



UMEÅ UNIVERSITY

An abstract, high-contrast image with swirling patterns in shades of blue, green, and orange, resembling a microscopic view or a stylized landscape. It serves as the background for the title and author information.

Volatility Forecasting Performance of GARCH Models:

A Study on Nordic Indices During COVID-19

Ludwig Schmidt

Master thesis, 15 credits

Master's Programme in Economics / Thesis II, 120 credits

Spring term 2021

Volatility Forecasting Performance of GARCH Models:

A study on Nordic indices during COVID-19

Abstract:

Volatility forecasting is an important tool in financial economics such as risk management, asset allocation and option pricing since an understanding of future volatility can help professional and private investors minimize their losses. The purpose of this paper is to investigate the volatility forecasting performance of symmetric and asymmetric GARCH models on Nordic indices during COVID-19. The models examined in this paper are GARCH, EGARCH, GJR and APARCH and the forecasting performance is evaluated by several statistical performance measures. The results of this paper are that the symmetric GARCH($1,1$) on average has the worst forecasting performance during a crisis. However, the difference between the predictability of the models is in practice small. The superior forecasting models are the GJR($1,1$) and EGARCH($1,1$) when forecasting a crisis on Nordic indices according to the evaluation measures. This is because the asymmetric extensions of these models enable them to capture the more present leverage effect in the event of a crisis.

Key words: *GARCH, Volatility, Leverage Effect, Covid-19*

Ludwig Schmidt
Master thesis in Economics, Spring term 2021
Umeå University,
Department of Economics

Acknowledgements

Thanks to Tomas Sjögren, professor in Economics at Umeå University, for his review of the paper and valuable inputs.

Tel: +46722525536
Email: ludwigg.schmidt@hotmail.com

Table of Contents

<i>1. Introduction</i>	4
<i>2. Literature review</i>	6
<i>3. Theoretical Approach</i>	8
<i>3.1. ARCH</i>	8
<i>3.2. GARCH</i>	9
<i>3.3. EGARCH</i>	9
<i>3.4. GJR model</i>	10
<i>3.5. APARCH</i>	11
<i>3.5. Conditional Mean</i>	11
<i>3.6. Distribution of the error term</i>	12
<i>3.7. Maximum Likelihood Estimation</i>	13
<i>4. Data & Methodology</i>	14
<i>4.1. Diagnostics tests</i>	15
<i>4.2. Out-of-sample forecasting method</i>	17
<i>4.3. Realized volatility</i>	18
<i>4.4. Forecast evaluation</i>	18
<i>5. Estimation results</i>	20
<i>6. Discussion</i>	22
<i>7. Conclusion</i>	24
<i>7.1. Suggestions of future research</i>	24
<i>References</i>	25
<i>APPENDIX</i>	28

1. Introduction

Volatility forecasting is an important tool in financial economics such as risk management, asset allocation and option pricing since an understanding of future volatility can help professional and private investors minimize their losses. Volatility is not directly observable in practice and thus needs to be estimated from the underlying price of an asset. The interesting part of estimating the volatility with asset return as underlying series is that the volatility has four commonly seen characteristics (Tsay, 2013). (1) The variance of the series differs between high and low in different periods, creating “volatility clusters”. (2) Volatility jumps are rare since volatility evolves in a continuous matter. (3) Volatility varies within some fixed range and does not diverge to infinity. (4) Big price drops seem to have a larger impact on the volatility than an equally large price increase, i.e., asymmetric impacts. The fourth characteristic is known as the “leverage effect” and is frequently encountered in financial time series (Mandelbrot, 1963; Black, 1976). Since these four phenomena have been found to characterize the movement of volatility in financial time series, they have played a significant role in the development of volatility forecasting models (Tsay, 2013).

The earlier theoretical models on volatility assumed constant variance (Merton, 1969; Black and Scholes, 1973). These could for example be homoscedastic regression models, which do not reflect the properties of volatility. To better reflect the characteristics of volatility in the models, Engle (1982) proposed the Autoregressive Conditional Heteroscedasticity (ARCH) model. Unlike the traditional models of constant variance, the ARCH process accounts for the time-varying conditional variance of financial time-series using lagged disturbances. The disadvantage of the ARCH was that it had to use many parameters to capture the dynamics of conditional variance. Due to this, Bollerslev (1986) proposed a generalized extension to the ARCH, the Generalized ARCH (GARCH), which allowed for a more flexible lag structure that could reduce the number of parameters in the model. Both the ARCH and GARCH can capture the commonly seen characteristics of volatility clustering and leptokurtosis. The disadvantage of these models is that they fail to capture the leverage effect due to being symmetric models. Many asymmetric extensions to GARCH have thereafter been proposed to address the issue that a negative shock in asset return will have a larger impact on the volatility of the series than an equally large positive shock (Tsay, 2013). Examples of asymmetric extensions are the Exponential GARCH (EGARCH) by Nelson (1991), Glosten-Jaganathan-Runkle (GJR) by

Glosten, Jaganathan and Runkle (1993) and Asymmetric Power ARCH (APARCH) by Ding et al. (1993).

The purpose of this paper is to investigate the volatility forecasting performance of symmetric and asymmetric GARCH models on Nordic indices during COVID-19. The delimitations of this paper are the use of GARCH, EGARCH, GJR and APARCH as forecasting models and OMX Stockholm PI, OMX Helsinki 25 and OMX Copenhagen 20 as a representation of Nordic indices. The evaluation of the volatility forecasting performance of the models is on the crisis COVID-19 and not on a series during “normal” circumstances. This is interesting since during a crisis, such as COVID-19, the variance of a financial series on average increase and there are big downturns in price which should make the leverage effect more central. The previous research, reviewed in section two, is generally forecasted on a series under “normal” circumstances and the results are quite contradictory. Furthermore, there are few existing studies on the forecasting performance of GARCH models on Nordic indices. The intended contribution of this paper is therefore to evaluate which GARCH model is superior when forecasting during a crisis on Nordic indices. The performance measures Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2 LOG, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to compare and evaluate the forecasting performance of the models, i.e., which of the models achieves a predicted volatility closest to the realized volatility.

The paper is structured in the following way. Section 2 presents a literature review of previous findings on volatility forecasting performance using various GARCH models. Section 3 describes the theoretical approach of the GARCH models included in this study. Section 4 introduces the data and methodology. Section 5 presents the estimation results, and the thesis ends with a discussion on the findings in section 6 and a conclusion in section 7. The findings are that asymmetric extensions to the GARCH are important when volatility forecasting during a crisis on Nordic indices and that the superior models with regard to the evaluation measures are the asymmetric GJR(I, I) and EGARCH(I, I).

2. Literature review

There are numerous previous papers on the forecasting performance of GARCH models. Various findings have been that QGARCH is the superior model forecasting the volatility of stock returns in Sweden, Italy, Germany and Spain (Franses and Van Dijk, 1996), GJR has been found to be the superior model when volatility forecasting an Australian index (Brailsford and Faff, 1996), it is difficult to pick one specification of GARCH for the DJ composite index (Brooks, 1998) and EGARCH has been found to might be more useful when forecasting two Tel Aviv indices (Alberg et al., 2008). The results are, as seen, contradictory. The conclusions are often that the asymmetric extensions of GARCH to capture the leverage effect improve the forecasting performance and that the model that is found to be optimal in the study is the “most promising”, not certainly the best performing model regardless. Discussions about what distribution of the error terms to use and the evaluation method are central since the evaluation methods penalize in different ways. Some insights to the conflicting results are discussed in a survey by Poon and Granger (2001) where they emphasize the importance of the underlying markets and assets since stocks, exchange rates, equity indices, etc. have contrasting historical variance and fluctuations in volatility. Volatility is also a stock variable¹ which needs to be accumulated over a period and thus the frequency (minutely, hourly, daily, etc.) of the data will affect the results. The frequency of the data can make extensive improvements to the volatility estimates and forecasts (Fung & Hsieh, 1991; Andersen & Bollerslev, 1998). The trade-off is that too high a frequency can lead to spurious regression². Poon and Granger conclude that some of the earlier research was done on nonstationary series, which violates the forecasts since the models assume a stationary series.

What is more relevant to this paper is the forecasting performance of GARCH models in the presence of a crisis. Angabini and Wasiuzzaman (2011) evaluates the forecasting performance of the GARCH, EGARCH and GJR on the Malaysian Stock market, specifically the Kuala Lumpur Composite Index (KLCI), on the financial crisis 2008. All financial markets were affected by the global financial crisis and a concern was how this changed the volatility in the stock market. The findings were that the KLCI exhibited the stylized characteristics of asymmetry, leverage effect and leptokurtosis. Furthermore, there was a significant increase in the volatility and leverage effect due to the crisis, however short-lived. In the comparison of

¹ A quantified variable measured at a particular point in time.

² Nonstationary and independent series.

how the GJR and EGARCH estimated the change in volatility, both the models produced similar result which was that the volatility had increased from 11.5 to 18.5% for the index. Despite these findings, the EGARCH and GJR with their nonlinear asymmetric extensions were outperformed by the GARCH which according to the evaluation measures predicted the volatility closest to the realized volatility. This further stands in contradiction to for example papers of Shamiri and Abu Hassan (2007) which found that the GJR was the superior forecasting model of the Malaysian stock market under “normal” circumstances.

Another interesting finding in previous research is that the model with the best in-sample fit does not necessarily produce the best out-of-sample forecast. It is a statistical necessity to correctly specify the lag order, conditional mean process and distribution of the error term to reflect the historical movements of the series, but since the future is unknown the best-fitted model does not necessarily produce the best forecast (Shamiri and Isa, 2009). Previous research has also found that the smallest lag order is suitable to capture the changing volatility and thus provide adequate results (Gokcan, 2000; Javed and Mantalos, 2013). Due to this, the smallest lag order is the common application when forecasting using GARCH models (Teräsvirta, 2006).

3. Theoretical Approach

The GARCH models reviewed in this paper are based on the Autoregressive Conditional Heteroscedasticity (ARCH) model by Engle (1982). Consequently, the ARCH will first be presented and then extended to the symmetric GARCH and the asymmetric extensions to the GARCH, i.e., the EGARCH, APARCH and GJR. Thereafter the conditional mean, distribution of the error term and Maximum Likelihood Estimation are presented.

3.1. ARCH

The first model that provided a systematic framework to model time-varying conditional variance was the Autoregressive Conditional Heteroscedasticity (ARCH) model introduced by Engle (1982). The idea behind the ARCH is that the dependence of the shock can be described as a quadratic function of its lagged values and that the shock of an asset is dependent but serially uncorrelated. The ARCH(q) is formally given by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad (1)$$

, where

$$\varepsilon_t = \sigma_t e_t \quad (2)$$

$$e_t \sim I.I.D.(0,1), \quad \alpha_0 > 0, \quad \alpha_i > 0$$

The σ_t^2 is the conditional variance, ε_t is the innovation of an asset, σ_t is the volatility of the series at time t and, e_t is a white-noise error term which should be identically and independently distributed with mean zero and variance one, given by $e_t \sim I.I.D.(0,1)$. α_0 and α_i are coefficients and the restrictions are imposed to avoid negative variance. By allowing the conditional variance to change over time as a function of past errors, the ARCH can capture volatility clustering and leptokurtosis. However, some of the weaknesses are that it often requires many parameters and a high order of the ARCH term q to capture the dynamic behavior of conditional variance, it tends to overpredict the volatility since it responds slowly to isolated shocks in returns, and it fails to capture the leverage effect (Tsay, 2013). To account for this, other models based on the ARCH framework were proposed.

3.2. GARCH

Bollerslev (1986) proposed a generalized ARCH, creating the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. The GARCH allows for a more flexible lag structure by imposing non-linear restrictions which enables to reduce the number of parameters in the model. The GARCH(p, q) model is given by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (3)$$

, where

$$q > 0, \quad p \geq 0, \quad \alpha_0 > 0$$

$$\alpha_i \geq 0 \text{ for } i = 1, \dots, q$$

$$\beta_j \geq 0 \text{ for } j = 1, \dots, p$$

The parameters to be estimated are α_0, α_i and β_j , q is the order of the ARCH terms and p is the order of the GARCH terms. To ensure a positive conditional variance σ_t^2 for all t , the restrictions $\alpha_0 > 0$, $\alpha_i \geq 0$ (for $i = 1, \dots, q$) and $\beta_j \geq 0$ (for $j = 1, \dots, p$) are imposed. For $p = 0$, the GARCH(p, q) reduce to an ARCH(q) process. The conditional variance in the ARCH is only specified as a linear function of past sample values, whereas the GARCH includes lagged conditional variance into the process in addition. Considering a GARCH($1, 1$) process, this means that the conditional variance today σ_t^2 is affected by the previous shock ε_{t-1}^2 like in the ARCH, but also on the conditional variance the day before σ_{t-1}^2 . Consequently, if the volatility at time $t - 1$ went far away from its average, this will be accounted for in the estimated volatility at time t making it remain more volatile until it eventually returns to its average. However, the limitation to the standard GARCH process is that it still fails to capture the leverage effect due to its symmetric distribution. Several nonlinear extensions to the GARCH have therefore been proposed to address this problem, such as the EGARCH, APARCH and GJR model.

3.3. EGARCH

Both the symmetric ARCH and GARCH can model volatility clustering and leptokurtosis, but they fail to capture the asymmetric relationship between asset returns and volatility changes which is needed when handling financial time-series data. Nelson (1991) proposed the Exponential Generalized Autoregressive Conditional Heteroscedasticity (EGARCH) model to overcome the weaknesses of the symmetric models, in particular the leverage effect. Nelson

imposed the natural logarithm of conditional variance σ_t^2 to guarantee a nonnegative conditional variance, instead of imposing restrictions. He also enables asymmetric effects between positive and negative returns by a weighted innovation $g(\varepsilon_{t-i})$. The EGARCH(p, q) is given by:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^q \alpha_i g(\varepsilon_{t-i}) + \sum_{j=1}^p \beta_j \ln(\sigma_{t-j}^2) \quad (4)$$

, where
$$g(\varepsilon_t) = \theta \varepsilon_t + \gamma [|\varepsilon_t| - E|\varepsilon_t|] \quad (5)$$

In the asymmetric response function $g(\varepsilon_t)$, θ and γ are parameters. The response function has two components, where the $\theta \varepsilon_t$ determines the sign and $\gamma [|\varepsilon_t| - E|\varepsilon_t|]$ the size of the effect. Both components are i.i.d. with zero mean. By rewriting $g(\varepsilon_t)$, the asymmetry is more easily shown. If $\varepsilon_t \geq 0$, the $g(\varepsilon_t)$ takes the form $(\theta + \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|)$ and in contrary if $\varepsilon_t < 0$, $g(\varepsilon_t)$ takes the form $(\theta - \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|)$. The $(\theta + \gamma)$ and $(\theta - \gamma)$ is what enables the asymmetric response of a positive versus a negative lagged innovation. The coefficient θ is expected to be negative to ensure asymmetry (Tsay, 2013). By this weighted innovation, a negative shock in asset return will have a larger impact on the conditional variance than an equally large positive shock.

3.4. GJR model

An alternative way to model volatility asymmetry is by the Glosten-Jagannathan-Runkle Generalized Autoregressive Conditional Heteroscedasticity model introduced by Glosten et al. (1993), commonly referred to as the GJR-GARCH or GJR model. The advantage of GJR to EGARCH is that it is easier to implement in practice since the variance is directly modeled instead of using the natural logarithm (Hayashi, 2000). The GJR(p, q) is formally given by:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q (\alpha_i + \gamma_i I_{t-i}) \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (6)$$

, where
$$I_{t-i} = \begin{cases} 1 & \text{if } \varepsilon_{t-i}^2 < 0 \\ 0 & \text{if } \varepsilon_{t-i}^2 \geq 0 \end{cases} \quad (7)$$

The restrictions $\alpha_0 > 0$, $\alpha_i \geq 0$ (for $i = 1, \dots, q$), $\gamma_i \geq 0$ (for $i = 1, \dots, q$) and $\beta_j \geq 0$ (for $j = 1, \dots, p$) are again imposed on the coefficients to guarantee a positive conditional

variance for all t . The binary variable I_{t-i} ensure that a negative shock on asset returns will have a stronger impact on the conditional variance since the coefficient associated with a negative shock is $(\alpha_i + \gamma_i)$.

3.5. APARCH

Another extension to the GARCH model is the Asymmetric Power Autoregressive Conditional Heteroscedasticity (APARCH) model introduced by Ding et al. (1993). The advantage of APARCH is that it can transform to different GARCH model (Ibid.). This is explained further below. The APARCH(p, q) is formally given by:

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta \quad (8)$$

, where $\delta \geq 0, \quad -1 < \gamma_i < 1, \text{ for } i = 1, \dots, q$

The δ is a positive real number and the coefficients $\alpha_0, \alpha_i, \beta_j$ and γ_i fulfills the same regularity conditions as before to ensure a positive conditional variance. It is difficult to find a good interpretation for δ according to Tsay (2013). The interesting thing about the APARCH is that by varying the parameters in the model can take the form of several different GARCH models. Example wise, when $\delta = 2$, the APARCH becomes the GJR model and, when $\delta = 0$, the APARCH reduce to an EGARCH. If $\delta = 1$, then the model estimates using the volatility straight in the equations. This type of power function of the APARCH aims to enhance the goodness of fit of the model, which can be argued is a reasonable approach when forecasting.

3.5. Conditional Mean

In addition to the conditional variance, the conditional mean of the GARCH models needs to be specified. The rugarch package in R enables to define the dynamics for the conditional mean from the Autoregressive Moving Average (ARMA) model for the univariate GARCH models. It is highly relevant to use the concept of ARMA models in volatility modeling and the GARCH model can be regarded as a type of non-standard ARMA model (Tsay, 2013). The ARMA(p, q) is given by:

$$X_t = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=1}^q \beta_j \varepsilon_{t-j} \quad (9)$$

Where X_t is the time-series of data, the Autoregressive (AR) part involves regressing on the own lagged values of the series X_{t-i} and the Moving Average (MA) part involves a linear combination of past errors e_{t-j} . The e_t is assumed to be i.i.d. with zero mean and variance one.

To examine the appropriate order of p and q in the conditional mean model ARMA, the syntax “auto.arima” is used in R. The syntax fits all the different combinations of AR and MA orders in the ARMA to find the best model for the data. The goodness-of-fit is evaluated using the Bayesian Information Criterion (BIC) which is closely related to the Akaike Information Criterion (AIC). The advantage of BIC to AIC is that it involves a penalty term that punish higher orders to account for over-fitting the parameters of the model (Schwarz, 1978). The BIC is formally given by:

$$BIC = -2\log L(\theta) + k\log(n) \quad (10)$$

Where $\log L(\theta)$ is the maximized value of the likelihood function of the model, k the number of parameters and n the sample size. The model with the lowest BIC value is the best. The preferred model by the BIC for all three indices are an ARMA(0,0) with non-zero mean, the results can be seen in appendix “A1. Results of fitting ARMA models using BIC”. Thus, the ARMA(0,0) with non-zero mean is used when forecasting the volatility.

3.6. Distribution of the error term

To specify the GARCH models properly, the distribution of the error term e_t must also be correctly specified to ensure the goodness-of-fit. This is to make sure that the sample data match the true population in the best possible way. Going back to the parametric model of ε_t presented in section “3.1. ARCH”, it is given by:

$$\varepsilon_t = \sigma_t e_t \quad (11)$$

Where the error term e_t should be identically and independently distributed with mean zero and variance one, given by $e_t \sim I.I.D. (0,1)$. However, the distribution must be specified. The three distributions that are considered in this paper are the Gaussian (normal) distribution, the Student’s t-distribution, and the Generalized Error Distribution (GED) which is standard procedure to test for in GARCH models (Chuang et al., 2007). An evaluation of the distribution

of the error term is performed in section “4.1. Diagnostics test”, where the Student’s t-distribution is found to be the appropriate distribution for all three indices. This is commonly the case when working with GARCH models (Ibid.). Since the Student’s t-distribution is found to be the best in this paper, this is what will be explained below and in section 3.7. of the ML Estimation. The Student’s t-distribution is formally given by:

$$f(\varepsilon) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})} \left(1 + \frac{\varepsilon^2}{v}\right)^{-\left(\frac{v+1}{2}\right)} \quad (12)$$

, where

$$-\infty < \varepsilon < \infty$$

The Γ is the gamma function and v the degrees of freedom. When $v \rightarrow \infty$, the student’s t-distribution approach a normal distribution. Hence, the lower degrees of freedom the fatter the tails of the distribution.

3.7. Maximum Likelihood Estimation

The GARCH family models are estimated using the Maximum Likelihood Estimation (MLE) provided by the rugarch package in R. The MLE is based on the standard likelihood function but instead it maximizes the logarithm of the likelihood function. The MLE with a student’s t-distribution is formally given by:

$$\log[L\theta|I_{n-1}] = n \log \left[\frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi(v-2)}\Gamma(\frac{v}{2})} \right] - \frac{1}{2} \sum_{i=1}^n \log(\sigma_i^2) - \frac{v+1}{2} \sum_{i=1}^n \log \left[1 + \frac{\varepsilon_i^2}{\sigma_i^2(v-2)} \right] \quad (13)$$

Where L denotes the function of parameters, θ the set of parameters that are to be estimated, I_{n-1} the information available at time t , the $[\cdot]$ in brackets captures the Student’s t-distribution where Γ is the gamma function and v the degrees of freedom, and the σ_i^2 is the conditional variance substituted recursively by regard to the respective model. Maximizing the logarithm of the likelihood function yields a good in-sample fit for the parameters of the GARCH models and it is the standard procedure for estimating GARCH models (Ghalanos, 2017).

4. Data & Methodology

The closing-, high- and low price is collected from Yahoo Finance for the indices OMX Stockholm PI, OMX Helsinki 25 and OMX Copenhagen 20. The time-period is ranging from 2015-01-05 to 2021-04-07, making a total of 1572 observations for the three indices respectively. The indices were chosen since they are the only ones with daily observations available for the time-period. Moreover, the time-period from 2015 is chosen to not capture any unnecessary leptokurtosis from previous crises and since it is common practice to use an in-sample that is at least as long as the forecast horizon (Alford & Boatsman, 1995; Figlewski, 1997). The data is then divided into two parts, a before-pandemic sample ranging from 2015-01-05 to 2020-02-19 referred to as the in-sample and a during-pandemic sample ranging from 2020-02-19 to 2021-04-07 referred to as the out-of-sample. The variable that will be used is the daily log-returns, which is specified:

$$r_t = \log(p_t^c) - \log(p_{t-1}^c) \quad (14)$$

Where p_t^c is the closing price at time t for OMXSPI, OMXH25 or OMXC20. The in-sample daily log-returns is used for estimation and forecasting the volatility in the out-of-sample, i.e., the during-pandemic period.

Figure 1. Daily closing price and log-returns for OMXSPI, OMXH25 and OMXC20

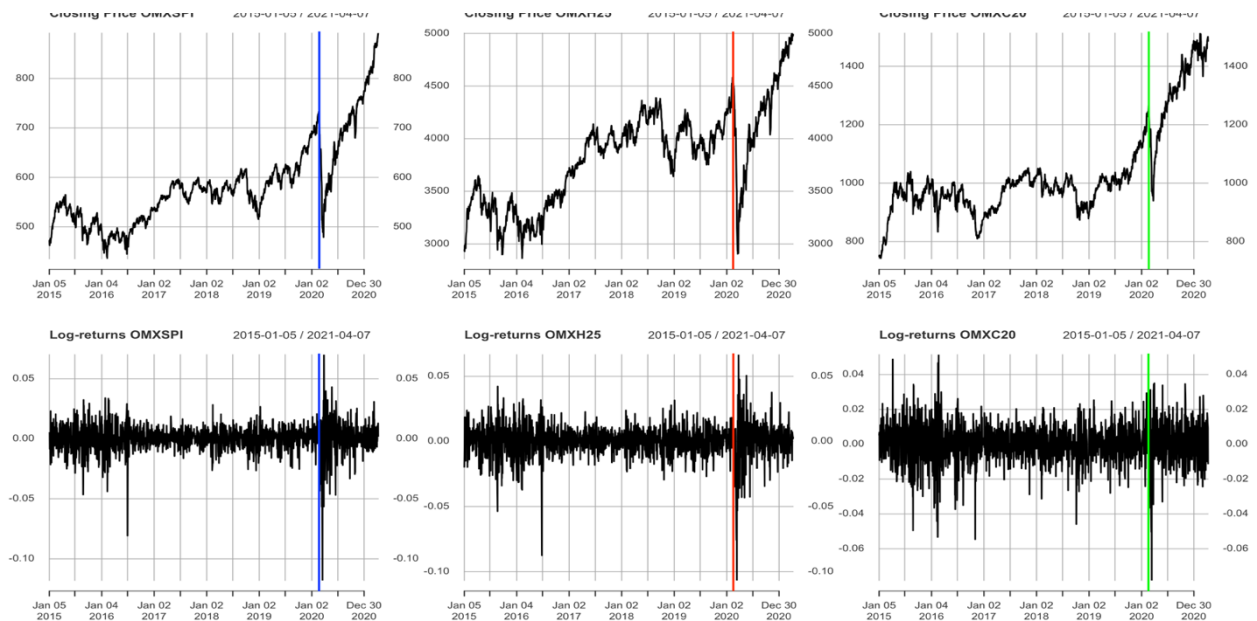


Figure 1 display the daily closing price and daily log-returns for the indices. The colored vertical lines represent the date 2020-02-19, which is when COVID-19 triggered a freefall in the prices

of the indices as seen in the plots of the closing prices. The observations before the colored lines are the in-sample and the observations after are the out-of-sample. One can graphically see the increase in the variance of the series by the plots of the daily log-returns. OMXSPI and OMXH25 seem to be more affected by COVID-19, causing a larger price drop and a bigger increase in variance that converges back towards the “normal” level during the out-of-sample observations. OMXC20, on the other hand, experience a smaller price drop and a variance that is more constant after 2020-02-19. However, a small increase in the variance. This might be explained by that OMXSPI includes all shares listed on the OMX Nordic Exchange Stockholm and OMXH25 includes a higher number of shares than OMXC20, therefore better represents how the total stock exchange is affected by the pandemic.

Table 1. Descriptive statistics OMXSPI, OMXH25 and OMXC20

In-Sample	Obs	Mean	Std	Min	Max	Skewness	Kurtosis
OMXSPI	1288	0.0003362898	0.009523298	-0.08072009	0.03399269	-0.6869146	5.25597
OMXH25	1288	0.0003131833	0.01055248	-0.08750944	0.04228911	-0.6103065	4.637616
OMXC20	1288	0.0004121263	0.01084418	-0.0545814	0.0514281	-0.3630825	2.716479

Table 1 presents the in-sample descriptive statistics. The mean and standard deviations are similar for the indices. There is a quite large difference between the minimum and maximum values, which is expected since indices sometimes experience a big down or upturn in price. OMXSPI and OMXH25 have experienced bigger price drops than OMXC20 by their larger minimum values. This also causes a higher kurtosis for these two indices. All samples have a negative skew and therefore skewed to the right-modal with an average positive return. Since the stock exchange in general makes a positive return over time and these indices are price indices of development on the respective stock exchange, this is expected. The data of OMXSPI and OMXH25 are moderately skewed and the data of OMXC20 is fairly symmetrical. For further graphical examination, see “A2. density plots” in appendix.

4.1. Diagnostics tests

The GARCH, EGARCH, GJR and APARCH were chosen due to the time frame of this thesis. Additional GARCH models would be preferred. All models are specified with the smallest lag order since this provides adequate results and is the common approach, as stated in section 2. This is also done to make the models more comparable. To examine the goodness-of-fit of the models, tests for serial correlation in the ARMA residuals, ARCH effects (autoregressive conditional heteroscedasticity) and distribution of the error term are performed.

The Ljung-Box test by Ljung and Box (1978) tests the lack of fit of times-series models by examining if there is serial correlation in the residuals of the ARMA(p, q) model. H_0 : The data is independently distributed, vs. H_a : The data is not independently distributed (the model exhibits serial correlation). Going back to the two basic ideas behind the ARCH and GARCH models presented in section “3.1. ARCH”, one was that the lagged values of return should be dependent, but serial uncorrelated. If H_a should be accepted, the model exhibits lack of fit due to serial correlation in the residuals. For a formal explanation of the Ljung-Box test, see appendix “A3. Ljung-Box test”. The results of the test are presented in table 2. The Ljung-Box test statistic Q_2 is not significant, hence the H_0 cannot be rejected at any significance level and the ARMA residuals exhibit no serial correlation.

The Autoregressive Conditional Heteroscedasticity-Lagrange Multiplier (ARCH-LM) test by Engle (1982) is used to test for the presence of ARCH effects in the residuals of the in-sample data. ARCH effects refer to Autoregressive Conditional Heteroscedasticity, i.e., volatility clustering. If ARCH effects are detected, the GARCH models are the appropriate framework for the specific time-series data. In figure 1, plots of log-returns, clusters of volatility can graphically be suspected since the series is more volatile in some periods and less in others. The hypotheses of the LM-test are H_0 : No ARCH(q) effects in the residuals, vs. H_a : ARCH(q) disturbances in the residuals. For a formal explanation of the test, see appendix “A4. Lagrange-Multiplier test”. The results can be seen in table 2. H_0 can be rejected for all three in-samples, stating that there are ARCH(q) disturbances in the residuals up to the twelfth lag at a 1% significance level.

Table 2. Lagrange-Multiplier test and Ljung-Box test

In-Sample	Q_2	ARCH-LM(12)
OMXSPI	0.41278	429***
OMXH25	0.23019	394***
OMXC20	0.56669	198***

Note: Q_2 presents the Ljung-Box test statistic and ARCH-LM(12) presents the LM test statistic.

* significant at 10% significance level, ** significant at 5% significance level, *** significant at 1% significance level.

To test for the appropriate distribution of the error term e_t , as described in “3.6. Distribution of the error term”, Quantile-Quantile (QQ) plots are performed for the Gaussian (normal) distribution, the Student’s t-distribution, and the Generalized Error Distribution (GED). Larger samples generally approach normality, but as stated by Chuang et al. (2007) the student’s t-

distribution is common for GARCH models. This is because the financial time series generally have fat tails, positive average mean and skewed to the right-modal. The student's t-distribution is found to be the best fit for all three indices by the QQ-plots in appendix "A4. Quantile-Quantile plots", where a further explanation of the QQ-plots can be found.

4.2. Out-of-sample forecasting method

To forecast the volatility in the out-of-sample, a one-step-ahead recursive window forecast is used. To explain the methodology, consider the GARCH(1,1). The estimated model is:

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\varepsilon}_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2 \quad (15)$$

, where

$$t = 1, \dots, T$$

The T denotes the last day in the in-sample, in this case 2020-02-19. The start of the one-step-ahead volatility forecast is:

$$\hat{\sigma}_{T+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\varepsilon}_T^2 + \hat{\beta}_1 \hat{\sigma}_T^2 \quad (16)$$

At time T all the past values of $\hat{\varepsilon}_T^2$ and $\hat{\sigma}_T^2$ are known. In the one-step-ahead forecast, the $\hat{\varepsilon}_T$ is observable since it is the last estimated residual in the GARCH. To continue forecast the volatility in the two-step-ahead period, $T+2$, the expectations of $\hat{\varepsilon}_{T+1}$ is used in the forecast (Alexander, 2008). This process proceeds up to the last observation in the out-of-sample. The same methodology is applied for all the GARCH models examined in this paper, with regard to their different specifications. There are two common approaches when forecasting, either the rolling or recursive approach (Stock and Watson, 2003; Fama and MacBeth, 1973). The rolling approach makes use of a fixed window of data to re-estimate the parameters, whereas the recursive specification makes use of an expanding window to re-estimate the parameters. In this paper, the recursive-window approach is used. Hence, the predictions in the prior time-period are used as inputs to predict the following period and the window of data that is used to re-estimates the parameters expand as the model proceeds to forecast the next day. An advantage of the recursive window strategy is that only one model is required and re-estimated which saves significant computational time for larger time-series, with a large in-sample and a long out-of-sample period to be predicted (Hyndman, 2012).

4.3. Realized volatility

To compare the predicted volatility by the GARCH models, a realized volatility proxy needs to be calculated. To calculate the most precise realized volatility, cumulative squared intra-day returns are needed which requires high-frequency data. Collecting high-frequency data is both time consuming and costly and it is therefore difficult to find since it is most often available for short time-periods or non-existent (Dacorogna et al., 2001). This paper makes use of the High-Low Range proxy first introduced by Parkinson (1980), which is given by:

$$v_t = \ln \left(\frac{H_t}{L_t} \right) \quad (17)$$

Where H_t is the daily high price at time t and L_t is the daily low price at time t . This range-equation incorporates how the price has fluctuated throughout the day. Another commonly used realized volatility proxy is the daily squared returns proxy by Patton (2006). The disadvantage of the daily squared return proxy is that it can be quite noisy for the true conditional variance (Patton and Sheppard, 2009). According to Andersen and Bollerslev (1998) the High-Low Range proxy attains an equally good proxy as an intra-daily high-frequency realized volatility proxy based on intervals of two to three hours. Hence, the decision to use the High-Low Range proxy was made.

4.4. Forecast evaluation

To evaluate the out-of-sample volatility forecasts, statistical loss functions are used. A loss function summarizes the forecast errors, the difference between the realized (actual) and the forecasted volatility, which provides a measurement of how well the prediction matches the observed data. This way the accuracy of the out-of-sample forecast can be examined since the smallest deviation from zero provides the most accurate forecast. According to Diebold and Lopez (1996) and Lopez (2001) selecting the most appropriate loss function is not obvious since different loss functions penalize differently depending on the equation of the measure. The forecasts obtained in this paper are examined by five different measures, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2 LOG and Mean Absolute Percentage Error (MAPE) suggested by Bollerslev et al. (1994) and Poon and Granger (2001) and the Mean Absolute Error (MAE) suggested by Hansen and Lunde (2005). The advantage of several measures resides in selecting the optimal forecast model (Andersen and Bollerslev, 1998). The loss functions are given by:

$$MSE = n^{-1} \sum_{t=1}^n (v_{t+1}^2 - \hat{\sigma}_{t+1}^2)^2 \quad (18)$$

$$RMSE = \sqrt{n^{-1} \sum_{t=1}^n (v_{t+1}^2 - \hat{\sigma}_{t+1}^2)^2} \quad (19)$$

$$R^2LOG = n^{-1} \sum_{t=1}^n \left(\log\left(\frac{v_{t+1}^2}{\hat{\sigma}_{t+1}^2}\right) \right)^2 \quad (20)$$

$$MAE = n^{-1} \sum_{t=1}^n |v_{t+1}^2 - \hat{\sigma}_{t+1}^2| \quad (21)$$

$$MAPE = 100n^{-1} \sum_{t=1}^n \left| \frac{(v_{t+1}^2 - \hat{\sigma}_{t+1}^2)}{v_{t+1}^2} \right| \quad (22)$$

Where n is the number of days in the out-of-sample, v_{t+1}^2 is the volatility proxy and $\hat{\sigma}_{t+1}^2$ is the one-day-ahead volatility forecast. The MSE is the average of the squared errors, which is a commonly used measurement (James et al., 2013). Due to the square of the errors that cancel out negative values, the MSE is sensitive to outliers and a prediction that differs substantially from the observed value will result in a large MSE . $RMSE$ is the square root of MSE , so that the effect of each error on $RMSE$ is proportional to the size of the squared error. Consequently, the $RMSE$ is sensitive to outliers but not as penalized as in the case of MSE . R^2LOG is similar to MSE (Hansen and Lunde, 2005). It still assigns a higher weighting to outliers but due to the logarithmic function it is not as penalized as in the case of MSE . MAE is the average of the absolute values of the errors. In this case, the negative values are made positive by taking the absolute value resulting in that each error influence MAE in direct proportion which makes it less sensitive to outliers. $MAPE$ is the average percentage of the absolute error and the advantages of $MAPE$ are a more intuitive interpretation and robustness to outliers. The disadvantages of $MAPE$ are for example singularity problem³ and scale sensitivity, i.e., it takes on extreme values when the realized volatility is close to zero (Hyndman and Athanasopoulos, 2008).

³ Singularity problem of the form “one divided by zero” for example.

5. Estimation results

The in-sample fit of the models will not be discussed since it is not the main focus of this paper and as stated in section 2 does not necessarily improve the out-of-sample forecast. For interested readers, the coefficients and their respective significance can be found in appendix “A5. In-sample coefficients”. The main focus of this paper is the out-of-sample forecasting performance. The forecasting performance is evaluated by *MSE*, *RMSE*, *R²LOG*, *MAE* and *MAPE* and thus the model with the smallest loss value indicates the model that has the closest predicted volatility to the realized volatility, which in this paper is represented by the High-Low Range proxy. The results are seen in tables 3, 4 and 5 respectively for the three indices, where the smallest loss value is highlighted in grey and the highest loss value is highlighted in red.

Table 3. Evaluation of forecast OMX Stockholm PI

Model	MSE	RMSE	R2LOG	MAE	MAPE
GARCH(<i>I,I</i>)	4,59839E-07	0,000678114	0,182518721	0,000260206	92,7163941
EGARCH(<i>I,I</i>)	4,33401E-07	0,000658332	0,169812995	0,000229373	61,121123
APARCH(<i>I,I</i>)	4,29881E-07	0,000655653	0,171965038	0,000244605	70,2202051
GJR(<i>I,I</i>)	4,09215E-07	0,000639699	0,181588051	0,000250325	79,0973623

Table 4. Evaluation of forecast OMX Helsinki 25

Model	MSE	RMSE	R2LOG	MAE	MAPE
GARCH(<i>I,I</i>)	3,36089E-07	0,000579732	0,166725721	0,000256516	78,6140964
EGARCH(<i>I,I</i>)	3,54709E-07	0,000595574	0,152302032	0,000249016	61,0609753
APARCH(<i>I,I</i>)	2,4285E-07	0,000492798	0,155135591	0,0002278	63,5648028
GJR(<i>I,I</i>)	2,34033E-07	0,00048377	0,156685546	0,000224628	66,8969049

Table 5. Evaluation of forecast OMX Copenhagen 20

Model	MSE	RMSE	R2LOG	MAE	MAPE
GARCH(<i>I,I</i>)	4,02333E-07	0,000634297	0,154371099	0,000255578	58,84902688
EGARCH(<i>I,I</i>)	3,93397E-07	0,000627214	0,154231049	0,000258091	58,45854324
APARCH(<i>I,I</i>)	3,48511E-07	0,000590348	0,158524177	0,00024771	55,62899828
GJR(<i>I,I</i>)	3,44153E-07	0,000586646	0,156969772	0,000250078	56,33920525

The difference between the best and worst fitted model is in general small. For instance, the difference between the forecasting performance of GARCH and GJR for OMXSPI by the evaluation measure MSE is 5,0614E-8. Most often the GARCH is worse off forecasting the

volatility. This indicates that the innovations of EGARCH, APARCH and GJR to capture the asymmetry, or leverage effect, achieves better predictions when forecasting during a crisis on Nordic indices. The GJR is preferred by the *MSE* and *RMSE* regardless of which index. Since *MSE* and *RMSE* are sensitive to outliers, the GJR is closer in its prediction of outliers than the other models according to these two measures. In contrary, EGARCH is the model preferred by *R²LOG* regardless of index. Comparing the results of each index, the EGARCH is the preferred model by the measures for OMXSPI, it is a tie between EGARCH and GJR for OMXH25 and a tie between GJR and APARCH for OMXC20. The two models that on average predict the realized volatility best according to the measures are GJR and EGARCH, which are preferred seven and six times respectively.

To examine the forecasting performance graphically the out-of-sample forecasts of the respective index are plotted, see appendix “A6. One-step-ahead recursive window forecast”. Noticeable is that all models tend to underpredict the realized volatility and all predictions follow the fluctuations of the realized volatility, but they react a bit delayed to the fluctuations. The GARCH and EGARCH are more cautious in their predictions during the more volatile period in comparison to GJR and APARCH, which on average are closer to the realized volatility in the more volatile period. The GJR and APARCH predictions also fluctuate closely together and in the tables the evaluation measures give approximately the same loss values for the APARCH and GJR, especially for the forecasts on OMXH25 and OMXC20. During the less volatile period, when the increased volatility of the crisis has started to settle, all predictions fluctuate closely together. This is most evident in the figure of OMXC20 with the lower variance.

To summarize the results, the difference between the best and worst predictions is on average small. The asymmetric extensions to the GARCH are important when volatility forecasting during a crisis on Nordic indices and the two models that on average predict closest to the realized volatility according to the evaluation measures are GJR and EGARCH, which are preferred seven and six times respectively.

6. Discussion

The result of this paper is somewhat aligned with the conclusions and discussions of the previous research in section two. The GARCH is regardless of index worse off in volatility forecasting during a crisis, such as COVID-19, for the Nordic indices. The previous research has had a more general approach of forecasting volatility under “normal” circumstances. Since this paper focuses on forecasting a crisis, the findings of this paper therefore further strengthen the importance of using GARCH models with asymmetric extensions such as EGARCH, APARCH and GJR. This is because the leverage effect is more present during a crisis and including the asymmetry when forecasting a crisis will result in better forecasting performance. The graphical examination with the increase in volatility due to the large price drops at the beginning of the series also gives evidence that the leverage effect is an important characteristic of volatility in financial time series.

According to the loss functions, the superior models forecasting the crisis in this study is the GJR and EGARCH. The GJR uses a binary variable to enable the asymmetric effects and the EGARCH uses an asymmetric response function. These asymmetric extensions make the predicted volatility on average closer to the realized volatility in comparison to the other two models. Interesting is the APARCH model’s ability of transforming to the different GARCH models depending on the value of the parameter δ . In the plots in appendix “A6. One-step-ahead recursive window forecast”, the GJR and APARCH predictions fluctuate closely together. Furthermore, the difference of the two models in loss value is significantly small, especially in table 4 and 5. It might therefore be suspected that the parameter δ of the APARCH is close to two, which is when APARCH becomes GJR. Nonetheless, the APARCH does not seem to be able to achieve a predicted volatility as close to the realized as in the case of GJR by regard to the included performance measures.

The results of this paper are different from the results found by Angabini and Wasiuzzaman (2011) which evaluated the forecasting performance of GARCH, EGARCH and GJR on the Kuala Lumpur Composite Index during the financial crisis of 2008. The authors found a significant increase in volatility and leverage effect due to the crisis but nonetheless the forecast of the symmetric GARCH was the superior one. This is a quite strange results, since the intention of the asymmetric extensions is to capture the important characteristic of leverage effect in asset returns volatility which is evident in the paper. Because the leverage effect is

more prominent in the event of a crisis it is furthermore odd that the authors find GARCH to be the superior model while Shamiri and Abu Hassan (2007) find GJR to be best model under “normal” circumstances. In contrary to the results of Angabini and Wasiuzzaman (2011), the results of this paper are more aligned with theory. The increase in volatility due to the crisis COVID-19 is evident both in Figure 1 and the graphical examination in appendix “A6. One-step-ahead recursive window forecast”. Due to this, the symmetric GARCH is on average outperformed by the asymmetric models. Moreover, it is difficult to compare the results of this paper to the results of the other previous research since they are done on a series when there is no crisis. These researchers found the asymmetric models to be the “most promising” when forecasting on a series under “normal circumstances” and the findings of this paper show that this is also the case when forecasting during a crisis.

Regardless of the findings, some things are important to highlight. The difference between the predictability of the models is small in practice and the difference between the best and worst model can be as small as $5,0614E-8$, as in the example of the MSE for GJR and GARCH on OMXSPI. It can also be the case that one evaluation measure suggests a particular model as the best while another measure suggests the same model to be the worst for the same index. The evaluation measures can disagree since they penalize the forecasts differently depending on the loss function. Consequently, this emphasizes the importance of choosing an adequate loss function that represents the intended use of the forecast. The intended use could for example be that one wants the forecast to be able to predict the large outliers, the peaks of the volatility, or if the forecast on average should predict the volatility fluctuations better but without the magnitude of the peaks in volatility. The graphical examination of the forecast shows that the GJR and APARCH on average predict the volatility closer to the realized volatility in the more volatile period of the series. If the intended use of the forecast was to find when the volatility peaks, one should rely on an evaluation measure that chooses a model that reflects this strategy. If the forecast should be of advantage to risk managers or private investors, one needs to know which loss function reflects which strategy. It can therefore be difficult to conclude which forecasting model outperforms another. This can be seen in the previous research since in some cases it is a subjective decision to choose a model that fits the strategy and therefore the conclusions that are drawn by the authors’ is that the model is the “most promising”. In other cases, it can be difficult to pick one model over another due to contradictory results of the evaluation measures such as in the paper by Brooks (1998).

7. Conclusion

This paper evaluates the volatility forecasting performance of the symmetric GARCH and asymmetric EGARCH, GJR and APARCH on the Nordic indices OMXSPI, OMXH25 and OMXC20. The purpose is to find which model achieves a predicted volatility closest to the realized volatility during a crisis. The realized volatility is in this paper represented by the High-Low Range proxy by Parkinson (1980). The performance of the forecasts is evaluated by the performance measures MSE, RMSE, R^2 LOG, MAE and MAPE and the superior forecasting model is chosen based on which model on average achieves the best volatility predictions.

All GARCH models in this paper are configured with the Student's t-distribution, the conditional mean process $ARMA(0,0)$ with non-zero mean and the smallest lag order. The graphical examination of the out-of-sample recursive forecasts shows that all models tend to underestimate the realized volatility. Furthermore, the predictions react a bit delayed but follow the fluctuations of the realized volatility. The results of this paper are that the symmetric $GARCH(1,1)$ on average has the worst volatility forecasting performance when forecasting a crisis on Nordic indices. However, the difference between the predictability of the models is in practice small and if one wants to know which forecasting model fits the intended use of the forecast, one should evaluate the forecasts with the corresponding evaluation measure. The superior forecasting models are the $GJR(1,1)$ and $EGARCH(1,1)$ when forecasting a crisis on Nordic indices according to the evaluation measures of this paper. The results therefore imply that the asymmetric extensions improve the forecasting performance of the GARCH models when forecasting a crisis and furthermore strengthens the leverage effect as an important characteristic to account for when forecasting asset returns volatility.

7.1. Suggestions of future research

Due to the time frame of this thesis, only the forecasting performance of four models on three Nordic indices was examined. Evaluating other symmetric and asymmetric GARCH models, additional Nordic indices, other data frequencies and forecast horizons could be of interest to achieve an even more reliable result. Furthermore, the evaluation measures penalize differently depending on the underlying loss function. Due to this, one could also further investigate which evaluation method provides the superior out-of-sample forecast model by regard to the intended use of the forecast. If the results should be of any advantage to risk managers or private investors, one needs to know which loss function reflects which strategy.

References

- Alberg, D., Shalit, H. and Yosef, R. (2008). *Estimating stock market volatility using asymmetric GARCH models*. Applied Financial Economics, 18:15, 1201-1208, DOI: 10.1080/09603100701604225.
- Alexander, C. (2008). *Market Risk Analysis*. Practical Financial Econometrics. John Wiley & Sons Ltd, 2.
- Alford, A.W. and Boatsman, J.R. (1995). *Predicting long-term stock return volatility: Implications for accounting and valuation of equity derivatives*. Accounting Review, 70, 4, 599-618.
- Andersen, T. and Bollerslev, T. (1998). *Answering the skeptics: yes, standard volatility models do provide accurate forecasts*. International Economics Review, 39, 4, 885-905.
- Angabini, A., and Wasiuzzaman, S. (2011). *GARCH models and the financial crisis-a study of the Malaysian*. The International Journal of Applied Economics and Finance, 5(3), 226-236.
- Black, F. (1976). *Studies of stock market volatility changes*. Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177-81.
- Black, F., Scholes, M. (1973). *The pricing of options and corporate liabilities*. Journal of political Economy, 81: 637 – 654.
- Brailsford, T.J. and Faff, R.W. (1996). *An evaluation of volatility forecasting techniques*. Journal of Banking and Finance, 20, 3, 419-438.
- Bollerslev, T. (1986). *Generalized Autoregressive Conditional Heteroscedasticity*. Journal of Econometrics, 31, p. 307-327.
- Bollerslev, T., Engle, R.F. and Nelson, D. (1994). *ARCH models*. Handbook of Econometrics, ed. by Engle, R.F. and McFadden D.L., vol. IV.
- Brooks, C. (1998). *Predicting stock market volatility: can market volume help?* Journal of Forecasting, 17, 1, 59-80.
- Chuang, I. Y., Lu, J. R., and Lee, P. H. (2007). *Forecasting volatility in the financial markets: A comparison of alternative distributional assumptions*. Applied Financial Economics, 17.
- Dacorogna, M.M., Gencay, R., Muller, U., Olsen, R.B., Olsen, O.V. (2001). *An introduction to high frequency finance*. Academic Press, New York.
- Diebold, F.X. and Lopez, J.A. (1996). *Forecast evaluation and combination*. Handbook of Statistics, Vol. 14, Statistical Methods in Finance, Maddala GS, Rao CR (eds). North-Holland: Amsterdam; 241-268.
- Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). *A long memory property of stock market returns and a new model*. Journal of Empirical Finance, 1, 83-106.

- Engle R.F. (1982). *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. *Econometrica*. Vol. 50(4).
- Franses, P.H. and Van Dijk, D. (1996). *Forecasting stock market volatility using (non-linear) Garch models*. *Journal of forecasting*, 15, 3, 229-235.
- Figlewski, S. (1997). *Forecasting volatility, Financial Markets, Institutions and Instruments*. New York University Salmon Center, 6, 1, 1-88.
- Fama, E., and Macbeth, J. (1973). *Risk, Return, and Equilibrium Empirical Tests*. *Journal of Political Economy*, 81: 607-636.
- Fung, H-G. and Hsieh, D.A. (1991). *Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies*. Working Paper, Dept of Finance, Fuqua School of Business.
- Ghalanos, A. (2017). *rugarch: Univariate GARCH Models*. R package version 1.3-8. Available online, <https://CRAN.R-project.org/package=rugarch> (05-05-2021)
- Glosten, L. R., Jagannathan, R., Runkle, D. E. (1993). *On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks*. *The Journal of Finance*, 48:1779–1801.
- Gokcan, S. (2000). *Forecasting volatility of emerging stock markets: Linear versus non-linear GARCH models*. *Journal of Forecasting*, 19.
- Hansen, P.R. and Lunde, A. (2005). *A forecast comparison of volatility models: does anything beat a GARCH(1,1)?*. *Journal of Applied Econometrics*, 20:873–889.
- Hayashi, F. (2000). *Econometrics*. New Jersey: Princeton University Press.
- Hyndman, R. J., and Athanasopoulos, G. (2008). *Forecasting: principles and practice*, 2.
- Hyndman, R.J. (2012). *Recursive and direct multi-step forecasting: the best of both worlds*. Department of Economics and Business Statistics, No 19/2012.
- Javed, F. and Mantalos, P. (2013). *GARCH-Type Models and Performance of Information Criteria*. *Communications in statistics. Simulation and computation*. 42(8), page 1917-1933.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to Statistical Learning with Applications in R*. Springer.
- Ljung, G. and Box, G. (1978). *On a Measure of Lack of Fit in Time Series Models*, *Biometrika*, 65, 297-303.
- Lopez, J.A. (2001). *Evaluation of predictive accuracy of volatility models*. *Journal of Forecasting*, 20(1).
- Mandelbrot, B. (1963). *The variation of certain speculative prices*. *Journal of Business*, 36, 394–419.

- Merton, R.C. (1969). *Lifetime portfolio selection under uncertainty: The continuous-time case*. Review of Economics and Statistics, 51: 247 – 257.
- Nelson, D.B (1991). *Conditional Heteroskedasticity in Asset Returns: A New Approach*. *Econometrica*, 59, 347-370
- Parkinson, M. (1980). *The Extreme Value Method for Estimating the Variance of the Rate of Return*. The Journal of Business, 53, 61-65.
- Patton, A. J. (2006). *Volatility forecast comparison using imperfect volatility proxies*. Quantitative Finance Research Centre, University of Technology Sydney, Research Paper 175.
- Patton, A. J., and Sheppard, K. (2009). *Evaluating volatility and correlation forecasts*. In T. Mikosch, J.-P. Kreiß, R. A. Davis & T. G. Andersen (Eds.), *Handbook of financial time series* (pp. 801–838). Berlin, Heidelberg: Springer.
- Poon, S. and Granger, C. (2001). *Forecasting financial market volatility: A Review*. Department of Economics.
- Schwarz, G. E. (1978). *Estimating the dimension of a model*, *Annals of Statistics*, 6 (2): 461–464, doi:10.1214/aos/1176344136, MR 0468014
- Shamiri, A., and M.S.N. Abu Hassan. (2007). *Modeling and forecasting volatility of the Malaysian and the Singaporean stock indices using asymmetric GARCH models and Non-normal densities*. *Malaysian J.Math.Sci.*, 83-102.
- Shamiri, A., Isa, Z. (2009). *Modelling and forecasting volatility of Malaysian stock markets*. *Journal of Mathematics and Statistics*, 3: 234 – 240.
- Stock, J.H. and Watson, M.W. (2003). *Forecasting output and inflation the role of asset price*. *Journal of Economic Literature*, 41: 788-829.
- Teräsvirta, T. (2006). *An Introduction to Univariate GARCH Models*. SSE/EFI Working Papers, Economics and Finance, No. 646.
- Tsay, R. T. (2013). *An introduction to analysis of financial data with R*. Johan Wiley & Sons Ltd.

APPENDIX

A1. Results of fitting ARMA models using BIC

OMXSPI

ARIMA(2,0,2) with non-zero mean : -8308.484
ARIMA(0,0,0) with non-zero mean : -8320.235
ARIMA(1,0,0) with non-zero mean : -8312.783
ARIMA(0,0,1) with non-zero mean : -8313.509
ARIMA(0,0,0) with zero mean : 3661.596
ARIMA(1,0,1) with non-zero mean : -8305.905

ARIMA(0,0,0) with non-zero mean : -8320.235

Best model: ARIMA(0,0,0) with non-zero mean

OMXH25

ARIMA(2,0,2) with non-zero mean : -8323.546
ARIMA(0,0,0) with non-zero mean : -8340.003
ARIMA(1,0,0) with non-zero mean : -8335.805
ARIMA(0,0,1) with non-zero mean : -8333.085
ARIMA(0,0,0) with zero mean : 3609.621
ARIMA(1,0,1) with non-zero mean : -8328.683

ARIMA(0,0,0) with non-zero mean : -8340.003

Best model: ARIMA(0,0,0) with non-zero mean

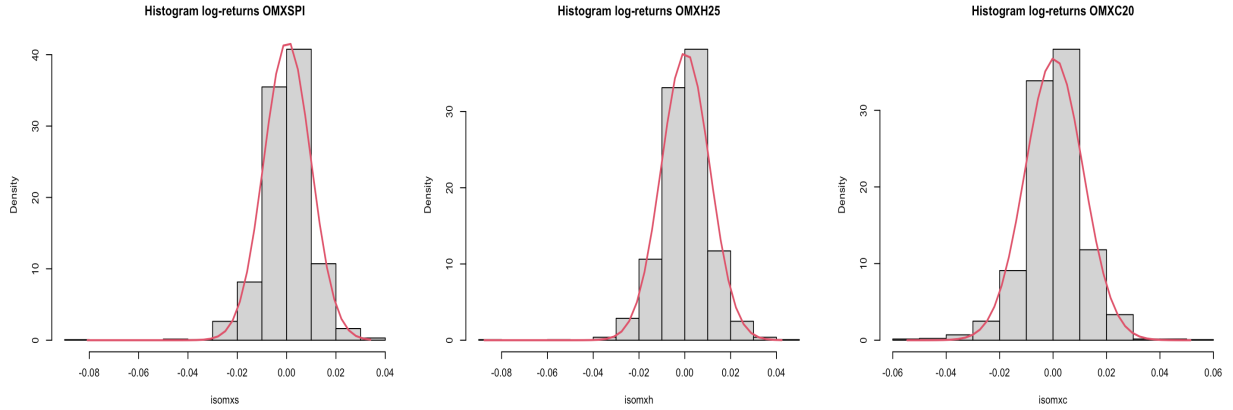
OMXC20

ARIMA(2,0,2) with non-zero mean : -7961.156
ARIMA(0,0,0) with non-zero mean : -7985.647
ARIMA(1,0,0) with non-zero mean : -7978.237
ARIMA(0,0,1) with non-zero mean : -7979.046
ARIMA(0,0,0) with zero mean : 3661.436
ARIMA(1,0,1) with non-zero mean : -7971.08

ARIMA(0,0,0) with non-zero mean : -7985.647

Best model: ARIMA(0,0,0) with non-zero mean

A2. Density plots



The density plots of log-returns for the three indices all have a negative skew, resulting in skewed to the right modal with average return above positive values. OMXSPI and OMXH25 are moderately skewed and OMXC20 is fairly symmetrical.

A3. Ljung-Box test

The Ljung-Box test by Ljung and Box (1978) measures the lack of fit of the model by testing for serial correlation in the ARMA residuals. The test is formally given by:

$$Q = n(n+2) \sum_{k=1}^h \frac{\widehat{\rho}_k^2}{n-k}$$

Where n is the sample size, $\widehat{\rho}_k^2$ the sample autocorrelation at lag k and h the lags that are tested. For a $ARMA(p,q)$ the lags should be set to $h - p - q$, where p and q is the order of the ARMA.

The Box-Ljung tests the hypotheses:

H_0 : The data are independently distributed

H_a : The data are not independently distributed (the model exhibits serial correlation)

A4. Lagrange-Multiplier test

The Lagrange-Multiplier test is configured using an auxiliary regression based on an $AR(q)$ process for the squared innovation of an asset, ε_t^2 . The auxiliary regression the LM-test is constructed from is given by:

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

The hypotheses are:

$H_0: \alpha_0 = \alpha_1 = \dots = \alpha_q = 0$

$H_a: \alpha_0 \neq 0$ or $\alpha_1 \neq 0$ up to $\alpha_q \neq 0$

More general the hypotheses are: No ARCH(q) effects in the residuals vs. H_a : ARCH(q) disturbances in the residuals, as given in the text.

The test statistic is:

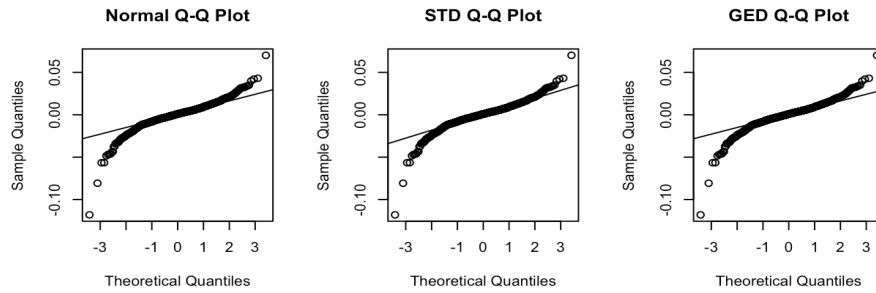
$$LM = T \cdot R^2 \sim X^2(q)$$

Where T is sample size and R^2 is computed from the auxiliary regression.

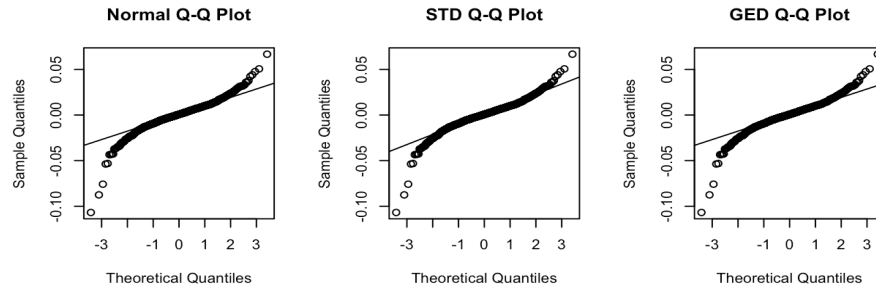
A4. Quantile-Quantile plots

The QQ-plots are constructed using the syntax “qqnorm” in R. Quantiles, or also referred to as percentiles, are points in the sample data below which a certain proportion of the data falls. In this case, the QQ-plots takes quantiles of the in-sample data, sorted in ascending order, and plots it against quantiles from a standard normal distribution.

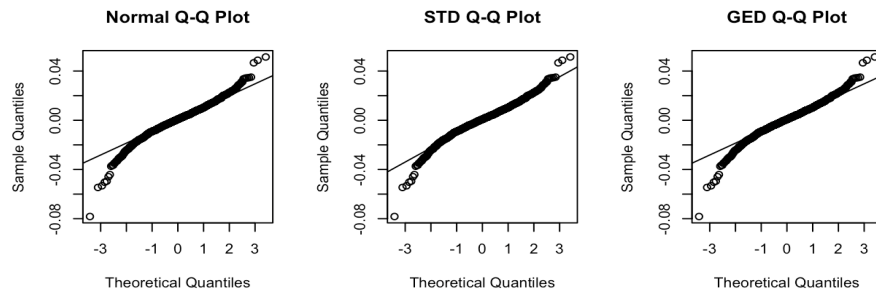
OMXSPI



OMXH25



OMXC20



A5. In-sample coefficients

OMXSPI

Model	α_0	α_1	β_1	γ_1	δ
GARCH	0,000	0,132***	0,831***		
EGARCH	-0,434***	-0,172***	0,952***	0,132***	
GJR	0,000***	0,000	0,83***	0,238***	
APARCH	0,000	0,063**	0,828***	2,65***	0,567***

OMXH25

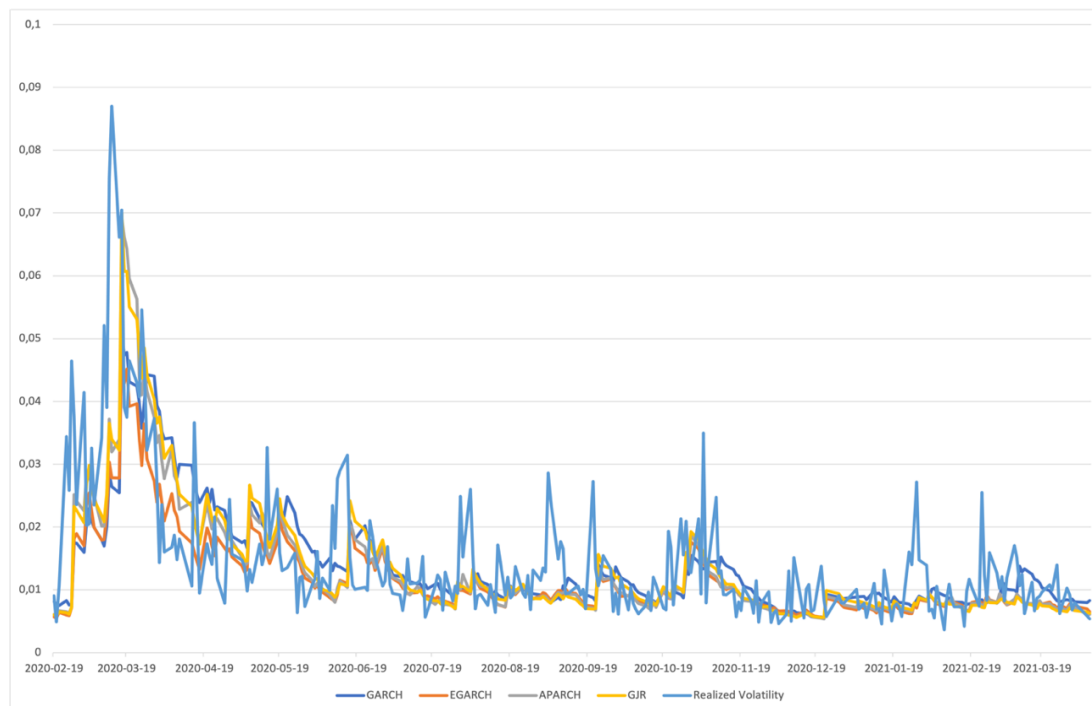
Model	α_0	α_1	β_1	γ_1	δ
GARCH	0,000	0,131***	0,833***		
EGARCH	-0,506***	-0,122***	0,946***	0,197***	
GJR	0,000***	0,031***	0,823***	0,191***	
APARCH	0,000	0,101***	0,786***	2,692***	0,344***

OMXC20

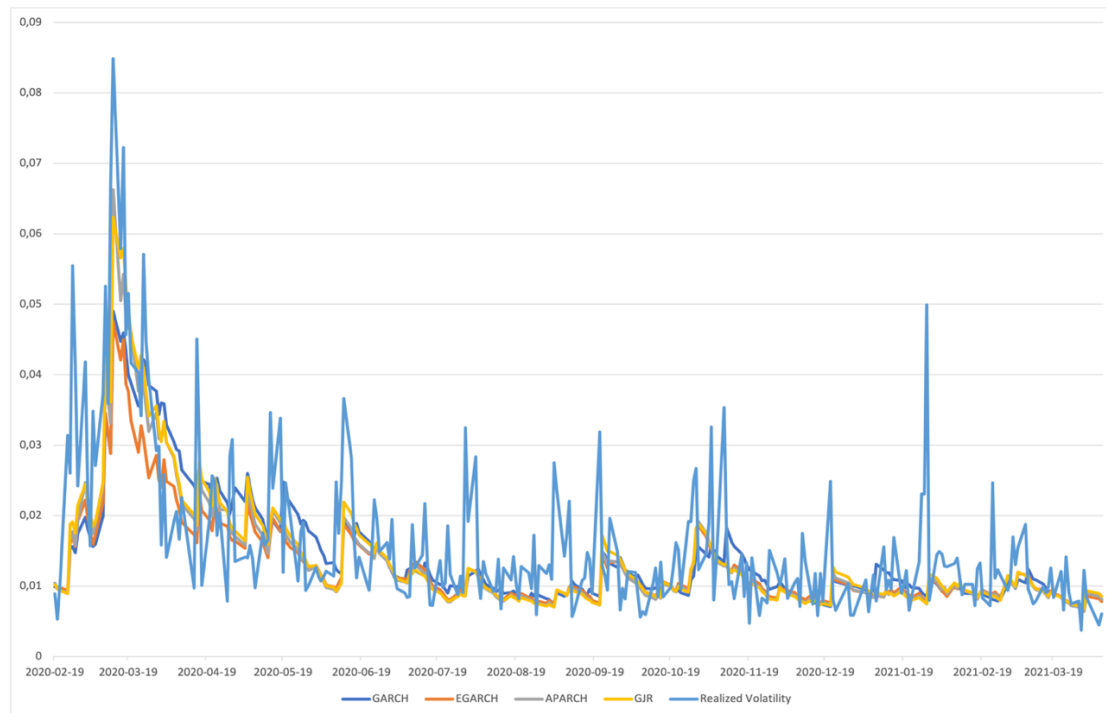
Model	α_0	α_1	β_1	γ_1	δ
GARCH	0,000	0,091***	0,866***		
EGARCH	-0,473***	-0,099***	0,949***	0,169***	
GJR	0,000***	0,0278**	0,824***	0,141***	
APARCH	0,000	0,046**	0,856***	2,976***	0,241**

A6. One-step-ahead recursive window forecast

Out-of-sample forecast OMX Stockholm PI



Out-of-sample forecast OMX Helsinki 25



Out-of-sample forecast OMX Copenhagen 20

