating statistically more accurate forecast of the best ranked ARIMA(0,0,0)-EGARCH(2,2). Under squared loss differentials, the forecasts of the best ranked ARIMA(0,0,0)-EGARCH(2,2) and following EGARCH(1,1) are statistically equally accurate. For all the other models, the forecast of the ARIMA(0,0,0)-EGARCH(2,2) is statistically more accurate. For the Shanghai data the result from the Diebold-Mariano test showed that the forecasts of the best performing EWMA and second ranked MA5 under MSE are statistically equally accurate whereas for the third ranked MA21 there's statistical evidence of the forecasts to be different in favor of the EWMA forecast. Based on MAE loss function the Diebold-Mariano test showed the forecasts to be equal holds for the best performing MA5, MA21 and the third EWMA, as the null hypothesis of equal predictive accuracy was not rejected.

According to Diebold-Mariano test with S&P500 data, the forecast of the first ranked ARIMA(1,0,1)-EGARCH(1,1) with normal distribution under MSE was statistically significant as the null hypothesis of equal predictive accuracy was rejected for all of the other model forecasts. The only exception to this was the MA5, indicating no statistical difference between the two forecasts. Based on MAE, the similar result for the forecasts to be statistically equal holds also for the best performing ARIMA(1,0,1)-EGARCH(1,1) with t-distribution and the second ranked ARIMA(1,0,1)-EGARCH(1,1) with normal distribution and the third ranked MA5 as the null hypothesis of equal predictive accuracy was not rejected. For all the other model forecasts, there's statistical difference indicating statistically more accurate forecast of the ARIMA(1,0,1)-EGARCH(1,1) with t-distribution. For the Dow Jones data the Diebold-Mariano test show that the forecast of the best performing ARIMA(1,0,1)-EGARCH(1,1) with normal distribution under MSE was statistically significant as the null hypothesis of equal predictive accuracy was rejected for all of the other model forecasts. However, with the forecasts of the ARIMA(0,0,0)-EGARCH(2,2) with normal distribution, ARIMA(1,0,1)-EGARCH(1,1) with t-distribution and MA5, the null was not rejected, indicating forecast accuracy to be equal in population for each of the models. Based on MAE, the forecasts of the best performing ARIMA(1,0,1)-EGARCH(1,1) with normal distribution and the second ARIMA(0,0,0)-EGARCH(2,2) also with normal distribution, and the third ARIMA(1,0,1)-EGARCH(1,1) with t-distribution, would be statistically equal in population. For all the other model forecasts there's statistical difference indicating statistically more accurate forecast of the ARIMA(1,0,1)-EGARCH(1,1) with normal distribution.

As the results from the Diebold-Mariano tests show, that in most cases, there's no statistical difference between the first ranked model forecast and following second or third ranked model forecasts. With monthly forecasts the exception to this are the results from DAX30 and FTSE100 where first and second ranked model forecasts are statistically significant with 10% and 5% level. Under absolute loss differentials with Shanghai data there's also significant difference in monthly forecasting accuracy favoring the first ranked GARCH(1,1) with normal distribution. With daily forecasts under squared loss differentials with S&P500 data there's also clear statistical difference in volatility forecasting accuracy in favor of the first ranked ARIMA(1,0,1)-EGARCH(1,1) with normal distribution. Similar kind result is also from the FTSE100 under absolute loss differentials indicating statistical difference in volatility forecasting accuracy in favor of the first ranked ARIMA(1,0,1)-EGARCH(1,1) with t-distribution. The Diebold-Mariano test confirm the good performance of the five-day moving average, as the model's forecasts are almost always statistically equally accurate with the best ranked models' forecasts. Lastly, based on the results from the Diebold-Mariano tests, it is evident that the results of the volatility forecasting models can be re-evaluated and the ones with no statistical difference can be considered as one of the best models. However, after stating this, it should be noted that the comparison of the performance of the models, should still also include the use of loss functions. The Diebold-Mariano test is for comparing forecasts and perhaps the most profound evaluation of the results can be done using loss functions and Diebold-Mariano test or some other similar test such as White's (2000) Reality Check and Hansen's (2005) Superior Predictive Ability tests.

SUMMARY

Volatility is perhaps one of the most important subjects in finance and having certain well-known characteristics, such as volatility clustering and leverage effects, modelling and forecasting volatility has raised a lot of attention among research and practitioners. GARCH-type models were created not only to capture the characteristics but also to forecast volatility. Beside of the sophisticated GARCH-type models, some more simpler models are also widely used to forecast volatility. Perhaps one of the well-known are the moving average (MA) and exponentially weighted moving average (EWMA). The results from previous volatility forecasting studies are mixed and there's no solid and consistent view of superior volatility forecasting model while some of the research give support to GARCH-type models and others to more simpler moving average models. Volatility forecasts are useful basically everyone involved in financial markets. Different volatility forecasts are used in practice e.g., by risk and portfolio managers to adjust their asset positions and to hedge portfolios. Volatility is also one of the parameters in Black-Scholes-Merton option pricing. Lately, perhaps one of the most interesting area has been around volatility-based trading strategies.

This thesis examined volatility forecasting in six global equity indices, namely DAX30 (Germany), FTSE100 (UK), Shanghai SE Composite (China), NIKKEI225 (Japan), S&P500 (US) and Dow Jones Industrial Average (US). The sample data from January 1, 2000 through December 31, 2020 was divided to January 1, 2000 to December 31, 2015 in-sample and January 1, 2016 to December 31, 2020 forecasting periods. The one-day ahead volatility forecasts of different ARIMA-GARCH, ARIMA-EGARCH, ARIMA-GJR-GARCH and moving average models were compared and ranked based on mean squared error and mean absolute error loss functions. Basic GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) forecasting models were also included and monthly forecasts were produced only with these models. Although fat tails and excess kurtosis are a well-known phenomenon in financial data, it is far from clear whether volatility forecasting models with normal distributions or distributions allowing more kurtoses are better. Therefore, to compare the volatility forecasting results the normally distributed and t-distributed models were included. The Diebold-Mariano test was used to test statistical difference of the models' forecasts.

The overall conclusion was that there was no single model outperforming across the indices. However, the results provided support for the asymmetric models, as at least one of the EGARCH model was constantly ranked in top three under both loss functions. The exception

to this was the mean squared error ranking with daily volatility forecasts for the German DAX30 where the GJR-GARCH and the five-day moving average (MA5) model ranked best. The performance of the MA5 was also relatively good and the model outperformed many of the GARCH-type models across the indices. Contrary to previous research the standard GARCH model was almost constantly outperformed by the asymmetric EGARCH and GJR-GARCH models. The results from Shanghai SE Composite index revealed also that some of the simple forecasting models can outperform more complex models as the moving average and exponentially weighted moving average ranked best volatility forecasting models. According to results from the monthly forecasts, the exception to top performance of the asymmetric models were the results from Shanghai as under MAE, the GARCH(1,1) with normal t-distribution ranked first and second place. Under MSE, the EGARCH(1,1) hold first place followed by the GARCH(1,1).

Beside of the good performance of the asymmetric and the five-day moving average models, the other key take away message from this thesis is that although the data showed signs of excess kurtosis and fat tails, models with t-distributions were not constantly outperforming their normal distributed counterparts. Also, the Diebold-Mariano test result indicated that in most cases, there's no statistical difference between the first ranked model forecast and following second or third ranked model forecasts. When evaluating these results, it is important to consider the dependence of methodological choices. The data in this thesis was stock market data from different equity indices. As previous research has also outlined, the performance of the volatility forecasting models can vary with different datasets, assets and evaluation techniques. Perhaps the biggest benefit is gained when there's possibility to choose different models that perform well in different markets and based on the objective of the forecast.

SUGGESTIONS OF FUTURE RESEARCH

The results from this thesis created new approaches to consider in future research. The use of intraday data and evaluating the forecasting results with different loss functions would be fruitful research area. Evaluating volatility forecasts in different market conditions would also be an interesting topic. Some volatility forecasting models could offer interesting results during rising markets, whereas it would be interesting to see which models perform well in market downturns. Lastly, longer forecasting horizon could also bring valuable information to strategic portfolio management.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on automatic control. Vol. 19, No. 6, pp. 716–723.

Akgiray, V. 1989. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. The Journal of Business. Vol. 62, No. 1, pp. 55–80.

Alizadeh, S., Brandt, M., W. and Diebold, F., X. 2002. Range-based estimation of stochastic volatility models. The Journal of Finance. Vol. 57, No. 3, pp. 1047–1091.

Andersen, T., G. and Bollerslev, T. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review. Vol. 39, No. 4, pp. 885–905.

Andersen, T., G., Bollerslev, T., Diebold, F., X. & Labys, P. 2003. Modeling and forecasting realized volatility. Econometrica. Vol. 71, No. 2, pp. 579–625.

Awartani, B., M., A. & Corradi, V. 2005. Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. International Journal of Forecasting. Vol. 21, No. 1, pp. 167–183.

Bera, A., K. and Higgins, M., L. 1997. ARCH and bilinearity as competing models for non-linear dependence. Journal of Business and Economic Statistics. Vol. 15, No. 1, pp. 43–50.

Black, F. and Scholes, M. 1973. The pricing of options and corporate liabilities. Journal of Political Economy. Vol. 81, No. 3, pp. 637–654.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics. Vol. 31, No. 3, pp. 307–327.

Bollerslev, T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics. Vol. 69, No. 3, pp. 542–547.

Box, G., E., P. and Jenkins, G., M. 1976. Time series analysis: Forecasting and control. Revised Edition. San Francisco: Holden-Day.

Brailsford, T., J. and Faff, R. 1996. An evaluation of volatility forecasting techniques. Journal of Banking & Finance. Vol. 20, No. 3, pp. 419–438.

Brooks, C. 2019. Introductory Econometrics for Finance. 4th Edition. Cambridge University Press.

Cao, C., Q. & Tsay, R., S. 1992. Nonlinear time-series analysis of stock volatilities. Journal of Applied Econometrics. Vol. 7, No. S1, pp. 165–185.

China Securities Index. 2021. SSE Composite Index [online document]. [Accessed 9 October 2021]. Available at: https://www.csindex.com.cn/en/indices/index-detail/000001#/indices/family.

Cumby, R., Figlewski, S. and Hasbrouck, J. 1993. Forecasting volatilities and correlations with EGARCH models. The Journal of Derivatives. Vol. 1, No. 2, pp. 51–63.

Deutsche Börse. 2021. 30 years of DAX [online document]. [Accessed 9 October 2021]. Available at: https://www.deutsche-boerse.com/dbg-en/our-company/30-years-dax/dax-facts.

Diebold, F., X. and Mariano, R., S. 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics. Vol. 13, pp. 253–263.

Diebold, F., X. and Yilmaz, K. 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. International Journal of Forecasting. Vol. 28, No. 1, pp. 57–66.

Ederington, L., H. & Guan, W. 2005. Forecasting volatility. The Journal of Futures Markets. Vol. 25, No. 5, pp. 465–490.

Engle, R., F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica. Vol. 50, No. 4, pp. 987–1007.

Engle, R., F. and Ng, V., K. 1993. Measuring and testing the impact of news on volatility. The Journal of Finance. Vol. 48, No. 5, pp. 1749–1778.

Figlewski, S. 1997. Forecasting volatility. Financial Markets, Institutions & Instruments. Vol. 6, No. 1, pp. 1–88.

Franses, P., H. and Van Dijk, D. 1996. Forecasting Stock Market Volatility Using (Non-Linear) GARCH Models. Journal of Forecasting. Vol. 15, No. 3, pp. 229–335.

FTSE Russell. 2021. FTSE 100 Index [online document]. [Accessed 9 October 2021]. Available at: https://www.ftserussell.com/analytics/factsheets/home/search.

Garman, M., B. and Klass, M., J. 1980. On the estimation of security price volatilities from historical data. The Journal of Business. Vol. 53, No. 1, pp. 67–78.

Glosten, L., R., Jagannathan, R. and Runkle, D., E. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance. Vol. 48, No. 5, pp. 1779–1801.

Hansen, P., R. 2005. A test for superior predictive ability. Journal of Business & Economic Statistics. Vol. 23, No. 4, pp. 365–380.

Hansen, P., R. and Lunde, A. 2005. A Forecast Comparison of Volatility Models: Does anything beat a GARCH(1,1)? Journal of Applied Econometrics. Vol. 20, No. 7, pp. 873–889.

Hansen, P., R. and Lunde, A. 2006. Consistent ranking of volatility models. Journal of Econometrics. Vol. 131, No. 1-2, pp. 97–121.

Knight, J., L. and Satchell S. 2007. Forecasting Volatility in the Financial Markets. 3rd ed. Amsterdam, Butterworth-Heinemann.

Kuen, T., Y. and Tung, S., H. 1992. Forecasting volatility in the Singapore stock market. Asia Pacific Journal of Management. Vol. 9, No. 1, pp. 1–13.

Kwiatkowski, D., Phillips, P., C., B., Schmidt, P. and Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? The Journal of Econometrics. Vol. 54, No. 1-3, pp. 159–178.

Liu, W. & Morley, B. 2009. Volatility forecasting in the Hang Seng Index using the GARCH approach. Asia-Pacific Financial Markets. Vol. 16, pp. 51–63.

Ljung, G., M. & Box, G., E., P. 1978. On a measure of lack of fit in time series models. Biometrika. Vol. 65, No. 2, pp. 297–303.

Loudon, G., F., Watt, W., H. & Yadav, P., K. 2000. An empirical analysis of alternative parametric ARCH models. Journal of Applied Econometrics. Vol. 15, No. 2, pp. 117–136.

Mandelbrot, B. 1963. The variation of certain speculative prices. The Journal of Business. Vol. 36, No. 4, pp. 394–419.

Mathworks. 2022. KPSS test for stationarity [online document]. [Accessed 10 January 2022]. Available at: https://www.mathworks.com/help/econ/kpsstest.html.

McMillan, D., G., Speight, A. and Gwilym, O. 2000. Forecasting UK stock market volatility. Journal of Applied Econometrics. Vol. 10, No. 4, pp. 435–448.

Merton, R., C. 1973. Theory of rational option pricing. Bell Journal of Economics and Management Science. Vol. 4, No. 1, pp. 141–183.

MSCI. 2021. RiskMetrics™ — Technical document [online document]. [Accessed 1 October 2021]. Available at: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

Nikkei. 2021. Index list [online document]. [Accessed 9 October 2021]. Available at: https://indexes.nikkei.co.jp/en/nkave/index.

Nelson, D., B. 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica. Vol. 59, No. 2, pp. 347–370.

Pagan, A., R. and Schwert G., W. 1990. Alternative models for conditional stock volatility. Journal of Econometrics. Vol. 45, Issues 1-2, pp. 267–290.

Parkinson, M. 1980. The extreme value method for estimating the variance of the rate of return. The Journal of Business. Vol. 53, No. 1, pp. 61–65.

Poon, S.-H. and Granger, C., W., J. 2003. Forecasting Volatility in Financial Markets: A Review. Journal of Economic Literature. Vol. 41, No. 2, pp. 478–539.

Poterba, J., M. and Summers, L., H. 1986. The persistence of volatility and stock market fluctuations. The American Economic Review. Vol. 76, No. 5, pp. 1142–1151.

Rogers, L., C., G. and Satchell, S., E. 1991. Estimating variance from high, low and closing prices. The Annals of Applied Probability. Vol. 1, No. 4, pp. 504–512.

Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics. Vol. 6, No. 2, pp. 461–464.

Sharma, P. and Sharma V. 2015. Forecasting stock index volatility with GARCH models: international evidence. Studies in Economics and Finance. Vol. 32, No. 4, pp. 445–463.

Shu, J. and Zhang, J., E. 2006. Testing range estimators of historical volatility. The Journal of Futures Markets. Vol. 26, No. 3, pp. 297–313.

S&P Dow Jones Indices. 2021a. S&P U.S. Indices [online document]. [Accessed 9 October 2021]. Available at: https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf.

S&P Dow Jones Indices. 2021b. Dow Jones Industrial Average [online document]. [Accessed 9 October 2021]. Available at: https://www.spglobal.com/spdji/en/indices/equity/dow-jones-industrial-average/#data.

Walsh, D., M. and Tsou, G., Y.-G. 1998. Forecasting index volatility: Sampling integral and non-trading effects. Applied Financial Economics. Vol. 8, No. 5, pp. 477–485.

White, H. 2000. A reality check for data snooping. Econometrica. Vol. 68, No. 5, pp. 1097–1126.

Wilhelmsson, A. 2006. Garch forecasting performance under different distribution assumptions. Journal of Forecasting. Vol. 25, No. 8, pp. 561–578.

Yang, D. and Zhang, Q. 2000. Drift independent volatility estimation based on high, low, open, and close prices. The Journal of Business. Vol. 73, No. 3, pp. 477–492.