



**LUND UNIVERSITY**  
School of Economics and Management

# Forecasting Volatility

- A Comparison Study of Model Based Forecasts and Implied Volatility

---

Course: Master thesis

Supervisor: Anders Vilhelmsson

Authors: Bujar Bunjaku 850803

Armin Näsholm 870319

## **Abstract**

- Title:** Forecasting Volatility – A Comparison Study of Model Based Forecasts and Implied Volatility
- Course:** Master Thesis – Finance Program
- Authors:** Bujar Bunjaku  
Armin Näsholm
- Supervisor:** Ph.D Anders Vilhelmsson  
Department of Economics, Lund University
- Keywords:** Volatility forecast, ARCH-family, ARMA, implied volatility, realized volatility, evaluation models
- Purpose:** The purpose is to investigate which of the selected models that forecasts the out-of-sample data most accurate and whether the model based estimators make better forecasts than the implied volatility.
- Methodology:** Trough in-sample data from a Swedish stock index return series and a exchange rate return series, different forecasting models are evaluated to see which one that predicts the out-of-sample realized volatility most accurate. The data is forecasted under different distribution assumptions and then evaluated against each other and the implied volatility.
- Results:** Trough this thesis, it can be concluded that the asymmetric EGARCH under general error distribution and under normal distribution most accurately describes the stock index return series and the exchange rate return series respectively. It can also be concluded that the implied volatility does not predict the volatility more accurate than the model based forecasts.

## Contents

1	Introduction .....	5
1.1	Problem discussion.....	6
1.2	Purpose.....	6
1.3	Target Audience .....	6
1.4	Outline .....	7
2	Prior research.....	9
3	Stylized facts of the financial time series.....	10
4	Theory .....	12
4.1	ARMA.....	12
4.2	ARCH .....	13
4.2.1	GARCH .....	14
4.2.2	EGARCH .....	15
4.2.3	Maximum likelihood .....	15
4.2.4	Properties when forecasting with GARCH .....	16
4.2.5	Test for ARCH-effects .....	16
4.2.6	Test for asymmetries in volatility.....	17
4.3	Jarque-Bera test for non-normality.....	18
4.4	Implied Volatility .....	18
4.4.1	Black-Scholes model and Implied Volatility .....	18
4.4.2	Volatility Smile and Implied Distribution.....	19
4.4.3	Market efficiency.....	22
4.4.4	Volatility term structure and volatility risk premium .....	22
4.5	Realized volatility .....	23
4.6	Evaluation models.....	24
5.	Method .....	27
5.1	Data and sampling procedure model based forecasts .....	27
5.1.1	In-, and out-of-sample length .....	27
5.1.2	Treatment of raw data .....	28
5.2	Data and sampling procedure Implied Volatility .....	28
5.3	ATM or Weighted Implied volatility.....	29
5.4	Building ARMA models.....	30
5.5	Decision of the most accurate ARMA models.....	30
5.6	Diagnostic checking of ARMA models .....	31
5.7	GARCH-model .....	32
5.8	Problems with ARCH-family models.....	33

5.9 Forecasting procedure .....	33
5.10 Calculating realized volatility.....	34
5.11 Evaluating with loss functions.....	34
6 Results.....	35
6.1 Descriptive statistics of data .....	35
6.2 Parameter estimates .....	36
6.3 ARMA.....	38
6.3.1 Descriptive statistics for chosen ARMA models.....	38
6.4 GARCH.....	39
6.4.1 Test for ARCH effects.....	39
6.4.2 Test for asymmetry in volatility .....	39
6.4.3 Descriptive statistics chosen ARCH-family models .....	40
6.5 Comparing forecasting performance .....	41
6.5.1 Evaluation models for OMXS30 .....	41
6.5.2 Evaluation models USD/EURO .....	43
6.6 Implied Volatility versus model based forecasts .....	45
7 Conclusion .....	47
8 References .....	49
9 Appendix.....	54
ARMA(1,1) without a constant.....	54
ARMA(2,2) with a constant.....	55
GARCH(1,1) normal distribution.....	56
GARCH(1,1) t-distribution .....	57
GARCH(1,1) General error distribution .....	57
EGARCH .....	58

## 1 Introduction

The uncertainty of asset returns has for a long time captured the interest of speculating investors and academic researchers. Volatility is an important topic to anyone involved in financial markets, and it has recently become an issue of great interest to academics. High volatility means deviations from the mean and deviation implies risk (Figlewski, 1997).

Volatility is the only unknown variable in Black and Scholes (1973) options pricing model and therefore it has to be forecasted. An options' value today depends, in part, on the volatility that will occur over the time until it is exercised or expired. This implied volatility has a central role in determining fair value in option pricing models (Hull, 2006). Even portfolio management models, such as Sharp (1964) and Linter (1965) CAPM, that are based on 'mean variance' theory has volatility as an important factor. Volatility is an important factor even in risk management where Value-at-Risk (VaR) models are common. Realized volatility, the volatility that will occur from now into a specific time in the future, can easily be computed from historical data. It can also be forecasted using 'model based' forecasting models such as ARCH family models (Zumbach, 2009).

When B-S was introduced in the 1970s there were only a few types options traded. These had short maturity dates, of only a few months, which made it easy to predict volatility. Over short horizons volatility is assumed to remain constant. However, the derivative market has since developed and contracts are now available with longer maturity periods. It has been more difficult to accurately determine the value of such products, as volatility tends to change over longer periods of time. This has brought about the need for more complex forecasting models which account for variation in volatility (Figlewski, 1997).

The derivative market has grown rapidly over recent years and today, an enormous amount of derivatives exist. Options are the most common and actively traded derivatives within the market. An option gives the right to buy 'go long', or to sell 'go short' on an underlying asset at a certain price and time to maturity (Hull, 2006). The options' 'implied volatility' is the volatility that is expected by market participants during its life. Implied volatility is therefore likely to be a good predictor of the future volatility. Implied volatility is calculated by backing it out from Black-Scholes option pricing model, which is based on an assumption of constant volatility. It is now established, both in practice and in theory, that volatility is not constant

(Poon, 2005).

## **1.1 Problem discussion**

This thesis evaluates different methods of forecasting volatility to get as good a comparison as is possible. There are several crucial factors that have to be discussed. The different factors have made prior research ambiguous and there are different conclusions dependent on how the different sample periods, forecast horizon and ex post variance, are chosen. The factors might be distribution assumptions, asymmetric effects in the data, serial correlation and heteroskedasticity. The models are tested and analyzed for the mentioned factors to produce accurate forecasting estimates of volatility.

Models of the ARCH-family have been proven to make good estimates of future variance, while some authors states that implied volatility is better due to the market's expectation of the future volatility is taken into consideration. There are plenty of ARCH-family models that could explain the future volatility, and here we test some of the models recommended in previous papers, such as GARCH(1,1), EGARCH, ARMA-models, and implied volatility. To be able to make the comparison between the models, it is important to have an accurate estimate of the true variance in the out-of-sample period. To investigate the phenomena of non-normal distribution in financial time-series data, the models are also analyzed under different distribution assumptions.

## **1.2 Purpose**

The purpose is to investigate which of the selected models that forecasts the out-of-sample data most accurate and whether the model based estimators make better forecasts than the implied volatility.

## **1.3 Target Audience**

This thesis is aimed mainly to students at Master level and with an interest in financial economics. The reader should have some basic knowledge of financial markets, derivative securities, such as Black-Scholes option pricing theory, and volatility methods like ARMA and ARCH-family models. The subject could also be of interest for financial institutions,

banks and risk managers whose prior interest lies in finding good models to predict volatility for their instruments.

## 1.4 Outline

Following the introductory part is section 2 with a description of previous research in the forecasting subject. It describes how different model based prediction measurements has developed and that the different results might be caused by that different authors test the models from different distribution assumptions and sample length. Section 3 describes some key definitions and empirical findings in time series data are stated. Section 4 is a theory section which starts with a description of some basic model based forecasts. Some of them are not used as forecasting models in this thesis, yet they are still important factors in understanding why the chosen models look like they do.

To be able to analyze the data at hand, section 4.2.5 until 4.3 defines and shows the calculations behind some different tests that are necessary. Section 4.4 presents a description of implied volatility and its characteristic. Here, a description of volatility smile, risk premium and market efficiency can be found. Section 4.5 and 4.6 defines the realized volatility, realized range and also includes a discussion of which evaluation models that are most appropriate.

Section 5 begins with a method part where the raw data is described. Here, the in-, and out-of-sample is defined and a discussion concerning data filtering is of importance. Also, a discussion of the sampling procedure for the implied volatility data is described. The procedure for how to determine and estimate the ARMA and ARCH-family models is described from section 5.4 to 5.8. The method part is fulfilled with a description of how the procedure for the forecasting is implemented and how the realized volatility and loss functions are calculated.

In section 6 the results are presented starting with the descriptive statistics from the raw return data. Then the parameter estimates from the different forecasting models are presented followed by results of the ARMA models. Section 6.4 shows the results from the different test conducted on the ARCH-family models and are accompanied by its descriptive statistics. The most accurate performing forecasting models are plotted against their respective realized

volatility and the section is finalized with a discussion about why the implied volatility was not the most accurate forecasting method.

In section 7 the conclusion from this thesis are presented followed by the reference list in section 8. Section 9 presents an Appendix where the EViews coding for the forecasting models can be found.



## 2 Prior research

The prediction of future volatility is, and has for several decades been, a topic of great interest. Early, autoregressive-and moving average models were used to try to capture the volatility dynamics, and later the Box-Jenkins-type ARMA-model gained attention as a method for capturing volatility movement. However, researchers and practitioners found that the volatility suffered from 'volatility clustering' and hence, other more sophisticated models were invented. The ARCH model was the earliest of these and was groundbreaking due to its sufficiency to, in a better way, explain the non-linear dynamics of financial data. However, the ARCH-model had limitations due to its non-negatively constrains and the GARCH model was introduced by Bollerslev (1986) and Taylor (1986). The GARCH-model is more parsimonious, avoids over-fitting and is less likely to breach the non-negativity constraint than the ARCH-model. (Brooks, 2008)

Empirically, findings suggested that negative shocks are likely to increase variance more than positive shocks of the same magnitude, this is known as the leverage effect. This effect was the origin to what become to be asymmetric GARCH-model. Examples of these models are the EGARCH by Nelson (1991) and the GJR-GARCH presented by Glosten, Jagannathan and Runkle (1993).

However, the model based forecasts were challenged by implied volatility as the best means of forecasting. The implied volatility is the volatility backed out from the Black and Scholes (1973) options pricing model and is said to incorporate all of the available information of the market (Poon, 2005). Therefore, it has been stated that implied volatility is a superior forecasting tool than model based forecasts. However, this has been questioned as the volatility risk premium might set a premium and thus, overestimate it (Mixon, 2007).

Empirical findings are ambiguous; Franses and van Dijk (1996) find that asymmetric GARCH models such as GJR model were unable to outperform standard GARCH when forecasting volatility on stock market indices, while authors like Pagan and Schwert (1990), Lee (1991), Cao and Tsay (1992) and Heynen and Kat (1994) find that models that capture the property of volatility asymmetry perform well in forecasting because of strong negative relationship between volatility and shock. These findings are confirmed by Poon and Granger (2003) who state, by summarizing 93 working papers on the subject, that GARCH models perform better than ARCH models, while asymmetric models outperform standard GARCH models. Still, Poon and Granger (2003) do not find the results homogenous. Vilhelmsson (2006) states that

one reason for the different results is due to that researcher uses different models and also, different sampling periods, sample frequency and forecast horizon. Further, Vilhelmsson (2006) also states that the proxy used for ex-post variance, loss function and distribution are of great importance. The proxy might be 'noisy' if not calculated through intraday data and hence, the evaluation models can produce inaccurate estimates.

### 3 Stylized facts of the financial time series

*Volatility clustering* – a phenomenon in financial time-series modeling is that one turbulent trading day tends to be followed by another and vice versa concerning tranquil periods (Poon, 2005).

*Leverage effects* – a fall in the stock price would shift a firm's debt to equity ratio upwards, meaning that their equity value decreases. This implies an increase in both the leverage and the risk of the firm. Poon (2005) and Christie (1982) acknowledge this effect as the 'leverage effect', which means that the stock price volatility increases more when a negative shock occurs, than if a positive shock of the same magnitude occurs.

*Skewness* – also known as 'the third moment', measures how the distribution deviates from its mean value (Dowd, 2005).

*Kurtosis* – also known as 'the fourth moment', measures how fat the tails of the distribution are (Brooks, 2008)

*Leptokurtosis* – is characterized by fatter tails and a greater peak at the mean than normal distribution, though it still has the same mean and variance (Brooks, 2008).

*Volatility* - is a measure of the spread of asset returns of all likely outcomes of an uncertain variable provided by the underlying asset. Usually volatility is measured as the sample standard deviation. To estimate the volatility of stock prices the stock price is observed at fixed intervals of time (e.g. minute, day, week, or month). This may be represented as:

$$\sigma = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \mu)^2}$$

where  $r_t$  is the return on the specific interval chosen at time t, and  $\mu$  is the average return over the period. Even Variance,  $\sigma^2$ , is used as a measure of volatility, however, variance is less stable and less desirable than volatility as forecast evaluation. Volatility is associated

with risk, but according to Poon (2005), it is a bad measure of risk because it does not say anything about the shape of the distribution.

## 4 Theory

### 4.1 ARMA

There is a lot of academic literature that suggests that ARMA models provide a good forecast of volatility. Despite the fact that they are marginally less accurate in capturing volatility than more complex models, they are relatively simple to implement and the payoff is usually beneficial when balanced against the cost of producing forecast (Harris and Sollis, 2003).

First, a simple first-order autoregressive model AR (1), given in the equation below. This model states that variable  $\sigma_t$  is generated by its own past, and an error term  $\mu_t$ . The error term represent the influence of all the other variables that should be in the model but are excluded. The error term  $\mu_t$  will be considered to be a 'white noise process', which is an important assumption for time series analysis. Such a process has a constant mean and is homoskedastic and does not allow for autocorrelation in the error term. The first order autoregressive model is one where  $\sigma_t$  depends on current and previous values of linear combinations of the white noise process of the error term (Verbeek, 2008).

The AR(1) model is expressed in equation below

$$\sigma_t = \mu + \phi_1 \sigma_{t-1} + \mu_t, \quad \text{where } |\phi_1| < 1 \text{ for stationary}$$

The dependence of  $\sigma_t$  on its own past as a moving average (MA) process is an alternative to the AR-process and is given in the equation below. Consequently an MA(1) structure implies that observations more than two lags apart are uncorrelated. There are no fundamental differences between an AR-, and MA-models since the AR(1) model can be rewritten terms of a moving average model.

A moving average model where the dependence of  $\sigma_t$  is on its own past error term is given below.

$$\sigma_t = \mu + \theta_1 \mu_{t-1} + \mu_t \quad \text{Where } |\theta| < 1$$

It is possible to combine autoregressive models with moving average processes, which gives the ARMA model. When mixing higher orders of AR and MA-models, we get the ARMA(p,q) model which is expressed below. The ARMA model observed value of today  $\sigma_t$ ,

depends on both previous values of  $\sigma_t$  and, current and previous ,white noise error terms of  $\mu_t$  .The advantage with ARMA models is that they are very flexible. They can describe many time series and are parsimonious, that is, they work well with small models. (Harris and Sollis, 2003)

ARMA model of higher order

$$\sigma_t = \phi_1\sigma_{t-1} + \dots + \phi_p\sigma_{t-p} + \mu_t + \theta_1\mu_{t-1} + \dots + \theta_q\mu_{t-q} + \mu_t, \quad |\phi| < 1, \quad |\theta| < 1$$

A shock in an MA (1) process affects  $\sigma_t$  in two periods only while a shock in the AR(1) process affects all future observations (Verbeek, 2008).

Models that are non-stationary will often lead to spurious regressions, whereby the results assume a relationship, when the results are in fact coincidental. The definition of a stationary process is that it tends to return to its mean value and fluctuate around it with a constant range. It has finite variance. The statistical properties of a stationary process are as following: Its expected value is zero  $[E(u_t)=0]$  and has a constant variance  $E(u_t^2)=\sigma^2$ . It is also uncorrelated with its own previous values  $[E(u_t u_{t-k})=0]$  where the latter condition measures the independences between observations (Harris and Sollis, 2003).

Stationarity is a desirable property of an AR model. In absence of a stationary process, the impact of previous values is non-declining. If a process contains a unit root that is non-stationary, and it cannot be modeled as an ARMA model, it instead has to be modeled as an autoregressive integrated moving average (ARIMA). Integrated means that the process has a unit root and has to be differentiated to be stationary and modeled. When testing for stationarity it is important to choose the right model. An Augmented Dickey-Fuller test for stationarity is thus important for choosing an appropriate model. (Brooks, 2008)

## 4.2 ARCH

The Autoregressive conditionally heteroskedasticity model, also known as ARCH, is useful when the data researched is of non-linear character. In financial time-series a problem known as heteroskedasticity might occur, which explains that the variance error term is not constant over time. Working with a model that assumes constant variance would then worsen the approximations and hence, the ARCH-model, that does not assume constant variance over

time, might be a more appropriate model to use. Also, volatility might appear in clusters, known as volatility clustering. The ARCH-family models are designed to capture these effects. The conditional variance of an ARCH(1) depends on one previous value of the squared error:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

The movement in the dependant variable,  $y_t$ , is given by the ARCH conditional mean equation and is specified after the researchers' preferences. One example might be

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \text{ where } u_t \sim N(0, \sigma_t^2)$$

The ARCH(1) can easily be extended to include infinite lags. This is expressed as the ARCH(q) which depends on q lags of the squared returns

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_q u_{t-q}^2$$

One problem is that there is not any clear approach when determining the number of lags in the variance equation. There might also be a problem with not having a parsimonious conditional variance when a large number of q is needed to capture all of the dependence. The non-negativity constraint might also be a problem as q gets large.

#### 4.2.1 GARCH

To get around the drawbacks with ARCH-type models, a more parsimonious model, that is less likely to breach the non-negativity constraint, was developed by Bollerslev (1986) and Taylor (1986). The ARCH-model is generalized to include, except for the squared error in the variance, previous own lags. The conditional GARCH-model, as it is called, both includes the squared errors in the variance, as in the ARCH, but also previous own conditional variance lags.

When including the fitted variance from the model during the previous period, a GARCH (1,1) is expressed as following

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2$$

The GARCH(1,1) is often used in empirical studies due to its sufficiency to capture volatility clustering in the data. The GARCH-models can be extended to include q lags of the squared

error and  $p$  lags of the conditional variance. Rarely any higher order model is used in empirical studies. Also, due to that the GARCH(1,1) only includes three variables, it is more parsimonious than the GARCH(p,q)-model. The GARCH(p,q) is expressed as follows (Brooks, 2008).

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \dots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \dots + \beta_p \sigma_{t-p}^2$$

The past squared residuals capture high frequency effects, while the lagged variance captures long term influences (Figlewski, 1997).

#### 4.2.2 EGARCH

In financial time-series, it has been stated that volatility behaves differently depending on if a positive or negative shock occurs. This asymmetric relationship is called *leverage effect*, and describes how a negative shock causes volatility to rise more than if a positive shock with the same magnitude had occurred. To capture this asymmetry, different models have been developed and the one used in this study is EGARCH. This model states the conditional variance as

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

The properties of this model define the variance as positive, even though the parameters might be negative. In contrast to the other ARCH-models, EGARCH does not need to constrain the model for negative parameters. If the gamma coefficient is less than zero, than negative shocks will increase the variance more than a positive shock of the same magnitude. Vice versa is the case if the gamma coefficient is larger than one (Brooks, 2008)

#### 4.2.3 Maximum likelihood

When estimating the GARCH-type models, it is not appropriate to use *ordinary least square*, instead *maximum likelihood* should be employed. This method maximizes, given a log-likelihood function, the most likely parameter values given in the data. The fundamentals are when estimating the maximum likelihood, is to first specify which distribution should be used and then to specify the conditional mean and variance (Brooks, 2008). In this thesis, the

GARCH-models are to be maximized under the normal-, student t-, and general error distribution. The normal distribution is the distribution that allows for less kurtosis, while the student-t distribution and GED is less restrictive (Hamilton, 1994). The student t-distribution converges to the normal distribution as the degrees of freedom increase, but the advantage of the t-distribution over the normal is that it can handle a reasonable amount of excess kurtosis (Dowd, 2005). The generalized error distribution is regarded as a more flexible distribution than the normal. Depending on the degrees of freedom, the GED can take different forms, and in a special case GED can also take the form of the normal distribution. Depending on desired kurtosis, the degrees of freedom can be adjusted and GED is said to be an improvement over normal distribution (Dowd, 2005). When maximizing in EViews an iterative procedure takes place to numerically find the values under the log-likelihood function (Brooks, 2008).

#### 4.2.4 Properties when forecasting with GARCH

GARCH-type models are convenient in forecasting due to that when working with financial time-series data, a forecast of  $\sigma_t^2$  will also be a forecast of the future variance of  $y_t$ . As Brooks (2008) expresses it

$$\text{var}(y_t | y_{t-1}, y_{t-2}, \dots) = \text{var}(u_t | u_{t-1}, u_{t-2}, \dots)$$

Given the lagged variables, the conditional variance of  $y$  will be equal to the conditional variance of  $u$ .

#### 4.2.5 Test for ARCH-effects

As stated before, financial time-series data are often assumed to be non-linear and to appropriately conclude the use of a non-linear model, a test for it should be conducted. A non-linear model should be used when it is needed and there are different ways to conduct its presence. Some argue for the use of a non-linear model when financial theory states that the data at hand requires a non-linear model, while, from a statistical point of view, some argue that the use of a model should depend on which one describes all of the important features of the data most appropriate (Brooks, 2008).

The first step in the test is to regress a linear equation as follows

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$



To test for ARCH-effects, the residual should be squared and regressed on q own lags

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

The test follows a  $\chi^2$  distribution with a test statistic defined as  $TR^2$  and the null hypothesis defined as

$$\gamma_0 = 0 \text{ and } \gamma_1 = 0 \text{ and } \gamma_2 = 0 \text{ and } \dots \text{ and } \gamma_q = 0$$

And the alternative hypothesis as

$$\gamma_0 \neq 0 \text{ and } \gamma_1 \neq 0 \text{ and } \gamma_2 \neq 0 \text{ and } \dots \text{ and } \gamma_q \neq 0$$

This test is pre-programmed in EViews as a residual test for heteroskedasticity.

#### 4.2.6 Test for asymmetries in volatility

To see whether an asymmetric GARCH-model is appropriate, Engle and Ng's (1993) sign and size bias test is to be performed. This test indicates if the residuals in an ordinary symmetric GARCH-model are sign-, or size-biased. To test for sign and size bias, the following formula is used

$$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- + v_t$$

The residuals from the symmetric GARCH are programmed to take the value 1 if  $u_{t-1} < 0$  and gives the slope dummy value  $S_{t-1}^-$ . If  $\phi_1$  is significant, negative and positive shocks impact differently on the conditional variance and hence, an asymmetric GARCH-model is justifiable.

A test for sign bias can also be conducted using the following test

$$\hat{u}_t^2 = \phi_0 + \phi_1 S_{t-1}^- u_{t-1} + v_t$$

As in the previous test, a significant coefficient  $\phi_1$  will indicate that the size of a shock will have an asymmetric impact on volatility.

A joint test can be conducted through defining  $S_{t-1}^+$  as  $1 - S_{t-1}^-$ , which indicates if positive size bias is present. The joint test for positive sign bias and positive or negative size bias can be expressed as

$$\hat{u}_t^2 = \emptyset_0 + \emptyset_1 S_{t-1}^- + \emptyset_2 S_{t-1}^- u_{t-1} + \emptyset_3 S_{t-1}^+ u_{t-1} + v_t$$

The test follows a  $\chi^2$  distribution with degrees of freedom equal to 3. The joint test statistic is expressed as  $TR^2$  under the null hypothesis of no asymmetric effects (Brooks, 2008).

### 4.3 Jarque-Bera test for non-normality

Both Bollerslev (1987) and Nelson (1991) early addressed characteristic of excess kurtosis in financial time-series data and hence, a normal distribution does not correctly describe the data. It is also known that stock index returns exhibit negative skewness (Glosten et al., 1993). To test the data for normality, and see if the same properties are present in this thesis data, a Jarque-Bera (1987) test was performed. A normal distribution is symmetric/ mesokurtic, when it has a coefficient of kurtosis equal to 3. According to Brooks (2008), financial time-series also often show tendencies to be leptokurtic. The test performed evaluates the third and fourth moment, which is expressed as

$$b_1 = \frac{E[u^3]}{(\sigma^2)^{3/2}} \text{ and } b_2 = \frac{E[u^4]}{(\sigma^2)^2}$$

Here,  $u$  is defined as the error and  $\sigma^2$  as the variance, which test that the third and fourth moment are jointly zero. The test statistic for a Jarque-Bera test is

$$W = T \left[ \frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24} \right] \sim \chi^2$$

where  $T$  is the sample size

The null hypothesis states that distribution of the series is symmetric and mesokurtic, which implies that a rejection is in place if the residuals from the model is skewed or leptokurtic (Brooks, 2008).

## 4.4 Implied Volatility

### 4.4.1 Black-Scholes model and Implied Volatility

To understand implied volatility, the Black-Scholes option-pricing model is first described. There are five parameters in the option pricing equation that give its fair value; the price (P) of the underlying asset, the option's strike price (K) and time to maturity ( $\Delta T$ ), the riskless

interest rate ( $r_f$ ), and the volatility. Volatility is the only variable that cannot be directly observed (Hull, 2006). The option's implied volatility is the volatility that is expected by the market participants during the life of the option. It is possible to solve the model backwards from the observed price to determine what implied volatility must be. The inverted option pricing formula to derive the unknown volatility is what is known as implied volatility. Due to the put-call parity, the implied volatility is the same for both call and put options with the same time to maturity and the same strike price. (Poon, 2005)

#### 4.4.2 Volatility Smile and Implied Distribution

The relationship between implied volatility and strike price at a given maturity is the volatility smile. The shape of implied volatility derived from options is anything but a straight line. It is well known that implied volatility  $\sigma_{iv}$  differs across different strike prices and that the shape is like a smile when plotting it against different strike prices. The implied volatility is usually low for at-the-money options, for in-the-money and out-of-the-money options it becomes progressively higher which explains why it is U-shaped. (Hull, 2006 and Poon, 2005)

First after the crash of 1987 the smile is downward sloping, which suggests that market participants started to incorporate the possibility of future crashes when pricing options. The smile is less dependent on time to maturity if it is expressed as a relationship between implied volatility and the ratio between strike price and spot price ( $K/S_0$ ). Time to maturity is the second parameter that affects the smile. The smile flattens out as it is approaching expiration. This is the term structure of implied volatility. The term structure is hence a function of both the strike price and time to maturity. Volatility surface is a combination of volatility smile and the term structure (Hull, 2006).

The smile tells us that there is a premium charged for in-the-money (ITM) options and out-of-the-money (OTM) options. The lognormal distribution fails to capture extreme outcomes in stock prices. For ATM options the implied volatility is equal to the market's constant volatility (Poon, 2005). Comparison between implied distribution and lognormal with same mean, and standard deviation, is shown in figure 4.1. The solid line is the lognormal and the dashed line is the implied.

The volatility smile implies that the distribution is not lognormal distributed; it understates the

probability of extreme outcomes. The figure therefore supports the existence of fat tails from extreme movements. The reason why the distribution of assets is lognormal is because of the assumption about volatility being constant, and because price changes are smooth without jumps. Many securities and equities exhibit more extreme outcomes than those consistent with the lognormal model (Hull, 2006).

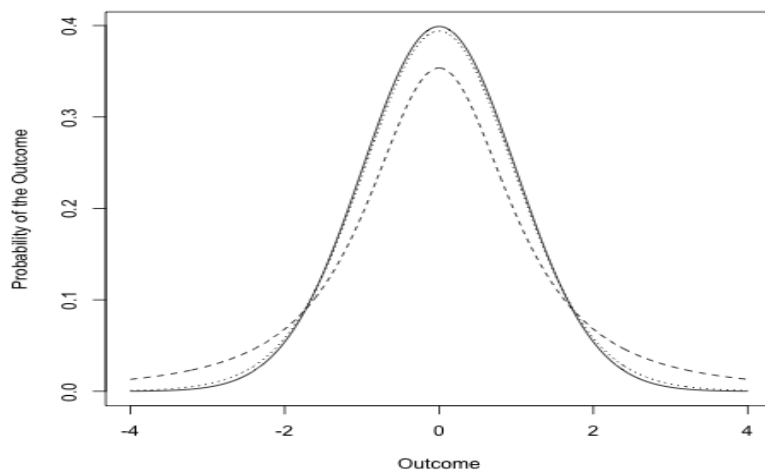


Figure 4.1 A comparison between implied distribution (dashed line) and lognormal (solid line) with same mean and standard deviation is illustrated.

The relationship between the smile figure 4.2 and implied distribution figure 4.1 are as follows. Consider a deep-out-of-the-money call option with a high strike price and an outcome at 3. This option is exercised only if the underlying asset is above the strike price. The implied distribution shows that the probability of this is higher than for lognormal distribution. This is exactly why we expect the price of an option to be higher when implied volatility is used. The same proof is valid for deep-out-of-the-money put options. According to the smile and implied distribution, lognormal distribution underestimates the probability of extreme outcomes. The volatility smile tells us that the implied volatility is relatively low for at-the-money options, whereas it becomes progressively higher when it moves either in-to-the money, or out-of-the money. The picture below illustrates when options are under and overestimated (Hull, 2006).

As mentioned the B-S option pricing model requires stock prices to follow a lognormal distribution. There is now widely documented empirical evidence that asset returns have leptokurtic tails. A leptokurtic right tail will give deep-out-of-the money higher probability to exceed the strike price and turn in-to-the money. This leads to higher call prices and higher

implied volatility at higher strike prices and the notion that options has intrinsic and time value. Time value is influenced by the uncertainty of volatility and intrinsic value reflects how deep in-to-the money the option is (Poon, 2005).

Leverage is one possible reason for the existence of smile; when a company's equity decreases in value the company's leverage increases, which imply that the equity becomes more risky and more volatile. Another reason could be the fear of another crash similar to 1987 (Hull, 2006).

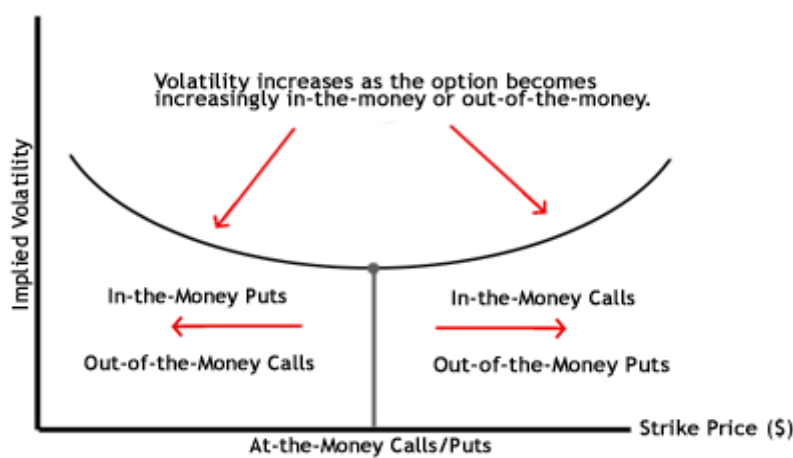


Figure 4.2. The figure describes the relationship between implied volatility and strike price of the option. Notice that the at-the-money option has the lowest implied volatility. The implied volatility increases as the option moves either into-, or out of the money.

At-the-money options have the lowest implied volatility and implied volatility rises monotonically as one moves to lower (in-the-money) or higher (out-of-the-money) strikes. This “classic” U-shaped relationship between IV and moneyness is known as the volatility “smile.” Although precise details vary from market to market, and over time within a given market, a smile is very common, to the point that it is unusual to find a market that does not exhibit something like it. In some cases, only one side will have a strong upward curvature, making a skew or smirk (Figlewski, 1997).

This must be considered as strong evidence that the market is valuing options using a different model from the one the analyst is assuming. If so, there is no reason to think that implied volatilities computed from the wrong model, whether examined individually or combined into a weighted average, will yield the market's true estimate of the volatility of the underlying asset (Figlewski, 1997).

### 4.4.3 Market efficiency

Financial economics believes that financial markets are efficient; market prices impound all available information that is relevant for valuing the underlying asset of an option. Historical volatility only takes into account past returns and is backward looking, whereas implied volatility is forward looking and also contains information about future (Hull 2006).

Statistical properties of implied volatility,

$$\sigma_{IV} = E_{MKT}[\sigma]$$

The equation above says that implied volatility is a precise representation of the markets expectations about future volatility. This requires that implied volatility has to be computed exactly the same way as the model the market uses in pricing options.

$$E_{MKT}[\sigma] = E[\sigma | \phi_{MKT}]$$

This equation says that given the markets information set,  $\phi_{MKT}$ , the markets expected value of volatility is the true conditional value. The hypothesis of market efficiency says that the market makers, makes efficient volatility forecasts from the available information.

$$[S_{t-1}, S_{t-2}, \dots, S_{t-n}] \subseteq \phi_{PUBLIC} \subseteq \phi_{MKT}$$

The set of historical prices includes a subset of public information and a subset of the markets information, which could include insider information. It would not be beyond belief if the markets expectations about future volatility may include the possibility of unexpected happenings, such as discrete price jumps, mean reversion, to fat-tailed distributions, which Black-Scholes do not take into consideration. Such a behavior will be impounded in implied volatility if it is computed by Black-Scholes model (Figlewski, 1997).

### 4.4.4 Volatility term structure and volatility risk premium

As mentioned in the previous section under the Black-Scholes model the term structure of volatility should be a flat line. In practice the slope could be upward or downward, but the

conduciveness causes are not given. Mixon (2007) finds that the term structure tends to be upwardly biased which contributes to an over-prediction bias as the forecasting maturity horizon increases. Hull and White (1987) show that the Black-Scholes model overprices ATM options and this bias tends to increase as the maturity lengthens.

According to an empirical study done by Mixon (2007) the expectation hypothesis fails. Consequently, the slope of volatility term structure has the capability to predict future implied volatility but not at the grade predicted by expectation hypothesis. The consequence is that the slope is supposed to have some information about where the market believes implied volatility to be in the future.

One factor that will lead to the failure of the expectation hypothesis is the probability of a crash, which causes the expectation hypothesis to be mispriced. Mixon (2007) did not find support for the expectation hypothesis to hold. The results agree with volatility risk premium. The risk premium makes the options volatility to deviate from the realized volatility as the future implied volatility is said to be overprice as a forecast. The gap between the realized and implied volatility, as Mixon (2007) calls it, is the volatility risk premium. The slope of the term structure is assumed to be a significant predictor of the future short-term implied volatility, even if the prediction is not consistent with the expectation hypothesis. Including a risk premium into the expectation hypothesis will improve the model. Thus implied volatility is overpriced, as forecast of future volatility and it will be obvious as the forecast horizon lengthens.

The volatility risk premium is caused by traders' liquidity risk, the uncertainty of companies' dividend policies, and the probability of a crash. Mixon (2007) finds support that the risk premium is highly correlated with the magnitudes of volatility. Fleming et al. (1998) finds support that the volatility index, VIX, contains a premium for risk, thus overstating future realized volatility.

#### **4.5 Realized volatility**

The realized volatility is an estimate of the true out-of-sample volatility and is defined by Zumbach (2009) as *the volatility that will occur between  $t$  and  $t + \Delta T$* . Since the realized volatility is used when evaluating the forecasts, it is crucial that the measure is properly calculated. The realized volatility, calculated from squared intraday returns is, according to

Anderson et al (2001b), an unbiased and efficient estimator of the variance. It has also been stated by Andersson and Bollerslev (1998) that if using intraday squared returns of no shorter than 5 minutes intervals, an accurate measure of the latent process that defines volatility can be estimated. If using too short frequency, market microstructure effects might appear in the realized volatility and according to Bandi and Russel (2005,2006), Aït-Sahalia et al. (2005) and Hansen and Lunde (2006b), the realized volatility becomes biased and inconsistent.

Since intraday data is hard to get a hold of, the following model proposed by Parkinson (1980) is used instead

$$RR_t^\Delta = \frac{(\ln H_t - \ln L_t)^2}{4 \ln 2}$$

The intraday high and low at time  $t$ ,  $H_t$  and  $L_t$ , gives the realized range for time  $t$ . One of the alternatives to this expression is to estimate the realized volatility through daily squared returns, but Parkinson (1980) shows that the realized range expression is five times more efficient and is also an unbiased estimator when bias adjusted. Martens and Van Dijk (2006) confirms the results that realized range is a better measure than the daily squared return. When the daily squared return is collected at high frequency, the realized range is still a more efficient estimator in theory. In practice though, market microstructure effects worsen the realized range estimator due to infrequent trading, a problem that high frequency realized volatility does not suffer from. However, the bid-ask bounce is a problem that both estimators suffer from. To get around this problem in realized range, Martens and Van Dijk (2006) propose a bias adjustment procedure. This bias-adjustment procedure is shown to improve significantly in the realized range over the intraday squared returns.

## 4.6 Evaluation models

The use of an imperfect volatility proxy can lead to undesirable outcomes when evaluating forecast volatility over different models. The evaluation models chosen have to be those most robust against the presence of noise. The impact from a few extreme outcomes may lead to a large influence on forecast evaluation and comparison test. The solution is to employ more robust forecast loss functions that are less sensitive to extremes. According to Patton (2006), a robust loss function is not only a function that is robust to noise in the proxy (Huber, 1981) but also to an expected loss ranking of between two volatility forecasts is the same. Hence,



the ranking should not differ with respect to the true conditional variance,  $\sigma_t^2$  or if some conditionally unbiased volatility proxy,  $\hat{\sigma}_t^2$ . This means that for the model to be robust against noise, the true conditional variance should be the optimal forecast (Patton, 2006). Tests conducted by Patton (2006) indicate that the only evaluation model that is robust, according to this criterion, is the mean squared error, MSE.

As discussed in the previous section, the realized range estimator is, when bias-adjusted, an unbiased estimator of conditional variance. It has been shown by Anderson and Bollerslev (1998, footnote 20) that, the adjusted realized range estimator produces comparable results with the 2-3 hours of intraday squared return.

Patton (2006) shows that, as the intraday frequency increases, realized volatility converges to the true conditional variance.

$$0,5625\sigma_t^2 \text{ for } m = 1$$

$$0,9616\sigma_t^2 \text{ for } m = 13$$

$$0,9936\sigma_t^2 \text{ for } m = 78$$

Where  $m$  is the number of intraday observations – as the number of observation increases, the variance rapidly converges to the true conditional variance. Thus, as the adjusted realized range is comparable to realized volatility at 2-3 hours intraday returns, the adjusted realized range produces a good approximation of the conditional variance. Patton (2006) shows that the range estimator is approximately the same as using 6 intra-daily observations in the realized volatility. The true conditional volatility approximated with the realized range estimator is shown by Patton (2006) to be  $0,9184\sigma_t^2$ .

The results presented by Patton (2006), shown above, indicate that robust MSE evaluation model should be used in this thesis when using the realized range as a proxy. However, MSE has limitations when forecasting variance. According to Vilhelmsson (2006) MSE as a loss function is sensitive to outliers. Instead, the mean absolute error is used by Vilhelmsson (2006) in the sense that it is more robust against outliers. It is also used in this thesis due to the fact that outliers might be present in the financial time-series data. As a third model to adjust for heteroskedasticity, the heteroskedasticity-adjusted mean absolute error is used (Andersen et al. 1999).

Mean square error is defined as

$$\frac{1}{N} \sum_{t=1}^N \varepsilon_t^2 = \frac{1}{N} \sum_{t=1}^N (RV_t - \sigma_t)^2$$

Mean absolute error is defined by,

$$\frac{1}{N} \sum_{t=1}^N |\varepsilon_t| = \frac{1}{N} \sum_{t=1}^N |RV_t - \sigma_t|$$

The third performance measure is HMAE, which is a heteroskedasticity-adjusted mean absolute error and is used to account for heteroskedasticity.

$$\frac{1}{N} \sum_{t=1}^N |\varepsilon_t| = \frac{1}{N} \sum_{t=1}^N \left| \frac{RV_t}{\sigma_t} - 1 \right|$$

The difference between using MAE and HMAE is that instead of using squared mean distance, as in MSE, MAE uses absolute distance. Hence, the optimization problem in the loss function changes to produce a median of the series, instead of an expected value, as the MSE does.

## 5. Method

### 5.1 Data and sampling procedure model based forecasts

We are going to investigate which of the Swedish Stock market index OMXS30, USD/EURO exchange rate and implied volatility, for the same series, that predicts the out-of-sample forecast most accurately. The reason for choosing these series is both that options trading with them is liquid and that it is interesting to compare the different series. The data is collected from DATASTREAM and processed in both EViews and Excel.

#### 5.1.1 In-, and out-of-sample length

Market microstructure problems, or as often called noise in the data, occurs in real markets due to bid-ask spreads (bounce), non-trading, and serial correlation. This makes most intraday data unusable for calculations (Figlewski, 1997). Instead returns from daily closing prices are calculated and used in the different GARCH-models. However, positive serial correlation is often found in daily closing prices for equities and other securities (Figlewski, 1997). Sampling at longer intervals is an easy way to limit the effect of serial dependence at high frequencies, but it also means using fewer data points, which increases sampling error. The best choice of sampling frequency must depend on the statistical properties of the particular price series under consideration (Figlewski, 1997). The choice of the length of the forecasting horizon has to be taken into consideration when deciding which historical data to elaborate on. Having a large data sample does not guarantee an accurate model that provides unbiased volatility, because volatility tends to change over time. There is a trade-off between trying to collect as much data as possible and trying to eliminate data that is obsolete. When the forecasting horizon is short, it is more appropriate to choose a short sample of the latest observations, which captures volatility clustering, thereby, capturing the abilities/phenomenon of the current market conditions (Figlewski, 1997).

By using a large number of daily observations, sampling error can be reduced, while for intraday data the deviations are more apparent. The choice of frequency at which the data is collected can have a large effect on volatility (Figlewski, 1997). Practitioners and researchers usually use very recent past data when forecasting. However, Figlewski (1997) shows that this might not produce an accurate forecast and he instead advocates usage of a longer horizon. Hence, the choice of in-sample data period is, in this case, set to five years of daily

observations – 2003-01-01 to 2007-12-31 – while the out-of-sample forecasted period is set to two years – 2008-01-01 to 2009-12-31. When forecasting short horizons, intraday data is needed to get more accurate estimates of the daily volatility. In lack of intraday data, the length of the pre-forecast data is extended to give as many observations as possible, without including obsolete observations.

### **5.1.2 Treatment of raw data**

Before performing any tests, the behavior of the raw data has to be analyzed. First, a statement is made regarding why no filtering action is performed on the outliers. The reasons for not filtering for outliers are due to the assumption that outliers are extreme events that might occur in the forecast period. Filtering out, for example extreme events like natural catastrophes and financial crisis, might give a better forecast of tranquil periods where no such events occur. Including them will give the opportunity to capture such an event. In previous years, there have been several extreme events, and in our opinion, there is little evidence that tells us that these extreme events will not occur again. Hence, we choose not to filter the data to give a better view of the recent events that deviate from normality and, in our opinion, give a better explanation of what might occur in the future.

Ordinary non-trading days, e.g. Easter and Christmas are excluded from the in-sample to not affect the out-of-sample volatility. The dates are then converted into figures instead of dates to represent upcoming future trading days in the out-of-sample forecasted period.

The raw data is then tested for normality through the earlier presented (section 4.3) Jarque-Bera test. This gives knowledge about if the data at hand is to be assumed normally distributed, is skewed or suffers from excess kurtosis. The raw data from the entire in-sample period, 2003-01-01 to 2007-12-31, is used to calculate the Jarque-Bera test.

## **5.2 Data and sampling procedure Implied Volatility**

The data, delivered by Thomson's to DATASTREAM, are one month's at-the-money European call options in both the OMXS30 and USD/EURO index. The raw data are gathered daily from the out of sample period 1 January 2008 to 31 December 2009. The values are the same as traders deal with from the market's Black-Scholes implied volatility. The implied volatility is backed out from the Black-Scholes model. Implied volatility is expressed on a

yearly basis at each time  $t$  and needs to be converted into daily volatility using the following scalar.

$$\sigma_{yearly} = \sqrt{252} \sigma_{daily}$$

### 5.3 ATM or Weighted Implied volatility

It is well known that options of different strikes provide different implied volatilities even though the Black and Scholes model assumes it to be constant. There are typically two strategies when deciding on which options that should be used, at-the-money options or a weighted scheme. The implied volatility derived from ATM options is most liquid and less exposed to measurement error when comparing with implied volatility at different strikes. Options that are traded with greater liquidity are expected to contain more information than less frequently traded options. Options that have longer maturities or are far away from the money are not traded as often as ATM options (Figlewski 1997). ATM options and near expiration options provides, on average, the volatility assumedly implied by the life of the option. Omitting or assuming volatility risk premium to be either zero or constant makes the implied volatility less likely to be biased. Volatilities derived from call and put options with different strike prices are combined to produce a weighted scheme composite of implied volatility. The composite volatility implied that favor ATM options are said to be less prone to measurement error (Poon, 2005).

There are different findings from authors whether an individual implied performed better than a weighted one. Beckers et al. (1981) found support for an implied to be the best, while Kroner, Kneafsey, and Claessens (1995) find that composite implied provide better forecasts (Poon and Granger 2003).

Since there are different opinions whether an individual ATM or a weighted scheme is preferred, we will select ATM options that are one month from maturity. As mentioned earlier ATM options are less prone to measurement error. They are traded at a higher volume and are believed to have more reliable information than less traded options.

## 5.4 Building ARMA models

Box and Jenkins (1976) were the first to estimate ARMA models using a three step procedure.

The steps involves following:

- 1) Identification
- 2) Estimation
- 3) Diagnostic testing

In the first step the dynamic features of the data are checked. Plotting a graph of the data and a correlogram of the ACF and PACF has to be done to determine if the series are correlated. A plot of the residuals is also important when checking for outliers. In the second step the estimation is done by OLS of the specified model from the first step. The third step involves model checking, determining whether the model is accurate or not (Brooks 230).

It is preferred to form a parsimonious model, which is one that captures all features of the data of interest with as few variables as possible. A model which has more parameters than needed will increase the standard error and lead to misleading inference. The identification stage is done by checking which information criterion gives the lowest value. The advantage of information criteria against R-squared is that it punishes models with irrelevant variables. The three most popular criteria are those of Akaike s (1974), Schwartz s (1978) and Hannan-Quinn (HQIC). When choosing between different AR and ARMA models, the best is the model that is significant and has the lowest AIC, SBIC or HQIC value. The residual diagnostic checking for serial correlation is done by a Ljung-Box test. Recall that for a series to be considered white noise, the autocorrelations should be zero (Brooks, 231).

## 5.5 Decision of the most accurate ARMA models

To forecast the future volatility through the ARMA-models, the historical realized volatility is used instead of the returns series. Through the 21 day overlapping rolling window forecast, an estimate will be received for each observation  $t + n$ .

In the examination of the OMXS30 index we found out that ARMA(1,1) was the most accurate model because both the AR(1) and MA(1) term where significant. The choice of picking a simple model is preferred to a more complex one, because too many lags reduce the

power of the test while too few lags could mean that we have remaining autocorrelation. Since the ARMA(1,1) does not show any sign of correlation in all of the tests performed, this would not be issue.

We also estimate several other combinations of ARMA(p,q) models and test the significance of the additional parameters. ARMA(2,2) and ARMA(2,1) also show significant values of the additional parameters but have marginally higher criteria values than the chosen ARMA(1,1). In some of the cases the three information criteria suggest different models, in these cases we choose the model that is suggested from two out of the three information criteria and, as mentioned above, a simple model is preferred because it has better statistic properties.

The examination of USD/EURO is similar to the OMXS30 index. In this case the most accurate model with significant values of all parameters is an ARMA(2,2) with a constant. One reason why the constant should be included is due to that it is a period where one exchange rate is appreciating against the other. The other candidates were ARMA(2,1) and ARMA(2,2) without a constant. Since Akaike, Schwarz and Hannan-Quinn all suggested ARMA(2,2) to be the most adequate model, there is no doubt in which model that is most accurate.

The model used is the one with most information criteria speaking for it. Also, to see whether the model must be integrated, the models are tested for unit roots. As a last step, the models distribution is also analyzed. The results of these tests are presented in the next section.

## 5.6 Diagnostic checking of ARMA models

As a last step, diagnostic checking is performed on ARMA(1,1) which indicates that it is the best series for OMXS30. Meanwhile, ARMA(2,2) with a constant is found to be the best model in association with the USD/EURO examination. The residual analysis provided by the Ljung-Box test, and the Dickey-Fuller unit root test are performed.

The test for unit root is done by a Dickey-Fuller, under the null hypothesis that  $\phi = 1$  implies a unit root. Accepting the null hypothesis of a unit root implies that the series is non-stationary and follows a random-walk (Verbeek 2008).

$$DF = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$$

The Dickey-fuller test for stationary rejects the null of a unit root in the ARMA(1,1) model. This is appropriate because otherwise the series would give misleading statistics. The unit root test for ARMA(2,2) with a constant is done by an augmented Dickey-Fuller test. The reason is that we have an autoregressive of higher order, namely AR(2) model with a constant in this case. The stationary condition requires that both  $\phi_1$  and  $\phi_2$  are less than one. The null-hypotheses of a unit root for the ARMA(2,2) is  $\phi_1 + \phi_2 = 1$ . According to the ADF test statistic the null of unit root is rejected even here. The residuals should though be white noise, and for white noise the autocorrelation should be zero.

The Ljung-Box test provides a Box-Pierce Q-statistic for testing whether the autocorrelations are zero (Brooks, 2008).

$$Q^* = T(T + 2) \sum_{k=1}^m \frac{\hat{t}_k^2}{T - k} \sim \chi_m^2$$

The Q-stat is approximately Chi-squared distributed with 2 degrees of freedom in the ARMA(1,1) and 4 degrees of freedom in the ARMA(2,2). We test the null hypothesis of the presence of autocorrelation against the alternative, of a white noise. The test indicates no presence of autocorrelation at lags less than three. When lags increase to more than three, there is presence of autocorrelation in the models. Since the other autocorrelation test indicates of no autocorrelation, and that the information criteria indicates that ARMA(1,1) and ARMA(2,2) are the most accurate. Therefore the chosen models are the ARMA(1,1) for OMXS30 and the ARMA(2,2) for exchange rate.

## 5.7 GARCH-model

The first step before concluding that an ARCH-family model could be appropriate is to perform an ARCH-test as described in the theory. The test for ARCH-effects conducted non-linearity in the data and an ARCH-type model is therefore used.

As stated in the 4.2.1 GARCH-model theory part, a GARCH(1,1) is often used due to that it is parsimonious. Figlewski (1997) describes that a GARCH-model with many parameters tends to fit the in-sample data better than a model with less parameters but, the model with several parameters tends to fall apart quicker in the out-of-sample. Figlewski (1997) also emphasizes the importance of a GARCH-model that is sufficiently stable and hold over time. Prior



research points towards that a GARCH(1,1) model is the most sufficient and stable, and is hence used as a forecast tool in this thesis.

Next, the test for sign and size bias (section 4.2.6) is performed on the residuals of the GARCH-model. The sign and size bias test conclude whether there is an asymmetry in the residuals of the model. The test finds a presence of sign bias and hence, an asymmetric GARCH-model is justifiable to use in the forecast. The earlier presented (section 4.2.2) asymmetric EGARCH-model is chosen due to our lack of programming skills and hence, pre-made programming is used when forecasting with EGARCH.

## 5.8 Problems with ARCH-family models

A common problem with ARCH-type models is that they require a large amount of data to get a robust estimation. Models that have many parameters tend to fit the in-sample data better, but also they tend to fall apart in the out-of-sample. This is a problem, as to produce an accurate forecast the model has to be sufficiently stable and continue to hold over time (Figlewski, 1997).

When maximizing the model, the coefficients might be non-positive or sum to values greater than one. The latter implies long run instability in the model and a long run steady state requires the parameters to sum to one. If the coefficients are negative or do not sum to one, the maximum might lie outside of the theoretically accepted region (Figlewski, 1997). According to Verbeek (2008), a parameter value that sums to more than one gives a non-stationary process, but is a typical finding in empirical studies.

## 5.9 Forecasting procedure

Figure 5.1 above describes the fundamentals behind how the forecast is performed. The out-of-sample period is forecasted through the in-sample data with a 21-day overlapping rolling window. This implies that the first forecasted period is done using the entire in-sample data. Moving one step ahead in the forecast, the in-sample data used for forecasting is moved one step forward, meaning that the oldest observation now is excluded from the forecast. The realized value in the out-of-sample, as the forecast is moved one step ahead, is used in the in-sample data to do the next forecast. This is done throughout the entire out-of-sample period giving the conditional variance for the forecast. Table 6.2 in the results section gives the

different parameter estimates for the models used to forecast the volatility expressed as an average of the forecasted period.

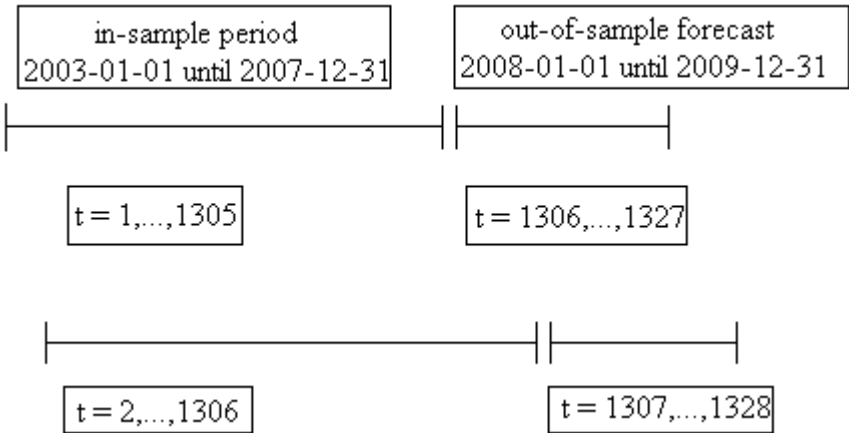


Figure 5.1. The 21 days ahead overlapping forecasting procedure

**5.10 Calculating realized volatility**

Due to the lack of intraday data, other approaches have to be considered when computing the realized volatility. As stated above, the high-low measure recommended Parkinson (1980) is used to minimize the impact of market microstructure effects and get an as good measure of the realized volatility as possible.

The realized range estimator is calculated at daily basis due to the conditional variances, and ARMA forecasts producing daily estimates. When evaluating the forecasts with the different loss functions, the daily realized range estimates would simplify the loss function calculations.

**5.11 Evaluating with loss functions**

The inputs in the loss functions are given by the values of the realized range estimate at time  $t$  and the forecasted variance from the different models at time  $t$ . The estimate is then calculated through mean square error, mean absolute error and heteroskedasticity-adjusted mean absolute error loss functions. They are summed from  $t = 1306$  until  $t = 1826$  giving an average for the out-of-sample loss which can be compared between the different models.

## 6 Results

### 6.1 Descriptive statistics of data

As has been described earlier, the Jarque-Bera test is  $\chi^2$  distributed with a skewness and kurtosis of zero respective three. At a five percent significant level, the normality assumptions of the error terms are rejected for both the OMXS30 and USD/EURO series. The exchange rate returns error has a slightly positive skewness while the OMXS30 has a negative. Both the USD/EURO and the OMXS30 series have heavier tails than a normally distributed series. The heavy tails indicate that the returns suffer from extreme events outside the normal distribution assumption.

	OMXS30	USD/EURO
Mean	0,000585	0,000263
Median	0,000684	0,000235
Maximum	0,053495	0,023981
Minimum	-0,048819	-0,021438
Std.dev	0,011288	0,005626
Skewness	-0,235133	0,006338
Kurtosis	5,224527	3,960231
Jarque-Bera	281,1009	50,14486
Probability	0,00000	0,00000

Table 6.1. The descriptive statistics are shown above for the two different series.

The results for the OMXS30 series are in line with Glosten et al. (1993) who has found, as stated in 4.2.6, that stock index returns often exhibit negative skewness. The excess kurtosis in the series is also expected since outliers are not removed in the raw data. This implies that the usage of different distributions that might capture extreme events in tails is justified. Also, as stated by Figlewski (1997), equities and many other securities suffer from fatter tails since the lognormal diffusion model is inconsistent with the large price changes.

## 6.2 Parameter estimates

Repeating the different models to ease the reading of the parameter estimates

GARCH(1,1)

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2$$

EGARCH under normal distribution

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[ \frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \right]$$

ARMA(p,q)

$$\sigma_t = \phi_1 \sigma_{t-1} + \dots + \phi_p \sigma_{t-p} + \mu_t + \theta_1 \mu_{t-1} + \dots + \theta_q \mu_{t-q} + \mu_t$$

---

### OMX

GARCH(1,1)normal  $\sigma_t^2 = 2.156e-06 + 0.067 * u_{t-1}^2 + 0.928 * \sigma_{t-1}^2$

GARCH(1,1) t-dist  $\sigma_t^2 = 2.434e-06 + 0.065 * u_{t-1}^2 + 0.929 * \sigma_{t-1}^2$

GARCH(1,1) GED  $\sigma_t^2 = 2.829e-06 + 0.067 * u_{t-1}^2 + 0.926 * \sigma_{t-1}^2$

EGARCH normal  $\ln(\sigma_t^2) = -0.123 + 0.077 * (|u_{t-1}| / \text{SQRT}(\sigma_{t-1}^2)) - 0.096 * u_{t-1} / \text{SQRT}(\sigma_{t-1}^2) + 0.992 * \ln(\sigma_{t-1}^2)$

EGARCH t-dist  $\ln(\sigma_t^2) = -7.699 + 0.261 * (|u_{t-1}| / \text{SQRT}(\sigma_{t-1}^2)) - 0.126 * u_{t-1} / \text{SQRT}(\sigma_{t-1}^2) + 0.033 * \ln(\sigma_{t-1}^2)$

EGARCH GED  $\ln(\sigma_t^2) = -0.128 + 0.077 * (|u_{t-1}| / \text{SQRT}(\sigma_{t-1}^2)) - 0.098 * u_{t-1} / \text{SQRT}(\sigma_{t-1}^2) + 0.991 * \ln(\sigma_{t-1}^2)$

ARMA(1,1)  $RR_t^\Delta = 0.988 * RR_{t-1}^\Delta + (-0.82 * u_{t-1})$

### USD/EURO

GARCH(1,1)normal  $\sigma_t^2 = 6.287e-07 + 0.043 * u_{t-1}^2 + 0.95 * \sigma_{t-1}^2$

GARCH(1,1) t-dist  $\sigma_t^2 = 1.144e-06 + 0.052 * u_{t-1}^2 + 0.933 * \sigma_{t-1}^2$

GARCH(1,1) GED  $\sigma_t^2 = 9.155e-07 + 0.045 * u_{t-1}^2 + 0.943 * \sigma_{t-1}^2$

EGARCH normal  $\ln(\sigma_t^2) = -0.209 + 0.127 * (|u_{t-1}| / \text{SQRT}(\sigma_{t-1}^2)) + 0.016 * u_{t-1} / \text{SQRT}(\sigma_{t-1}^2) + 0.988 * \ln(\sigma_{t-1}^2)$

EGARCH t-dist  $\ln(\sigma_t^2) = -0.235 + 0.132 * (|u_{t-1}| / \text{SQRT}(\sigma_{t-1}^2)) + 0.009 * u_{t-1} / \text{SQRT}(\sigma_{t-1}^2) + 0.986 * \ln(\sigma_{t-1}^2)$

EGARCH GED  $\ln(\sigma_t^2) = -0.208 + 0.12 * (|u_{t-1}| / \text{SQRT}(\sigma_{t-1}^2)) + 0.007 * u_{t-1} / \text{SQRT}(\sigma_{t-1}^2) + 0.988 * \ln(\sigma_{t-1}^2)$

ARMA(2,2)  $RR_t^\Delta = 7.328e-05 + 1.646 * RR_{t-1}^\Delta + (-0.65 * RR_{t-2}^\Delta) + (-1.406 * u_{t-1}) + 0.442 * u_{t-2}$

---

Table 6.2. Forecast output for the different parameter estimates at an average of the daily 21 days overlapping rolling window forecast.

Analyzing the coefficients of the GARCH(1,1) models under different distribution assumptions and underlying series, it can be seen that the sum of  $\alpha$  and  $\beta$  is less than one. This indicates, as stated by Verbeek (2008), that the GARCH(1,1) is close to being non-stationary and could suffer from instability in the models. This also implies that, when maximizing, the problem with getting a sample outside the accepted region is possible. The  $\alpha$  and  $\beta$  close to one also indicates, as stated by Harris and Sollis (2003) that the persistence of conditional variance is high. This means that the volatility tomorrow is highly dependent on the volatility today. A GARCH model that provides persistent forecasts does not work very well due to its inability to capture periods with highly fluctuating volatility (Schmalensee and Trippi, 1978).

Analyzing EGARCH, the gamma parameter indicates, as described in 4.2.2, that if it is less than zero, as in the case of the OMXS30 series, negative shocks increases the volatility more than positive shocks. This finding is in line with the empirical findings of leverage effects. For the exchange rate series, the gamma parameter is greater than zero, indicating that a positive shock increases volatility more than a negative shock of the same magnitude.

Interesting to notice is that the EGARCH estimate for the t-distribution for the OMXS30 deviates from the other results. The beta value is equal to approximately 0,033 which indicate that tomorrow's volatility is not that dependant of today's. One reason why the beta-value for the EGARCH under t-distribution might differ from the normal distribution could be a consequence of that the t-distribution allows for heavier tails. Although, comparing to the EGARCH under GED, the t-distributions beta almost differs as much as in the normal distribution case. The possible reason might be that the forecasting model continuously underestimates the impact of the volatility in  $t - 1$ . This might be a consequence of why this model did not explain the forecast in the most accurate way.

As in the case of the GARCH(1,1), the EGARCH model at all distributions also follows a high persistence. Only the beta value effects the volatility at time  $t$  with more than  $0.98 * \ln(\sigma_{t-1}^2)$ . The same follows for the ARMA models, where the realized volatility at time  $t - 1$  is given a lot of weight, indicating high volatility persistence. Although, according to Brooks (2008), there is no real value in trying to interpret the coefficients individually. Since the ARMA model is not constructed from any economic or financial theory, the model should be interpreted as a forecasting tool and not by its parameters. What is interesting is whether

the model is stationary or not. In section 6.3 and 6.4, the best performing models are presented.

### 6.3 ARMA

For the OMXS30, the three information criteria showed unambiguous results. The ARMA(1,1) performs best according to all of the information criteria, while still being significant in all parameters. For USD/EURO, the ARMA(2,1) has lower SBIC than the ARMA(2,2) with a constant, while the latter has both better AIC and HQIC. As the ARMA(2,2) has the lowest value in two out of three information criteria, this model is the most appropriate. Neither of the models have a unit root according to the performed Dickey-Fuller test, which implies that the models are stationary. The model is also analyzed for autocorrelation through a Durbin Watson test and a plot of the correlogram is performed to investigate whether ACF and PACF are autocorrelated. The Ljung-Box test investigates whether the model follows a white noise or not. The test indicates no presence of autocorrelation at lags less than three. Though, when lags increase to more than three, presence of autocorrelation is found in the models. According to Figlewski (1997) it is not uncommon in financial time series. The correlation might come from bid-ask bounce or from index series that are less liquid (Figlewski, 1997). However, the Durbin-Watson statistic, ACF and PACF does not show any sign of autocorrelation in the series. Since the information criteria indicates that the ARMA(1,1) and the ARMA(2,2) are most accurate, these models are chosen.

#### 6.3.1 Descriptive statistics for chosen ARMA models

The chosen ARMA models are presented in Table 6.3 with diagnostic checking. The OMXS30 indicates that both the AR(1) and the MA(1) term are highly significant. The Durbin-Watson statistic indicates no presence of autocorrelation when analyzing the residuals from the model. The presented values for Ljung-Box test up to lag 3 indicates that the chosen ARMA(1,1) is white noise, and therefore indicates no presence of autocorrelation. The Dickey-Fuller test for stationary implies that there is no unit root in our series.

Analyzing the exchange rate ARMA(2,2) forecast it can be seen in the same table that all the coefficients are significant. Even in this case, there is no presence of autocorrelation up to the

third lag and the model thus follows a white noise process. The Augmented Dickey-Fuller test for unit root implies that our series are stationary.

	Coefficient	P-value
<b>ARMA(1,1) OMXS30</b>		
AR(1)	0.987733	0.0000
MA(1)	-0.820341	0.0000
Q-stat	7,8089	0,0990
Durbin-Watson	1.842611	
D-F test	-23,08949	0.0000
<b>ARMA(2,2) USD/EURO</b>		
C	0.0251	0.0251
AR(1)	1.646146	0.0000
AR(2)	-0.649509	0.0005
MA(1)	-1.406411	0.0000
MA(2)	0.442373	0.0154
Q-stat	2.28322	0,2430
Durbin-Watson	2.026558	
ADF test	-23,06723	0,0000

Table 6.3. Shows descriptive statistics for the chosen ARMA models.

## 6.4 GARCH

### 6.4.1 Test for ARCH effects

To properly motivate the usage of an ARCH-family model, there has to be evidence of ARCH-effects in the returns. The test for ARCH-effects, described in 4.2.5 and presented in Table 6.4, conducts presence of ARCH-effects. What can be concluded from this test is that the squared residuals from previous lags are correlated with the squared residual at time  $t$ . This indicates that there is presence of heteroskedasticity in the return data and hence, a model that does not assume constant variance might make better forecasts.

ARCH-effects	OMX	USD/EURO
P-value	0.0000	0.0000
F-statistic	80.25	134.78

Table 6.4. The p-value and the F-statistic from the performed ARCH-test.

### 6.4.2 Test for asymmetry in volatility

The joint test for asymmetries in the volatility indicates that there is neither any positive-, nor negative size bias. However, a sign bias is present and it can thus be concluded that positive

and negative shocks impact the volatility differently. The usage of an asymmetric GARCH-model is hence justified by the test.

MODEL	SIGN BIAS,	NEGATIVE SIZE	POSITIVE SIZE	JOINT p-value
	p-value	BIAS, p-value	BIAS, p-value	
GARCH(1,1) NORMAL DIST	0.4621	0.0000	0.0002	0.0000

Table 6.5. The sign and size bias test indicates of a sign bias, but not any neither positive nor negative size bias.

### 6.4.3 Descriptive statistics chosen ARCH-family models

Gamma is the indicator of asymmetry effects in the model and it is interesting to analyze whether it is significant or not. In the OMXS30 case, gamma is negative and significant, meaning that there is an asymmetry in the forecast of the future volatility. In the case of the exchange rate, gamma is positive but insignificant. This is in line with the result Verbeek (2008) got from forecasting the volatility in exchange rate with an EGARCH model where the gamma also was insignificant.

In exchange rates, the asymmetric effects are not as obvious as in the case of stock indices. The test of the GARCH residuals indicates presence of asymmetry and the loss function also indicates that the EGARCH model under normal distribution produces the most accurate forecast. Therefore, the statistical rejection of the model is overlooked in this case due to that the economical properties suggests that the chosen model produces most accurate forecasts. According to Verbeek (2008) statistical arguments are not always certain and sometimes practical economical selected variables might be more accurate.

Coefficients	EGARCH GED		EGARCH norm dist	
	OMXS30	p-value	USD/EURO	p-value
Omega	-0.128	0.0099	-0.209	0.0033
Beta	0.991	0.0000	0.988	0.0000
Gamma	-0.098	0.0000	0.016	0.2954
Alpha	0.077	0.0194	0.127	0.0000

Table 6.6. The above table gives descriptive statistics and p-values of the chosen ARCH-family models. The descriptive statistics is expressed as an average of the out-of-sample forecasted period.



## 6.5 Comparing forecasting performance

### 6.5.1 Evaluation models for OMXS30

As mentioned in section 4.7, a robust loss function according to Patton (2006) is one where the true conditional variance is the optimal forecast for the model. The mean square error loss function is the only one that fulfills this criterion of the three loss functions chosen. Hence, the MSE is the function that should carry most weight in our decision on which forecasting model is the most accurate. Vilhelmsson (2006) though, states that the MAE is more robust against outliers and should therefore be of great importance since the outliers are not removed from the raw data. The last evaluation model takes the heteroskedasticity into account and adjusts for it. This implies that when the volatility is high, the HMAE does not punish the model for forecasting high errors (Koehler, 2009).

<b>OMXS30</b>	MSE	MAE	HMAE
<i>GARCH(1,1)</i>			
Normal dist	0,00168	2,5743	<b>0,5332(1)</b>
T-dist	0,00167	2,5654	0,5347(3)
GED-dist	0,00165(3)	2,5569(2)	0,5334(2)
<i>EGARCH</i>			
Normal dist	0,00159(2)	2,5641(3)	0,5510
T-dist	0,00189	2,8716	0,6540
GED dist	<b>0,00157(1)</b>	<b>2,5333(1)</b>	0,5516
<i>ARMA(1,1)</i>	0,00371	3,4581	194,0038
<i>Implied volatility</i>	0,01101	9,9018	0,7836

Table 6.7. Under each loss function, a ranking of the forecasted models from 1 to 3 can be seen to the right of the figures. The bold ones indicates of the most accurate forecasting model according to that specific loss function.

Since the loss functions are calculated in completely different ways, no comparison can be made between the functions. Instead the comparison is done for each loss function against the different forecasting models. According to the loss functions, the MSE and MAE states that the EGARCH under the GED makes for the most accurate forecast. However, when adjusting for heteroskedasticity, the normally distributed GARCH(1,1) model has the most accurate performance. Thus, both the robust loss function according to Patton (2006) and MAE, that is robust to outliers, points towards the EGARCH. Under GED this model is assumed to produce most accurate forecasts. This is not a surprise since the test for asymmetry in the data indicates ARCH-effects.

The ARMA(1,1) evaluated by the HMAE loss function produce a notable estimate while it for the other forecasted model produce quite similar results. The remarkable deviation can be explained trough the characteristics of the HMAE loss function. Here, the realized volatility is divided by the forecasted and if the inputs deviate a lot, the loss function estimate will be quite large. This is not the case in the other loss functions used in the thesis. The outliers might be a factor that could produce high forecasted values that deviates a lot. However, since an average is calculated, this should not produce an estimate that deviates at this magnitude. When the forecasted volatility is a product of several deviating inputs, the problem is instead a consequence of the way the loss function is calculated.

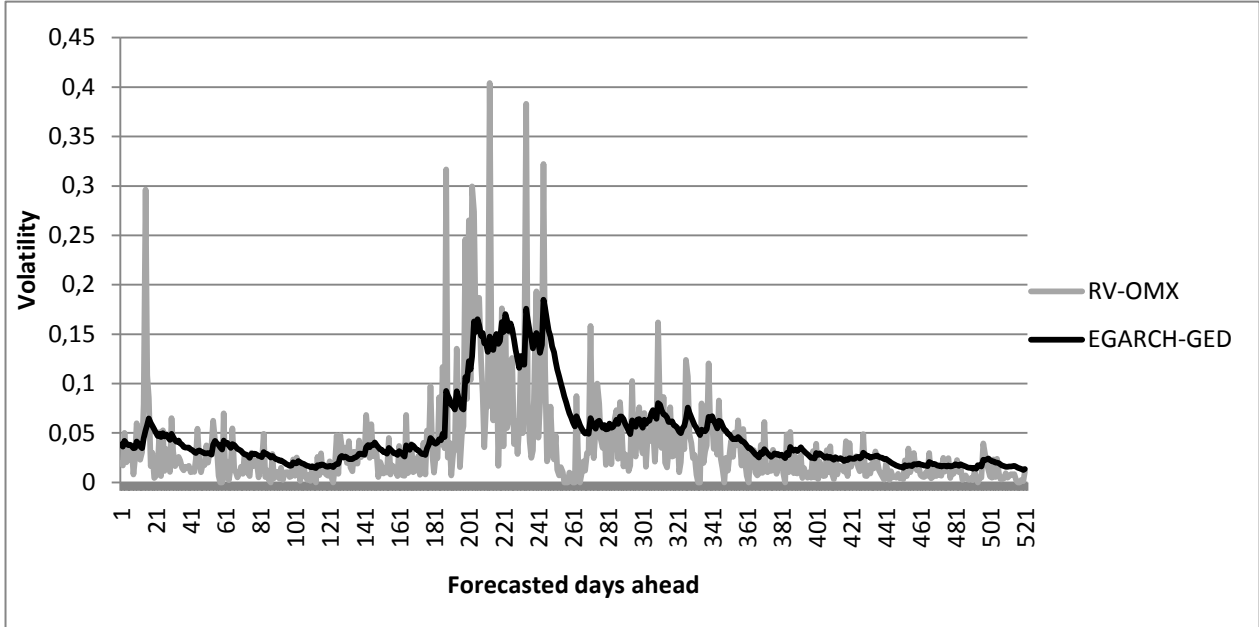


Figure 6.1. Plotting the daily realized volatility against the daily forecasted with a 21 day overlapping rolling window of the most accurate model describing OMX.

Except for the peak in volatility around the 15<sup>th</sup> observation, the volatility of the forecast of EGARCH clusters in a tranquil period between the 20<sup>th</sup> observation and 170<sup>th</sup> observation. It is followed by a more turbulent period for the OMXS30 where the volatility tends to fluctuate at a higher rate than in the previous period. The clustering effect of the volatility can easily be distinguished through a simple overview of Figure 6.1. The turbulent period is followed by one less fluctuating, but still clustering. At the end of the forecasted horizon the volatility is moving towards a cluster in a more tranquil period.

As Figure 6.1 shows, the forecasted trend follows the realized volatility under tranquil periods, but under periods where the volatility tends to fluctuate more, the model is not able to capture these fully, even though the distribution is excepting for more in the tails. This can be explained by the parameter estimates where it has been concluded that the EGARCH model experience persistence. As the figure indicates, the trend is able to be followed when the next day's volatility is near that of the previous. Though, when the volatility tends to fluctuate, the parsimonious EGARCH model is not able to capture the movements. This problem is mitigated due to that the volatility tends to cluster over time. Hence, due to persistence in the volatility it is still able to capture some of the movements, but still not as much as if the model did not suffers from persistence.

### 6.5.2 Evaluation models USD/EURO

The same statement is done here as in the discussion of robust loss functions in the part of evaluating the OMXS30 forecasting models (section 6.5.1). The robust MSE in Patton's (2006) sense, the robust MAE against outliers and the MAE that adjusts for heteroskedasticity indicate unanimously that, the normally distributed EGARCH model produces the most accurate forecasts for the exchange rate series.

<b>USD/EURO</b>	<b>MSE</b>	<b>MAE</b>	<b>HMAE</b>
<i>GARCH(1,1)</i>			
Normal dist	0,000051(2)	0,424139(2)	0,559353(2)
T-dist	0,003108	2,912885	4,181000
GED-dist	0,003108	2,907513	4,205031
<i>EGARCH</i>			
Normal dist	<b>0,000047(1)</b>	<b>0,407200(1)</b>	<b>0,547664(1)</b>
T-dist	0,003096	2,900437	4,200073
GED dist	0,003097	2,898162	4,211903
<i>ARMA(1,1)</i>	0,000140(3)	0,651881(3)	1,977421
<i>Implied volatility</i>	0,002404	4,675655	0,879222(3)

Table 6.8. Under each loss function, a ranking of the forecasted models from 1 to 3 can be seen to the right of the figures. The bold ones indicate of the most accurate forecasting model according to that specific loss function.

Even though the returns are not normally distributed according to the normality test performed, the loss function indicates that the normal distribution is explaining the

forecasting models most accurate. Comparing the kurtosis, the exchange rate suffers less than the OMXS30. Still, the exchange rate suffers from excess kurtosis in the in-sample, but not to the extent of the OMXS30. The test also indicates that the exchange rate series is closer to being normally distributed than the OMXS30 since the Jarque-Bera statistic is closer to being significant. This can be an explanation why the normal distribution is the most accurate in the exchange rate forecast and not a distribution that accounts for more extreme events. Since the EGARCH captures the asymmetry in the data it has, as in the case of OMXS30, been shown to most accurately produce the forecast. This is not a surprise since the test for asymmetry in the data indicates ARCH-effects.

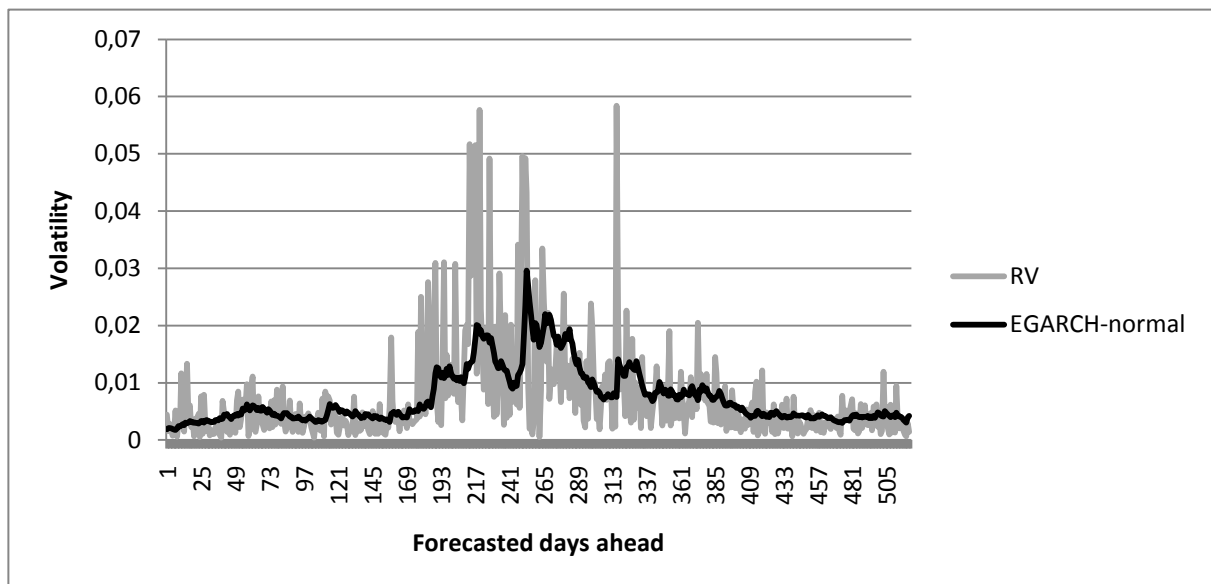


Figure 6.2. Plotting the daily realized volatility against the daily forecasted with a 21 day overlapping rolling window of the most accurate model describing the exchange rate.

As in the OMXS30 stock index return series, the clustering effects of volatility can be easily distinguished. During the first period it can be seen that the exchange rate volatility has low fluctuation through a tranquil period. It then turns into a much more turbulent period, as in the OMXS30 case. This could be due to the financial crisis that started during the 2008. Since the EGARCH model is persistent, it is not able to capture the drastic changes in realized volatility under fluctuating periods. The turbulent period is followed by one less fluctuating but still, tendencies of clustering effects can be distinguished. At the end of the forecasted horizon, a more tranquil period can be observed where the volatility is clustering with relative low fluctuations.

As the Figure 6.2 shows, the forecasted trend follows the realized volatility under tranquil periods, but under periods where the volatility tends to fluctuate more, the model is not able to capture all. As the OMXS30 series shows, this can be explained by the persistence shown in the parameter estimates. As the figure indicates, the EGARCH model under normal distribution is able to follow the realized volatility when the next day's volatility is near that of the previous days. Though, when the volatility tends to fluctuate more, the EGARCH model is not able to capture the movements. Since the volatility tends to cluster over time, this problem is mitigated. Hence, even though there is some persistence in the volatility, it is still able to capture some of the movements, but still not as much as if the model did not suffer from persistence.

## 6.6 Implied Volatility versus model based forecasts

Implied volatility is the volatility that should have the richest, most up-to-date, information set and it should be the most accurate model. Poon (2005) summarized numerous articles where the conclusions drawn were that implied volatility outperformed historical price volatility models. In the stock market index case most of the studies indicated that implied volatility performed better than historical volatility (Poon, 2005). It is surprising that Canina and Figlewski (1993) conclude that implied volatility does not have any further information about future volatility. The same conclusion is drawn by Frennberg and Hansson (1996) who study the Swedish OMXS30 market which is assumed to be small. According to Poon (2005), implied volatility tends to outperform historical volatility models even in the currency market. The most studies were done in USD/EURO as in our case. The conclusions drawn were that implied volatility did not outperformed historical volatility models in stock index markets and exchange rate markets.

Since the authors have different forecasting horizons, different frequency in the data, evaluate using different loss functions, investigate different markets and apply different historical models, it is complicated to conclude which market is most efficient under the circumstances. According to Mixon (2007) the risk premium is found to be highly correlated with the magnitude of volatility. Meanwhile, Hull and White (1987) show that the Black-Scholes model overprices ATM options and this bias tends to increase as the maturity lengthens.

The risk premium makes the option's volatility deviate from the realized volatility as the future implied volatility is, by contrast, "overpriced" as a forecast. The gap between the

realized and implied volatility, as Mixon (2007) calls it, is the volatility risk premium. This volatility risk premium is caused by traders facing liquidity risk, companies facing uncertain dividend policies and the possibility of a crash. Fleming (1998 et. al) finds support that volatility index, VIX, contains a premium for risk, thus it overstates future realized volatility. Including a risk premium into the expectation hypothesis improves the model. Thus implied volatility is overpriced in contrast to forecast of future volatility and this becomes increasingly obvious as the forecast horizon lengthens.

## 7 Conclusion

As most previous research has shown, the raw data from financial time series suffers from skewness and excess kurtosis. This has been taken into consideration by adopting two other distributions, in addition to the normal. This allows for more observation in the tails. Tests conclude that the volatility reacts asymmetrically dependant on the shocks imposed on it. Furthermore, the volatility also tends to be very persistent, meaning that the lags of previous volatility tend to die away slowly.

To examine the different models forecasting performance, three loss functions with different characteristics are used: MSE that is robust against noise; MAE that is robust against outliers; and, the HMAE that is adjusted for heteroskedasticity. To properly perform the loss functions, an as good as possible measure of the realized volatility has to be calculated. The realized volatility is usually, and most accurately, calculated through high frequency data with intervals of no less than 5 minutes. Since the intraday data is hard to get a hold of, the realized range estimator is instead used in this thesis. Through bias-adjusting the realized range, it performs as well as the realized volatility calculated through intraday data with an interval of 2-3 hours. It has also been shown that as the intraday observations increase from 1, the realized volatility quickly moves towards the true conditional variance.

The ARMA models in the thesis were chosen in line with Box-Jenkins procedure. The results indicate that the ARMA(1,1) model and ARMA(2,2) are the most appropriate when forecasting the OMXS30 index and the exchange rate USD/EURO respectively. The conclusion is that the ARMA models do not produce the most accurate forecast nor do they produce the worst.

Since the asymmetry test evidenced heteroskedasticity in the data, the ARCH-family models are used as one predicting tool. The GARCH(1,1) is first used, as it has been found to be the most accurately performing model in previous research, while still being parsimonious. Empirical studies also conclude that there is a leverage effect in financial data and hence, a sign and size bias test is performed to confirm this statement. Since there is presence of sign bias, the EGARCH model is used as another tool to predict volatility. All forecasting is done under normal-, student t-, and general error distribution. From all tests, it can be concluded that the EGARCH outperforms the other models to forecast volatility.

Implied volatility is supposed to outperform historical volatility models since it is assumed to contain not only historic information, but also the markets expectation of future volatility.

This is not the case in our thesis; the conclusions drawn are that implied volatility does not outperform any of the historical volatility models in both the stock index market and exchange rate market. The reason may be that Black-Scholes' model overprices ATM options and that this bias tends to increase as the maturity increases. The risk premium makes the options volatility deviate from the realized volatility as the future implied volatility is said to be "overpriced" as a forecast. This volatility risk premium could be caused by traders facing liquidity risk, companies facing uncertain dividend policies, or by the possibility of a crash.



## 8 References

- Aït-Sahalia, Y., Mykland, P.A., Zhang, L., (2005) *How often to sample a continuous-time process in the presence of market microstructure noise*, *Review of Financial Studies* 18, 351–416.
- Anderson, Toben G and Bollerslev Tim (1998) *Yes, standard volatility models do provide accurate forecasts*, *International Economic Review*, vol 39, nr 4
- Andersen, T.G., Bollerslev, T., Lange S (1999) *Forecasting financial market variance: sample frequency vis-à-vis forecasting horizon*, *Journal of Finance*, 52, 975-1005
- Anderson, Torben G., Bollerslev, Tim, Diebold, Francis X. and Labys, Paul (2001b) *The distribution of realized exchange rate volatility*, *Journal of the American Statistical Association*, 96, 42-55
- Bandi, F.M., Russell, J.R., (2005) *Microstructure noise, realized variance, and optimal sampling*, Graduate School of Business, University of Chicago, Working Paper.
- Bandi, F.M., Russell, J.R., (2006) *Separating microstructure noise from volatility*, *Journal of Financial Economics* 79, 655–692.
- Beckers, S. (1981) *Standard deviations implied in option prices as predictors of future stock price variability*, *Journal of Banking and Finance*, 5, 363-382.
- Black, Fischer; Myron Scholes (1973). *The Pricing of Options and Corporate Liabilities*, *Journal of Political Economy*, 81 (3): 637–654
- Bollen, B and Inder, B (2002), *Estimating daily volatility in financial markets utilizing intraday data*, *Journal of Empirical Finance*, 9, 551-562
- Bollerslev, T (1986) *Generalized Autoregressive Conditional Heteroskedasticity*, *Journal of Econometrics* 31, 307-327
- Bollerslev, T (1987) *A conditionally heteroskedasticity*, *Journal of Econometrics*, 31, 307-328

- Box, G. E. P and Jenkins, G. M (1976) *Time series analysis: Forecasting and control*, 2<sup>nd</sup> edition, Holden-Day, San Francisco
- Brooks C., 2008, *Introductory Econometrics for Finance 2nd edition*, Cambridge University Press
- Cao, C.Q., and R.S. Tsay (1992) *Nonlinear time-series analysis of stock volatilities*, Journal of Applied Econometrics, December, Supplement,1, S165-S185
- Christie, A.A (1982) *The stochastic behaviour of common stock variances: Value, leverage, and interest rate effect*, Journal of Financial Economics, 10, 407-432
- Davidian, M and Carroll, R.J (1987) *Variance function estimation*, Journal of American Statistical Association, 82, 1079-1091
- Dowd, Kevin (2005) *Measuring market risk 2<sup>nd</sup> edition*, John Wiley & sons Ltd, England
- Engle, R.F and Ng, Victor K (1993) *Measuring and Testing the Impact of News on Volatility*, The Journal of Finance, vol 48, No 5, 1749-1778
- Figlewski, S (1997) *Forecasting volatility*, Financial Markets, Institutions & Instruments, Vol. 6, No. 1, p. 1-88
- Flemming, 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 and van Dijk, D (1996) *Forecasting stock market volatility using non-linear GARCH models*, Journal of Forecasting 15, 229-235
- Frennberg, P., and B.Hansson (1996) *An evaluation of alternative models fore predicting stock volatility*, Journal of International Financial Markets, Institutions amd Money, 5, 117-134
- Glosten, L. R, Jagannathan, R. and Runkle, D. E (1993) *On the relation between the expected value and the volatility of the nominal excess return on stocks*, The Journal of Finance 48(5), 1779-1801

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

Hamilton, James. D (1994) *Time series analysis*, Princeton University Press, New Jersey

Hansen, P.R., Lunde, A., (2006b) *Realized variance and market microstructure noise*, *Journal of Business and Economic Statistics* 24, 127–218 (with comments and rejoinder).

Harris, Richard and Sollis, Robert (2003) *Applied time series modeling and forecasting*, John Wiley and Sons Ltd, England

Heynen, R.C (1995) *Essays on Derivatives Pricing Theory*, Thesis Publishers, Amsterdam.

Huber, P.J (1981) *Robust Statistics*, Wiley, New York, USA

Hull J.C., White A., (1997) *The pricing of options on assets with stochastic volatilities*, *Journal of Finance*, 42,2,281-300.

Hull, J.C, (2006) *Options, Futures and Other Derivatives 6th edition*, Prentice Hall

Jarque, C. M, Bera, A. K, (1987) *An efficient large-sample test for normality of observations and regression residuals*, Australian National University Working papers in Econometrics 40, Canberra

Koehler, Paul A. (2009) *Predicting Realized Volatility: The Effects of Extended Hours Trading*, Boston University Department of Economics, USA

Kroner, Kenneth., Kneafsey, Kevin and Claessens, Steijn (1995) *Forecasting Volatility in Commodity Markets*, *Journal of Forecast*, 14, 77-95

Lee, K.Y (1991) *Are the Garch best in out-of-sample performance?*, *Economics Letters*, 37,3, 305-308.

Lintner, John. (1965) *The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets*, *Review of Economics and Statistics*, 47:1, pp. 13–37

Martens, Martin and van Dijk, D (2006) *Measuring volatility with the realized range*, Journal of Econometrics, 138, 181-207

Mixon, Scott (2007) *The implied volatility term structure of stock index options*, Journal of Empirical Finance, Washington

Nelson, D. B (1991) *Conditional Heteroskedasticity in Asset Returns: A new approach*, Econometrica 59(2), 347-370

Pagan, A. R. and Schwert, G. W (1990) *Alternative Models for Conditional Stock Volatilities*, Journal of Econometrics

Parkinson, M., (1980) *extreme value method for estimating the variance of the rate of return*, Journal of Business 53, 61–65.

Patton, Andrew, (2006) *Volatility Forecast Comparison using Imperfect Volatility Proxies*, University of Technology, Sydney

Poon, S-H., and C.W.J, Granger (2003), *Forecasting financial market volatility: A Review*, *Journal of Economic Literature*, Vol. 41, No. 2, p. 478-539

Poon, Ser-Huang (2005) *A practical guide to Forecasting financial market Volatility*, Wiley Finance, England

Verbeek, M., (2008) *A guide to modern econometrics*, John Willy and Sons

Vilhelmsson, Anders (2006) *Garch Forecasting Performance under Different Distribution Assumptions*, Journal of Forecasting, 25, 561-578

Schmalensee, R., and R.R Trippi (1978) *Common stock volatility expectations implied by option premia*, Journal of Finance, 33, 1, 129-147

Sharpe, William F. (1964) *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*, Journal of Finance, 19:3, pp. 425–42

Taylor, S.J (1986) *Forecasting the volatility of currency exchange rates*, International Journal of Forecasting 3, 159-170

Zumbach, Gilles (2009) *Volatility and the at-the-money implied volatility: a mult-components ARCH approach and its relation with market models*, RiskMetrics Group, New York

## 9 Appendix

The forecasting script for the different models used in EViews is presented below. Only the ARMA models that are used as a forecasting tool is presented.

### ARMA(1,1) without a constant

```
scalar noobs=1826
scalar nw=521
matrix(1826,21)armaforecast
vector(21)forecastresult
for !i=nw to noobs
  smpl !i-nw+1 !i
  equation arma_n
  arma_n.ls rvdaily ar(1) ma(1)
  for !j=1 to 21

  if !j=1 then
    forecastresult(1)=arma_n.@coefs(1)*rvdaily(!i)+arma_n.@coefs(2)*resid(!i)
  endif

  if !j=2 then
    forecastresult(2)=arma_n.@coefs(1)*forecastresult(1)+arma_n.@coefs(2)*rvdaily(!i)
  endif

  if !j=3 then
    forecastresult(3)=arma_n.@coefs(1)*forecastresult(2)+arma_n.@coefs(2)*forecastresult(1)
  endif

  if !j>3 then
    forecastresult(!j)=arma_n.@coefs(1)*forecastresult(!j-1)+arma_n.@coefs(2)*forecastresult(!j-2)
  endif

next !j
rowplace(armaforecast,@transpose(forecastresult),!i)
```

```
next !i
```

## ARMA(2,2) with a constant

```
scalar noobs=1826
scalar nw=521
matrix(1826,21)armaforecast
vector(21)forecastresult
for !i=nw to noobs
  smpl !i-nw+1 !i
  equation arma_n
  arma_n.ls rvdaily c ar(1) ar(2) ma(1) ma(2)
  for !j=1 to 21

  if !j=1 then
    forecastresult(1)=arma_n.@coefs(2)*rvdaily(!i)+arma_n.@coefs(3)*rvdaily(!i-1)+arma_n.@coefs(4)*resid(!i)+arma_n.@coefs(4)*resid(!i-1)
  endif

  if !j=2 then
    forecastresult(2)=arma_n.@coefs(2)*forecastresult(1)+arma_n.@coefs(3)*rvdaily(!i)
  endif

  if !j=3 then
    forecastresult(3)=arma_n.@coefs(2)*forecastresult(2)+arma_n.@coefs(3)*forecastresult(1)
  endif

  if !j=4 then
    forecastresult(4)=arma_n.@coefs(2)*forecastresult(!j-1)+arma_n.@coefs(3)*forecastresult(!j-2)
  endif

  if !j>4 then
    forecastresult(!j)=arma_n.@coefs(2)*forecastresult(!j-2)+arma_n.@coefs(3)*forecastresult(!j-3)
```

```
endif
```

```
next !j
```

```
rowplace(armaforecast,@transpose(forecastresult),!i)
```

```
next !i
```

## **GARCH(1,1) normal distribution**

```
smp1 @all
```

```
scalar noobs=1826
```

```
scalar nw=521
```

```
matrix(1826,21) garchforecast
```

```
vector(21) forecastresult
```

```
for !i=nw to noobs
```

```
  smp1 !i-nw+1 !i
```

```
  equation garch_n
```

```
  garch_n.arch(1,1) series02 c
```

```
  garch_n.makegarch garchcondvar
```

```
  for !j=1 to 21
```

```
    if !j=1 then
```

```
      forecastresult(1)=garch_n.@coefs(2)+garch_n.@coefs(3)*resid(!i)^2+garch_n.@coefs(4)*garchcondvar(!i)
```

```
    endif
```

```
  if !j>1 then
```

```
    forecastresult(!j)=garch_n.@coefs(2)+garch_n.@coefs(3)*@SQRT(forecastresult(!j-1))^2+garch_n.@coefs(4)*forecastresult(!j-1)
```

```
  endif
```

```
next !j
```

```
rowplace(garchforecast,@transpose(forecastresult),!i)
```

```
next !i
```



## GARCH(1,1) t-distribution

```
smp1 @all
scalar noobs=1826
scalar nw=521
matrix(1826,21) garchforecast
vector(21) forecastresult
for !i=nw to noobs
smp1 !i-nw+1 !i
equation garch_n
garch_n.arch(1,1, tdist) series02 c
garch_n.makegarch garchcondvar

for !j=1 to 21
if !j=1 then
forecastresult(1)=garch_n.@coefs(2)+garch_n.@coefs(3)*resid(!i)^2+garch_n.@coefs(4)*garchcondvar(!i)
endif

if !j>1 then
forecastresult(!j)=garch_n.@coefs(2)+garch_n.@coefs(3)*@SQRT(forecastresult(!j-1))^2+garch_n.@coefs(4)*forecastresult(!j-1)
endif

next !j

rowplace(garchforecast,@transpose(forecastresult),!i)

next !i
```

## GARCH(1,1) General error distribution

```
smp1 @all
scalar noobs=1826
scalar nw=521
matrix(1826,21) garchforecast
vector(21) forecastresult
for !i=nw to noobs
smp1 !i-nw+1 !i
```

```

equation garch_n
garch_n.arch(1,1, GED) series02 c
garch_n.makegarch garchcondvar

for !j=1 to 21
  if !j=1 then
forecastresult(1)=garch_n.@coefs(2)+garch_n.@coefs(3)*resid(!i)^2+garch_n.@coefs(4)*garchcondvar(!i)
endif

  if !j>1 then
forecastresult(!j)=garch_n.@coefs(2)+garch_n.@coefs(3)*@SQRT(forecastresult(!j-1))^2+garch_n.@coefs(4)*forecastresult(!j-1)
endif

next !j

rowplace(garchforecast,@transpose(forecastresult),!i)

next !i

```

## EGARCH

```

scalar noobs=1826
scalar nw=521
matrix(1826,21) garchforecast
vector(21) forecastresult
for !i=nw to noobs
  smpl !i-nw+1 !i
  equation garch_n
  garch_n.arch(egarch) series02 c
  garch_n.makegarch garchcondvar

  for !j=1 to 21

  if !j=1 then

```

```

forecastresult(1)=@EXP(garch_n.@coefs(2)+garch_n.@coefs(3)*@ABS(resid(!i)/@SQRT(garchcondvar(!i)))+garch_n.@coefs(4)*(resid(!i)/@SQRT(garchcondvar(!i)))+garch_n.@coefs(5)*@LOG(garchcondvar(!i)))
endif

if !j>1 then
forecastresult(!j)=@EXP(garch_n.@coefs(2)+garch_n.@coefs(3)*@LOG(forecastresult(!j-1)))
endif
next !j
rowplace(garchforecast,@transpose(forecastresult),!i)
next !i

```

The forecasts for the different distribution is done the some way in EGARCH as in the GARCH(1,1) programming.