Vaasan yliopisto
UNIVERSITY OF VAASA

Emma Kontsas

Volatility Forecasting in Emerging Markets

School of Accounting and Finance
Master's Thesis in Finance
Master's Programme in Finance

Vaasa 2020

UNIVERSITY OF VAASA
School of Accounting and Finance
Author: Emma Kontsas
Title of the Thesis: Volatility Forecasting in Emerging Markets
Degree: Master of Science in Economics and Business Administration
Programme: Master's Degree Programme in Finance
Supervisor: Anupam Dutta
Year: 2020 **Sivumäärä:** 89

ABSTRACT:
This thesis examines the forecasting accuracy of implied volatility and GARCH(1,1) model volatility in the context of emerging equity markets. As a measure of risk volatility is a key factor in risk management and investing. Financial markets have become more global and the importance of volatility forecasting in emerging markets has increased. Emerging equity markets have more different risks than developed stock markets. As risk affects the potential return it is important to test and study how volatility models are able to forecast future volatility in emerging markets. The purpose of this thesis is to study the forecasting abilities and limitations of option implied volatility and GARCH(1,1) in the riskier emerging market environment.

The majority of previous studies on volatility forecasting are focused on developed markets. Previous results suggest that in developed equity markets implied volatility provides an accurate short-term future volatility forecast whereas GARCH models offer a better long-term volatility forecast. The previous results in emerging market context have been in rather inconclusive. However, there is more evidence of GARCH(1,1) volatility being the most accurate future volatility forecaster. The main motivation behind this thesis is to examine which models is best suited for volatility forecasting in emerging equity markets.

The forecasting accuracy of option implied volatility and GARCH(1,1) volatility is tested with an OLS regression model. The data consist of MSCI Emerging Market Price index data and corresponding option data from 1.1.2015 to 31.12.2019. In this thesis the daily closing prices of the index and option are used to compute daily and monthly implied volatility and GARCH(1,1) model volatility forecasts. Loss functions are applied to test the fit of the models.

The results suggest that both models contain information about one-day future volatility as the explanatory power of both models is statistically significant for daily and monthly forecasts. The GARCH(1,1) volatility is a more accurate future volatility estimate than implied volatility for both daily and monthly volatilities. The monthly volatility forecast is more accurate for both models than the daily forecast. The results indicate that in both daily and monthly values GARCH(1,1) volatility is a more accurate estimate for future volatility than implied volatility. The GARCH(1,1) monthly volatility offers the best fit for future volatility with the highest predictive power and lowest error measures, suggesting that it is the most appropriate fit for future volatility forecasting in emerging equity markets.

KEYWORDS: Volatility forecasting, Implied volatility, GARCH, Emerging markets

Contents

Pictures

Figures

Tables

Abbreviations

ARCH	Autoregressive Conditional Heteroscedasticity
GARCH	Generalised Autoregressive Conditional Heteroscedasticity
IV	Implied volatility
MAE	Mean Absolute Error
OSL	Ordinary Least Squares
RMSE	Root Mean Square Error
RV	Realised volatility
VBA	Visual Basic for Applications
VIX	Implied volatility of the S&P500 index

1 Introduction

Forecasting equity market risk has held the attention of finance professionals and researchers for over two decades. An accurate estimate of future volatility is a key input in investing and risk management. As financial markets have become increasingly global and efficient, there are multiple models that can be applied to volatility forecasting. However, current research of these models is more focused on developed equity markets. The application of volatility forecasting models to emerging markets has not been widely researched. As emerging equity markets have more risks that affect stock returns, it is important to test and study how volatility models are able to forecast future volatility.

This thesis focuses on two most commonly used volatility forecasting methods, the option implied volatility and Generalized Autoregressive Conditional Heteroscedasticity. Implied volatility is calculated from the Black-Scholes (1973) option pricing formula when other model inputs, such as option price and underlying price, are known. As a measure implied volatility has its drawbacks. The model assumes volatility to be constant over option's life when in reality volatility changes over time and exhibits clustering. Implied volatility is also affected by option moneyness.

Although option implied volatility is a widely used model in finance, Engle (1982) developed Autoregressive Conditional Heteroscedasticity in order to predict time-varying volatility. The Generalized model, known as GARCH, recognises that volatility changes over time and exhibits clustering where volatility tends to be high or low for extended time periods. Bollerslev (1986) introduced the GARCH(1,1) volatility model which includes lag-factors for previous return and volatility, taking the clustering effect of volatility into consideration.

This thesis examines the forecasting accuracy of option implied volatility and GARCH(1,1) in emerging equity markets. The MSCI Emerging Market Price Index and the corresponding index option are used to compute these model's volatility estimates which are then compared to one-day ahead realised volatility on a daily and monthly level during 1.1.2015–31.12.2019.

1.1 Motivation and purpose

Emerging markets experience more volatility than developed markets. The main motivation and purpose of this thesis is to examine whether the most commonly used volatility forecasting models, implied volatility and GARCH(1,1), have informational content over future volatility in a riskier environment. A major interest in this thesis is to test volatility forecasting models in a market environment that is more volatile than developed markets. Risk and volatility forecasting have been widely studied in developed markets and in that environment these models have provided accurate estimates of future stock market volatility. However, the research of these models in emerging markets has been inconclusive.

Emerging economies have become increasingly significant in global financial growth. This makes it crucial to understand the risks in emerging equity markets and how to forecast future volatility. Emerging markets are an interesting topic in risk research as these market experience risks that are not as present in developed markets. These include political, financial and environmental risk factors. These risks are drivers to higher volatility levels than in developed markets. The main focus is to study how accurately volatility forecasting models can predict future volatility in a more risky environment.

1.2 Research question and hypothesis

This thesis aims to analyse option implied volatility and GARCH(1,1) volatility as forecasting methods in emerging equity markets. The main research question is whether these models contain one-day ahead information about future volatility on a daily and monthly level. This research question leads to following null hypothesis:

H_0: Implied volatility and GARCH(1,1) volatility do not contain information over realised volatility in emerging equity markets

The alternative hypotheses are then analysed for both models in terms of daily and monthly volatilities:

H_1: Daily implied volatility accurately predicts future realised volatility in emerging equity markets
H_1: Daily GARCH(1,1) volatility accurately predicts future realised volatility in emerging equity markets
H_2: Monthly implied volatility accurately predicts future realised volatility in emerging equity markets
H_2: Monthly GARCH(1,1) volatility accurately predicts future realised volatility in emerging equity markets

The research question is analysed through MSCI Emerging Market Price index and corresponding option price data during the time period of 1.1.2015–31.12.2019. The hypotheses are tested with Ordinary Least Squares regressions and two loss functions are applied to test the fitting accuracy of these models.

1.3 Previous studies

Figlewski (1994) defines volatility as a statistical risk measure that describes the dispersion of asset returns around the mean. It is measured with variance or standard deviation. Abken and Nandi (1996) used logarithmic returns to measure realised volatility from market data and Parkinson (1980) presents a range based model that uses the change between highest and lowest observed price. Realised volatility or historical volatility can also be used as a long-term future volatility level estimate but research by Canina and Figlewski (1993) suggest that it offers an inaccurate measure for short-term forecasting.

In order to better forecast future volatility, Black and Scholes (1973) introduced the option implied volatility. Implied volatility can be derived from the Black-Scholes option pricing model when other inputs of the model, such as market price of the option and underlying stock, are observable in the market. As a measure implied volatility is theoretically a good estimate since the option price should contain information about future price levels until the end of maturity.

A drawback to implied volatility was first described by Mandelbrot (1963). While implied volatility model assumes a constant volatility over the option's life, in reality volatility changes over time. In addition to that, volatility has a clustering tendency, which means period of high/low volatility are followed by extended period of high/low volatility. Mandelbrot (2009) notes that in a well-functioning market stock returns should be uncorrelated with previous returns. However, there appears to exist autocorrelation between absolute periodic returns. Abken and Nandi (1996) suggest there is also another drawback to implied volatility as a forecaster: implied volatility changes in accordance with option's moneyness and maturity.

To correct for the implied volatility model's assumption of constant volatility over option's maturity, Engle (1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model. This stochastic model assumes that volatility changes over time and that it experiences autocorrelation with previous volatility. Bollerslev (1986) presented the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The GARCH models include more flexible lag component structure and are adaptive to different volatility levels which enables the calculation of long-term future volatility estimates. The GARCH(1,1) is a widely used adaptation of the model that has one lag-component for both past return and past volatility.

Implied volatility and GARCH(1,1) have been widely studied in developed equity markets. Poon and Granger (2003) suggest that implied volatility is more accurate as a short-term forecast and that at-the-money options are less affected by the implied volatility skew.

A review by Poon and Granger (2005) concludes that implied volatility dominates historical volatility, ARCH and GARCH models as a future volatility forecaster. However, Bentes (2015) suggests that while implied volatility provides the best short-term forecast, GARCH(1,1) model offers the best long-term forecast when data from US and emerging markets was compared.

According to Easterly, Islam and Stiglitz (2001) emerging markets experience more risk than developed markets as emerging markets have lower trading volume and lower levels of liquidity as well as more risk factors that are country specific, such as political risk. These risks make emerging markets an interesting research topic in volatility forecasting. The previous results on the accuracy of volatility forecasting models in emerging equity markets are inconclusive. Yang and Liu (2012) compared historical volatility, implied volatility and GARCH models in Taiwanese stock market and suggest that implied volatility is the most accurate forecast for monthly volatility. Gokcan (2000) compared GARCH based models in forecasting emerging market volatility. The results suggest that GARCH(1,1) volatility offers the most accurate future volatility forecast in these markets. As suggested by Bentes (2015), the best suited model for volatility forecasting depend on the forecasting time period.

1.4 Structure of the thesis

This thesis is structured in a manner that theories and terminology of volatility models is presented first and that is followed by examination of previous studies results. The following section of this thesis introduces the concept of volatility as a risk measure and presents the calculation method of realised volatility as well as presenting some results on historical volatility forecasting.

The third section of the thesis focuses on volatility forecasting models, most importantly implied volatility and stochastic GARCH based models. The advances and drawbacks of each model are analysed through existing literature and by reviewing previous studies in forecasting with these models. Based on previous research some conclusion of the research field are drawn. The fourth section introduces the emerging equity markets and the risks that arise especially in these markets. Previous results in volatility forecasting are presented and analysed. The section also compares the emerging market forecasting results to the ones presented in previous section for developed markets.

Data and methodology used in this thesis are described in the fifth section. Based on existing literature, appropriate models are chosen to analyse the research question and hypotheses. Section six presents the descriptive statistics, empirical results and offers topics for future research.

2 Volatility as a risk measure

Volatility is a statistical measure that describes the dispersion of observations around a mean. In finance volatility is commonly defined as dispersion of returns around expected mean. Volatility is therefore used to measure the amount of uncertainty as to size of changes in a security's price. Forecasting volatility is a crucial part of investment process when it comes to asset pricing and managing investment's risk. Volatility forecasting is a useful tool for investors and financial professionals. It has also held the attention of researchers for over two decades, and the research around volatility forecasting is still an evolving field of study. An accurate future volatility forecast is a key input in asset pricing and investment risk evaluation.

This section of the thesis defines volatility as a measure of risk and introduces calculation methods for realised and past volatility. Realised or historical volatility can be measured for a sample and it can also hold information about future volatility. This section also describes features of volatility that are observed in equity markets. Volatility has a clustering tendency, which is periods of high or low volatility that follow each other. Another feature in volatility forecasting is an observed volatility smile, which is described later in the thesis.

2.1 Definition of volatility

Volatility is a statistical measure and in finance it is defined as the dispersion of returns of a security. It is measured by standard deviation or variance of returns around a mean and can be interpreted as the amount of uncertainty in markets as to size of changes in a security's price. A higher volatility indicates a greater uncertainty of a security's future value. Higher volatility means there is a larger spread of security prices, which indicates there is a higher risk of price change. An asset with high volatility is more likely to experience larger price fluctuations during a short time period. A lower volatility indicates that the asset price is relatively stable over a short time period. (Figlewski, 1997; Poon & Granger, 2003)

Depending on the information and data available, the variance of a security can be calculated in two different ways. When the probability distribution of returns and expected return of a security can be defined, variance is calculated from the stock returns as the sum of averaged squared deviations of expected return as follows:

$$\sigma^2 = \sum p(s)[E(r) - r(s)]^2 \tag{1}$$

where σ^2 is the variance of security's return, $\sum p(s)$ is the sum of probabilities for each possible return and $[E(r) - r(s)]^2$ is the squared difference between expected and possible return. A larger variance σ^2 indicates a larger deviation of possible returns from expected return and the risk of price change is greater. A zero variance would indicate that there is no risk of price change. (Hull, 2015, pp. 210; Poon & Granger, 2003)

Variance can also be calculated from a data sample. Sample variance measures the spread of returns of a security from the data sample's mean return. It is defined as the sum of squared differences between each data point and the sample's mean:

$$\sigma^2 = \frac{\sum(X - \mu)^2}{N} \tag{2}$$

where σ^2 is the sample variance, $\sum(X - \mu)^2$ the sum of squared differences of each data point X from the sample mean μ, and N the number of data points in the data set. A sample variance of zero would indicate that all the values in the data set are equal and there is no price variation. A positive value indicates that there is variance of returns in the data sample. The larger the differences of prices from the mean, the larger the sample variance. (Zhang, Wu & Cheng, 2012)

In calculating sample variance, the arithmetic average of squared deviations is often multiplied by a factor of $N/(N-1)$, where N is the sample size. This is due to the use of sample mean μ in place of the expected value $E(r)$. The use of average causes a downward bias in sample variance calculation as in formula 2, which is referred to as *degrees of freedom bias*. By using the multiplying factor of $N/(N-1)$, the sample variance is commonly expressed as follows:

$$\sigma^2 = \left(\frac{N}{N-1}\right) \times \frac{\sum(X-\mu)^2}{N} = \frac{\sum(X-\mu)^2}{N-1} \tag{3}$$

Standard deviation is another measure of volatility. It is defined as the square root of variance:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum p(s)[r(s)-E(r)]^2} \tag{4}$$

or

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(X-\mu)^2}{N-1}} \tag{5}$$

where σ is the standard deviation. The interpretation of standard deviation is the same as for variance: the higher the standard deviation, the higher the chance of price change. (Hull, 2011, pp. 521–522; Poon & Granger, 2003)

Both standard deviation and variance are simple risk measures. Poon and Granger (2003) mention a drawback to these measures which is that both tend to put too much weight on outliers in the given data set. Outliers are observations that are far from the sample mean and may cause the variance to be abnormally large when calculating historical or future volatility. Outliers in the data set may lead to an upward or downward bias in sample variance.

2.2 Historical volatility

As presented by Abken and Nandi (1996), historical or realised volatility is the observable value that can be calculated as the average deviation of realised security returns from the realised average return for a time period. It is the simplest measure that can be used in estimating or forecasting future volatility of a security. Another method for calculating realised volatility is to use the underlying security's returns of a futures or option contract for a time period and changing the underlying security's logarithmic price changes into yearly volatility. In terms of volatility forecasting, a higher historical volatility would indicate a higher expected future volatility.

However, Abken and Nandi (1996) state that historical volatility as an estimate for future volatility does not have any indication of the security's price trend's direction. Another drawback to using historical volatility is that it is a measure of past price movements. To determine the correct historical time period that best reflects the future volatility of the stock price is difficult and the measure can be deceiving as it only reflects past trends.

2.2.1 Calculation of realised volatility

Mathematically realised volatility is calculated as the annualised standard deviation of returns. Hull (2015, pp. 201) has defined that realised volatility is calculated from the natural logarithm of daily stock returns:

$$R_t = \ln\left(\frac{S_t}{S_{t-1}}\right) \tag{6}$$

where R_t is the daily stock return of day t, S_t is the stock price or option or future contract's underlying price at day t and S_{t-1} is the stock price or option or future contract's underlying price at day $t-1$. Like the mathematical definition of sample volatility, realised volatility is measured by the variance or standard deviation of averaged squared deviations from the data sample's mean:

$$\sigma^2 = \frac{\sum_{t=1}^{T}(R_t - \bar{R})^2}{T-1} \tag{7}$$

where σ^2 is day t realised volatility, $\sum_{t=1}^{T}(R_t - \bar{R})^2$ is the sum of squared logarithmic return's deviation from the sample mean $\bar{R}$ and $T-1$ is the number of days in the sample period minus one. Daily realised volatility is often annualised by multiplying the daily value with $\sqrt{252}$ as there are approximately 252 trading days in a year. A monthly value can be computed by multiplying the daily realised volatility with$\sqrt{22}$. (Hull, 2015, pp. 201–203)

Parkinson (1980) suggests that realised volatility is more accurate when calculated with a range based method. A range based method utilises the highest and lowest value of the day as follows:

$$RV_t = \sqrt{\frac{\sum_{i=1}^{T}\ln(h_i - l_i)^2}{4ln(2)}} \tag{8}$$

where RV_t is the index' realised volatility, $\sum_{i=1}^{T}\ln(h_i - l_i)^2$ is the sum of natural logarithm of the difference between highest and lowest price during the sample period and *T* is number of days in the sample period.

When selecting a sample period for calculating realised volatility, it is relevant to consider the duration of the sample period and the frequency of observation. According to Bodie, Marcus and Kane (2014, pp. 737–743) an increased frequency of observations does not lead to a more accurate estimation of the data sample's mean. Lengthening the duration of sample period does however improve the accuracy of the mean, which suggests that a longer sample period would improve the realised volatility measure. Increasing the data observation frequency does in contrast improve the accuracy of the standard deviation estimate. Standard deviation increases at the rate of square root of time ($\sqrt{T}$). However, in practice it is usually complicated and not necessarily meaningful to obtain and use a long sample period. Older data may be less accurate and less informative, making it not representative of current volatility or future volatility estimates.

2.2.2 Forecasting with historical volatility

As Abken and Nandi (1996) suggest, it is complicated to evaluate whether a historical realised volatility value could contain information about future volatility. Poon and Granger (2005) indicate several issues with forecasting volatility based on historical volatility. Historical volatility is measured with squared standard deviations of realised returns from the sample period's mean. According to Poon and Granger (2005) this model is not robust to outliers in the data set which contribute to a biased volatility estimate. Outliers are abnormally high or low values in the sample period which may, depending on the data sample length and frequency, cause a biasness in realised volatility. Another issue with using historical volatility in forecasts is defining the correctly representative sample period. Does a longer sample period improve the accuracy of historical volatility forecast or would a shorter period be more describing of recent volatility expectations and market events?

There are several market phenomena that cause outliers in price data. Microstructure noise created by extremely high trading frequency is one cause of outliers in stock market data. Chan, Cheng and Fung (2010) examined whether the data frequency affects historical volatility's predictive power over future volatility. The results suggest that very high frequency market price data, such as 1-minute frequency, causes instability in historical volatility measure. Research results by Aït-Sahalia, Mykland and Zhang (2005) suggest that an optimal data frequency is 5-minutes as this sampling frequency eliminates market's microstructure noise. The ideal sampling frequency was further studied by Andersen, Bollerslev, Francis and Diebold (2007) who found that 5-minute sampling interval is robust enough to microstructure noise. When using data with 5-minute frequency, it also increases accuracy of historical volatility to use only open hours' data and eliminate the closed market data.

Volatility jumps are another cause for outliers in market price data. Volatility jumps are large changes in volatility that are caused by price shocks to stocks. Both firm-specific and market events can cause a jump in volatility. Andersen et al. (2007) adjusted their historical volatility forecasting model to include a volatility jump component. The results suggest that historical volatility is not a robust forecasting method when the data contains outliers caused by volatility jumps. The future predictability of volatility is higher in the non-jump component and jumps lead to a biased future volatility estimate. Historical volatility models are mean reverse and volatility jumps affect the mean and cause a bias in future volatility estimates.

When using historical volatility as an estimate for future volatility, the length of the sample period is another thing to consider. Figlewski (1994) examined the affects that sample period length has to the accuracy of historical volatility. Examining different time periods of historical volatility values for the S&P 500 index, the results indicate clearly that the longer the time period, the more accurate the historical volatility forecast. The results suggest that a five year sample period produces the most accurate future volatility estimate. As long-term volatility exhibits mean reversion, a longer time period (over 1 year) leads to an increase in historical volatility's accuracy as a future volatility forecaster.

When observation frequency and sample length are selected appropriately, historical volatility can provide a useful and accurate enough estimate of future volatility. Using UK FTSE100 stock returns during 1993–1995, Gwilym and Buckle (1999) compared the forecasting accuracy of historical volatility to option implied volatility. The results suggest that a one-year historical volatility provide an unbiased estimate of future volatility at 1%

significance level. The R^2-value of the model is low and indicates only 3% explanation to data variation. The results also indicate that a shorter than one-year sample period provides a non-reliable estimate for future volatility. In a more recent study Wang (2010) studied historical volatility of S&P 500 stocks during 1998–2008. A 60-day historical volatility had only 6.1% explanatory power over a next-day volatility forecast. However, the mean square error of 2.49 was lower than for a moving average model.

Fleming's (1998) results suggest that historical volatility models are inefficient when multiple lag-components are used. A 28-day historical volatility of S&P 500 stocks during 1985–1992 was an inaccurate one-day future volatility forecast with R^2 of 2%. Canina and Figlewski (1993) report similar results with S&P 100 stocks for time period of 1983–1987. Concluding that historical volatility is a poor estimate of future volatility, both studies also suggests that using historical volatility as a future estimate does not provide any value to investors when examining trading strategies.

Results by Alford and Boatsman (1995) suggest that taking industry and firm size into consideration improve historical volatility's accuracy as a future volatility forecast. Brous, Ince and Popova (2010) found supporting evidence examining S&P 100 stocks during 1996–2006. The study suggests that historical volatility outperforms implied volatility as a future volatility estimate for less-liquid stocks. The results indicate that taking industry, firm-size and liquidity into consideration lead to a more accurate historical volatility forecast. A more recent research by Chan, Jha and Kalimipalli (2009) also examined the economic benefits of S&P 500 historical volatility as a future volatility forecast. Results suggest no significant economic gains with historical volatility even when model is combined with option implied volatility forecast.

As previous results by Figlewski (1994) and Gwilym and Buckle (1999) indicate, the number one benefit of using historical volatility as a future volatility indicator is that it is easily calculated and price data is available almost from every market. Historical volatility interpreted as a long-run volatility level, when computed from a long time series, can provide a fairly accurate estimate of future volatility levels. A longer time period mitigates the effects of microstructure noise and volatility jumps. However, as results by Canina and Figlewski (1993) indicate, historical volatility is not an efficient estimate for one-day volatility forecasting. It is more useful for estimating a benchmark-level for long-run average volatility as it does not offer economical gain when used in investment strategies. For forecasting short-term volatility, a more accurate forecast is calculated with option implied volatility or a stochastic volatility model.

3 Volatility forecasting

Forecasting the volatility of equity returns is an important part of both investment process and management of risk. In addition to using historical volatility as a benchmark for future volatility, there are several approaches to future volatility forecasting. This section of the thesis presents two of the most commonly used forecasting methods: option implied volatility and Generalized Autoregressive Conditional Heteroscedasticity (GARCH).

The purpose is to describe the theory and assumptions behind these forecasting models and their calculation as well as summarising previous results on forecasting with implied volatility and GARCH. By examining previous studies this thesis aims to present the strengths and possible limitations of these forecasting approaches. The forecasting abilities and shortcomings of implied volatility and GARCH have held the attention of financial market researchers and professionals for over two decades and is still an evolving field of study.

3.1 Implied volatility

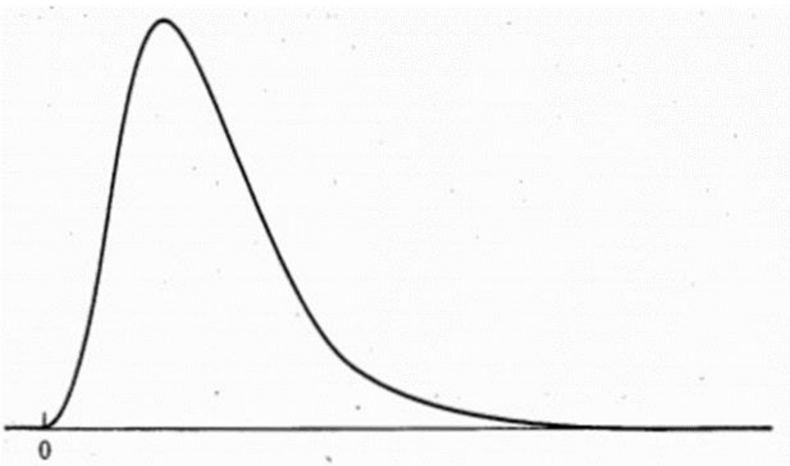
The most widely used and well-known measure to future volatility forecasting is option implied volatility. It is an option-based model and it is calculated from the Black-Scholes option pricing formula. Implied volatility is defined as the volatility level implied by option's price. It is calculates from the option pricing model when other factors of the model, such as option price and underlying asset's price, are known. This makes implied volatility a forward-looking measure rather than a historical model as it differs from historical volatility in the sense that calculated implied volatility is not based on historical information. (Hull, 2015, pp. 203)

Hull (2015, pp. 203–204) views implied volatility as the one variable in Black-Scholes option pricing model that cannot be observed directly. Derived from observable option prices, implied volatility is an estimate of the option's underlying stock's volatility. The

Chicago Board Option Exchange (CBOE) provides an implied volatility indexes for major equity indexes. The VIX index, which is the implied volatility index of the S&P 500 index, is commonly used by investors and risk managers to assess stock market volatility. During a bullish market implied volatility tends to be low as asset prices are expected to rise in a short time period. In a bearish market situation stock prices are expected to fall and implied volatility tends to rise due to greater price uncertainty.

3.1.1 Calculation of implied volatility

The Black-Scholes option pricing model was introduced by Fischer Black and Myron Scholes in 1973. Implied volatility can be computed through this option pricing model when other model variables are known. The Black-Scholes option pricing model is based on the assumption that the underlying stock's price approaches a lognormal distribution at the time of the option's expiration. A lognormal distribution is more skewed to the right than a normal distribution. As presented in Picture 1, it can have any value between zero and infinity.



Picture 1. Lognormal distribution. (Hull, 2011, pp. 323)

According to Hull (2011, pp. 303–304, 313) stock prices follow in a very short time period a *Wiener Process*, which is a continuous-variable stochastic process with a normal distribution and mean of zero and a variance rate of 1.0 per year. The Wiener Process is used in physics to characterise multiple small shocks to a particle. In option pricing this process is used to describe small price shocks to the underlying stock. The derivative's price is a function of stochastic underlying stock's price. This definition is known as *Itô's lemma*, and it denotes that at the option's expiration time the underlying stock's price, when given its price today, is lognormally distributed.

The Black-Scholes option pricing model can be computed for European and American call and put options. There are two significant assumptions in the model. First, the risk-free interest rate is assumed to be constant over the option's life. The second assumption is that the volatility of the stock price is constant over the life of the option. The Black-Scholes option pricing formulas for European call and put options are defined by Black and Scholes (1973) and Hull (2015, pp. 603–604) as follows:

$$c_0 = S_0 N(d_1) - Ke^{-rT} N(d_2) \tag{9}$$

$$p_0 = Ke^{-rT} N(-d_2) - S_0 N(-d_1) \tag{10}$$

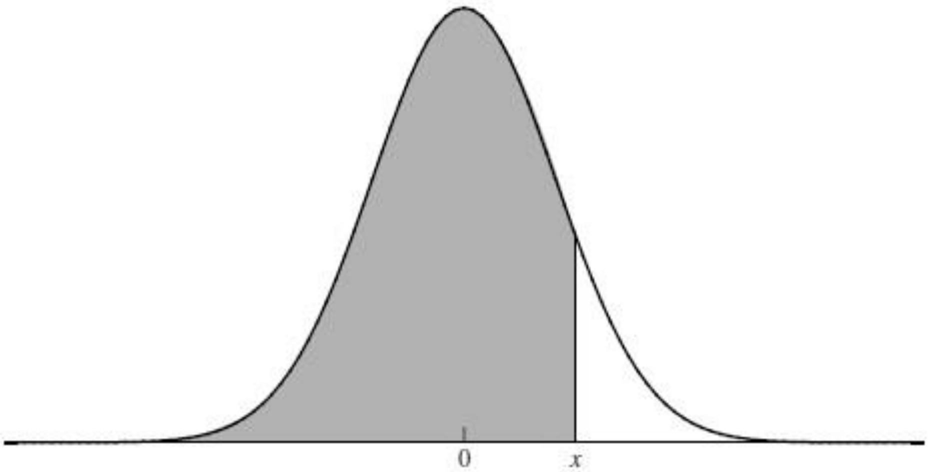
where

$$d_1 = \frac{ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{11}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{12}$$

where c_0 and p_0 are the current call and put option values, S_0 is the current price of the underlying stock, $N(d_1)$ is a factor by which the present value of a stock's random price exceeds the current stock price, $N(d_2)$ is the probability of the option being exercised, K is the option exercise price, e is Napier's constant that is the base of natural logarithm function $\ln$, r is the risk-free interest rate, T is time to option's expiration in years and σ is the standard deviation of the underlying stock's annualised, continuously compounded rate of return.

Picture 2 further demonstrates the cumulative probability distribution function of $N(d_2)$ factor. In the option pricing model the distribution describes the probability of the option being exercised. In Picture 2 the shaded area is the probability that the option is exercised.



Picture 2. The $N(d_2)$ function's cumulative probability distribution. (Hull, 2011, pp. 336)

Implied volatility can be calculated from the option pricing formula by finding the standard deviation that is consistent with the formula when option price is observed in the market. Implied volatility is computed by iteration when all the other inputs of the option pricing formula are known. This can be done by using a goal-seeking function that calculates the option implied volatility. (Black and Scholes, 1973; Hull, 2011, pp. 302–315, 321–343)

In practice, option implied volatility can be complex to calculate. Li (2005) presents several formulas for calculating an approximation of implied volatility for circumstances when the option meets certain properties. When the option is at-the-money, that is when the underlying stock price is equal to the discounted strike price of the option, implied volatility can be calculated for a call option using a model first presented by Brenner and Subrahmanyan (1988):

$$\sigma \approx \sqrt{\frac{2\pi}{T}} \times \frac{C}{S} \tag{13}$$

where σ is the approximation of standard deviation, π is mathematical constant pi, T is time to option's expiration, C is the call option's price and S is the spot price. This formula gives a representative approximation of volatility when the option is at-the-money. However, when the option is not at-the-money, an approximation formula by Corrado and Miller (1996) can be used to compute implied volatility:

$$\sigma \approx \sqrt{\frac{2\pi}{T}} \times \frac{1}{S+K}\left[C - \frac{S-K}{2} + \sqrt{\left[\left(C - \frac{S-K}{2}\right)^2 - \frac{(S-K)^2}{\pi}\right]}\right] \tag{14}$$

where σ is the approximation of standard deviation, , π is mathematical constant pi, T is time to option's expiration, C is the call option's price and S is the spot price. This model can be used to calculate an approximation of implied volatility for in-the-money or out-of-the-money options. Li's (2005) research suggest that the formula gives a fairly accurate benchmark for option implied volatility.

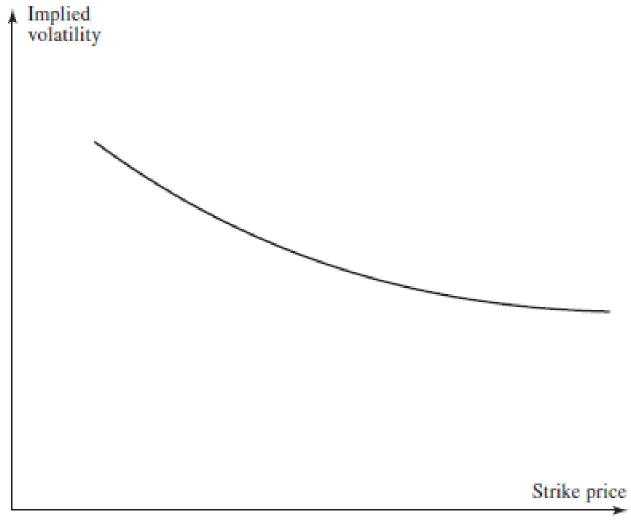
3.1.2 Features of implied volatility

As first described by Mandelbrot in 1963, large price changes of stocks tend to be followed by large price changes whereas small asset price changes tend to be followed by small changes. This phenomenon observed in equity markets is referred to as *volatility clustering*. There are extended periods of relatively high levels of volatility in markets that are then followed by an extended period of relatively low volatility levels. This clustering feature of volatility is an effect that is difficult to capture in volatility forecasting as the variance of daily returns can be high in one month and low in the following month. (Mandelbrot, 2009)

Mandelbrot (2009) specifies that in a well-functioning market stock returns are considered to be uncorrelated with previous returns. However, there appears to exist autocorrelation between absolute periodic returns. Volatility clustering is a market characteristic that is caused by market's slow reaction to new information and with large movements in price. This suggests that after a market shock that leads to high volatility, more high volatility levels can be expected for an extended time period.

Similarly to forecasting with historical volatility, when forecasting with implied volatility, clustering of high and low level volatility periods raises the question of how to choose a time period that best describes the expected future conditions for which the volatility forecast is modelled. As high volatilities tend to be followed by high volatilities and low volatilities by low volatilities, should the data time period include observations from the recent past or should it include both lower and higher volatility periods? Volatility clustering also raises the question whether the calculated forecast of volatility represents the future volatility conditions accurately. It can be complex to determine an appropriate volatility forecast that accurately describes future volatility since there is autocorrelation between returns during certain time periods.

Another feature of volatility to be considered when forecasting future volatility is the implied volatility skew. Abken and Nandi (1996) indicate that when implied volatility is calculated from an option pricing model such as the Black-Scholes model, it appears that implied volatility changes in accordance with option's moneyness and maturity. When implied volatility is plotted as a function of the option strike price with option maturity, the figure represents a *volatility smile* or *volatility skew*. This is displayed in Picture 3 where implied volatility as a function of strike price is shown to have a degreasing skew as the strike price increases.



Picture 3. Implied volatility skew. (Hull, 2011, pp. 436)

In the Black-Scholes option pricing formula implied volatility is assumed to be independent of the option strike price for a fixed time to maturity. As a function of strike price, implied volatility should yield a flat curve and not a skewed shape. However, Ederington and Guan (2002) suggest that in reality option implied volatility has a skew when for options of equal maturity the implied volatility of a deeply in-the-money call or out-of-the-money put is greater than the implied volatility of a deeply out-of-the-money call or in-the-money put.

The volatility skew has been observable in equity markets since the market crash of 1987. As suggested by Jackwerth and Rubinstein (1996), the skew or smile gives indication about investor's concerns about the possibility of market crashing. Therefore investors price options in accordance to expectations of another crash. This theory of crashophobia is supported by evidence that declines in the S&P 500 index are followed by a steepening in the skew and increases are correspondingly followed by a less steepening volatility skew.

Hull (2015, pp. 532–533) suggests that another cause of volatility skew are changes in company's leverage. A decline in company's equity increases leverage, which causes an increase in equity risk and thus an increase in volatility. Vice versa, an increase in equity reduces leverage, which results in a lower volatility. This implies that volatility is a decreasing function of asset price, which is consistent with the appearance of skewness.

The lognormality assumption is another factor that may cause bias in implied volatility calculation. Markets usually allow implied volatility to depend on the option time to maturity and the strike price. In reality, volatility skew is often less steep as the option's time to maturity increases. This real market phenomenon is referred to as *the volatility term structure.* Liu, Zhang and Xu (2014) examined the skewness of implied volatility. The results suggest that the skew is nearly flattened or less steep when investors are less informed and becomes steeper when investors have more information and behave more collectively.

Similarly to using historical volatility as an indicator of future volatility level, Abken and Nandi (1996) depict issues with the model's assumptions. One considerable assumption in the Black-Scholes formula is the presumption of volatility being constant over the option's life. Both in theory and in practice this assumption is false. However, Christensen and Prabhala (1998) suggest that implied volatility is a good estimate of short-term future volatility since it is likely for volatility to stay close to constant during few trading days.

Bollen and Whaley (2004) present another issue with using implied volatility in volatility forecasting. There is more demand on the market to some options than others, which causes demand pressure that leads to a price premium in option prices. The increase in demand raises the option price and thus raises the implied volatility. This can cause an upward bias in the future volatility estimate.

When forecasting future volatility with option implied volatility, clustering and skew are issues that need to be taken into consideration as well as choosing an appropriate time period of data. Stochastic volatility models have been created to correct the autocorrelation between absolute returns and possible biases in the Black-Scholes model. The stochastic models, including Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) were developed to solve these shortcomings of option implied volatility forecasting. These models are introduced in chapter 3.2 of the thesis which examines their application to volatility forecasting. (Abken & Nandi, 1996)

3.1.3 Forecasting volatility with implied volatility

Implied volatility has dominated other models in volatility forecasting and research on volatility. Theoretically it is the assumed future volatility for the remaining time to maturity of option making it by definition a forward looking measure. Previous studies have shown implied volatility to be an efficient and accurate forecast of future short term volatility and it is easy to compute from option pricing formula when appropriate data is available. It is also a key input in both option and stock pricing when interpreted as the level of price uncertainty. (Poon and Granger, 2003)

Implied volatility is calculated from the Black-Scholes option pricing model when other inputs of the model, such as option price and underlying stock price, are given. According to Christensen and Prabhala (1998) in an efficient market implied volatility should contain information about future volatility over the option's remaining maturity and at least all the information that is given by historical volatility. As the maturity of stock options is usually relatively short (<1 year), implied volatility should accurately forecast short-term future volatility.

Christensen and Prabhala (1998) studied the information content of monthly implied volatility calculated from S&P 100 index options in 1983–1995. The study uses non-overlapping data and a long time series and captured a regime shift after 1987 market crash. The results suggest that before the crash implied volatility is a biased estimate of future volatility due to poor signal-to-noise ratio during the crash and improved market information of investors after the crash. Since the crash, the results indicate with an adjusted R^2 of 62% that implied volatility is an accurate estimate of future volatility and outperforms historical volatility as a future volatility forecaster.

Poon and Granger (2003) and Blair, Poon and Taylor (2010) have studied the accuracy of implied volatility in forecasting future volatility for S&P 100 stocks after the crash from 1987 to 1992. The results suggest that implied volatility has the explanatory power of 12.9% – 35.6% for a future period of 1–20 days. The 20-days forecast provides the most accurate volatility estimate and 1-day forecast the least accurate. Poon and Granger (2005) concludes that implied volatility calculated from at-the-money options results in the most accurate estimates of future volatility. This is due to at-the-money options being less affected by the implied volatility skew and also having the highest trading volume.

Mayhew and Stivers (2003) examined the predictive power of implied volatility from the 50 most traded CBOE individual stock options and of the VIX index using daily option data from 1988 to 1995 with 22 days to maturity. The findings suggest that implied volatility contains almost all future information for the options with high trading volume. The implied volatility of the VIX index serves as a sufficient future volatility estimate for stocks with no options. A pre-crisis and after-crisis comparison revealed that the information content of implied volatility as a future volatility measure depends on option's trading volume. High trading volume options provide the most accurate forecasts and as the trading volume decreases, the accuracy of implied volatility forecast also decreases. As the trading volume increases after crisis, so does the informational content of implied volatility. Shaikh and Padhi (2015) report similar results around the market crash of 2007–2009. Studying the S&P CNX Nifty Index option's implied volatility, the results suggest that after the crisis high trading volume options provide a more reliable future volatility estimate.

Taylor, Yadav and Zhang (2010) did a comparison study of at-the-money S&P 100 index options and individual stock options during 1996–1999. The explanatory power of implied volatility for the index options is 43% whereas it is between 13–38% for the individual stock options. The results indicate that the higher explanatory power of the index options compared to individual stock options is due to higher trading volume. Also Han and Park (2013) suggest that the VIX index provides the most accurate estimate of future volatility since it has the highest trading volume.

Busch, Christensen and Nielsen (2011) examined how implied volatility is able to predict future realised volatility and volatility jumps. Using implied volatility calculated from at-the-money call option data of S&P 500 options from 1990 to 2002, the informational content of the measure is compared to realised volatility and volatility jump factors. The results suggest that implied volatility has the explanatory power (adjusted R^2) of 68% at a 5% significance level. Implied volatility contains high amount of information of future volatility for the option's life and the results indicate that it contains most of the information of volatility jumps.

Bentes (2015) studied implied volatility's accuracy in volatility forecasting for several volatility indexes. The research data consist of observations from the US (VIX), India (INVIXN), Hong Kong (VHSI) and Korea (KIX) from 2003 to 2012. The results indicate that implied volatility has an explanatory power of 45%-62% over historical volatility at 1% significance level. The results suggest that for these markets implied volatility is an accurate and unbiased estimate of future volatility. In comparing with historical volatility forecast implied volatility outperforms the historical measure.

Implied volatility is a commonly used measure to predict future volatility. It provides a more accurate estimate for future volatility than a historical measure. Previous research results suggest that the predictive power of implied volatility increases when the option's trading volume is higher. Blair et al. (2010) suggest that a forecasting period of 20 days provides the estimate with highest explanatory power and Christensen and Prabhala (1998) defines implied volatility as a short-term volatility forecaster. The informational content and accuracy of implied volatility as a future forecast is higher in the short-run as options usually mature in the near future. The assumption of constant volatility in the Black-Scholes option pricing model is more accurate for a short-term period. Busch et al. (2011) conclude that since option prices contain information about investor's expectations, implied volatility should capture the future expectations of volatility level and even volatility jumps. Poon and Granger (2005) suggest that using at-the-money options improves implied volatility's accuracy since option moneyness may cause skew and trading volume may cause biasness in the measure. When the forecasted period, trading volume and option's moneyness are taken into consideration, implied volatility provides an accurate and useful measure of future volatility.

3.2 Stochastic Volatility Models

A prominent issue with option implied volatility in future volatility forecasting is the Black-Scholes option pricing model's assumption of constant volatility over the life of the option. In order to correct this issue there are several developed stochastic volatility forecasting models. These models are more complicated to compute than implied volatility or historical volatility. An advantage of these models is, however, that they resolve most of the biases in implied volatility. This chapter of the thesis focuses on the calculation of Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH). With an emphasis on GARCH models, this chapter also presents previous results on stochastic volatility forecasting.

Stochastic model is a term for a model that includes a variable that changes over time. Stock price and volatility are stochastic continuous variables. Stochastic variables follow the *Markov process*, which indicates that in future forecasting only the variable's current value is relevant and historical values are assumed to be irrelevant. The ARCH and GARCH volatility forecasting models assume non-constant and varying volatilities and correlations. The models recognise volatility clustering where volatility tends to be high or low for extended time periods. (Engle, 1982)

3.2.1 Autoregressive Conditional Heteroscedasticity

Robert F. Engle first introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model in 1982. The model is specifically developed in order to model time-varying volatility. The basis of ARCH modelling is the least squares estimation model that is widely used in time series analysis. The least squares model assumes that the expected values for all squared error terms are equal at any given time point in the data. This assumption is referred to as *homoscedasticity*. However, volatility clustering is a phenomenon that causes *heteroscedasticity* in data when analysing future volatility. Heteroscedasticity means that the squared variances of error terms are not equal and there is autocorrelation of volatility between time points. (Engle, 1982)

Bollerslev, Chou and Kroner (1992) describe ARCH model treating heteroscedasticity in data as the variance to be modelled. The ARCH model's approach uses maximum likelihood estimation to correct the standard error caused by heteroscedasticity in the least squares estimation. The model provides a volatility forecast that is conditional on previous values as there appears to be autocorrelation between the volatility of returns. The maximum likelihood estimation method allows the data to be used to determine appropriate weight parameters to past variances in the ARCH model that best forecast the future volatility. The ARCH model has several extensions and applications to it. This thesis presents the ARCH(1) and ARCH(q) versions of the model.

Engle (1982) first introduced the simplest ARCH model which is the ARCH(1) model that consist of one lag-factor. The ARCH(1) is a regression model that consists of two different equations. The mean equation computes the mean return for the time series and the variance equation describes the error term variance. The mean equation is computed as follows:

$$y_t = \beta x_{t-1} + \varepsilon_t \tag{15}$$

where the dependent variable y_t is asset returns, x_{t-1} is a lagged return variable with β as a weight parameter and ε_t is the error term called w*hite noise*. White noise refers to random shocks to a variable that follow the Gaussian process. The error term ε_t is the second equation that describes the error term variance:

$$\varepsilon_t = u_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \tag{16}$$

where u_t is the white noise shock effect, α_0 and α_1 are stochastic process weight parameters and ε_{t-1}^2 is the squared lagged error term. The variance of the error term represents the time-varying volatility of the ARCH(1) model and is defined as follows:

$$\sigma^2(\varepsilon_t) = \alpha_0/1 - \alpha_1 \tag{17}$$

The ARCH(1) model has one-lag variable. A more general and usable model for volatility forecasting is the ARCH(q) model that is a qth order moving average process. According to Bollerslev, Engle and Nelson (1994) it differs from ARCH(1) in the sense that it has longer lag-variables and it can be computed form different time periods. In ARCH(q) the error term of returns is defined as follows:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + u_t \tag{18}$$

where ε_t^2 is the squared error term of the return equation, α_0 is a weight parameter, $\sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$ is the sum of lagged error terms at time point $t - i$ and u_t is the white noise shock effect. The volatility formula of ARCH(q) can be computed as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{19}$$

where σ_t^2 is the estimate of future variance, α_0 is a weight parameter and $\sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$ is the sum of lagged error terms at time point $t - i$. The ARCH(q) models focuses on the error term returns and is designed to forecast future volatility. (Bollerslev et al., 1994)

Engle and Mustafa (1992) examined the limitations of ARCH models. The ARCH model equations are fitted to returns and despite considering the heteroscedasticity the approach assumes the market environment to be relatively stable over the forecasting period. The model is not able to capture irregularities in the market such as new information effects, crashes, opening and closing of the markets or an option's price changes close to maturity. The results suggest that when markets experience unexpected price changes, the ARCH model is too conditional to past volatilities. During the market crash of 1987 the model's assumption of persistence of conditionality fails. These issues that arise when markets experience unexpected change are taken into account in Generalized ARCH models.

3.2.2 Generalized Autoregressive Conditional Heteroscedasticity

Tim Bollerslev (1986) first introduced the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. Based on Engle's ARCH model, the GARCH model includes more flexible lag component structure. The GARCH model has a learning mechanism that makes it more adaptive to different volatility levels and scenarios while still maintaining an easy adaptation and interpretation of results. The developed models allow more lag components with declining weights to be included in the calculation which creates a memory of past variances. GARCH models are mean reverting with constant unconditional variances that enable a longer forecasting period.

Bollerslev (1986) introduced the GARCH(1,1) model which is the simplest form of GARCH models. The formula has one autoregressive lag term and one moving average lag terms. The purpose of the model is to create a one period ahead forecast but also a two-period forecast can be made on the basis of the one-period forecast. The GARCH(1,1) approach is based on the ARCH process in equations 16 an 17. The variance rate is computed from the GARCH(1,1) model as follows:

$$\sigma_t^2 = \gamma V_L + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{20}$$

where σ_t^2 is the variance at time point t, V_L is the long-run average variance rate, u_{t-1}^2 is the lag-term of return at time point $t-1$ and σ_{t-1}^2 is the lag-term of variance time point $t-1$. The parameters γ, α and β are weights assigned to the long-run average variance, return lag-term and variance lag-term. These weights sum to one ($\gamma + \alpha + \beta = 1$). The term γV_L which is the long-run average variance can be also expressed as ω. According to Hull (2011, pp. 525) the GARCH(1,1) model is also commonly written as:

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{21}$$

Bollerslev's (1986) GARCH(1,1) approach is usually used to calculate daily volatility and it is compounded with daily information. The time point t in the equation is interpreted as today's volatility or the next day ($t+1$) volatility depending how the time point and lag-terms are determined. A volatility forecast can be computed from the model when market information on past returns and volatility are known factors. As mentioned by Hull (2011, pp. 526–529) the weightsγ, α and β are usually calculated with *maximum likelihood estimation* method and have different optimal values depending on the market situation. Usually β which is the weight assigned to the lagged variance term has a significantly larger value that the other weights. Since the lagged variance has a heavy weight, the GARCH(1,1) model is theoretically an appropriate forecasting method when there is volatility clustering.

In order to create forecasts that consider longer time intervals, Bollerslev (1986) designed the more general GARCH(p,q) process, which allows for multiple autoregressive and average lag terms to be included. The GARCH(p,q) model is computed as follows:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^{p} \gamma_j v_{t-j} + v_t \tag{22}$$

where ε_t^2 is the squared error term, $\sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2$ and $\sum_{j=1}^{p} \beta_j \varepsilon_{t-j}^2$ are the sum of lagged squared error terms for time points $t-i$ and $t-j$ with corresponding weights of α_i and β_j, $\sum_{j=1}^{p} \gamma_j v_{t-j}$ is the sum of long run variance term at time point $t-j$ with the weight of γ_j and v_t is the long run variance at time point t. From this formula Bollerslev et al. (1994) compute the volatility forecast as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{23}$$

where σ_t^2 is the variance at time point t, ω is the long-run average variance rate, $\sum_{i=1}^{q} \alpha_i u_{t-i}^2$ is the sum of lagged returns at time point $t-i$ with a weight α_i and $\sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$ is the sum of lagged variance at time point and $t-j$ with a weight β_j.

When calculating volatility forecasts with ARCH and GARCH models, the volatility estimate is usually calculated using a statistical programme. These stochastic approaches make adjustments to those assumptions made in implied volatility forecasting and create a simple solution to correct the models false assumptions about the behaviour of market volatility. When information about past returns and volatility are available, it is possible to use ARCH or GARCH models to calculate a future volatility estimate. (Hull, 2011, pp. 540)

3.2.3 Forecasting volatility with stochastic models

The key factor in stochastic volatility models in volatility forecasting is that the models allow a non-constant volatility. Thus, ARCH and GARCH models are more robust to some features of volatility that may cause biasness in implied volatility or historical volatility forecasts. The models are more robust to volatility clustering, skewness and outliers in the data. Although these stochastic models incorporate the natural behaviour of volatility, they are more complicated models than implied volatility and historical volatility. However, the stochastic models include historical and recent information of market returns and volatility and tend to provide a more accurate long-term volatility forecast than other forecasting methods. A review by Poon and Granger (2005) concludes that research on ARCH and GARCH volatility forecasting has provided varying results depending on market situation and data properties. This chapter examines the previous results on the accuracy of volatility forecasts when using these models.

Blair, Poon and Taylor (2001) examined the forecasting accuracy of simple ARCH models with S&P 100 index data from 1987–1992. The results indicate that for a 1 to 20 day forecast, the ARCH model has an explanatory power (R^2) of 30.7% at 5-10% significance level. However, a newer study by Yu (2002) compared volatility forecasting models predictive power in the New Zealand market with NZSE 40 Index from 1980–1998. The results for ARCH(q) model has a Theil's U of 1.1 which suggest that the model provides a weaker estimate than random walk.

Alam, Siddikee and Masukujjaman (2013) studied ARCH(1) model's ability to forecast Dhaka Stock Exchange General index' volatility from 2001–2011. The results indicate that ARCH(1) provided the best estimate of future volatility when compared to other ARCH-based models.

Nelson and Foster (1995) found evidence that ARCH models forecasting accuracy increases when the sample frequency approaches a continuous time series. High-frequency data includes information of conditional variances which leads to a more accurate future volatility estimate. Andersen and Bollerslev (1998) suggest that using at least 5-minute data frequency leads to a more accurate ARCH volatility forecast.

Akgiray (1989) examined the volatility of CRSP index data from 1963 to 1986. The study concludes that the index returns exhibit significant correlation. When volatility forecasts are compared, the GARCH(1,1) model results in the most accurate estimate of monthly volatility with lowest mean average error and root mean square error. The GARCH(1,1) model results in the best fit especially during periods of high volatility. The results suggest that the ARCH(q) method results in the second best volatility forecast while exponentially weighted moving average and historical volatility are more biased estimates.

Bera and Higgins (1997) studied GARCH(1,1) models performance in comparison to bilinear models in volatility forecasting. Using daily S&P 500 data from 1988–1993 the models are fitted to data sample and both in-sample and out-of-sample forecasts are computed. The results indicate that GARCH(1,1) has lowest root mean square error terms (0.489) for both in-sample and out-of-sample volatility forecasts.

Ederington and Guan (2005) studied the predictive power of GARCH(1,1) model for S&P 500 stocks from 1962 to 1995. The results suggest that even though GARCH(1,1) tends to put too much weight to the most recent observation in the data, it still provides better future volatility forecasts than historical volatility or exponentially weighted moving average model.

Andersen, Bollerslev and Meddahi (2004) examined whether adding lag-components to GARCH(p,q) model improves its accuracy in forecasting volatility. The results indicate that the accuracy of GARCH forecast improves with increasing the number of lag components. A model with 39 lag factors provided the most accurate estimate with the predictive power of 33-34% for a 1-20 week volatility forecast.

A more recent study by Bentes (2015) compared GARCH(1,1) model to other volatility forecasting models for US, India, Hong Kong and Korea market indexes with observations from 2003 to 2012. The results suggest that GARCH(1,1) outperforms implied volatility in long-term future volatility forecasting accuracy. The explanatory power of the GARCH model is 81%-89.3% at a 5% significance level. The results indicate that GARCH(1,1) provides an accurate future volatility forecast for a 24-month future period.

The forecasting abilities of ARCH and GARCH based models have been a continuous research topic that has resulted in different results depending on sample period length, data frequency and market volatility levels. Figlewski (1997) suggests that GARCH models' performance increases when at least daily data is available and the sample period included at least five years of data. When examining GARCH(1,1) forecasts of S&P 500 index from 1959–1993, the results indicate that GARCH(1,1) forecasts have lower root mean square error than historical volatility forecasts.

Day and Lewis (1992) studied the GARCH(1,1) model's accuracy in volatility forecasting with S&P 100 index data from 1983 to 1989. The results suggest that the GARCH(1,1) model provides a low quality future volatility estimate with the explanatory power (R^2) of only 3.9%. However, the results indicate that there is no biasness in GARCH(1,1) forecasts. Hansen and Lunde (2005) compared the GARCH(1,1) to multivariate GARCH models in forecasting volatility to IBM stocks from 1990–1999 . The results suggested that a multivariate approach is superior to the GARCH(1,1) model.

Glosten, Jagannathan and Runkle (1993) found evidence that monthly conditional volatility is not as persistent as previous research shows. By examining CRSP index monthly returns from 1951–1989, the results indicate that positive unanticipated returns result in a downward adjustment of conditional volatility whereas negative unexpected returns cause a rise in conditional volatility. When GARCH models are modified to allow the effects of unexpected returns on the conditional volatility (referred to as GJR-GARCH), the model provides more accurate monthly volatility forecasts. Further research on news effects by Engle and Ng (1993) suggests that negative return shocks have a greater impact on volatility than positive return shocks when examining Japanese stock returns during 1980–1988. When these shocks were included in forecasting data, the GJR-GARCH outperformed GARCH(1,1) model in volatility forecasting. Franses and Ghijels (1999) studied the effect of extreme values in GARCH modelling. The results suggest that removing outliers and extreme values from the data result in lower mean square error and a more accurate future forecast.

Engle and Patton (2007) examined the predictive power of GARCH(1,1) for Dow Jones Industrial Index from 1988–2000 for different data frequencies. The results indicate that GARCH(1,1) provides a good estimate of future volatility but the model is affected by sampling frequency. The coefficients in GARCH(1,1) while well suited for one sampling frequency are misspecified for another sampling frequency. These results suggest that GARCH model coefficients should be adjusted accordingly to the sample frequency.

3.3 Comparison of volatility forecasting models

As previous research results indicate, each volatility forecasting model has its own advantages and drawbacks. The accuracy of the forecasts depends on several market, return and volatility features. This chapter of the thesis compares historical volatility, implied volatility and GARCH models in terms of forecasting accuracy and suitable forecasting environment. The aim is to capture previous results on which model is best suited for different volatility levels and how the data sample affect the forecasting accuracy.

Historical volatility is a model based on previous realised volatilities. Results by Figlewski (1994) indicate that a longer sample period (over one year) increases the forecasting accuracy of historical volatility. Research by Aït-Sahalia et al. (2005), Andersen et al. (2007) and Chan et al. (2010) suggest that increasing the sample frequency up to 5-minute frequency improves the historical forecast. Canina and Figlewski (1993) concludes that for individual stocks historical volatility offers a poor future volatility estimate and it is best used as an overall volatility benchmark.

A survey by Poon and Granger (2005) indicates that historical volatility is not robust to outliers in the data. From 66 studies 76% found option implied volatility to have more predictive power over future volatility than historical volatility. Gwilym and Buckle (1999) suggest that implied volatility contains more information than historical volatility but implied volatility is also a more biased estimator. However, Christensen and Prabhala (1998) conclude that implied volatility is more accurate than historical volatility when using a long sample period. As options usually have less than a year to expiration, implied volatility should provide a stable short term future estimate.

Poon and Granger (2003) suggest that implied volatility is best suited for 1–20 day forecasts as the Black-Scholes model makes the assumption of constant volatility. Poon and Granger (2005) also conclude that at-the-money options are less affected by the implied volatility skew and demand pressure. Mayhew and Stivers (2003) and Shaikh and Padhi (2015) suggest that a higher trading volume improves the accuracy of implied volatility.

Busch et al. (2011) conclude that implied volatility contains all the information included in historical volatility and it captures volatility jumps. For S&P 500 index options implied volatility had an explanatory power of 68%. Bentes (2015) reports similar results for multiple markets. Implied volatility outperformed historical volatility as a future estimate with an explanatory power of 45%–62% depending on the market.

Bentes (2015) also compared GARCH(1,1) to implied volatility. Using market index data from US, India, Hong Kong and Korea, the results suggest that GARCH(1,1) outperformed implied volatility as a long-run forecast with an explanatory power of 81%–89%. However, implied volatility proved to be a better estimate for a short-period forecast. The explanatory power of both implied volatility and GARCH(1,1) has improved significantly in more recent observations.

Blair et al. (2010) suggest that implied volatility (VIX) provides more accurate estimates than ARCH models. Results by Andersen and Bollerslev (1998) and Nelson (1995) indicate that using high-frequency data improve the accuracy of ARCH models. In a review by Poon and Granger (2005) historical volatility outperformed ARCH models in 56% of studies. The results by Bentes (2015) indicate that ARCH volatility forecast for S&P 500 index is less explanatory than an implied volatility forecast.

Figlewski (1997) concludes that GARCH(1,1) is more accurate in forecasting volatility than historical volatility. The performance of GARCH(1,1) increases when the sample period contains daily data for at least a five year period. Similarly, Ederington and Guan (2005) suggest that GARCH(1,1) model outperforms historical volatility as a future volatility forecaster but the model puts too much weight on more recent observations. Andersen et al. (2004) indicate that a GARCH(p,q) model with multiple lag factors provides the most accurate future estimate. Engle and Patton (2007) suggest that the key to an accurate GARCH forecast is to adjust the model according to data frequency.

A review by Poon and Granger (2005) concludes that ARCH and GARCH volatility forecasting has provided varying results depending on market situation and data properties. The study suggest that implied volatility dominates historical volatility, ARCH and GARCH. Comparison between GARCH models ad historical volatility is less conclusive. Bartunek and Chowdhury (1995) found no statistically significant difference between historical volatility, implied volatility and GARCH(1,1) model when comparing volatility forecasts for different forecasting period for individual stocks from 1983–1984. However, Bentes (2015) concluded that implied volatility provided the best short-term forecast and GARCH(1,1) the best long-term forecast when data from US and emerging markets was compared.

It is evident that each model has its own benefits and drawbacks when it comes to forecasting future volatility. The next chapter of the thesis examines features that differentiate emerging markets from developed markets. Focusing on features of volatility present in emerging market environment, the next chapter discusses the previous results on forecasting volatility in emerging equity markets.

4 Volatility in emerging markets

For the past two decades, emerging market economies have grown rapidly and significantly. Globalisation of capital and trading has increased the importance of understanding emerging economies as the world economic growth is now more affected by emerging markets. Over the last decade emerging markets have accounted for three fourths of world economic growth. While return and risk have been widely studied in developed markets, the research in emerging markets is an ongoing subject that changes over time. (Knoop, 2013, pp. 32–34)

Previous results on volatility forecasting give examples of appropriate uses of each model. However, majority of the studies are focused in developed markets, especially the US market. A major motivation behind this thesis is to examine volatility forecasting models in a market environment that is more volatile than developed markets. The forecasting methods examined in this thesis have been studied in high volume and liquid market environments and the studies lack testing in a more volatile market. Emerging markets have risks that are not as present in developed markets which makes them an interesting subject for studying volatility forecasting in a riskier market setting. Whereas developed markets are often describes as well-functioning, the emerging markets lack features that are associated with functioning financial markets. Emerging markets have lower trading volumes and lower levels of liquidity as well as more risk factors that are country specific, such as political risk. These factors give motivation to study risk forecasting in emerging markets. (Easterly, Islam & Stiglitz, 2001)

This chapter describes the behaviour of volatility in emerging equity markets. There are several factors that are typical to emerging equity markets that are drivers to higher levels of volatility compared to developed stock markets. These factors include economic, financial, political and environmental risks that are more prominent in emerging economies and affect volatility levels of those markets. This chapter also examines previous findings in the field of volatility studies in emerging market context as well as previous results in volatility forecasting. This thesis defines emerging market economies based on existing definitions, more specifically, MSCI Emerging Markets Index. The MSCI Emerging Markets Index (MSCI, 2019) includes Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Russia, Qatar, South Africa, Taiwan, Thailand, Turkey and United Arab Emirates.

4.1 Features of volatility in emerging equity markets

Volatility is used as a risk measure in finance. Emerging stock markets experience economic and company specific risks that are not as present in developed stock markets. Bekaert, Erb, Harvey and Viskanta (1998) show that whereas developed market stock returns follow a relatively normal distribution, the returns of emerging equity do not necessarily follow a normal distribution. This is due to the fact that there are risks that are more present in emerging markets which leads to more outliers in return data. This increased chance of outliers causes the return distribution to have a larger kurtosis and skewness. Not only is there skewness and kurtosis in return distributions of emerging markets, these distributions change over time. A newer study by Adcock and Shutes (2005) shows that more skewness and kurtosis are still present when examining individual emerging economies. These outliers and more frequent extreme values in price changes raise the level of volatility in emerging stock markets.

The risks that are more present in emerging markets than developed markets arise often from lower levels of regulation of financial markets as well as political environment. According to Bilson, Brailsford and Hooper (2002) a different political environment that is less regulated by laws and more prone to extreme conditions, such as limitation of economic activities, changes in country leaders, wars and terrorism. These risks effects the market conditions more significantly in emerging markets than in developed markets. Political uncertainty causes insecurity in equity markets, which may slow down the economy and market activity.

A country's political environment also affects the laws and regulations concerning the equity market. Klomp and De Haan (2014) state that generally stricter regulation is associated with lower uncertainty and risk levels. Loose regulation allows for more freedom in all economic activity and also questionable actions. Most emerging economies are less regulated than developed economies which raises the risk level associated with emerging market stocks. Prasad (2010) concludes that while the need to develop financial market regulation is not only unique to emerging markets, it is more heightened in emerging economies. Emerging markets experience rapid financial development and inclusion via globalisation of financial markets while lacking in financial regulation.

Jun, Marathe and Shawky (2003) suggest that assets in emerging markets are less liquid compared to developed markets. This means that emerging market stocks have lower trading volumes than developed market stocks. The higher liquidity risk raises the overall risk level and thus emerging market volatility. Lesmond (2005) found evidence that when trading difficulty increases and raises illiquidity of emerging market stocks, the bid-ask spread also increases. However, liquidity of emerging market stocks increases when financial markets become more and more globalised. A newer study by Lischewski and Voronkova (2012) found no evidence of liquidity risk premium in Polish stock market. Another form of illiquidity in emerging markets is the illiquidity of domestic banks. Chang and Velasco (2001) suggest that illiquidity of domestic banks is a source of market uncertainty and even crashes.

Das, Papaioannou and Trebesch (2009) find evidence that in emerging economies firms experience more difficulty raising capital via external credit or equity issuance than in developed markets. Although market efficiency should drive funds to productive companies globally, emerging markets may not attract as many investments as developed markets. Hasan, Jackowicz, Kowalewski and Kozlowski (2017) suggest that local banks have a significant role in facilitating access to financing in Polish small- and medium-sized companies. However, Carvalho (2014) found evidence in Brazil that government control over banks increases the unbalance in capital flows as politics influence the allocation of bank lending to specific firms. Fernandes (2011) concludes that as emerging markets become more integrated with developed markets, the financial choices of emerging market firms increase when global capital becomes more available.

Khan, Sharif, Golpîra and Kumar (2019) suggest that environmental, social and governance risks have extensive effects to emerging economies. These risks are related to current trends in societies and economies and effect on both country-level and firm-level. Environmental risks arise from extreme environmental conditions and climate change. For example large floods and earthquakes may slow down the functioning of an entire country's economy. Social risks rise from responsibility issues. If a country starts regulating the labour market to increase social wellbeing, it might cause large changes in the entire economy. Regulation also applies to governance risks. Poor corporate governance increases the risk of inequality of investors.

ESG risks also affect the firm specific risk as emerging markets have less regulation and reporting demands than developed markets. Risks related to corporate governance can also be firm specific. Sherwood and Pollard (2017) suggest that taking ESG factors into account reduces firm specific risk in emerging markets. Firm specific governance issues, corruption and excessive risk taking are risks that exist in every company in every market. As emerging market economies are less regulated and supervised by independent authorities, there is a greater chance for firm specific risks than in developed economies.

Emerging market risk and return are linked to developed markets due to globalisation of financial markets, currency and capital structure. Sarwar and Khan (2017) studied the spill-over effects of US equity market risk to emerging equity markets before, during and after the financial crisis of 2008. Using data from 2003−2014, the study examines the contemporaneous and delayed risk transition from US markets to emerging markets. A multivariate regression analysis is used to determine the relationship between changes in VIX Index and MSCI Emerging Market Index. The results indicate that there is a contemporaneous negative relationship between emerging market stock returns and changes in VIX before, during and after the financial crisis. The negative relationship is also significant for a lag-period, which suggests that the changes in VIX continue to affect the emerging market returns on following day. The results show that higher US stock market risk transitions to emerging market returns by lowering the mean return and increasing the variance of returns. The study uses GARCH model coefficients to indicate that increases in VIX increase the emerging market volatility during all sub-periods and that the periods of increased volatility are extended suggesting higher future volatility.

Firm specific difficulties, lack of funding, poor regulation, political environment and economy wide phenomena increase the chance of risk in emerging market companies. These aforementioned risks increase the volatility in emerging stock markets. As firms have even a higher probability of bankruptcy, the emerging stock markets experience more volatility than developed stock markets. The next chapter further examines the ability to forecast volatility in emerging market environment. The focus of forecasting results is on implied volatility and GARCH models.

4.2 Volatility forecasting in emerging markets

The forecasting accuracy of historical volatility, implied volatility and GARCH was examined by Yang and Liu (2012) in Taiwanese stock market in 2006−2010. The results suggest that the option implied volatility outperforms historical volatility and GARCH model as 30-day volatility forecasts. All models are shown to have statistically significant positive correlation with realised volatility at 1% level with implied volatility having the highest t-statistic (21.7287). These findings suggest that option implied volatility is a simple and accurate model also in emerging market context where volatility levels are higher than in developed markets. In a contradicting study, Filis (2009) analysed the relationship between implied volatility forecasts and realised volatility in Greek option market from 2000–2003. The results suggest that implied volatility is a biased estimate of realised volatility.

Shaikh and Padhi (2015) studied the Indian S&P CNX Nifty Index option's implied volatility during 2007–2013. The study compared before and after the financial crisis forecasts. The results indicate that after the crisis high trading volume options provide a more reliable future volatility estimate. The explanatory power (R^2) varies between 0.10% and 0.45% for the whole sample period.

Pati, Barai and Rajib (2018) investigate the information content of implied volatility index in contrast to GARCH models in forecasting volatility of Indian, Australian and Hong Kong stock markets during 2008–2016. The study computes one-day forecasts using 5-minute frequency intraday data. The results suggest that while GARCH(1,1) provides statistically significant forecasts of future volatility, the model is improved when implied volatility is included in the regression. Results Bentes (2015) indicate that while implied volatility provides the most accurate short-term forecast, GARCH(1,1) model offers the best long-term forecast in emerging market context.

Gokcan (2000) examined GARCH models in forecasting emerging market volatility. By comparing linear stochastic models, ARCH(1,1) and GARCH(1,1), to non-linear EGARCH the study examines the predictive power of these models in Argentina, Brazil, Colombia, Malaysia, Mexico, Philippines and Taiwan during 1988–1996. The results indicate that the linear GARCH(1,1) outperforms other models even when the stock market return distribution is skewed. For all countries except for Brazil the mean square error of GARCH(1,1) forecast is smaller than the error of non-linear EGARCH. This suggests that GARCH(1,1) model is sufficient in forecasting future volatility in emerging markets.

Bley and Saad (2015) approached volatility forecasting in emerging markets by comparing the forecasting accuracy of history-based and conditional volatility models in Saudi Arabian stocks between 2004−2013. Despite the shortcomings of historical volatility as a future volatility forecaster, the results suggest that history-based models outperform ARCH and GARCH models especially when using an exponentially smoothed historical returns. The history-based models experience lowest mean absolute error measures

(MAE and MAPE) as well as lowest root mean square error (RMSE). This indicates that exponentially smoothed historical values provide a more accurate forecast on future volatility than conditional volatility models.

Pu, Chen and Ma (2016) examined volatility forecasting methods in Chinese stock market using high-frequency Shanghai Stock Exchange Composite Index data from 2000–2013. The results indicate that GARCH(1,1) and historical volatility are inferior to heterogeneous autoregressive models and have rather large mean square error terms.

Huang (2011) compared volatility forecasting models' results on developed markets to emerging markets. Using daily data of 24 emerging market stock indexes and seven developed market indexes from 2000–2006, the study computes historical volatility, GARCH(1,1), Monte Carlo simulation, stochastic model and quantile regression models to each market. For multiple emerging countries a historical volatility forecast of 20–60 provides the best future estimate while GARCH(1,1) is never the superior model. Balaban, Bayar and Faff (2006) did a similar study where several markets, both developed and emerging, were compared. While ARCH(1) provided the worst fitting volatility estimate in terms of mean absolute error for all emerging countries, historical volatility and GARCH(1,1) performed better. However, an exponentially smoothed model outperformed all GARCH based models.

Miah and Rahman (2016) studied the volatility of four Bangladeshi companies from 2000 to 2014. The study computes several GARCH(p,q) models with lag components and a GARCH(1,1) model. The GARCH(1,1) model outperformed GARCH(p,q) models by having the smallest error statistics.

5 Data and methodology

This section of the thesis introduces basic features of the emerging market data and the methodology used to model volatility forecasts and measure the predicative power of implied volatility and Generalized Autoregressive Conditional Heteroscedasticity. The data consists of MSCI Emerging Market Price Index daily closing prices from 1.1.2015 to 31.12.2019 and a time series of European December and June call option daily closing prices of the index for the same time period.

The methods used in this research are aimed to capture the predicative power of both volatility forecasting models. Realised volatility is calculated from the index closing price data, option implied volatility is calculated from the index price and option price data, and GARCH(1,1) forecasts are computed with index price data. To examine the predicative power, an OLS regression against realised volatility is used to measure the accuracy of volatility forecasts. Mean absolute error (MAE) and Root mean squared error (RMSE) are computed to measure the error terms of both forecasting methods.

5.1 Emerging Market Data

The data of this thesis consists of MSCI Emerging Market Price Index daily closing prices from 1.1.2015 to 31.12.2019 and a series of daily closing prices of December and June call options on MSCI Emerging Market Price Index with a strike price of USD 1100 from 1.1.2015 to 31.12.2019. All price data is obtained from Bloomberg database. The time period of 1.1.2015–31.12.2019 is chosen based on the existing option data and motivation to use most recent available information. Options with USD 1100 strike price has been available the longest and thus it enables the longest sample period.

5.1.1 Index Closing Prices

The MSCI Emerging Market Price Index (ticker: MXEF) captures large-cap and mid-cap companies in 24 emerging market countries. The index is quoted in US dollars. It represents approximately 85% of the free float-adjusted market capitalization of each country featured on the index. The emerging market countries included in the index are Brazil, Chile, China, Colombia, Czech Republic, Egypt, Greece, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Pakistan, Peru, Philippines, Poland, Russia, Qatar, South Africa, Taiwan, Thailand, Turkey and United Arab Emirates. (MSCI, 2019)

The index follows the MSCI Global Investable Market Indexes Methodology (GIMI), an approach that allows for adjustments based on market capitalization size, country and sector combinations. The MSCI Emerging Market Index is rebalanced in May and November and aims to reflect the underlying equity market in a liquid way. The largest sectors are information technology (27.38%) and financials (24.27%). The largest country weights are China (30.3%), South Korea (14.45%), Taiwan (11.39%), India (8.17%) and Brazil (7.49%). Other countries equal a weight of 28.2%. (MSCI, 2019)

In this research the MSCI Emerging Market Index daily closing prices from 1.1.2015 to 31.12.2019 are used to compute daily realised volatility levels for this time period as well as implied volatility and GARCH(1,1) model volatility forecasts.

5.1.2 Option Closing Prices

MSCI Emerging Market Price Index Options (ticker: OMEF) have been traded in Eurex Exchange starting from March 2014. The longest available option type for the entire research period is European December Call Option with a strike price of USD 1100. European June Call Options with the same strike price of USD 1100 are available from June 2015. The options are rolled so that December Call prices are used between December 31 and June 30, and June Call prices are used between June 31 and December 30. This creates a continuous daily option closing price data where options have between 169-353 days until expiration.

The daily option closing prices and daily index closing prices are used to compute implied volatility forecasts for the time period of 1.1.2015–31.12.2019. Implied volatility is calculated with Microsoft Excel using the Black-Scholes option pricing model with given option prices. Implied volatility is then computed with a VBA macro that uses the Excel's goal seek function that iterates the value of implied volatility for each day.

5.1.3 Descriptive statistics and market situation

Five years of MSCI Emerging Market Price Index data is used to analyse the daily volatility of emerging markets. Figure 1 presents the daily values and movements of the index. As Figure 1 presents, the lowest value of the index was on January 1st 2016 at USD 688.52. The index reached its highest value of USD 1273.07 on January 26th 2018.

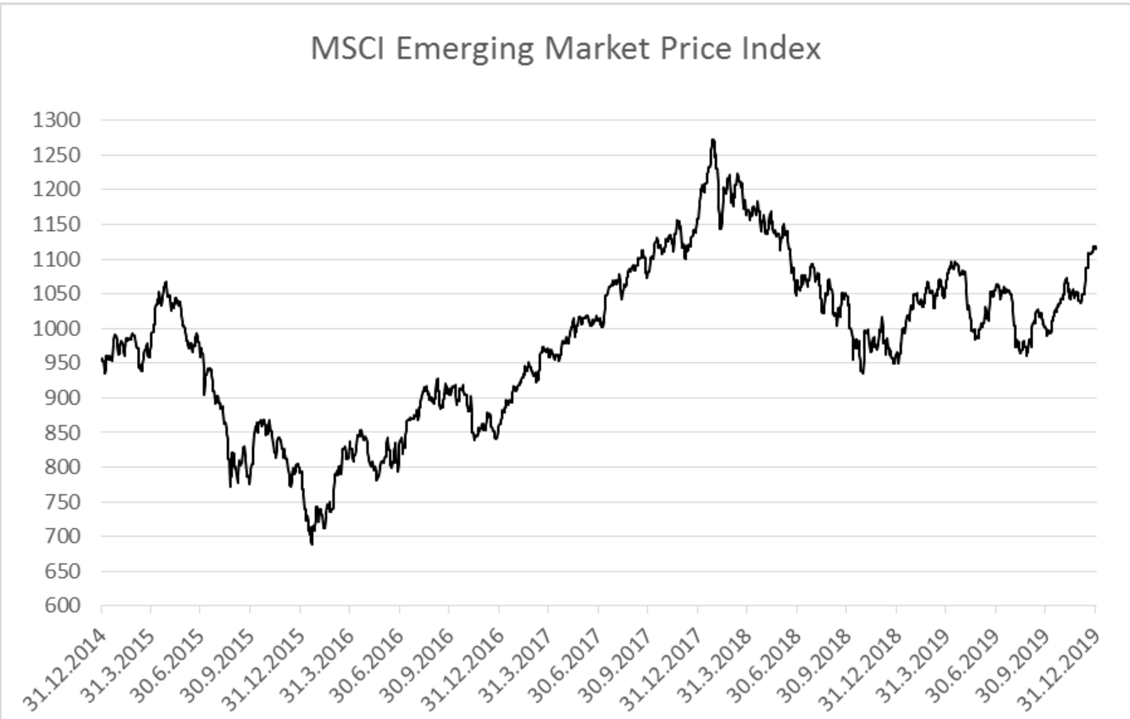


Figure 1. MSCI Emerging Market Price Index (Bloomberg).

The lowest daily return of the index was -5.00% on April 24th 2018. Highest daily return of 3.27% was experienced on August 27th 2015. During the time period of 1.1.2015–31.12.2019 there are 1305 days of which 687 days have a positive return and 618 days have a negative return. The daily returns are presented in Figure 2 and Figure 3 displays the distribution of returns.

As indicated by the number of days with positive and negative returns and Figure 2, it is evident that the returns are quite evenly distributed to positive and negative returns. Figure 3 further demonstrates the distribution of returns. As Figure 3 presents, the daily returns are distributed around a mean return of 0.02%.

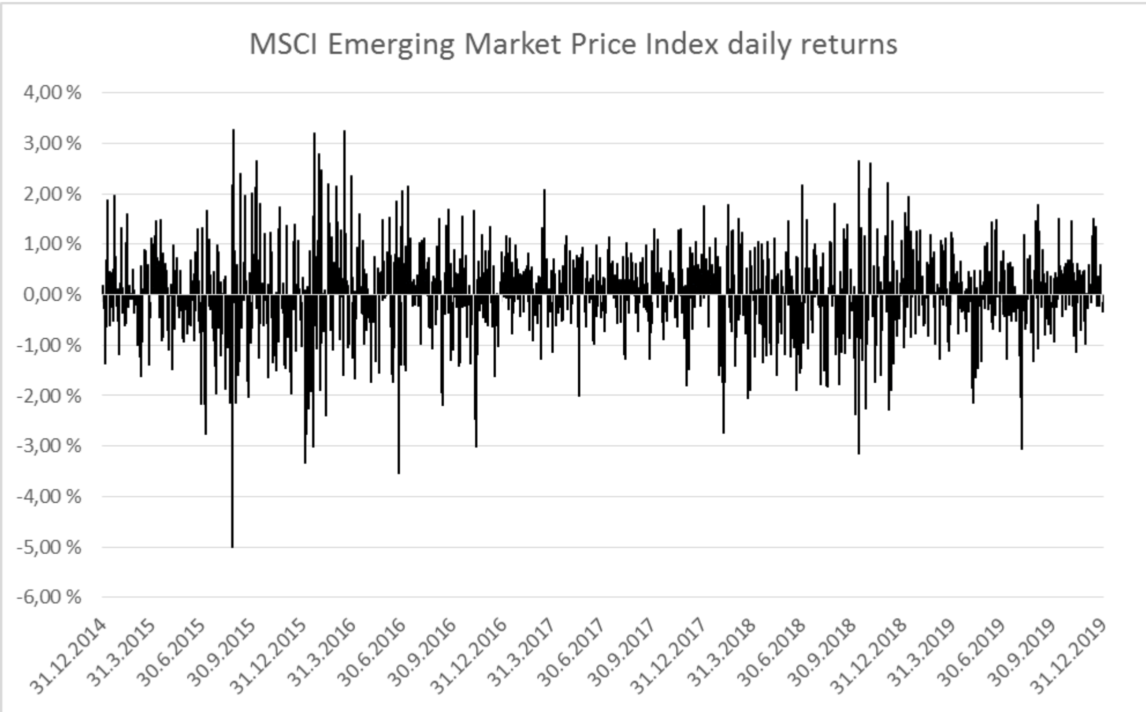


Figure 2. MSCI Emerging Market Price Index daily returns (Bloomberg).

Figure 3. Distribution of daily returns (Bloomberg).

The yearly returns of the index are presented in Table 1. As Table 1 shows, years 2016, 2017 and 2019 had positive returns and years 2015 and 2018 had negative returns. The most positive year was 2017 with the yearly return of 34.35%. This is also evident in Figure 1 that presents the index prices.

Table 1. Index yearly returns (Bloomberg).

Year	Return
2015	-16.81 %
2016	8.58 %
2017	34.35 %
2018	-16.63 %
2019	15.42 %

Notes: The yearly returns are calculated from the end-of-year closing prices.

During the whole sample period the MSCI Emerging Market Price Index's return is 16.77%. An annualised return for the sample period is 3.15%.

To calculate the option implied volatility of the MSCI Emerging Market Price Index, this thesis uses European call options on the index. The MSCI Emerging Market Price Index options used are December and June call options with strike price of USD 1100. The December and June options are rolled in order to avoid implied volatility spiking when the option reaches its maturity. December Call prices are used between December 31 and June 30, and June Call prices are used between June 31 and December 30. This creates a continuous daily option closing price data where options have between 169-353 days until expiration.

Figure 4 presents the rolled option prices for the sample period 1.1.2015–31.12.2019. The option's price in Figure 4 also reflects the option moneyness. In Figure 5 the strike price of USD 1100 is plotted against the index's spot price.

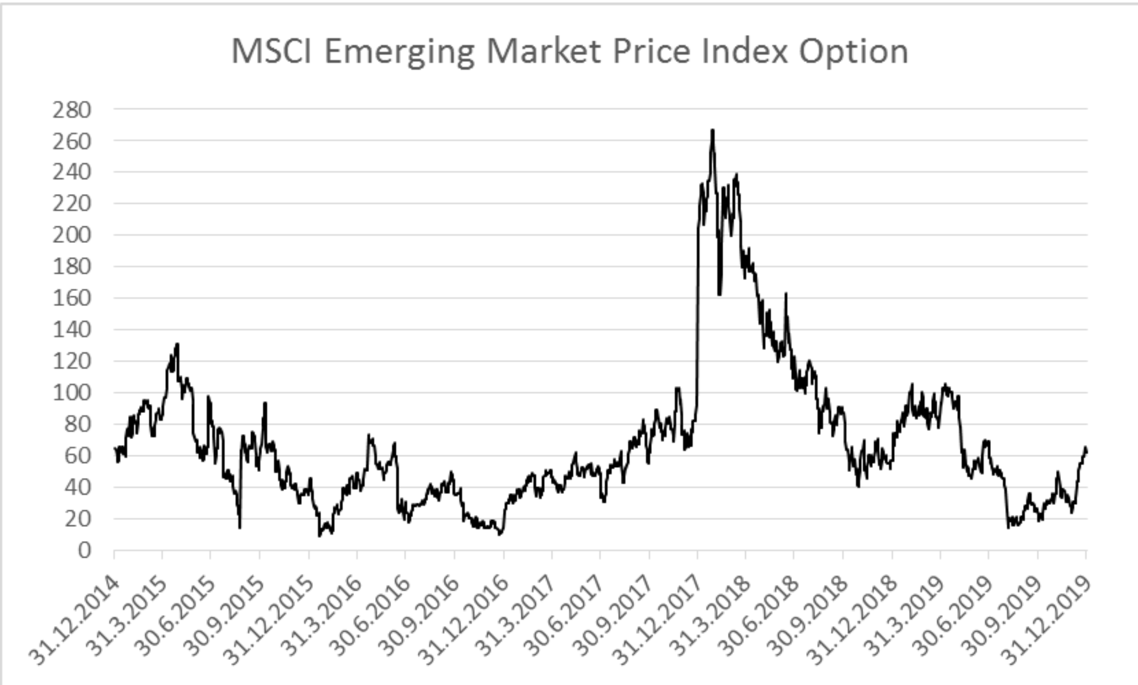


Figure 4. MSCI Emerging Market Price Index Option (Bloomberg).

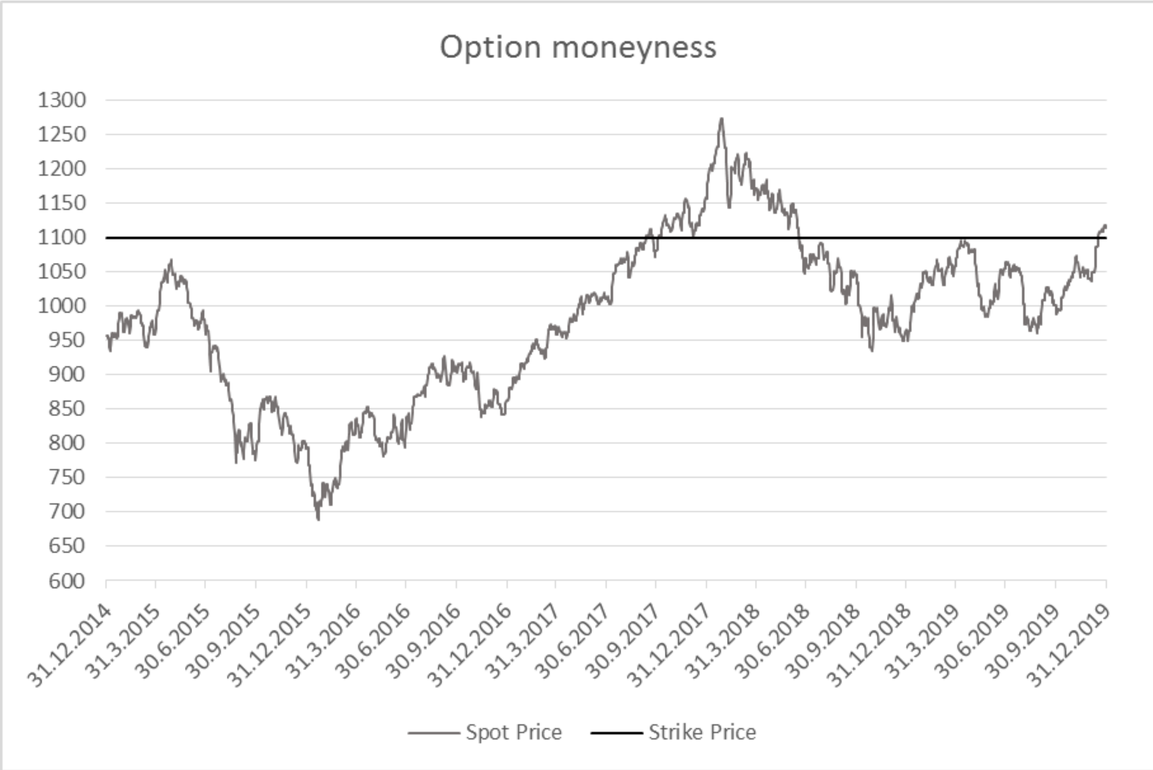


Figure 5. Option moneyness (Bloomberg).

When Figures 4 and 5 are compared, it is evident that option price in Figure 4 rises when the option is at-the-money and in-the-money. For most of the sample period the option is out-of-the-money. Figure 6 further demonstrates the linkage between option's moneyness and option price that is seen in Figures 4 and 5.

The vertical lines on Figure 6 demonstrate the time points when the option turns in-the-money. On September 12th 2017 the index reaches a price of USD 1102.26 which exceeds the strike price of USD 1100. Beginning from there the option is in-the-money until June 18th 2018. On December 17th 2019 the index price reaches USD 1102.62 and the option is in-the-money until the end of sample period. The spot price is USD 1114.66 at December 1st 2019. During other time points in the sample period the option is out-of-the-money when the index price is under USD 1100. (Bloomberg)

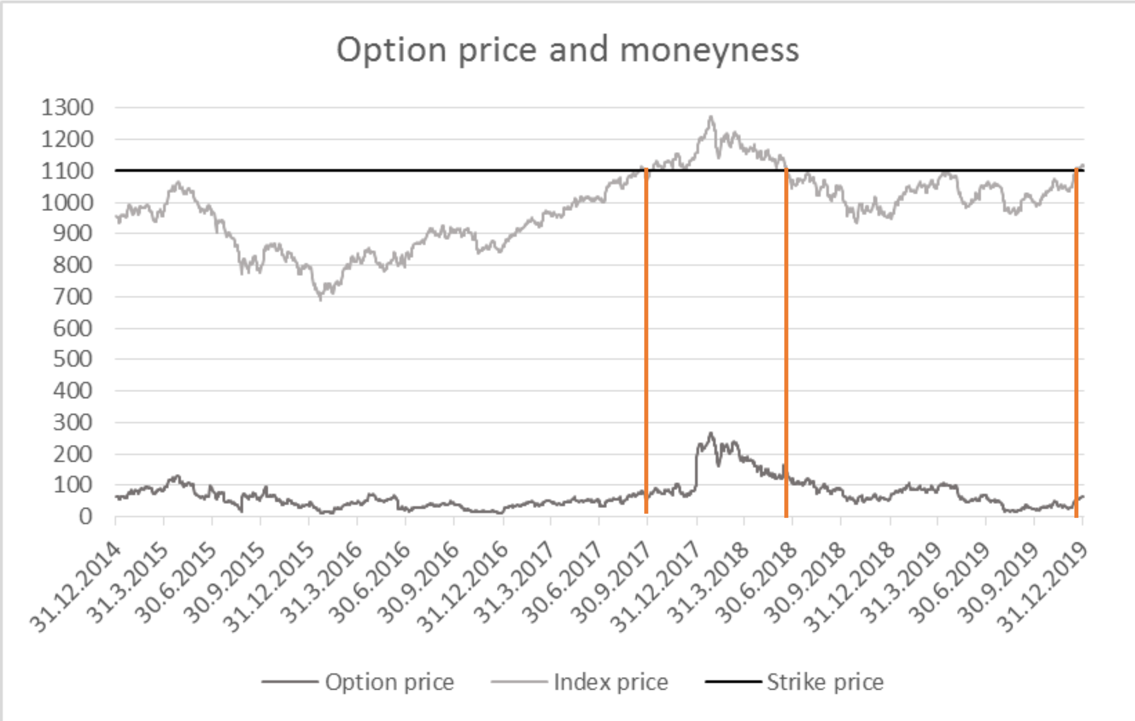


Figure 6. Option price and moneyness (Bloomberg).

The option price in Figure 6 seems to follow the option moneyness. When the index reaches its highest price of USD 1273.07 on January 26th 2018, the option also reaches its highest value of USD 266.5 during the sample period. That is the time point when the option is most in-the-money. On September 11th 2017 the index is very close to being at-the-money at USD 1099.18 and the option price experiences a jump of 11.41%.

Figure 7 further illustrates the difference between spot price and option strike price. As evident, the option is out-of-the-money for majority of the sample period. There are 202 days when the option is in-the-money and 1103 days when it is out-of-the-money.

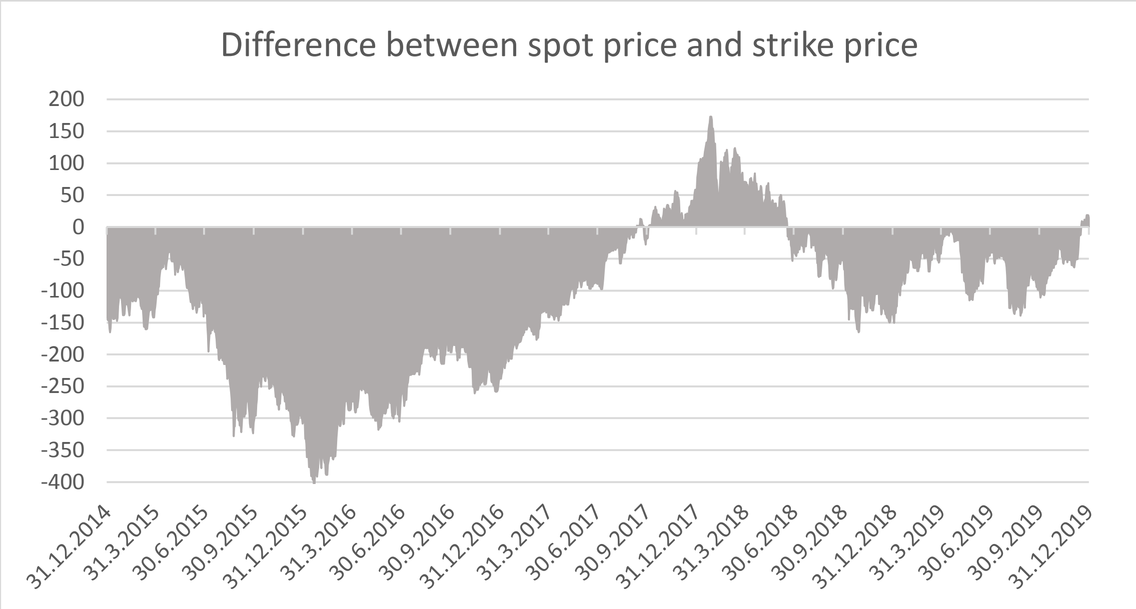


Figure 7. Difference between spot price and strike price (Bloomberg).

5.2 Methodology

The methodology of this thesis follows previous researches' methodology on volatility models' predicative power. Following the methods of Christensen and Prabhala (1998) and Dutta (2017), OLS regressions are used to measure the predicative power of implied volatility and GARCH(1,1) for one-day volatility and a monthly, 22-day volatility. Realised volatility, option implied volatility and GARCH(1,1) model's results are computed with Microsoft Excel. The OLS regressions and error terms are computed with EViews.

5.2.1 Measures of volatility

Following research by Parkinson (1980) a measure of Realised Volatility (RV) is computed using sample standard deviation of MSCI Emerging Market Price Index daily returns. As presented in chapter 2.2.1, the measure is calculated daily and monthly as follows:

$$RV_t = \sqrt{\frac{\sum_{i=1}^{T} \ln(h_i - l_i)^2}{4ln(2)}}$$

where RV_t is the index' realised volatility, $\sum_{i=1}^{T} \ln(h_i - l_i)^2$ is the sum of natural logarithm of the difference between highest and lowest price during the sample period and *T* is number of days in the sample period, which in this research is two for daily volatility and 22 for monthly volatility.

Following Christensen and Prabhala (1998), the option implied volatility is calculated using Black and Scholes (1973) option pricing formula for European call options. The option implied volatility is calculated daily for the MSCI Emerging Market Price Index Options as the option prices, underlying index price and risk-free rate are known. As presented in chapter 3.1.1, the Black-Scholes model the option implied volatility is calculated from call option formula as follows:

$$c_0 = S_0 N(d_1) - Ke^{-rT} N(d_2)$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

where c_0 is current call option value, S_0 is current stock price, $N(d_1)$ is the factor by which the present value of a random price of the stock exceeds the current stock price, $N(d_2)$ is the probability of the option being exercised, K is the option exercise price, *e* (Napier's constant) is a constant and the base of the natural logarithm function, *r* is the risk-free interest rate which is the 3-month T-bill rate, *T* is the time to option's expiration in years, *ln* is the natural logarithm function. The implied volatility is represented by σ, which is the standard deviation of the annualized and continuously compounded rate of return on the underlying stock:

$$IV_t = \sigma_t \tag{24}$$

where IV_t is the option implied volatility measure for day t and σ_t is the calculated value. The daily implied volatility is computed for the entire data period. The value calculated from the Black-Scholes option pricing formula provides an annualised volatility. In order to calculate the daily volatility, the value is divided with $\sqrt{252}$ as follows:

$$IV_{daily} = \sigma_{\mathrm{t}} / \sqrt{252} \tag{25}$$

$$IV_{monthly} = IV_{daily} * \sqrt{22} \tag{26}$$

The implied volatility measure is computed using a VBA macro (Learn365 Club, 2019) that loops Excel's goal seek function for each row:

```
Sub Goal_Seek_Range_MultipleGoal()
'Defining variable k
Dim k As Integer
'Looping through each row of the table
 For k = 2 To 1829
'Replicate the Goal Seek function via VBA
      Cells(k, "M").GoalSeek Goal:=Cells(k, "B"), Chang-
ingCell:=Cells(k, "F")
'Go to next iteration
 Next k
End Sub
```

Following Gokcan (2000), the GARCH(1,1) model is computed as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{23}$$

where σ_t^2 is the GARCH modelled variance (GV) at time t, ω is the long-run average variance rate, u_{t-i}^2 is the lagged time point *t-i* return and α_i is the weight assigned to each value in the sum, σ_{t-j}^2 is the lagged time point *t-j* variance term and β_j is the weight assigned to each value in the sum. The ω, *α* and *β* are positive constant parameters and *ω* + *α* + *β* = 1.

The daily and monthly GARCH(1,1) volatilities are then calculated with $t-i$ and $t-j$ being one-day lag and monthly forecast is computed by multiplying the daily volatility with $\sqrt{22}$.

$$\sigma_{daily} = \sqrt{\omega + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2} \tag{27}$$

$$\sigma_{monthly} = \sigma_{daily} * \sqrt{22} \tag{28}$$

Following the methods of Hull (2011, pp. 528), the coefficients *ω, α* and *β* are defined for the whole sample period using maximum likelihood estimation method. By maximising the likelihood of data occurring, the following equation is maximised using Excel Solver:

$$\sum_{i=1}^{m} \left[-\ln(v_i) - \frac{u_i^2}{v_i} \right] \tag{29}$$

where v_i is the GARCH(1,1) calculated variance, u_i is the index's daily return and m is number of observations occurring. The Excel Solver is used to maximise the GARCH(1,1) coefficients with restrictions that 0 ≥ *ω* ≥ 1, 0 ≥ *α* ≥ 1 and 0 ≥ *β* ≥ 1. The maximum likelihood estimation method provides the coefficient values in Table 2 that are used to calculate the GARCH(1,1) volatility values daily for the entire sample period:

Table 2. GARCH(1,1) coefficient values.

Coefficient	Value
ω	0.000000998
α	0.083
β	0.91

Note: The values of ω, α and β represent the values of maximum likelihood estimation of GARCH(1,1) parameters.

5.2.2 OLS Regressions

Following Dutta (2017), the OLS regressions to examine the predictive power of implied volatility (IV) and GARCH(1,1) modelled volatility (GV) for daily and monthly volatility forecasts are computed as follows:

$$RV_{t+1} = \alpha_0 + \beta_1 IV_t + \varepsilon_{t+1} \tag{30}$$

$$RV_{t+1} = \alpha_0 + \beta_1 GV_t + \varepsilon_{t+1} \tag{31}$$

where RV_{t+1} indicates the 1-day ahead realised volatility. The OLS models are tested for the entire time period of 1.1.2015–31.12.2019.

By testing the null hypothesis, whether $H_0 : \beta_1 = 0$, the results indicate if the volatility models include information on future realised volatility. If the coefficient β_1 is statistically different from zero, the results show evidence that implied volatility and GARCH(1,1) have significant predictive power over future emerging stock market volatility. (Dutta, 2017)

5.2.3 Error Terms (RMSE & MAE)

The examination of forecasting accuracy includes also the examination of residuals that are the difference between the forecasted value and the observed value. Following Dutta (2017), two error statistics, Root Mean Square Error (RMSE) and Mean Average Error (MAE), are computed to determine the error between the regression line and realised volatility.

The Root Mean Square error (RMSE) is the standard deviation of the residuals. It measures how far the residuals are from the regression line's data points when regression line is the best fit among the forecasted values. RMSE indicates the mean of squared differences between the actual volatility values and forecasted volatility values as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(RV_{a,t} - RV_{f,t})^2} \tag{32}$$

where $\sum_{i=1}^{N}(RV_{a,t} - RV_{f,t})^2$ is the sum of squarer between the forecasted values $RV_{f,t}$ and actual values $RV_{a,t}$ and N is the number of observations. (Barnston, 1992; Dutta, 2017)

Another error measure is the Mean Absolute Error (MAE), which measures the difference between two continuous variables. In this study, following Dutta (2017), MAE is used to measure the error between the observed volatility value and forecasted value. MAE is calculated for each predicted and realised value. In this case the error of the forecast is computed the following way:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|RV_{a,t} - RV_{f,t}\right| \tag{33}$$

where $\sum_{i=1}^{N}\left|RV_{a,t} - RV_{f,t}\right|$ is the sum of absolute difference between the actual value $RV_{a,t}$ and the forecasted value $RV_{f,t}$ and N is the number of observations. (Dutta, 2017; Wilmott & Matsuura, 2005)

6 Empirical results

This section of the thesis describes the results of implied volatility and GARCH(1,1) forecasts in emerging market environment. These volatility forecasting models have been widely studied in developed markets and previous results suggest that GARCH(1,1) offers better long-term forecasts than implied volatility. The previous results conclude that implied volatility has been the more used and superior model for a shorter 1–20 day forecasting period. However, there is a lack of consensus of the forecasting abilities of these model in emerging equity markets. Some previous studies have shown that implied volatility and GARCH(1,1) have significant information of future volatility while some have found the models to be biased and non-informative.

The next chapter presents an analysis of the data, including sample specific descriptive statistics. The following chapters present the forecasting results computed with the OLS regressions described in the previous section and the error measures of RMSE and MAE. There are four OLS regressions that present the forecasting accuracy of implied volatility and GARCH(1,1) for both one-day and 22-day forecasts. Regression results are compared to the research questions and hypotheses. The final chapter discusses the results critically and offers suggestions for future research topics.

6.1 Data analysis

The data that are used covers the daily prices of MSCI Emerging Market Price Index and call options of the index. Realised volatility is computed with the Parkinson (1980) range based formula, implied volatility is calculated from the Black-Scholes (1973) option pricing formula and GARCH(1,1) following Bollerslev's (1986) formula. Volatilities are measured with standard deviation and all values are presented as a one-day volatility and monthly, 22-day volatility.

The computed volatility time series covers observations from 1.1.2015 to 31.12.2019. Figures 8 and 9 present the calculated time series of daily and monthly values of realised volatility, implied volatility and GARCH(1,1) volatility during this time period. From these Figures it is evident that GARCH(1,1) volatility appears to be closer to the realised volatility values. Overall the volatility measures appear to follow a similar pattern.

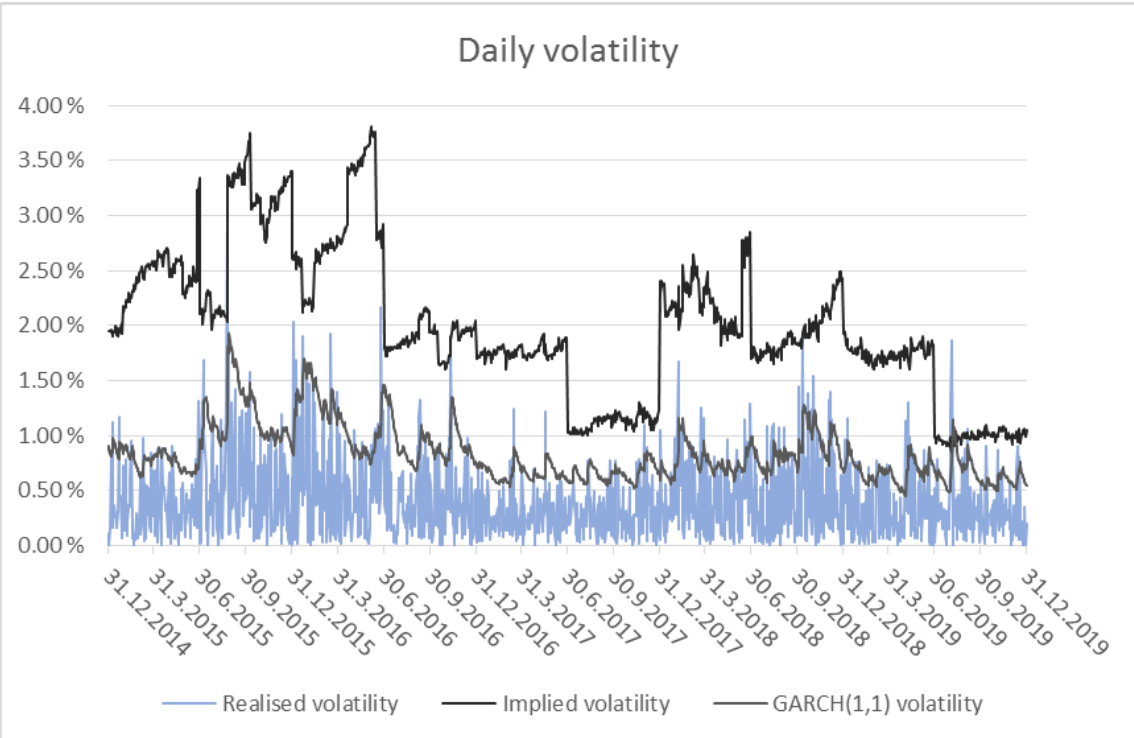


Figure 8. Daily volatilities during 1.1.2015–31.12.2019.

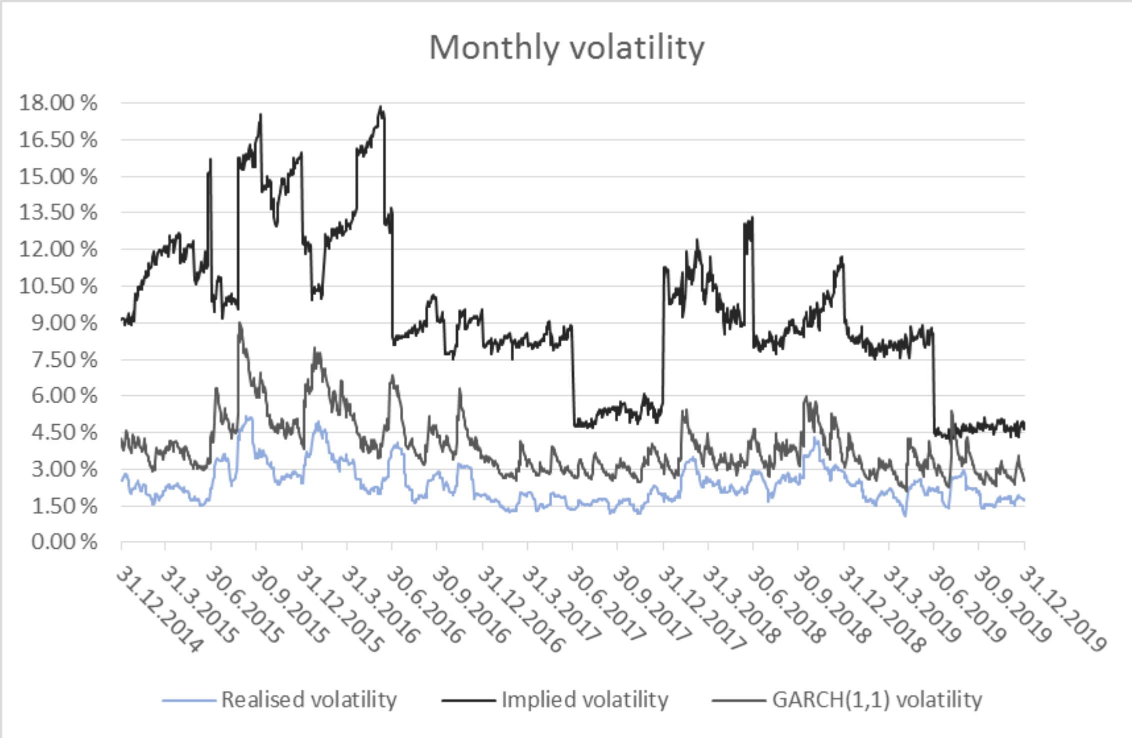


Figure 9. Monthly volatilities during 1.1.2015–31.12.2019.

The implied volatility also shows more visible jumps when compared to realised volatility and GARCH(1,1). Based on the time series, GARCH(1,1) daily volatility is on average 0.44% higher than realised volatility and on monthly level 1.53% higher. Implied volatility is 1.59% higher on daily level and 6.96% higher on monthly level.

The mean daily realised volatility is 0.40%, mean daily implied volatility is 1.99% and mean daily GARCH(1,1) volatility is 0.83%. The mean daily volatilities as well as skewness and kurtosis of the volatility distributions are presented in Panel A of Table 3 below. All daily volatility measures are skewed right, which indicates that the distribution has a tail on higher volatilities. The realised volatility skewness is 1.795 and GARCH(1,1) volatility skewness is 1.368 which suggests that these measures are more skewed to the right than implied volatility which has skewness of 0.460. Table 3 also presents kurtosis for all volatility measures. Realised volatility and GARCH(1,1) volatility show positive kurtosis with values of 4.997 and 1.888 correspondingly. This suggest that the distributions are leptokurtic with more values close to the mean and fat-tails indicating fluctuations. However, the GARCH(1,1) appears to be closer to normal distribution. The daily implied volatility has a negative kurtosis of -0.147 which indicates a more platykurtic distribution that has more values on a broader spectrum of the mean.

Table 3. Summary statistics.

Panel A. Daily forecasts

Model	*Mean*	*Skewness*	*Kurtosis*
Realised volatility	0.40 %	1.759	4.997
Implied volatility	1.99 %	0.460	-0.147
GARCH(1,1) volatility	0.83 %	1.368	1.888

Panel B. Monthly forecasts

Model	*Mean*	*Skewness*	*Kurtosis*
Realised volatility	2.83 %	1.095	1.121
Implied volatility	9.34 %	0.460	-0.147
GARCH(1,1) volatility	3.91 %	1.368	1.888

Notes: Panel A of Table 3 shows the descriptive statistics of daily volatility measures.
Panel B of Table 3 shows the descriptive statistics of monthly volatility measures.

The summary statistics of monthly values are presented in Panel B of Table 3. Realised volatility is on average 2.83% whereas implied volatility is 9.34% and GARCH(1,1) volatility is 3.91% on a monthly level. The skewness and kurtosis display similar values as in the daily volatility distributions. This further indicates that the realised volatility and GARCH(1,1) volatility are more closely similar distributions than implied volatility.

6.2 Regression results

An OLS regression analysis is performed for each volatility model. The OLS regression for implied volatility and GARCH(1,1) as explanatory variables are computed against realised volatility as a dependent variable as follows:

$$RV_{t+1} = \alpha_0 + \beta_1 IV_t + \varepsilon_{t+1}$$

$$RV_{t+1} = \alpha_0 + \beta_1 GV_t + \varepsilon_{t+1}$$

Both regressions are executed for daily and monthly values. The null hypothesis in all cases is that implied volatility or GARCH(1,1) do not have statistically significant informational content over realised volatility. The alternative hypotheses are that implied volatility and GARCH(1,1) do have predictive power over $t + 1$ realised volatility. The hypotheses are presented below in Table 4 in numeric format where H_1 is the alternative hypothesis for daily volatility and H_2 is the alternative hypothesis for monthly volatility:

Table 4. Hypotheses of the thesis.

Implied volatility forecasts	*GARCH(1,1) volatility forecasts*
H_0: $\beta_1 = 0$	H_0: $\beta_1 = 0$
H_1: *daily* $\beta_1 \neq 0$	H_1: *daily* $\beta_1 \neq 0$
H_2: *monthly* $\beta_1 \neq 0$	H_2: *monthly* $\beta_1 \neq 0$

Notes: The null and alternative hypotheses are further explained in chapter 1.2.

Testing of these hypotheses is executed with OLS regression analysis. The regression results are presented in Table 5. The Panel A includes the daily forecasting results and Panel B the monthly results. Coefficients, standard errors, F-statistics and a R^2 measure are presented for all regressions.

As presented in Panel A of Table 5, the regression of daily implied volatility as a $t + 1$ realised volatility forecast results in significant coefficient of β_1 at 1% level. The positive coefficient indicates a positive relationship between realised volatility and implied volatility. The t-statistic of β_1 is 7.0576 and the p-value is 0 which suggest the rejection of null-hypothesis as β_1 is statistically different from zero. However, the R^2 measure of daily implied volatility forecast is only 3.68% which indicates that the model explains only a small amount of $t + 1$ realised volatility.

The daily forecasting results of GARCH(1,1) volatility presented in Panel A of Table 5 are similar to the results given by the implied volatility forecast. Again, the β_1 coefficient is statistically significant at 1% level which indicates a significant positive relationship between GARCH(1,1) and $t + 1$ realised volatility. The t-statistic is 8.9046 and p-value is 0 which suggest the rejection of null hypothesis as β_1 is statistically different from zero. The R^2 of GARCH(1,1) volatility is slightly higher than implied volatility's at 5.74% which suggests that GARCH(1,1) model includes more information on one-day ahead forecast of realised volatility than the implied volatility model. However, the model only explains 5.74% of variation in realised volatility. The alternative hypothesis H_1 is accepted for both daily implied volatility and GARCH(1,1) volatility forecasts.

The Panel B of Table 5 presents the regression results of monthly volatility forecasts. The monthly volatility forecasts offer more explanatory results than daily forecasts. The monthly implied volatility has a positive β_1 coefficient that is significant at 1% level, which indicates a statistically significant positive relationship. With a t-statistic of 20.6726 and p-value of 0 the null hypothesis can be rejected as β_1 is statistically different from zero. The explanatory power of the forecast has improved from the daily level. The R^2 measure is 24.70% which indicates that the model explains over 24% of variation in $t + 1$ realised volatility.

Table 5. OLS regression results.

Panel A. ***Daily forecasts***

IV	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
α_0	0.001980***	0.0003	6.604541	0.0000
β_1	0.100699***	0.014267	7.057934	0.0000
R-squared	0.036823			
Number of observations	1305			
GARCH(1,1)	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
α_0	0.001144***	0.000333	3.43827	0.0006
β_1	0.340405***	0.038228	8.904615	0.0000
R-squared	0.057363			
Number of observations	1305			

Panel B. ***Monthly forecasts***

IV	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
α_0	0.012352***	0.000584	21.14105	0.0000
β_1	0.122547***	0.005928	20.67257	0.0000
R-squared	0.246975			
Number of observations	1305			
GARCH(1,1)	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
α_0	0.000065***	0.000317	0.206093	0.0004
β_1	0.606304***	0.007759	78.14380	0.0000
R-squared	0.824144			
Number of observations	1305			

Notes: Panel A of Table 5 shows the regression results of daily forecasts for IV and GARCH(1,1).
Panel B of Table 5 shows the regression results of monthly forecasts for IV and GARCH(1,1).
*The *, ** and *** refer to significance at 10%, 5% and 1% levels.*

The monthly GARCH(1,1) volatility offers even more promising results. The β_1 is again positive and significant at 1% level which suggests there is a positive and statistically significant relationship between the variables. The t-statistic of β_1 is 78.1438 with a p-value of 0 which indicates that the null hypothesis can be rejected at 1% level. The R^2 measure indicates that GARCH(1,1) volatility has an explanatory power of 82.41% over $t + 1$ realised volatility which suggests that the model is a highly suitable fit. The alternative hypothesis H_1 is accepted for both monthly implied volatility and GARCH(1,1) volatility forecasts.

A comparison of the regression results suggests that while the daily values of implied volatility and GARCH(1,1) volatility are statistically significant and unbiased estimators of one-day ahead future volatility, the models are lacking in explanatory power. The monthly volatility forecasts offer better explanatory power and appear to be well suited for 22-day volatility forecasts. On both daily and monthly levels the GARCH(1,1) model results in more accurate fitting forecast. The explanatory powers of daily forecasts do not suggest that there is a large difference in the appropriateness of the models. However, on a monthly level the GARCH(1,1) volatility dominates implied volatility in forecasting accuracy of realised volatility.

Overall, the regression results of implied volatility are similar to those reported by Blair et al. (2010) in US market context, although the daily forecast is weaker. In emerging market context the implied volatility shows weaker results than suggested by Bentes (2015). Especially as a daily forecast, the explanatory power of implied volatility calculated from MSCI Emerging Market index options is weaker than what previous studies reported. The monthly implied volatility forecast is similar to what Shaikh and Padhi (2015) reported on the Indian market.

The GARCH(1,1) results are quite similar to those reported by Bentes (2015) in Hong Kong, India and Korea. While both models show informational content over future volatility, the GARCH(1,1) model performs better than implied volatility. The results are opposing to those by Yang and Liu (2012) who suggest that an implied volatility index outperforms GARCH(1,1) as a monthly volatility forecast in Taiwanese stock market. However, the results of this thesis offer clarification of volatility model selection in emerging equity markets as GARCH(1,1) volatility appears to be the superior forecasting method on both daily and monthly volatilities.

6.3 Error measures

To further test the prediction performance of implied volatility and GARCH(1,1) volatility, two loss function measures, RMSE and MAE, are computed as described in the previous chapter. An in-sample estimation of these error measures are produced for both model in daily and monthly volatility forecasts at time $t + 1$. The results of the loss functions are documented in Table 6 where Panel A presents the values for daily forecasts and Panel B for monthly forecasts.

Table 6. Error measures of the models.

Panel A. ***Daily forecasts***

Model	*RMSE*	*MAE*
IV	0.003465	0.002589
GARCH(1,1)	0.003428	0.002572

Panel B. ***Monthly forecasts***

Model	*RMSE*	*MAE*
IV	0.006752	0.005230
GARCH(1,1)	0.003263	0.002545

Notes: Panel A of Table 6 shows the RMSE and MAE loss function values of daily forecasts for IV and GARCH(1,1). Panel B of Table 6 shows the RMSE and MAE loss function values of monthly forecasts for IV and GARCH(1,1).

The results for daily forecasts show similar values to both model. The RMSE for implied volatility is 0.0035 and for GARCH(1,1) volatility 0.0034. The MAE measures are 0.0026 for both models. These values of error measures suggest that both implied volatility and GARCH(1,1) volatility are well fitted to describe realised volatility. The daily GARCH(1,1) volatility has a slightly lower RMSE and MAE value which indicates that it is more accurate than implied volatility. This is consistent with the OLS regression results.

The results in Panel B display the RMSE and MAE for monthly volatilities. Implied volatility has a RMSE of 0.0068 while GARCH(1,1) volatility has a value of 0.0033. The MAE for monthly implied volatility is 0.0052 and 0.0025 for GARCH(1,1) volatility. Consistent with the OLS regression results, the difference in fit of the models is clearer in monthly volatility forecasts than in daily volatility forecasts. Both loss function results indicate that GARCH(1,1) offers a better estimate of future volatility.

6.4 Criticism of results and further studies

The results presented in this thesis offer clarifying results to previous studies varying conclusions. The data period length is sufficient and the observations are relevant and include recent. The use of MSCI Emerging Market Index includes an extensive amount of emerging market countries making it a good sample of emerging equity markets. While it offers interesting results, it is also difficult to conclude whether these results are consistent in single emerging countries or for a single asset.

The accuracy of implied volatility is also dependent on option moneyness. The option was chosen for being the closest to at-the-money, however, it is still out-of-the-money for majority of the forecasting period. The estimates provided by implied volatility could be improved if a more at-the-money option was available. While the GARCH(1,1) volatility provides a more accurate forecast for both daily and monthly values, an out-of-sample forecasting period should be tested to further validate the results. Out-of-sample testing, testing on different emerging market economies and adjusting the option selection is left to further research.

A topic for further research is also the comparison of growing research in emerging market context to the existing literature of developed market volatility. Although implied volatility seems to be a popular volatility forecaster, more recent evidence in emerging market research and this thesis suggests that a well-fitted GARCH model is able to provide a more accurate future volatility forecast. A combination model could also provide new informational content on emerging market volatility.

7 Conclusions

The ability to accurately forecast volatility is an evolving field of study in finance as volatility is a key feature in investing and management of risk. Previous studies have shown that implied volatility and GARCH based models have dominated volatility forecasting both in terms of forecasting accuracy and popularity in using the models. In developed market environment, the previous results suggest that implied volatility is an accurate short-term forecaster and GARCH models offer a good long-term future forecast.

However, previous studies on volatility forecasting in emerging market environment have been inconclusive in terms of forecasting accuracy and model selection. Emerging equity markets experience more risks than developed equity markets. These risks arise from economic and political uncertainties that are more present in emerging than developed markets. This makes volatility forecasting in emerging equity markets an interesting field of study, since the significance of emerging economies has grown as financial markets are more globalised than ever.

This thesis examined the forecasting accuracy of two models, implied volatility and GARCH(1,1) in the context of emerging equity markets. MSCI Emerging Market Price index and an index option were used to calculate implied volatility and GARCH(1,1) volatility forecasts for the time period of 1.1.2015–31.12.2019. A one-day forecast was calculated for both models in terms of daily and monthly volatility. A regression analysis was computed in order to determine whether implied volatility and GARCH(1,1) volatility contain information of future volatility in emerging markets. Error terms were also computed in order to assess the fitness of both models.

The results indicate that both daily and monthly implied volatility and GARCH(1,1) volatility contain significant information about one-day ahead future volatility. However, the predictive power of monthly values is higher than daily values for both models. The results suggest that in both daily and monthly values GARCH(1,1) volatility is a more accurate estimate for future volatility. The GARCH(1,1) monthly volatility offers the best fit for future volatility with the highest predictive power and lowest error measures, suggesting that it is the most appropriate fit for future volatility forecasting in emerging equity markets.

The results presented in this thesis contribute to the study of volatility forecasting in emerging equity markets. The GARCH(1,1) model offers the most accurate future volatility estimate and offers support to some of the existing studies in emerging market context. The effects of option moneyness and out-of-sample testing is left for further research.

References

Abken P.A. & Nandi, S. (1996). Options and volatility. *Economic Review* 81:3–6, 21.

Adcock, C. J., & Shutes, K. (2005). An analysis of skewness and skewness persistence in three emerging markets. *Emerging Markets Review*, 6(4), 396-418.

Ait-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The review of financial studies*, 18(2), 351-416.

Akgiray, V. (1989). Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *Journal of business*, 55-80.

Alam, M. Z., Siddikee, M. N., & Masukujjaman, M. (2013). Forecasting volatility of stock indices with ARCH model. *International Journal of Financial Research*, 4(2), 126.

Alford, A. W., & Boatsman, J. R. (1995). Predicting long-term stock return volatility: Implications for accounting and valuation of equity derivatives. *Accounting Review*, 599-618.

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

Andersen, T. G., Bollerslev, T., & Meddahi, N. (2004). Analytical evaluation of volatility forecasts. *International Economic Review*, 45(4), 1079-1110.

B.J. Christensen & Prabhala, N.R. (1998). The relation between implied and realised volatility. *Journal of Financial Economics* 50, 125-150.

Balaban, E., Bayar, A., & Faff, R. W. (2006). Forecasting stock market volatility: Further international evidence. *The European Journal of Finance*, 12(2), 171-188.

Barnston, A., (1992). "Correspondence among the Correlation [root mean square error] and Heidke Verification Measures; Refinement of the Heidke Score." *Notes and Correspondence*, Climate Analysis Center.

Bartunek, K. S., & Chowdhury, M. (1995). Implied volatility vs. GARCH: a comparison of forecasts. *Managerial Finance*.

Bekaert, G., Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1998). Distributional characteristics of emerging market returns and asset allocation. *Journal of portfolio man agement*, 24(2), 102-+.

Bentes, S. R. (2015). A comparative analysis of the predictive power of implied volatility indices and GARCH forecasted volatility. *Physica A: Statistical Mechanics and its Applications*, 424, 105-112.

Bera, A. K., & Higgins, M. L. (1997). ARCH and bilinearity as competing models for nonlinear dependence. *Journal of Business & Economic Statistics*, 15(1), 43-50.

Bilson, C. M., Brailsford, T. J., & Hooper, V. C. (2002). The explanatory power of political risk in emerging markets. *International Review of Financial Analysis*, 11(1), 1-27.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3), 637-654.

Blair, B. J., Poon, S. H., & Taylor, S. J. (2001). Modelling S&P 100 volatility: The information content of stock returns. *Journal of banking & finance*, 25(9), 1665-1679.

Blair, B., Poon, S. H., & Taylor, S. J. (2002). Asymmetric and crash effects in stock volatility for the S&P 100 index and its constituents. *Applied Financial Economics*, 12(5), 319-329.

Bley, J., & Saad, M. (2015). Idiosyncratic volatility forecasting in the stock market of Saudi Arabia. *Emerging Markets Finance and Trade*, 51(6), 1342-1357.

Bloomberg database (2020).

Bollen, N. P., & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions?. *The Journal of Finance*, 59(2), 711-753.

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

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1-2), 5-59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. *Handbook of econometrics*, 4, 2959-3038.

Brenner, M., & Subrahmanyan, M. G. (1988). A simple formula to compute the implied standard deviation. *Financial Analysts Journal*, 44(5), 80-83.

Brous, P., Ince, U., & Popova, I. (2010). Volatility forecasting and liquidity: Evidence from individual stocks. *Journal of Derivatives & Hedge Funds*, 16(2), 144-159.

Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realised volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics*, 160(1), 48-57.

Canina, L., & Figlewski, S. (1993). The informational content of implied volatility. *The Review of Financial Studies*, 6(3), 659-681.

Carvalho, D. (2014). The real effects of government-owned banks: Evidence from an emerging market. *The Journal of Finance*, 69(2), 577. Retrieved from https://search-proquest- com.proxy.uwasa.fi/docview/1510781006?accountid=14797

Chan, W. H., Cheng, X., & Fung, J. K. (2010). Forecasting volatility: Roles of sampling frequency and forecasting horizon. *Journal of Futures Markets*, 30(12), 1167-1191.

Chan, W. H., Jha, R., & Kalimipalli, M. (2009). The economic value of using realised volatility in forecasting future implied volatility. *Journal of Financial Research*, 32(3), 231-259.

Chang, R., & Velasco, A. (2001). A model of financial crises in emerging markets. *The Quarterly Journal of Economics*, 116(2), 489-517.

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realised volatility. *Journal of financial economics*, 50(2), 125-150.

Corrado, C. J., & Miller Jr, T. W. (1996). A note on a simple, accurate formula to compute implied standard deviations. *Journal of Banking & Finance,* 20(3), 595-603.

Das, U. S., Papaioannou, M. G., & Trebesch, C. (2009). Sovereign default risk and private sector access to capital in emerging markets. *Debt Relief and Beyond. Lessons Learned and Challenges Ahead*, 141-80.

Day, T. E., & Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52(1-2), 267-287.

Dutta, A. (2017). Modeling and forecasting oil price risk: The role of implied volatility index. *Journal of Economic Studies*, 44(6), pp. 1003-1016. doi:10.1108/JES-11-2016-0218

Easterly, W., Islam, R., & Stiglitz, J. E. (2001). Volatility and Macroeconomic Paradigms for Rich and Poor. *Advances in Macroeconomic Theory* (pp. 352-372). Palgrave Macmillan, London.

Ederington, L. H. (2002). Why are those options smiling?. *The Journal of Derivatives*, 10(2), 9-34.

Ederington, L. H., & Guan, W. (2005). Forecasting volatility. *Journal of Futures Markets*: Futures, Options, and Other Derivative Products, 25(5), 465-490.

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

Engle, R. F., & Mustafa, C. (1992). Implied ARCH models from options prices. *Journal of Econometrics*, 52(1-2), 289-311.

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

Engle, R. F., & Patton, A. J. (2007). What good is a volatility model?. *Forecasting volatility in the financial markets* (pp. 47-63). Butterworth-Heinemann.

Fernandes, N. (2011). Global convergence of financing policies: Evidence for emerging-market firms. *Journal of International Business Studies*, 42(8), 1043-1059. Retrieved from: doi: http://dx.doi.org.proxy.uwasa.fi/10.1057/jibs.2011.27

Figlewski, S. (1994). Forecasting volatility using historical data. *NYU Working Paper* No. FIN-94-032.

Figlewski, S. (1997). Forecasting volatility. *Financial markets, institutions & instru ments*, 6(1), 1-88.

Filis, G. (2009). An analysis between implied and realised volatility in the Greek derivative market. *Journal of Emerging Market Finance*, 8(3), 251-263.

Fleming, J. (1998). The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of empirical finance*, 5(4), 317-345.

Franses, P. H., & Ghijsels, H. (1999). Additive outliers, GARCH and forecasting volatility. *International Journal of forecasting*, 15(1), 1-9.

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

Gokcan, S. (2000). Forecasting volatility of emerging stock markets: Linear versus non-linear GARCH models. *Journal of Forecasting*, 19(6), 499-504. Retrieved from https://search-proquest- com.proxy.uwasa.fi/docview/219165913?ac countid=14797.

Gwilym, O. A., & Buckle, M. (1999). Volatility forecasting in the framework of the option expiry cycle. *The European Journal of Finance*, 5(1), 73-94.

Han, H., & Park, M. D. (2013). Comparison of realised measure and implied volatility in forecasting volatility. *Journal of Forecasting*, 32(6), 522-533.

Han, H., & Park, M. D. (2013). Comparison of realised measure and implied volatility in forecasting volatility. *Journal of Forecasting*, 32(6), 522-533.

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

Hansen, Peter R., and Asger Lunde. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?. *Journal of applied econometrics*
20, 873-889.

Hasan, I., Jackowicz, K., Kowalewski, O., & Kozłowski, Ł. (2017). Do local banking market structures matter for SME financing and performance? New evidence from an emerging economy. *Journal of Banking & Finance*, 79, 142-158.

Huang, A. Y. (2011). Volatility forecasting in emerging markets with application of stochastic volatility model. *Applied Financial Economics*, 21(9), 665-681.

Hull, J. C. (2015). *Options, futures, and other derivatives* (9th ed.). Harlow: Pearson Education.

Hull, J. C. (2015). *Risk management and financial institutions : Risk management and financial institutions* (4th edition). ProQuest Ebook Central https://ebookcentral-proquest-com.proxy.uwasa.fi

Jackwerth, J. C., & Rubinstein, M. (1996). Recovering probability distributions from option prices. *The Journal of Finance*, 51(5), 1611-1631.

John C. Hull (2011). *Options, Futures and Other Derivatives*. 9th Edition. Pearson. pp. 302–315, 321–343, 431–442, 521–540, 624–650. ISBN 978-0133456318.

Jun, S. G., Marathe, A., & Shawky, H. A. (2003). Liquidity and stock returns in emerging equity markets. *Emerging Markets Review*, 4(1), 1-24.

Khan, S. A. R., Sharif, A., Golpîra, H. & Kumar, A. (2019). A green ideology in Asian emerging economies: From environmental policy and sustainable develop ment. *Sustainable Development*, 27(6), pp. 1063-1075.

Klomp, J., & De Haan, J. (2014). Bank regulation, the quality of institutions, and banking risk in emerging and developing countries: an empirical analysis. *Emerging Markets Finance and Trade*, 50(6), 19-40.

Knoop, T. (2013). *Global finance in emerging market economies*. ProQuest Ebook Central https://ebookcentral-proquest-com.proxy.uwasa.fi

LearnClub 365 (2019). Goal Seek Analysis for Multiple Cells (Excel VBA). https://www.youtube.com/watch?v=D7Hi6rrYfD4

Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of financial eco nomics*, 77(2), 411-452.

Li, S. (2005). A new formula for computing implied volatility. *Applied mathematics and computation*, 170(1), 611-625.

Lischewski, J., & Voronkova, S. (2012). Size, value and liquidity. Do they really matter on an emerging stock market?. *Emerging Markets Review*, 13(1), 8-25.

Liu, Y. F., Zhang, W., & Xu, H. C. (2014). Collective behavior and options volatility smile: An agent-based explanation. *Economic Modelling*, 39, 232-239.

Mandelbrot, B. B. (2009). "New Methods of Statistical Economics," revisited: Short versus long tails and Gaussian versus power-law distributions. *Complexity*, 14(3), 55-65.

Mayhew, S., & Stivers, C. (2003). Stock return dynamics, option volume, and the information content of implied volatility. *Journal of Futures Markets*: Futures, Options, and Other Derivative Products, 23(7), 615-646.

Miah, M., & Rahman, A. (2016). Modelling Volatility of Daily Stock Returns: Is GARCH (1, 1) Enough?. *American Scientific Research Journal for Engineering, Technology, and Sciences* (ASRJETS), 18(1), 29-39.

MSCI (2019). https://www.msci.com/our-solutions/index/emerging-markets.

Nelson, D. B., & Foster, D. P. (1995). Filtering and forecasting with misspecified ARCH models II: making the right forecast with the wrong model. *Journal of Econometrics*, 67(2), 303-335.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of business*, 61-65.

Pati, P. C., Barai, P., & Rajib, P. (2018). Forecasting stock market volatility and information content of implied volatility index. *Applied Economics*, 50(23), 2552-2568.

Poon, S. H., & Granger, C. (2005). Practical issues in forecasting volatility. *Financial analysts' journal*, 61(1), 45-56.

Poon, S. H., & Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2), 478-539.

Poon, Ser-Huang, and Clive WJ Granger. "Forecasting volatility in financial markets: A review." *Journal of economic literature* 41.2 (2003): 478-539.

Prasad, E. S. (2010). Financial sector regulation and reforms in emerging markets: An overview (No. w16428). *National Bureau of Economic Research*.

Pu, W., Chen, Y. & Ma, F. (2016). Forecasting the realised volatility in the Chinese stock market: Further evidence. *Applied economics*, 48(33), pp. 3116-3130. doi:10.1080/00036846.2015.1136394

Sarwar, G., & Khan, W. (2017). The effect of US stock market uncertainty on emerging market returns. *Emerging Markets Finance and Trade*, 53(8), 1796-1811.

Shaikh, I., & Padhi, P. (2015). On the Relationship of Ex-ante and Ex-post Volatility: A Sub-period Analysis of S&P CNX Nifty Index Options. *Journal of Emerging Market Finance*, 14(2), 140-175.

Sherwood, M. W., & Pollard, J. L. (2018). The risk-adjusted return potential of integrating ESG strategies into emerging market equities. *Journal of Sustainable Finance & Investment*, 8(1), 26-44.

Suley Gokcan 2000: Forecasting volatility of emerging stock markets: linear versus non-linear GARCH models, *Journal of Forecasting*, 19:6, pp. 499-504.

Taylor, S. J., Yadav, P. K., & Zhang, Y. (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance*, 34(4), 871-881.

Wang, K. (2010). Forecasting volatilities in equity, bond and money markets: A market-based approach. *Australian Journal of Management*, 35(2), 165-180.

Willmott, Cort J.; Matsuura, Kenji (December 19, 2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance". *Climate Research*. 30: 79–82. doi:10.3354/cr030079.

Yang, M. J., & Liu, M. Y. (2012). The forecasting power of the volatility index in emerging markets: Evidence from the Taiwan stock market. *International Journal of Economics and Finance*, 4(2), 217-231.

Yu, J. (2002). Forecasting volatility in the New Zealand stock market. *Applied Financial Economics*, 12(3), 193-202.

Zhang, Y., Wu, H., & Cheng, L. (2012, June). Some new deformation formulas about variance and covariance. In 2012 *Proceedings of International Conference on Modelling, Identification and Control* (pp. 987-992)

Zvi Bodie, Kane, Alex & Marcus, Alan J. (2014). *Investments*. 10th Edition. New York: McGraw-Hill Education. pp: 737–743. ISBN 978-0-07-786167-4.