LUND UNIVERSITY
School of Economics and Management

# Predicting Stock Index Volatility Using Artificial Neural Networks

An empirical study of the OMXS30, FTSE100 & S&P/ASX200

by

Ola Johnsson

June 2018

Master's Program in Economics

Supervisor: Dag Rydorff

# Abstract

In this thesis I study the performances of artificial neural networks (ANNs) and three various ARCH-type models to predict weekly volatility of the Swedish (OMXS30), the British (FTSE100) and the Australian (S&P/ASX200) major stock indices. The three various ARCH-type models are the GARCH(1,1), the EGARCH(1,1) and the TGARCH(1,1). The purpose is to investigate if ANNs outperform the more traditional ARCH-type models in predicting weekly stock index volatility. An out-of-sample testing methodology is applied to the most recent 20 percent of the data observations, which fully range from 8th February 2008 to 29th December 2017. The metrics used to evaluate the volatility-predicting performances of the different models are the RMSE, the MAE, the MAPE and the out-of-sample sample $R^2$. The results show no evidence of ANN predicting superiority for any of the three stock indices.

Keywords: artificial neural networks, volatility, ARCH-type models

# Acknowledgements

I wish to thank my supervisor Dag Rydorff for his guidance and inputs at the time of writing this thesis. I would also gratefully acknowledge the help of Adrian Mehic with the computing efforts in the software programs.

# Table of Contents

# Glossary

**Algorithm…** is a step by step method of solving a (class of) problem(s).

**Artificial Neural Networks (ANNs)…** are universal and highly flexible function approximators that can map complex and nonlinear functions. This technique is part of the scientific discipline of machine learning.

**Bayesian regularization…** is a mathematical process that updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights and then determines the correct combination so as to produce a network that generalizes well.

**Dynamic network…** is a type of network which allows the output to depend not only on the current input, $y(t)$, but also on the previous inputs, as for example $y(t-1)$.

**Error back-propagation (EBP)…** , or just back-propagation, is a training algorithm which compares the forecast to the actual result and uses the difference to adjust the weights of the connections between the neurons. The goal is to adjust the weights in order to minimize the errors

**Homoscedasticity…** (meaning "same variance") describes a situation in which the error term (i.e., the "noise" or random disturbance in the relationship between the independent variables and the dependent variable) is the same across all values of the independent variables.

**Latent variable…** is a variable that is not directly observed but is rather inferred (e.g. through a mathematical model) from a (class of) variable(s) that is observed (directly measured).

**Levenberg-Marquardt optimization…** is a standard technique for solving nonlinear least squares problems. Least squares problems arise in the context of fitting a parameterized function to a set of measured data points by minimizing the sum of the squares of the errors between the data points and the function.

**Proxy…** is an easily measurable variable that is used in place of a variable that cannot be measured or is difficult to measure.

**Three-layer feed-forward neural network…** is a type of an ANN. It contains three layers of neurons, which are the input, hidden and output layers. These are arranged in a feed-forward manner. Every neuron in the input layer is connected to every neutron in the hidden layer, and the neurons in the hidden layer are similarly connected to the neuron in the output layer. See Figure 2, pp. 11.

**Stationarity…** refers to a stationary process which has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but can briefly be explained as a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations.

**Supervised Learning…** is a method that presents the algorithm with both input data and the subsequent output data. The input data targets the specific output data and the algorithm learns the mapping procedure.

**Unsupervised learning…** is a method, unlike supervised learning, that has no corresponding target output of the inputs. The model, e.g. an ANN, is expected to determine the output by itself.

# 1 Introduction

Volatility refers to the spread of all likely outcomes of an uncertain variable. Typically, in financial markets, it is measured by the standard deviation or variance of returns. Volatility is often used as a crude measure of total risk of financial assets. It is, however, not exactly the same as risk, but it is related. Risk is associated with undesirable outcome, whereas volatility as a measure strictly for uncertainty could be due to a positive outcome. This important difference is often overlooked (Poon, 2005, pp. 1).

Returns on financial assets are often thought of as being nearly unpredictable. However, this "unpredictability" refers to the mean and not to the variance. In many situations the volatility carries valuable information that can be used in estimation and forecasting. It is, for example, very common that financial time series exhibit periods of unusually large volatility followed by periods of relative tranquility. This stylized fact is referred to as volatility clustering (Enders, 2015, pp. 123). Typically, the most volatile periods correspond to major (economic) events such as oil crises, stock market breakdowns or technological changes.

Being able to accurately estimate and predict volatility are crucial in various fields of finance such as option pricing (e.g. Black-Scholes formula), risk management (e.g. Value-at-Risk calculations) and portfolio management (e.g. CAPM). Nowadays, volatility also has become the subject of trading itself. There are now exchange-traded contracts written on volatility (e.g. the VIX index). Financial market volatility, as well, has a wider impact on financial regulation, monetary policy and macroeconomy (Poon, 2005, pp. 15).

The current consensus is that volatility can be predicted using particular time series models. Within this class of models, the GARCH model proposed by Bollerslev (1986) appears to be among the most successful that are currently available. For example, Hansen and Lunde (2005) compares 330 ARCH-type models, and shows that the GARCH(1,1) is not outperformed by more sophisticated models when analyzing exchange rates, although more complex models (i.e. GARCH-extensions) are better when using stock returns data. A potential reason to the underperformance of the GARCH(1,1) model when using stock returns data might be due to another stylized fact, namely that the distributions of returns often are skewed. Franses & van Dijk (1996) states that for some stock indices, returns are skewed to the left. In such scenarios there are higher amounts of negative than positive outlying observations which the symmetric GARCH model cannot cope with. Two modifications to the GARCH model, which explicitly take account of skewed distributions, are the EGARCH model proposed by Nelson (1991) and the TGARCH model proposed by Zakoian (1994).

Among the large array of "new" approaches available for estimating and forecasting volatility of stock market returns, artificial neural networks (ANNs) are gaining in popularity. ANNs are one of the most recently examined techniques within the scientific discipline of machine learning. They are universal and highly flexible function approximators that can map complex and nonlinear functions. As such flexible function approximators they have proven to be powerful methods for pattern recognition, classification and forecasting (Kaastra & Boyd, 1996).

In this thesis I study the performances of artificial neural networks (ANNs) and three various ARCH-type models to predict weekly volatility of the Swedish (OMXS30), the British (FTSE100) and the Australian (S&P/ASX200) major stock indices. The stock indices are chosen based on geographical location and different numbers of stocks included in the indices. The back-propagation, dynamic, three-layer feed-forward neural network represent the ANN structure in this study and the three various ARCH-type models are the GARCH(1,1), the EGARCH(1,1) and the TGARCH(1,1).

*The purpose of this thesis is to investigate if ANNs outperform the more traditional ARCH-type models in predicting weekly stock index volatility.*

The data sample used in this study consists of weekly price index observations of the OMXS30, the FTSE100 and the S&P/ASX200, which ranges from 8th February 2008 to 29th December 2017. The weekly price index observations are converted into logarithmic returns, $r_t$, and due to the fact that volatility is a latent variable, the square of the week's $r_t$ is used as a proxy for volatility. There are in total 516 weekly observations of logarithmic returns, $r_t$, and equally many volatility observations for each of the indices.

To be able to evaluate the volatility predicting performances of the ANNs and the ARCH-type models, an out-of-sample testing methodology is applied. The first 80 percent of the three indices' volatility observations (413 weeks per each) are used to train the networks and estimate the various model parameters. The estimated models are then used to make volatility predictions of the stock indices for the out-of-sample/testing sets which range from 15th January 2016 to 29th December 2017. The out-of-sample/testing sets contain the most recent 20 percent of the volatility observations (103 weeks per each). Subsequently, four types of metrics (RMSE, MAE, MAPE and out-of-sample $R^2$) are used to evaluate the volatility predicting performance of the ANNs and the ARCH-type models by comparing the forecasts of the models to the actual volatility observations.

The results show no evidence of ANN predicting superiority to the ARCH-type models for any of the three stock indices. In fact, ARCH-type models clearly outperform ANNs in predicting the volatility for the OMXS30 and the FTSE100, and are slightly better for the S&P/ASX200.

The remainder of this thesis is structured as follows. Chapter 2 provides a review of the applicable research to this thesis. In chapter 3, the relevant theories regarding the objective of this thesis are presented. Chapter 4 presents the data this study is based on. Chapter 5 introduces the chosen methodology. In Chapter 6, the results of the models are presented and analyzed. Chapter 7 provides a conclusion as well as proposals of further research in the area.

# 2 Previous Research

Modelling and forecasting volatility of financial time series have been the subject of empirical and theoretical research for quite some time. ARCH-type models have been empirically very important for analyzing volatility, starting with the journal article by Engle (1982) on ARCH, and the extension to the GARCH by Bollerslev (1986). Researchers who evaluate the volatility forecasting performance often find the GARCH model to be at least as good as the ARCH model. Akgiray (1989), for example, shows that the GARCH model is superior to the ARCH model, the EWMA model, and historical mean models for forecasting monthly US stock index volatility. Similar results are presented by West & Cho (1995) when using five bilateral weekly exchange rates for the US Dollar. Although, for longer forecasting horizons the models are equivalent.

In general, models that allow for volatility asymmetry come out well in forecasting contests because of the strong negative relationship between volatility and shock (Poon, 2005, pp. 43). Lee (1991), Cao & Tsay (1992) and Heynen & Kat (1994) highlight the EGARCH model for predicting the volatility of stock indices and exchange rates, whereas Brailsford & Faff (1996) and Taylor (2004) show that the GJRGARCH model, which is similar to the TGARCH, outperforms the GARCH model in predicting the volatility of stock indices. The GJRGARCH model is proposed by Glosten, Jagannathan & Runkle (1993).

In 1943 Warren McCullock and Walter Pitts published “A Logical Calculus of Ideas Immanent in Nervous Activity” (McCullock & Pitts, 1943), which laid the theoretical basis for the subsequent development of current ANNs. Since then, the rapid creation of new neural network architectures and training algorithms combined with the rapid increase in power of personal computers have led to the development of much more robust neural networks for financial forecasting (Gately, 1996, pp. 3-11). For certain types of problems, such as learning to interpret complex real-world sensor data, ANNs are among the most effective learning methods currently known (Mitchell, 1997, pp. 81). There has been a considerable amount of research conducted, in various fields, using ANNs for purposes of prediction. Sharda (1994) presents a survey of the research on applications of ANNs, from bank failure prediction to commodity trading, and illustrates that ANNs perform better than traditional statistical techniques in 30 of 42 studies (71%).

Prior research studies concerning ANNs for volatility forecasting purposes show various results. Shaikh & Iqbal (2004) shows that ANNs outperform the Barone-Adesi and Whaley (BAW) model in forecasting the volatility of S&P500 Index futures. The BAW-model is an implied volatility model. Hu & Tsoukalas (1999) examines the forecasting performance of different types of ARCH-type models, an ordinary least squares model, and an ANN applied to the European Monetary System (EMS) exchange rates. The results support the EGARCH specification especially after the foreign exchange crisis of August 1993, whereas the ANN model performed better during the August 1993 crisis especially in terms of root mean absolute prediction error. Brooks (1998) evaluates the forecasting performance of various linear, GARCH, EGARCH, TGARCH and ANN models for daily stock return volatility of an aggregate of all stocks traded on the NYSE. The results indicate that ANN models give a reasonable performance but are hardly worth the effort of producing.

More advanced types of research studies have proposed different types of hybrids between ARCH-type models and ANNs. For example, Roh (2007) shows that a hybrid model of an ANN and EGARCH model clearly outperforms other models for smaller forecasting periods than 10 days for the KOSPI200 index.

# 3 Theoretical Review

This chapter presents the models used in the modelling and forecasting exercise. Section 3.1 describes the ARCH-based models and section 3.2 displays the ANN in a broader context and its architecture.

## 3.1 The ARCH-type Models

### 3.1.1 Symmetric GARCH Models

In financial time series one often observes what is referred to as volatility clustering. In such circumstances, the homoscedasticity assumption is inappropriate. Big shocks (residuals) tend to be followed by big shocks in either direction, and small shocks tend to follow small shocks (Verbeek, 2012, pp. 325). The autoregressive conditional heteroscedasticity (ARCH) model is based on these types of observations. To illustrate this, returns ($r_t$) are first written as

$$r_t = \mu + \varepsilon_t$$

$$\varepsilon_t = v_t\sqrt{h_t} \tag{1}$$

where $\mu$ is the mean, $\varepsilon_t$ has the form described in equation $(1)$ and $v_t \sim D(0,1)$ is white-noise. The distribution $D$ is often taken as normal. The process $v_t$, is scaled by $h_t$, the conditional variance, which in turn is a function of past squared residual returns. In the ARCH($q$) process proposed by Engle (1982),

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 \tag{2}$$

where $\omega > 0$ and $\alpha_1, \dots, \alpha_q \geq 0$ ensure $h_t$ is strictly positive variance. Also, $\alpha_1 + \cdots + \alpha_q < 1$ needs to hold in order to ensure stationarity.

A problem with the ARCH($q$) model is that in practice a large $q$ is often required to capture the phenomenon of volatility persistence in financial markets and $\alpha_1 + \cdots + \alpha_q$ tend to be very close to one (Poon, 2005, pp. 37-38). In order to increase efficiency, one can apply the generalized-ARCH (GARCH) model. In the GARCH model the conditional variance is allowed to depend upon its own lags, which typically reduces the number of ARCH lags required. The GARCH(1,1) due to Bollerslev (1986) is given by:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{3}$$

where $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$ to ensure stationarity (Enders, 2015, pp.128-130). Furthermore, all odd moments of $\varepsilon_t$ in equation (1) equal zero, and hence $\varepsilon_t$ and $r_t$ are symmetric time series with fat tails (Franses & van Dijk, 1996).

Volatility forecasts from a GARCH(1,1) are made by repeated substitutions. From equation (1) we find that the estimate for the expected value for squared residuals is

$$E[\varepsilon_t^2] = h_t E[v_t^2] = h_t \tag{4}$$

The conditional variance, $h_{t+1}$, and the one-step-ahead forecast is known at time $t$,

$$\hat{h}_{t+1} = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 h_t \tag{5}$$

The forecast of a further step ahead makes use of the fact from equation (4). Notice that $E[\varepsilon_{t+1}^2] = h_{t+1}$. This gives

$$\hat{h}_{t+2} = \omega + \alpha_1 \varepsilon_{t+1}^2 + \beta_1 h_{t+1} = \omega + (\alpha_1 + \beta_1)\, h_{t+1} \tag{6}$$

As the forecast horizon, $n$, lengthens the general pattern is

$$\hat{h}_{t+n} = \frac{\omega}{1 - (\alpha_1 + \beta_1)} + (\alpha_1 + \beta_1)^n [\alpha_1 \varepsilon_t^2 + \beta_1 h_t] \tag{7}$$

As long as $\alpha_1 + \beta_1 < 1$, the second term on the RHS of equation (7) slowly dies out and $\hat{h}_{t+n}$ converges to the unconditional variance, $\omega / [1 - (\alpha_1 + \beta_1)\,]$ (Poon, 2005, 38-39).

### 3.1.2 Asymmetric GARCH Models

A problem with a standard GARCH model is that positive and negative shocks have an identical effect upon the conditional variance since their signs are lost upon taking the square. For stock market time series, however, $r_t$ may display significant negative skewness. Franses

& van Dijk (1996) argues that this empirical stylized fact can be explained by observing that stock market breakdowns occur more quickly than stock market booms and because the absolute size of breakdowns are much larger. It could also be explained as “bad” news increase volatility more than “good” news. Two models which allow for the asymmetric effect of news are the exponential-GARCH (EGARCH) model proposed by Nelson (1991) and the threshold-GARCH (TGARCH) model proposed by Zakoian (1994).

The EGARCH$(1, 1)$ model of Nelson (1991) is given by

$$ln(h_t) = \omega + \beta_1 ln(h_{t-1}^2) + \alpha_1 \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \gamma_1 \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} \tag{8}$$

where $\alpha_1$, $\beta_1$ and $\gamma_1$ are constant parameters. $h_t$ depends on both the size and the sign of $\varepsilon_t$. Because the level $\varepsilon_{t-1}/\sqrt{h_{t-1}}$ is included, the EGARCH model is asymmetric as long as $\alpha_1 \neq 0$. Also in this model, when $\alpha_1 < 0$, negative shocks (“bad” news) generate more volatility than positive shocks (Verbeek, 2012, pp. 328). An advantage of modeling $ln(h_t)$ as opposed to $h_t$ is that

$$\hat{h}_t = exp\left(\widehat{ln(h_t)}\right) > 0 \tag{9}$$

even if some of the estimated parameters are negative. Hence, the model does not require nonnegativity constraints. Even if there are some obvious advantages for the EGARCH model, forecasting with the model is a bit involved because of the logarithmic transformation (Enders, 2015, pp. 155-157). For simplicity, the one-step-ahead forecast for the EGARCH$(1, 0)$ model is given by Tsay (2002, pp. 105-106) as

$$\hat{h}_{t+1} = h_t^{2\alpha_1} exp[(1-\alpha_1)\omega] exp[g(\epsilon)] \tag{10}$$

$$g(\epsilon) = \alpha_1 \epsilon_{t-1} + \gamma\left(|\epsilon_{t-1}| - \sqrt{2/\pi}\right) \tag{11}$$

where

$$\epsilon_t = \varepsilon_t / \sqrt{h_t} \tag{12}$$

For multi-step forecast

$$\hat{h}_{t+n} = h_t^{2\alpha_1}(n-1)exp(\tau)\{exp[0{,}5(\alpha_1+\gamma_1)^2]\varphi(\alpha_1+\gamma_1) + exp[0{,}5(\alpha_1-\gamma_1)^2]\varphi(\alpha_1-\gamma_1)\} \tag{13}$$

where

$$\tau = (1-\alpha_1)\omega - \gamma_1\sqrt{2/\pi} \tag{14}$$

and $\varphi(\cdot)$ is the cumulative density function of the standard normal distribution.

The TGARCH model is similar to the GJRGARCH model proposed by Glosten et al. (1993). Brooks (2008, pp. 406) states that the TGARCH model is the same as the GJRGARCH model. This is not true.

The TGARCH$(1,1)$ model of Zakoian (1994) is given by

$$h_t = \omega + \alpha_1|\varepsilon_{t-1}| + \beta_1 h_{t-1} + \delta_1 D_{t-1}|\varepsilon_{t-1}| \tag{15}$$

$$D_{t-1} = \begin{cases} 1 & if\ \varepsilon_{t-1} < 0 \\ 0 & if\ \varepsilon_{t-1} \geq 0 \end{cases}$$

when $\omega > 0$, $\alpha_1 \geq 0$, $\alpha_1 + \delta_1 \geq 0$ and $\beta_1 \geq 0$, the conditional volatility is positive. If $\delta_1 > 0$, negative shocks will have larger effects on volatility than positive shocks (Enders, 2015, pp. 156). For a TGARCH$(1,1)$ model, the one-step ahead forecast is

$$\hat{h}_{t+1} = \omega + \alpha_1|\varepsilon_t| + \beta_1 h_t + \delta_1|\varepsilon_t|D_t \tag{16}$$

and the multi-step forecast is

$$\hat{h}_{t+n} = \omega + \left(\frac{1}{2}(\alpha_1+\delta_1) + \beta_1\right)h_{t+n-1} \tag{17}$$

where repeated multi-step substitution is needed for $h_{t+n-1}$ (Poon, 2005, pp. 42).

## 3.2 Artificial Neural Networks (ANNs)

### 3.2.1 From Artificial Intelligence (AI) to ANNs

Artificial Intelligence (AI) is the broadest term applied to any technique that enables computers to mimic human intelligence. Machine Learning (ML) is a subset of AI that includes the process of automatically discovering patterns and trends in data that go beyond simple analysis. Sophisticated mathematical algorithms are used to segment the data and to predict the likelihood of future events based on past events. The algorithms adaptively improve their performance as the number of input data samples for "learning" increases (Ghatak, 2017, pp. 8). Machine learning techniques are fairly generic and can be applied in various settings and used for various applications, such as web search, spam filters, credit scoring, fraud detection, stock trading and drug design (Silva & Zhao, 2016, pp. 71).

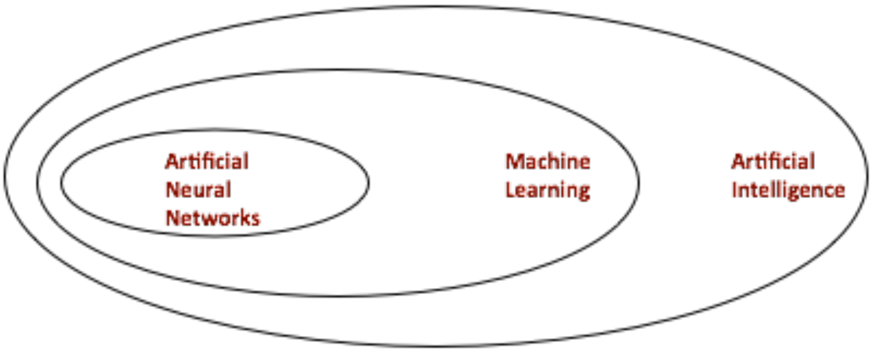



**Figure 1:** This figure is intended to display ANNs in a broader context. The ANN is one of the most recently examined techniques within the scientific discipline of machine learning. Machine learning is itself a subset of artificial intelligence. Artificial intelligence is the broadest term applied to any technique that enables computers to mimic human intelligence. Figure source: Intel Corporation.

There are two different concepts of machine learning, namely unsupervised learning and supervised learning. This thesis focuses on supervised learning which can be explained accordingly. Assume we have input data $x$ and want to find the response $y$. It can be represented by the function $y = f(x)$. Since we do not know the function $f$, given the data and the response, we try to approximate $f$ with a function $g$. The process of trying to arrive at the best approximation to $f$ is through a process known as machine learning (Ghatak, 2017, pp. 8-10).

Within the scientific discipline of machine learning there are various classification techniques in order to predict movements in financial markets. One of the most recently examined techniques is the artificial neural network (ANN). The technique of ANNs can briefly be explained as constructing algorithms that solve problems and make predictions. This is done by learning characteristics and testing a certain algorithm on a sample data to enable entering static equilibrium to a specific output given a new specific data set of inputs (Bishop, 2006, pp. 225-226). The primary advantage for ANNs over more conventional econometric techniques lies in their ability to model complex, possibly nonlinear processes without assuming any prior knowledge about the underlying data generating process (Brooks, 1998). The architecture of ANNs is explained in section 3.2.2.

### 3.2.2 The Architecture of ANNs

The structure of an ANN can look different depending on its task. One of the most popular types of ANNs, and the one used for weekly volatility prediction in this study, is the three-layer feed-forward neural network. It consists of three layers of processing units (also termed neurons or nodes), which are arranged in a feed-forward manner. The input values (input data) are fed to the neurons in the so-called input layer. The inputs (akin to regressors in a linear regression model) are connected to the output(s) (the regressand) via a hidden layer (Brooks, 1998). There can be one or more hidden layers which are critical for ANNs to identify the complex patterns in the data (Indro, Jiang, Patuwo, & Zhang, 1999). The hidden layer's job is basically to transform the inputs into something that the output layer can use. Figure 2 illustrates a three-layer feed-forward neural network.

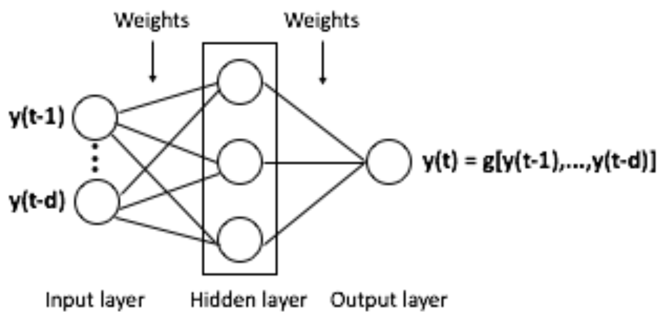



**Figure 2.** One of the most popular types of ANNs is known as a three-layer feed-forward neural network. It consists of three layers of neurons. The first layer of neurons, called the input layer, has one neuron for each input to the network. Each neuron in the input layer is connected to every neuron in the hidden area. The neurons in the hidden layer are similarly connected to the neuron in the output layer. The weights, or strengths, of the connections between neurons vary with the purpose of the network. Strong connections have greater weights than weaker connections, which means that important inputs are given larger weighting values. The weighting values ranging from zero to one. This figure illustrates, as an example, how $y(t)$ is predicted given $d$ past values of $y(t)$.

Each connection between neurons in different layers has an associated parameter indicating the strength of the connection, the so-called weight. By changing the weights in a specific manner, the network can "learn" to map patterns presented at the input layer to target values on the output layer. It is the weights that make each network unique. The process of weight adaption is called learning or training algorithm (Maciel & Ballini, 2008). The most popular training algorithm is the error back-propagation (EBP), or just back-propagation, which compares the forecast to the actual result and uses the difference to adjust the weights of the connections between the neurons. The goal is to adjust the weights in order to minimize the errors (Gately, 1996, pp. 15-16).

From an econometric perspective, the network model with the goal of optimally adjusting the weights between the layers can be written accordingly

$$\hat{Y}_{N,d}(Y;\beta,w) = \sum_{j=1}^{N} \beta_j \varphi\left(\sum_{i=1}^{d} w_{ij} y_i\right) \tag{18}$$

where the number of hidden neurons in the hidden layer is $N$. The inputs are selected as own lagged values of the series from $t-1$ to $t-d$, where $d$ is the number of inputs. $\hat{Y}$ is a vector of fitted values, $y_i$ is the input $i$, $\beta_j$ represents the hidden to output weight $j$, and $w_{ij}$ represents for input $i$ the input to hidden weight $j$. Important inputs are given large weighting values, and less important inputs are given small weighting values. The weighting values range from zero to one, $w, \beta \in [0,1]$. The activation function for the hidden layer is the sigmoid

$$\varphi(p) = \frac{1}{1+exp(-p)} \tag{19}$$

Activation functions are basically the mathematical transfer function that determines which level of the data is given the most emphasis (Gately, 1996, pp. 78).

Now, let

$$Y_t^d = (y_{t+d-1}, y_{t+d-2}, \dots, y_t)' \tag{20}$$

and the multivariate nonlinear least squares minimization problem is given by

$$\min_{\beta,w} \sum_{t=0}^{T-d-1} \left[Y_{t+d} - \hat{Y}_{N,d}\left(Y_t^d; \beta, w\right)\right]^2 \qquad (21)$$

where the objective is to minimize the sum of the squared differences between output values produced by the ANN and the desired (correct) values, for all training examples (Brooks, 1998).

In order to first be able to train an ANN and later evaluate its performance, the data sample should be divided into three non-overlapping periods, the so-called training set, valuation set and testing set. The training set is the largest set and is used to "teach" the ANN, i.e. find patterns and optimal interrelationships. The valuation set is used to provide an unbiased evaluation of a model fit while tuning model parameters. It can be regarded as a part of the training set and is usually used to halt training when generalization stops improving and thereby avoid overfitting. If the ANN is not able to generalize, but instead learns the individual properties of the training patterns without recognizing the general features of the data, it is said to be overfitted. Basically, what happens is that the ANN produces correct results for training patterns but has a high error rate in the test set. In order to improve generalization, the back-propagation training algorithm can be used together with a regularization technique, such as the Bayesian regularization. Kayri (2016) states that by incorporating the Bayesian regularization technique into the back-propagation training algorithm, the ANNs are less likely to overfit to training patterns.

The testing set is data reserved to validate the trained ANN using information the ANN has not seen during training. Only by testing the trained network with fresh data can the user be sure the network has learned to do the task within the accuracy expected, for example predict the volatility of a stock index. Normally 10-30 percent of the data is put aside in the testing set (Maciel & Ballini, 2008).

# 4 Data

The data used for this study are weekly observed indices for the stock markets in Sweden, the UK and Australia. The data span almost 10 years, with the first observation being the $8^{th}$ February 2008 and the last observation being the $29^{th}$ December 2017. After data conversion, which is explained below, there are 516 weekly observations included for each stock index in this study. The stock indices are chosen based on geographical location and different numbers of stocks included in the indices. The reason for using the frequency of weekly data points is to reduce the effect of different closing times of the different stock indices, compared to daily time stamps. The time period was determined arbitrarily, as it covers important events such as the aftermath of the financial crises 2007-2008, the “dotcom”-bubble 2010 and the latest year’s boom in financial markets. Gately (1996, pp. 69) states that the performance of an ANN would improve with the addition of more patterns, which is consistent with the general statistical rule that more data is better than less data. Short cuts often lead to incomplete or erroneous results. Brooks (1998) uses data running from $17^{th}$ November 1978 to $30^{th}$ June 1988 (approx. 10 years), when performing a similar study to mine, on solely the New York Stock Exchange (NYSE).

The Thomson Reuters Datastream database is used to collect data. The data series and the data type used in this study are presented in Table 1.

**Table 1:** The data series and the data type used in the study are collected from the Thomson Reuters Datastream database over the period from $8^{th}$ February 2008 to $29^{th}$ December 2017. For the data series and the data type, the Datastream ticker is presented as well as an individual description. The data points of the data type are collected at weekly frequency in order to reduce the effect of different closing times of the different stock indices.

| ***Data Series*** | ***Datastream ticker*** | ***Description*** | |
|---|---|---|---|
| OMXS30 | LSWEDOMX | A capitalization-weighted stock market index that consists of the 30 most actively traded stocks on the Stockholm Stock Exchange. | |
| FTSE100 | FTSE100 | A capitalization-weighted stock market index that consists of 100 companies listed on the London Stock Exchange with the highest market capitalization. | |
| S&P/ASX200 | ASX200I | A capitalization-weighted stock market index that consists of the 200 companies listed on the Australian Securities Exchange with the highest market capitalization. | |
| ***Data type*** | ***Datastream ticker*** | ***Description*** | ***Frequency*** |
| Price Index | PI | Index value. Computed by weighted average market capitalization without re-investment of dividends. Described in units of local currency. | Weekly |

For each of the stock indices, the continuously compounded return during week $t$ (between the end of week $t-1$ and the end of week $t$) are constructed as

$$r_t = ln\left(\frac{I_t}{I_{t-1}}\right) \quad (22)$$

where $I_t$ is the closing price index at Friday week $t$ and $I_{t-1}$ is the closing price index at Friday week $t-1$.

Volatility, on the other hand, is a latent variable (Poon, 2005, pp. 11). Due to this fact one has to rely on proxies when specifying, estimating and evaluating volatility models. The most common approach of estimating volatility, $\sigma_t^2$, is to set it equal to the variance of the $r_t$'s (Hull, 2015, pp. 209). When the most recent $m$ observations of the $r_t$ are used in conjunction with the usual formula for variance, this approach gives

$$\sigma_t^2 = \frac{1}{m-1}\sum_{i=1}^{m}(r_{t-i} - \bar{r})^2 \quad (23)$$

where $\bar{r}$ is the mean of the $r_t$'s

$$\bar{r} = \frac{1}{m}\sum_{i=1}^{m} r_{t-i} \quad (24)$$

This approach is, however, not entirely satisfactory, for example when the sample is small. Instead, an alternative approach is simply to use the squared $r_t$ as a volatility proxy. By assuming that $\bar{r}$ is zero, Lopez (2001) shows that the squared $r_t$ is an unbiased but very noisy volatility estimator.

In this study, I use the squared $r_t$ as a volatility proxy. Hence, I assume that $\bar{r}$ is zero for the different stock indices, which I believe is justifiable since Table 2 (pp. 17) shows that the means are fairly close to zero. There are two reasons why I use squared $r_t$ as the volatility proxy. First, it is quick to implement. Second, it will put the ARCH-type models and the ANNs to an extra test due to the noisiness. In studies similar to mine, squared $r_t$ are used as volatility proxies (see for example; West & Cho, 1995; Brooks, 1998).

In certain phases, all three stock indices tend to have a similar trending behavior. In other phases, they tend to differ. Figure 3 below shows the development of the price indices during the time period for this study. One can observe that the OMXS30 and the FTSE100 seem more correlated compared to the S&P/ASX200. Figure 4 shows the comparing development for logarithmic returns. Figure 5 shows the development of the volatilities. Notice the volatility clusters.

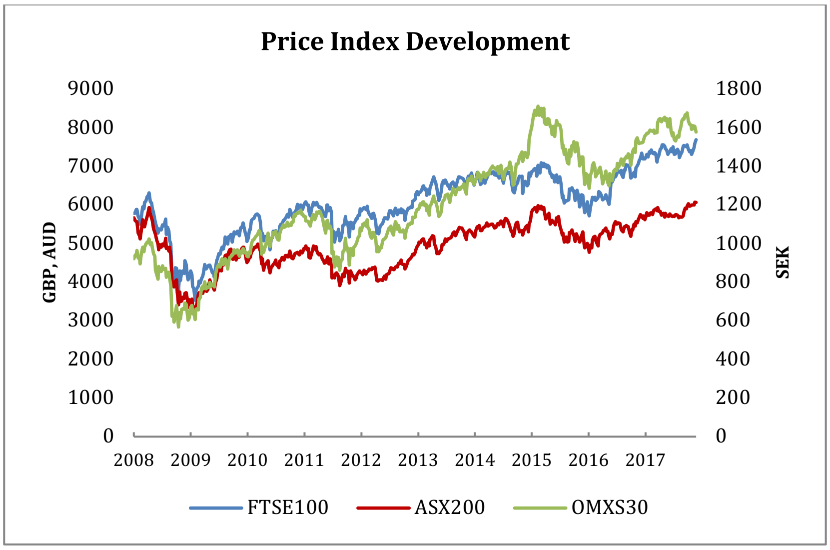



**Figure 3:** The figure shows the weekly price index development of the OMXS30, the FTSE100 and the S&P/ASX200 over the period from 8th February 2008 to 29th December 2017 (517 weeks). Each stock index is described in units of local currency.

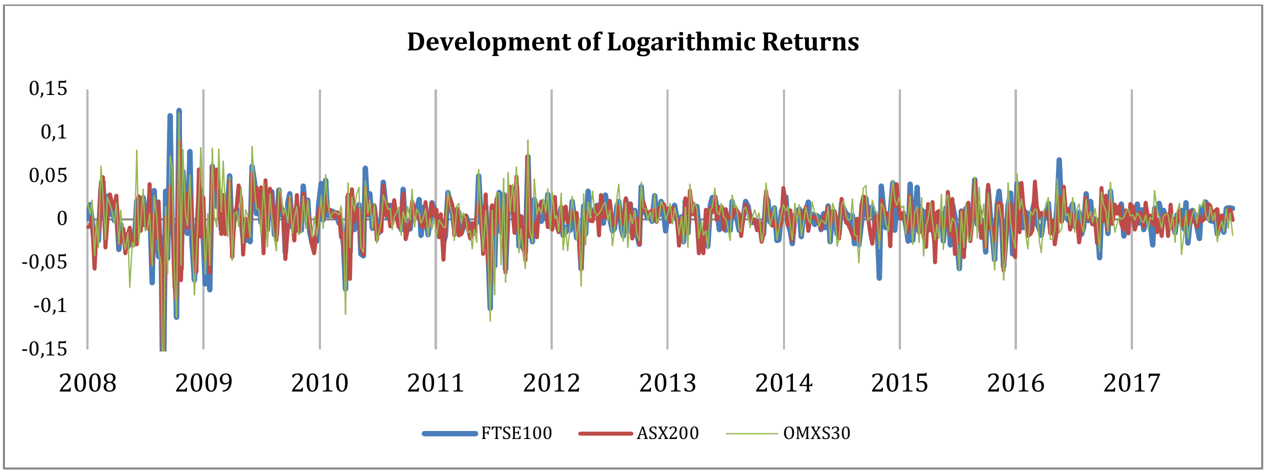



**Figure 4:** The figure shows the development of the weekly logarithmic returns of the OMXS30, the FTSE100 and the S&P/ASX200 over the period from 15th February 2008 to 29th December 2017 (516 weeks). The largest weekly drops, for all three indices, occur 10th October 2008 but are not fully viewable in this figure. During that week the log return observed for the FTSE100 is -0.236 (-23.6%), for the S&P/ASX200 -0.170 (-17%), and for the OMXS30 -0.225 (-22.5%).

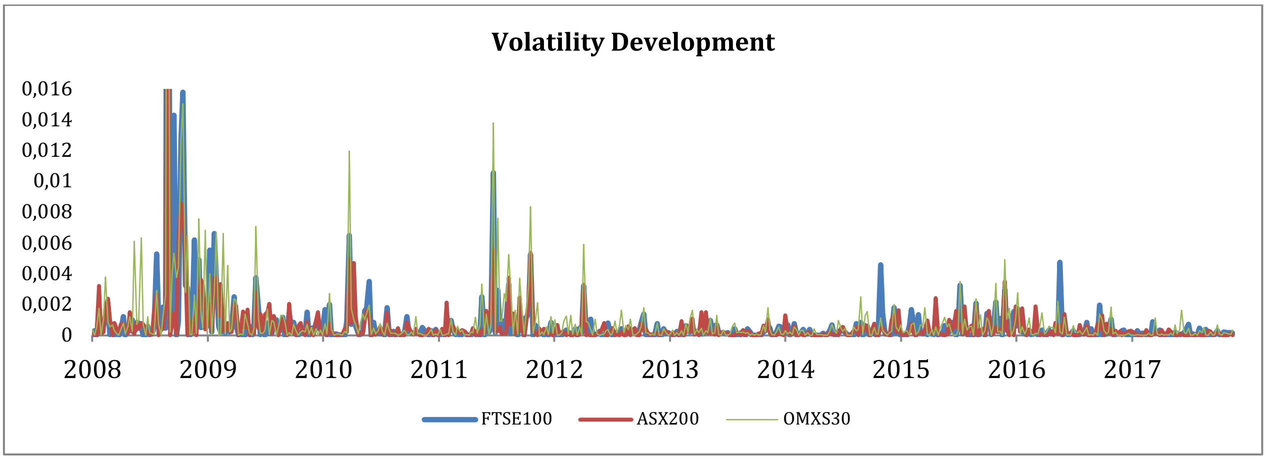



**Figure 5:** The figure shows the weekly volatility development of the OMXS30, the FTSE100 and the S&P/ASX200 over the period from 15th February 2008 to 29th December 2017 (516 weeks). The proxy for the weekly volatility, used in this study, is simply the square of each week's $r_t$. One can observe that certain periods exhibit unusually large volatilities followed by periods of relative tranquility. This is referred to as volatility clustering. The largest weekly volatility observations, for all three stock indices, occur the 10th October 2008. For that week, the volatility for the FTSE100 peaks at 0.056, for the S&P/ASX200 at 0.029, and for the OMXS30 at 0.051. This corresponds to the largest weekly drops in logarithmic returns, which can be observed in Figure 4.

A summary of some descriptive statistics of the $r_t$ series is given in Table 2. The number of weekly observations equals 516 for all three stock indices. The means are fairly close to zero and the sample variances are all quite small. For all the three series, the kurtoses exceed 3, indicating the necessity of fat-tailed distributions to describe these variables since the data has higher kurtosis than the normal distribution. An interesting feature of Table 2 is that the estimated measures of skewness are negative for all three series. This means that the distributions will have the mean to the left of the median and a longer tail in the negative direction. Especially for the FTSE100, skewness is quite large in an absolute sense.

**Table 2:** This table provides a summary of some descriptive statistics of the weekly logarithmic returns, $r_t$, for each of the stock indices. The samples cover weekly observations of $r_t$ over the period from 15th February 2008 to 29th December 2017 (516 weeks). The notation $(\times 10^{-4})$ means that the reported values should be multiplied with $10^{-4}$.

| **Stock market** | **Stock index** | $n$ | **Mean** $(\times \mathbf{10^{-4}})$ | **Variance** $(\times \mathbf{10^{-4}})$ | **Skewness** | **Kurtosis** |
|---|---|---|---|---|---|---|
| Sweden | OMXS30 | 516 | 10,427 | 8,777 | -1,020 | 8,176 |
| UK | FTSE100 | 516 | 5,514 | 6,986 | -1,413 | 14,796 |
| Australia | S&P/ASX200 | 516 | 1,347 | 5,700 | -0,992 | 5,877 |

# 5 Methodology

This chapter presents the methodology of the proposed models to satisfy the purpose of this thesis. Section 5.1 describes the approach taken in order to predict weekly stock index volatility using ARCH-type models. Section 5.2 describes the approach taken in order to predict weekly stock index volatility using ANN models. Section 5.3 explains how the predictions are to be evaluated.

## 5.1 Predicting Using ARCH-type Models

One way to be able to evaluate the volatility predicting performance of the ARCH-type models (GARCH, EGARCH and TGARCH) on historical data is to divide the data sample into two sets. This is commonly called "out-of-sample testing". The first data set, the so-called in-sample set, is used to estimate the various model parameters. The estimated models are then used to make predictions of the volatility for the period that is left out. The period that is left out is simply the second data set, the so-called out-of-sample set. This set consists of the most recent data observations which are used to see how accurate the predictions are compared to the ex post realizations and to determine whether the statistics of their errors are similar to those that the models made within the set of data that was fitted.

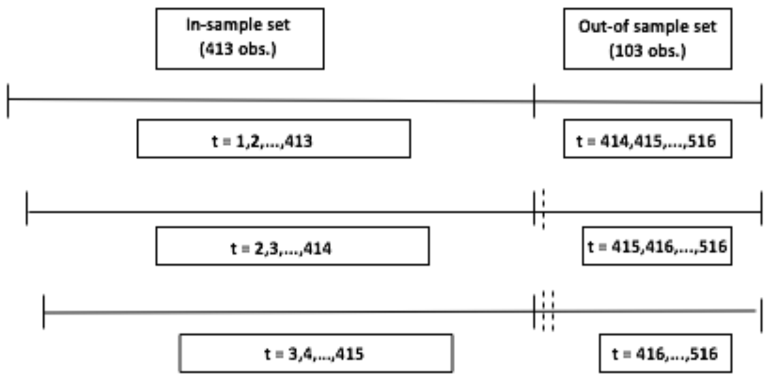
In this study, the data sample is from the start divided 80/20. The first 80 percent of the data observations (413 weeks) are used as an in-sample set and the most recent 20 percent of the data observations (103 weeks) are put aside as the out-of-sample set. The reason for this partition is so the out-of-sample set equals the testing set for the ANNs in number of weeks. The chosen testing set for the ANNs is further discussed in section 5.2.

This study focuses on only the $\text{GARCH}(1,1)$, $\text{EGARCH}(1,1)$ and $\text{TGARCH}(1,1)$. Brooks (2008, pp. 394) states that a $\text{GARCH}(1,1)$, in general, will be sufficient to capture the volatility clustering in financial data, and rarely is any higher order model estimated or even entertained in the academic finance literature. Hence, I decide to not perform any higher order estimations for neither of the three ARCH-type models.

For these three models both normal distributed errors and student $t$ distributed errors are considered. I focus on the latter type of distribution because it is shown in Chapter 4 that for each of the three index logarithmic return series, the kurtoses exceed 3. This is indicating the necessity of fat-tailed distributions to describe the logarithmic returns of the stock indices since the data have higher kurtosis than the normal distribution.

Furthermore, since it is not a priori assumed that one of the ARCH-type models necessarily dominates the other models over the whole sample, the modelling and predicting exercise are repeated for different subsets. This means that the models are first estimated to the in-sample of 413 weeks and a one-step-ahead forecast is generated. Subsequently, the first observation from the 413-week in-sample set is deleted and the next one is added (the first observation in the out-of-sample set), and again a one-step-ahead forecast can be generated. Figure 6 illustrates this procedure.

**Figure 6:** This figure illustrates the prediction methodology for the ARCH-type models. The data sample is from the start divided 80/20. The in-sample set ranges from 15th February 2008 to 8th January 2016 (approx. 8 years), which equals 413 weeks. The out-of-sample set ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years), which equals 103 weeks. After the first one-step-ahead forecast is generated for the out-of-sample set, the first observation from the in-sample set is deleted, and the next one is added (the first observation in the out-of-sample set). This procedure continues until a one-step-ahead forecast is generated for 29th December 2017.




In order to save space, the relevant parameter estimates are reported together with the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for the starting in-sample set. This should give an indication of the typical estimated parameter values. The AIC and BIC are reported to compare the estimated ARCH-type models for this period.

The predicting performances of the ARCH-type models for the out-of-sample set, ranging from 15th January 2016 to the 29th December 2017, are evaluated using four types of metrics.

The four types of metrics are RMSE, MAE, MAPE and out-of-sample $R^2$ which are further discussed in section 5.3.

For the process of model parameters estimation and weekly volatility prediction with the ARCH-type models the statistical software Eviews, version 9.5, is utilized.

## 5.2 Predicting Using ANNs

In regards to ANN terminology the data sample must be partitioned into three non-overlapping data sets, compared to two sets for the ARCH-type models, in order to produce and later evaluate weekly volatility predictions for ANN models. As explained in section 3.2.2, the first data set is called the training set which is used to "teach" the ANN to find patterns and optimal interrelationships. The second data set, the valuation set, can be seen as a part of the training process and is used to provide an unbiased evaluation of a model fit while tuning model parameters. The third data set, the testing set, consists of data reserved to validate the trained ANN using information the ANN has not seen during training. In this study, the testing set includes 20 percent of the most recent data, which equals the out-of-sample set for the ARCH-type models. Maciel & Ballini (2008) states that normally 10-30 percent of the data is put aside in the testing set for ANNs. My decision is to pick the percentage in-between, which means reserving 20 percent of the data for testing. Figure 7 illustrates the three partitioned data sets.

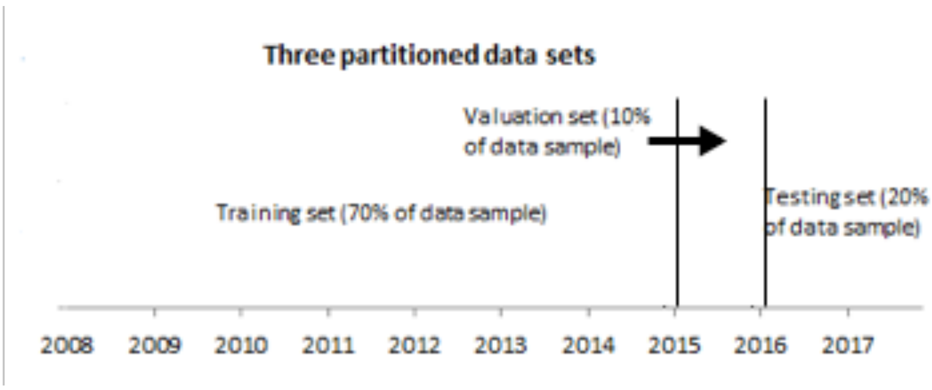



**Figure 7:** The figure illustrates how the data sample is partitioned into three non-overlapping data sets for the ANNs. The training set, which is allocated to model selection purposes, consists of weekly volatility observations from 15th February 2008 to 2nd January 2015 (361 weeks). The valuation set, which provides an unbiased evaluation of a model fit while tuning model parameters, consists of weekly volatility observations from 9th January 2015 to 8th January 2016 (52 weeks). The testing set, which is used to do the predictive performance evaluation, consists of weekly volatility observations from 15th January 2016 to 29th December 2017 (103 weeks).

The type of ANN model used for prediction in this study is a back-propagation, dynamic, three-layer feed-forward neural network. It consists of three layers of neurons which are arranged in a feed-forward manner. Figure 8 illustrates the ANN model. During the training process, the input data (i.e. current volatility values) is presented to the network one data point at a time from the spreadsheet. The network then makes a guess as to the correct volatility value one-step-ahead and compares it to the column that contains the correct answer. The network is allowed to be dynamic in the sense that its response may depend not only on the current input, but also on earlier inputs. In this study, the networks are allowed to depend on two delays. It means that the output, $y_{t+1}$, is allowed to depend not only on the current input, $y_t$, but also on the two previous inputs, $y_{t-1}$ and $y_{t-2}$. Only two delays are considered since it is likely that these will have the largest effect upon the current volatility value. Brooks (1998) makes a similar decision.

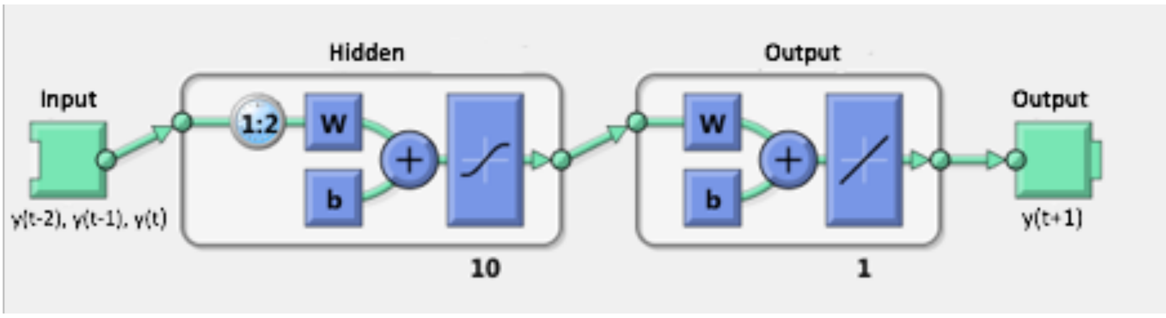



**Figure 8:** This figure illustrates the back-propagation, dynamic, three-layer feed-forward neural network which is the type of ANN used in this study. The circular 1:2 symbol in the hidden layer indicates that output, $y_{t+1}$, is allowed to depend not only on the current input, $y_t$, but also on the two previous inputs, $y_{t-1}$ and $y_{t-2}$. In the illustrated ANN model, the hidden layer contains 10 hidden neurons. For each stock index, four different ANN models are produced, with the difference of varied numbers of hidden neurons. In this study the numbers of hidden neurons tested are one, two, five and ten. Figure source: Matlab R2016a.

The hidden layers' job is basically to transform the inputs into something that the output layer can use. It is there to help the network make good guesses as to the correct volatility value one-step-ahead. There is no consensus regarding the optimal number of hidden neurons in the hidden layer. One has to use a trial-and-error method to find the most appropriate number and always bear in mind that including too many can result in overfitting. Gately (1996, pp. 72-73) shows that the improving effect of including more than 10 hidden neurons in a back-propagation, three-layer feed-forward neutral network is just marginal. Therefore, I decide that the potential improvement of including more than 10 hidden neurons are not worth the increasing overfitting-risk in my study. Instead, for each stock index I test four different versions of ANNs with different numbers of hidden neurons in the hidden layer. The arbitrarily chosen numbers of hidden neurons are 1, 2, 5 and 10. The reason for not testing more than four different numbers of hidden neurons is my belief that it is unlikely that one more or less hidden neutron would add any significant predictive power. In a similar study as mine, Brooks (1998) evaluates ANNs with the fixed number of 8 hidden neurons in the hidden layer.

After the network has made a guess, it records the error, adjusts the weight of the connections between the neurons and then goes on to the next data point and repeats this process. When all data has been scanned, it starts back at the beginning and goes through the data again, and again, and again. This learning process is called back-propagation. In this study, the back-propagation training algorithm is used together with a Bayesian regularization technique in order to improve generalization. Kayri (2016) explains that by using a regularization technique, such as Bayesian regularization (BR), with the back-propagation training algorithm smaller errors can be obtained. The BR assigns a probabilistic nature to the network weights, allowing the network to automatically and optimally penalize excessively complex models. The BR technique reduces the potential for overfitting and overtraining, improving the prediction quality and generalization of the network by making the algorithms more robust. Hence, the training algorithms are constructed in this way in this study. When further training fails to improve the network, the network configuration is saved and the training is done. Once the final model is decided upon, the generalization ability is tested on the untouched testing set.

For each stock index, four different ANNs are produced because of the different numbers of hidden neurons. Hence, in total twelve ANNs are produced. This results in an evaluation of 12 times 103 predictions. The same four metrics, as for the ARCH-type models, are used for evaluating the volatility-predicting performances of the ANNs. These are further discussed in section 5.3.

For the process of weekly volatility-predicting using ANNs, the software program Matlab, version R2016a, is utilized. The ANN models are generated using the Neural Network Toolbox, version 9.0, which is an interactive Matlab program one can use without writing any code. The implementation of a back-propagation training algorithm with a Bayesian regularization technique is simply a "click-in" option.

## 5.3 Quality Evaluation of Predictions

Generally, there is no guidance to find the appropriate choice of model for prediction. There are several ways to measure the accuracy of the predictions, all of which are based on a comparison of the generated predictions with the ex post realizations. Verbeek (2012, pp. 79-80) lists four types of metrics that are acceptable from a statistical point of view and can be used to choose from alternative models. Hence, these four are used in this study. The four metrics are the following: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and out-of-sample $R^2$. Except for the out-of-sample $R^2$, the lower the measure value the better the forecast.

### 5.3.1 Root Mean Square Error (RMSE)

The RMSE is, together with the Mean Square Error (MSE), a relatively popular approach for evaluating predictions. It is based on a quadratic loss function, where larger forecast errors, either positive or negative, are punished more heavily. This leads to the RMSE, given by

$$RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{y}_{T+h} - y_{T+h})^2} \tag{25}$$

where the series of predictions are denoted by $\hat{y}_{T+h}$, the series of ex post realizations are denoted $y_{T+h}$, $h = 1, 2, \dots, H$, where $T$ reflects the final period of the estimation sample and $H$ the number of forecasting periods.

### 5.3.2 Mean Absolute Error (MAE)

The MAE, also known as the Mean Absolute Deviation (MAD), is appropriate if the cost of making a wrong forecast is proportional to the absolute size of the forecast error. Compared to the RMSE, it is more sensitive to outliers. The MAE is given by

$$MAE = \frac{1}{H}\sum_{h=1}^{H} |\hat{y}_{T+h} - y_{T+h}| \quad (26)$$

where the inputs have the same meanings as for equation (25).

### 5.3.3 Mean Absolute Percentage Error (MAPE)

If the relative forecast error is more relevant, compared to MAE, the MAPE is appropriate. A feature that differs the MAPE from the RMSE and MAE is that it is expressed in percentage terms. It is given by

$$MAPE = \frac{100}{H}\sum_{h=1}^{H} \frac{|\hat{y}_{T+h} - y_{T+h}|}{|y_{T+h}|} \quad (27)$$

where the inputs again have the same meanings as for equation (25).

### 5.3.4 Out-of-sample $R^2$

In the stock return forecasting literature, an often used benchmark is the historical average return (Verbeek, 2012, pp. 80). In this study, I find a similar benchmark for the historical average volatility. This allows me to define an out-of-sample $R^2$ statistic which summarizes the models' forecasting performances. The out-of-sample $R^2$ is given by

$$R_{os}^2 = 1 - \frac{\sum_{h=1}^{H}(\hat{y}_{T+h} - y_{T+h})^2}{\sum_{h=1}^{H}(\bar{y} - y_{T+h})^2} \quad (28)$$

where $\bar{y}$ is the historical average volatility, in this study estimated over the period up to $T$ (15th February 2008 to 8th January 2016). The rest of the inputs have the same meanings as for equation (25). A positive out-of-sample $R^2$ indicates that the predictive regression has a lower mean squared prediction error than the historical average volatility.

# 6 Results and Analysis

In this chapter, results using the stated methodology are presented. Section 6.1 presents and evaluates the in-sample estimation results for the ARCH-type models. Section 6.2 presents and evaluates the out-of-sample predictions for the ARCH-type models and the ANN models. In section 6.3, the results and analysis are summarized and further discussed.

## 6.1 In-sample Estimation for ARCH-type Models

In this section I present some in-sample estimation results to give an idea of the possible usefulness of the ARCH-type models. The reason why I do not present any training set estimation results from the ANNs is because Matlab does not generate any. Furthermore, AIC and BIC are not applicable to ANNs.

As explained in section 5.1, the modelling exercise for the ARCH-type models are repeated for different subsets. To save space I consider the starting in-sample set covering the period from 15th February 2008 to 8th January 2016 (approx. 8 years/413 weeks), which should give an indication of the typical estimated parameter values. Notice that this set contains a period of the stock market breakdown in 2007-2008 and also the "dot-com" bubble in 2010, which affect the skewness of the data.

In section 5.1 it is also mentioned that my focus is to evaluate the GARCH$(1,1)$, the EGARCH$(1,1)$ and the TGARCH$(1,1)$ with Student-t errors. Estimations are performed for these models with normal errors as well. The general difference between the models with normal errors and the models with Student-t errors are that the latter are better fitted, due to lower AIC and BIC values. Hence, the results for the models with normal errors are presented in Appendix A.

In Table 3 the in-sample estimation results from the GARCH$(1,1)$ models with Student-t errors are presented. It is clear that all the estimated parameters $\hat{\alpha}_1$ and $\hat{\beta}_1$ are significant at least at the 5% level. Only for the FTSE100, the estimated intercept term $\hat{\omega}$ is significant at the 10% level. For the other two indices this term is not statistically different from zero. Furthermore, for all three stock indices the sum of the $\hat{\alpha}_1$ and $\hat{\beta}_1$ parameters are close to unity. According to Verbeek (2012, pp. 332) this is a typical finding in empirical applications. It means that conditional volatility is highly persistent. As such, I should anticipate that any shock creating uncertainty in the Swedish, British and/or Australian market should show little tendency to dissipate. For no index it is observed that $\hat{\alpha}_1 + \hat{\beta}_1 \geq 1$.

**Table 3:** Estimation results from GARCH(1,1) models with Student-t errors, for the in-sample period from 15th February 2008 to 8th January 2016 (413 weeks). The relevant parameter estimates are reported together with the robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5%, and 1% levels, respectively. Next to the parameter estimates, the values of AIC and BIC are reported in order to compare models. The notation ($\times 10^{-4}$) means that the reported value should be multiplied with $10^{-4}$.

| *GARCH(1,1)* | **Parameter estimates** | | | **Diagnostics** | |
|---|---|---|---|---|---|
| **Index** | $\hat{\boldsymbol{\omega}}$ ($\times \mathbf{10^{-4}}$) | $\hat{\boldsymbol{\alpha}}_1$ | $\hat{\boldsymbol{\beta}}_1$ | $\boldsymbol{AIC}$ | $\boldsymbol{BIC}$ |
| OMXS30 | 0,294 (0,179) | 0,118** (0,046) | 0,854*** (0,051) | -4,39 | -4,34 |
| FTSE100 | 0,374* (0,213) | 0,108** (0,048) | 0,836*** (0,066) | -4,66 | -4,61 |
| S&P/ASX200 | 0,167 (0,124) | 0,110*** (0,039) | 0,866*** (0,049) | -4,73 | -4,68 |

In Table 4 the in-sample estimation results from the EGARCH(1,1) models with Student-t errors are presented. For all indices, the estimated parameters $\hat{\alpha}_1$ and $\hat{\beta}_1$ are significant at the 1% level and the estimated intercept terms, $\hat{\omega}$, are at least significant at the 5% level. For all indices, the estimated parameters $\hat{\alpha}_1$ are negative, which mean that negative shocks (or "bad" news) generate more volatility than positive shocks (or "good" news).

**Table 4:** Estimation results from EGARCH(1,1) models with Student-t errors, for the in-sample period from 15th February 2008 to 8th January 2016 (413 weeks). The relevant parameter estimates are reported together with the robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5%, and 1% levels, respectively. Next to the parameter estimates, the values of AIC and BIC are reported in order to compare models.

| *EGARCH(1,1)* | **Parameter estimates** | | | | **Diagnostics** | |
|---|---|---|---|---|---|---|
| **Index** | $\hat{\boldsymbol{\omega}}$ | $\hat{\boldsymbol{\alpha}}_1$ | $\hat{\boldsymbol{\beta}}_1$ | $\hat{\boldsymbol{\gamma}}_1$ | $\boldsymbol{AIC}$ | $\boldsymbol{BIC}$ |
| OMXS30 | -0,321** (0,113) | -0,178*** (0,037) | 0,968*** (0,013) | 0,111* (0,059) | -4,43 | -4,38 |
| FTSE100 | -0,528*** (0,183) | -0,218*** (0,047) | 0,944*** (0,022) | 0,136** (0,061) | -4,71 | -4,65 |
| S&P/ASX200 | -0,290** (0,114) | -0,116*** (0,035) | 0,974*** (0,013) | 0,114** (0,051) | -4,75 | -4,69 |

This indicates that the data are asymmetric. Furthermore, the large coefficients for $\hat{\beta}_1$ reflect high persistence in volatility for the stock indices. Also observe that the AIC and BIC values are generally smaller compared to those generated from the GARCH$(1,1)$ and the TGARCH$(1,1)$ models.

In Table 5 the in-sample estimation results from the TGARCH$(1,1)$ models with Student-t errors are presented. The estimated parameters $\hat{\beta}_1$, for all indices, are significant at the 1% level and the intercept terms, $\widehat{\omega}$, are all at least significant at least at the 10% level. As for the EGARCH$(1,1)$ models, I find evidence of data asymmetry as the estimated $\hat{\delta}_1$coefficients, for all indices, are at least positive significant at the 5% level. Negative shocks seem to have larger effects on volatility than positive shocks for all indices. However, the TGARCH$(1,1)$ models are unsatisfactory from the point of view that the estimated $\hat{\alpha}_1$ coefficients are not present, meaning that they are not statistically different from zero. For a TGARCH$(1,1)$, as for a GARCH$(1,1)$, the $\alpha_1$ must be present in order for the model to be identified. If $\alpha_1 = 0$, the TGARCH$(1,1)$ instead becomes a TGARCH$(0,1)$. Except from that, the AICs are slightly lower compared to what the GARCH$(1,1)$ models generated. Also, two out of three estimates of BIC are lower compared to what the GARCH$(1,1)$ models generated. For the S&P/ASX200, the BIC estimates indicate that the TGARCH$(1,1)$ and the GARCH$(1,1)$ perform equally.

**Table 5:** Estimation results from TGARCH(1,1) models with Student-t errors, for the in-sample period from 15th February 2008 to 8th January 2016 (413 weeks). The relevant parameter estimates are reported together with the robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5%, and 1% levels, respectively. Next to the parameter estimates, the values of AIC and BIC are reported in order to compare models. The notation $(\times 10^{-4})$ means that the reported value should be multiplied with $10^{-4}$.

| *TGARCH(1,1)* | **Parameter estimates** | | | | **Diagnostics** | |
|---|---|---|---|---|---|---|
| **Index** | $\widehat{\boldsymbol{\omega}}$ $(\times \mathbf{10^{-4}})$ | $\widehat{\boldsymbol{\alpha}}_1$ | $\widehat{\boldsymbol{\beta}}_1$ | $\widehat{\boldsymbol{\delta}}_1$ | $\boldsymbol{AIC}$ | $\boldsymbol{BIC}$ |
| OMXS30 | 0,292** (0,127) | -0,022 (0,044) | 0,866*** (0,046) | 0,230*** (0,067) | -4,42 | -4,36 |
| FTSE100 | 0,636*** (0,214) | -0,041 (0,047) | 0,754*** (0,060) | 0,379*** (0,114) | -4,70 | -4,64 |
| S&P/ASX200 | 0,126* (0,686) | -0,014 (0,047) | 0,912*** (0,034) | 0,140** (0,057) | -4,74 | -4,68 |

## 6.2 Out-of-Sample Predictions from ARCH-type Models and ANNs

To evaluate the ability of the ARCH-type models and the ANNs to adequately predict the weekly volatility for the OMXS30, the FTSE100 and the S&P/ASX200, I use the four metrics explained in section 5.3. The metrics are all based on comparisons of the weekly generated predictions with the ex post weekly realizations for the out-of-sample period which ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years). This period includes 103 weeks.

Table 6 presents the models' RMSE, MAE, MAPE and Out-of-sample $R^2$ for the OMXS30. All three ARCH-type models clearly outperform all four versions of the ANNs. Green, underlined values indicate the best measure values for each of the metrics. Notice that none of the green, underlined values belong to any of the ANNs. Fifty percent of the green, underlined values belong to GARCH(1,1), which is the highest ratio among the ARCH-type models. Orange, italic values indicate the best measure values, among the ANNs, for each of the metrics. Apparently, it is enough for ANNs to only have one hidden neuron in the hidden layer in order to produce weekly volatility-predictions of the OMXS30 for the out-of-sample period.

**Table 6:** Out-of-sample predicting performance of GARCH(1,1), EGARCH(1,1), TGARCH(1,1) and four ANN models with 1, 2, 5 and 10 hidden neurons, respectively, in the hidden layer, for the weekly volatility of OMXS30. The ARCH-type models have Student-t errors. The out-of-sample period ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years/103 weeks). The four metrics used to compare the volatility-predicting performance are the Root Mean Squared Error (RMSE), the Mean Average Error (MAE), the Mean Average Percentage Error (MAPE) and the Out-of-sample $R^2$. The notations $(\times 10^{-3})$ and $(\times 10^{3})$ mean that the reported values should be multiplied with $10^{-3}$ and $10^{3}$, respectively.

| | ***Metrics*** | | | |
|---|---|---|---|---|
| **Model** | **RMSE** $(\times \mathbf{10^{-3}})$ | **MAE** $(\times \mathbf{10^{-3}})$ | **MAPE** $(\times \mathbf{10^{3}})$ | **Out-of-sample** $\boldsymbol{R^2}$ |
| GARCH(1,1) Student-t errors | <u>0,530</u> | 0,541 | 13,3% | <u>0,612</u> |
| EGARCH(1,1) Student-t errors | 0,551 | 0,449 | <u>11,7%</u> | 0,551 |
| TGARCH(1,1) Student-t errors | 0,544 | <u>0,433</u> | 11,9% | 0,592 |
| ANN (1 hidden neuron) | ***0,650*** | ***0,558*** | ***16,9%*** | ***0,417*** |
| ANN (2 hidden neurons) | 0,701 | 0,649 | 26,6% | 0,322 |
| ANN (5 hidden neurons) | 0,703 | 0,655 | 27,6% | 0,318 |
| ANN (10 hidden neurons) | 0,671 | 0,610 | 23,7% | 0,379 |

Figure 9 illustrates the weekly volatility-predictions of the OMXS30 from the GARCH(1,1) and the ANN (1 hidden neuron), compared to the ex post weekly realizations, for the out-of-sample period. According to the measure values in Table 6, the GARCH(1,1) and the ANN (1 hidden neuron) seem to be the best weekly volatility-predictors for OMXS30 from each family of models. Notice that the weekly volatility-predictions from the GARCH(1,1) deviate less from the correct weekly volatilities compared to the ANN (1 hidden neuron). It seems that the volatility-predictions of the ANN (1 hidden neuron) are lagging one or two weeks compared to the ex post realizations. Notice also that this seems not to be a problem only in the end of the out-of-sample period, it starts already in the beginning. The lagging problem most certainly affects the measure values of the ANN (1 hidden neuron) and is one explanation why they are worse compared to the measure values of the GARCH(1,1).

**Figure 9:** This figure illustrates the weekly volatility-predictions from the GARCH(1,1) and the ANN (1 hidden neuron), compared to the ex post realizations, for the out-of-sample period of the OMXS30. The out-of-sample period ranges from 15th January to 29th of December 2017 (approx. 2 years/103 weeks). According to the measure values in Table 6, the GARCH(1,1) seems to be the best volatility-predictor for OMXS30 among the ARCH-type models and the ANN (1 hidden neuron) seems to be best among the different types of ANNs.

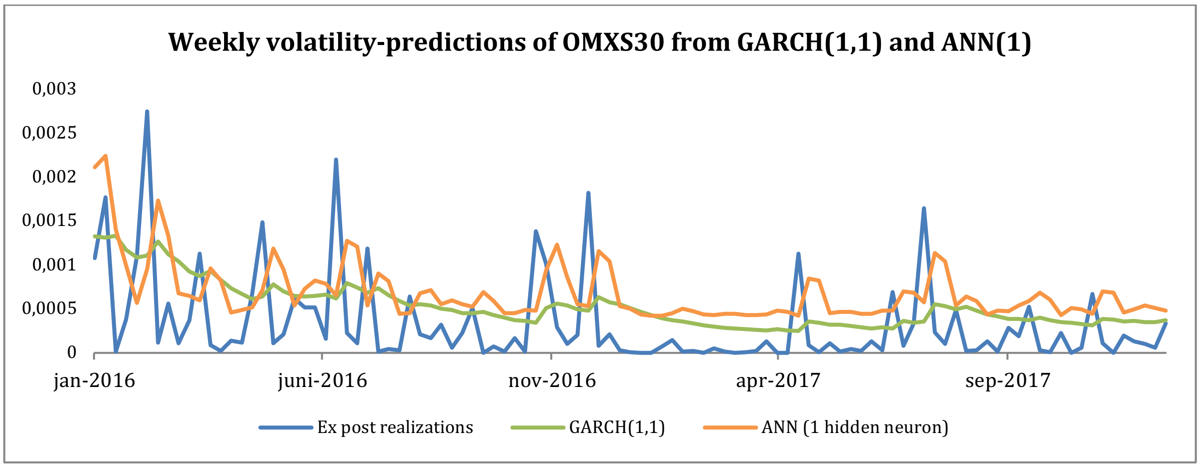



Table 7 presents the models' RMSE, MAE, MAPE and Out-of-sample $R^2$ for the FTSE100. Again, all the three ARCH-type models clearly outperform all four versions of the ANNs. Green, underlined values and orange, italic values have the same meaning in Table 7 as in Table 6. Again, notice that none of the green, underlined values belongs to any of the ANNs. Instead, all green, underlined values belong to the EGARCH(1,1). Hence, the EGARCH(1,1) seems to dominate as a weekly volatility-predictor for the FTSE100. The ANN (1 hidden neuron) has three out of four (75%) orange, italic values, which indicate that it is enough for

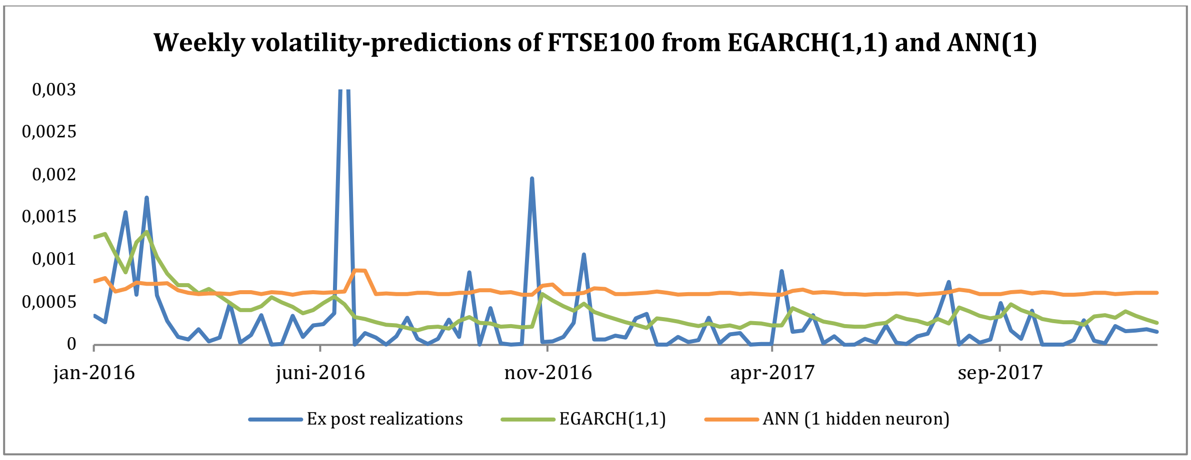
ANNs to only have one hidden neuron in the hidden layer in order to produce weekly volatility-predictions of the FTSE100.

**Table 7:** Out-of-sample predicting performance of GARCH(1,1), EGARCH(1,1), TGARCH(1,1) and four ANN models with 1, 2, 5 and 10 hidden neurons, respectively, in the hidden layer, for the weekly volatility of FTSE100. The ARCH-type models have Student-t errors. The out-of-sample period ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years/103 weeks). The four metrics used to compare the volatility-predicting performance are the Root Mean Squared Error (RMSE), the Mean Average Error (MAE), the Mean Average Percentage Error (MAPE) and the Out-of-sample $R^2$. The notations $(\times 10^{-3})$ and $(\times 10^{3})$ mean that the reported values should be multiplied with $10^{-3}$ and $10^{3}$, respectively.

| | ***Metrics*** | | | |
|---|---|---|---|---|
| **Model** | **RMSE** $(\times \mathbf{10^{-3}})$ | **MAE** $(\times \mathbf{10^{-3}})$ | **MAPE** $(\times \mathbf{10^{3}})$ | **Out-of-sample** $\boldsymbol{R^2}$ |
| GARCH(1,1) Student-t errors | 0,587 | 0,368 | 313% | 0,428 |
| EGARCH(1,1) Student-t errors | **0,564** | **0,330** | **219%** | **0,471** |
| TGARCH(1,1) Student-t errors | 0,587 | 0,352 | 316% | 0,429 |
| ANN (1 hidden neuron) | ***0,665*** | ***0,528*** | 588% | ***0,267*** |
| ANN (2 hidden neurons) | 0,676 | 0,530 | 560% | 0,242 |
| ANN (5 hidden neurons) | 0,676 | 0,545 | 602% | 0,241 |
| ANN (10 hidden neurons) | 0,695 | 0,449 | *245%* | 0,198 |

Figure 10 illustrates the weekly volatility-predictions of the FTSE100 from the EGARCH(1,1) and the ANN (1 hidden neurons), compared to the ex post weekly realizations, for the out-of-sample period.

**Figure 10:** This figure illustrates the weekly volatility-predictions from the EGARCH(1,1) and the ANN (1 hidden neuron), compared to the ex post realizations, for the out-of-sample period of the FTSE100. The out-of-sample period ranges from 15th January to 29th of December 2017 (approx. 2 years/103 weeks). According to the measure values in Table 7, the EGARCH(1,1) seems to be the best volatility-predictor for FTSE100 among the ARCH-type models and the ANN (1 hidden neuron) seems to be best among the different types of ANNs. The weekly volatility for the FTSE100 spiked 2016-07-01, which was the week after UK voted to leave EU.



According to the measure values in Table 7, the EGARCH(1,1) and the ANN (1 hidden neuron) seem to be the best weekly volatility-predictors for FTSE100 from each family of models. The weekly volatility-predictions from the EGARCH(1,1) are closer in line with the correct weekly volatilities, compared to the weekly volatility-predictions from the ANN (1 hidden neuron). The weekly volatility-predictions from the ANN (1 hidden neuron) seem to be almost constant during the period and are not capturing any of the correct volatility fluctuations. Neither the EGARCH(1,1) nor the ANN (1 hidden neuron) are even close of capturing the volatility spikes in the beginning of July 2016, after UK voted to leave EU, or in the beginning of November 2016 when the FTSE100 reacted to the increasing uncertainty leading up to the U.S. election.

Table 8 presents the models' RMSE, MAE, MAPE and Out-of-sample $R^2$ for the S&P/ASX200. Again, the ARCH-type models outperform all four versions of the ANNs, but not as clear as for the OMXS30 and FTSE100.

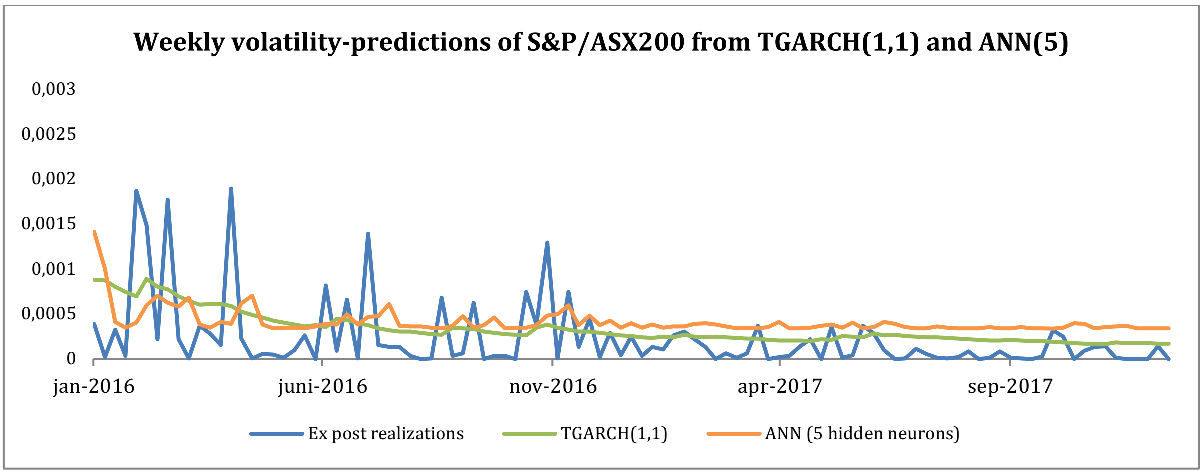
**Table 8:** Out-of-sample predicting performance of GARCH(1,1), EGARCH(1,1), TGARCH(1,1) and four ANN models with 1, 2, 5 and 10 hidden neurons, respectively, in the hidden layer, for the weekly volatility of S&P/ASX200. The ARCH-type models have Student-t errors. The out-of-sample period ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years/103 weeks). The four metrics used to compare the volatility-predicting performance are the Root Mean Squared Error (RMSE), the Mean Average Error (MAE), the Mean Average Percentage Error (MAPE) and the Out-of-sample $R^2$. The notations $(\times 10^{-3})$ and $(\times 10^{3})$ mean that the reported values should be multiplied with $10^{-3}$ and $10^{3}$, respectively.

| | ***Metrics*** | | | |
|---|---|---|---|---|
| **Model** | **RMSE** $(\times \mathbf{10^{-3}})$ | **MAE** $(\times \mathbf{10^{-3}})$ | **MAPE** $(\times \mathbf{10^{3}})$ | **Out-of-sample $R^2$** |
| GARCH(1,1) Student-t errors | 0,398 | 0,310 | 116% | 0,524 |
| EGARCH(1,1) student | 0,377 | 0,294 | 121% | 0,572 |
| TGARCH(1,1) Student-t errors | <u>0,370</u> | <u>0,281</u> | <u>114%</u> | <u>0,589</u> |
| ANN (1 hidden neuron) | 0,467 | 0,413 | 246% | 0,343 |
| ANN (2 hidden neurons) | 0,493 | 0,395 | 212% | 0,269 |
| ANN (5 hidden neurons) | *0,434* | *0,347* | *189%* | *0,434* |
| ANN (10 hidden neurons) | 0,492 | 0,384 | 203% | 0,271 |

Green, underlined values and orange, italic values have the same meaning in Table 8 as in Table 6 and Table 7. Again, notice that none of the green, underlined values belong to any of the ANNs. Instead, all green, underlined values belong to the TGARCH(1,1). In section 6.1, it is shown that the TGARCH(1,1) can not be identified since the estimated parameter $\hat{\alpha}_1$ is not statistically different from zero. Despite this, and even if the TGARCH(1,1) is in fact the TGARCH(0,1), it seems to dominate as a weekly volatility-predictor for the S&P/ASX200. The ANN (5 hidden neurons) dominates as volatility-predictor in the ANN family, since all orange, italic values belong to this version of ANNs.

Figure 11 illustrates the weekly volatility-predictions of the S&P/ASX200 from the TGARCH(1,1) and the ANN (5 hidden neurons), compared to the ex post weekly realizations, for the out-of-sample period. According to the measure values in Table 7, the TGARCH(1,1) and the ANN (5 hidden neurons) seem to be the best weekly volatility-predictors for S&P/ASX200 from each family of models. The weekly volatility-predictions from the TGARCH(1,1) are just slightly more in line with the correct weekly volatilities, compared to the volatility-predictions from the ANN (5 hidden neuron). This gives reason to believe that ANNs perform better as volatility predictors in relatively calm periods. Notice that the out-of-sample period for the S&P/ASX200 is less volatile compared to the OMXS30 and FTSE100. A potential reason to that might be the higher amount of stocks included in the S&P/ASX200 which creates a higher degree of diversification compared to the OMXS30 and FTSE100.

**Figure 11:** This figure illustrates the weekly volatility-predictions from the TGARCH(1,1) and the ANN (5 hidden neuron), compared to the ex post realizations, for the out-of-sample period of the S&P/ASX200. The out-of-sample period ranges from 15th January to 29th of December 2017 (approx. 2 years/103 weeks). According to the measure values in Table 8, the TGARCH(1,1) seems to be the best volatility-predictor for S&P/ASX200 among the ARCH-type models and the ANN (5 hidden neuron) seems to be best among the different versions of ANNs.




The volatility-predicting performance of the GARCH(1,1), EGARCH(1,1) and TGARCH(1,1) with normal errors are presented in Appendix B. For all indices, these models are outperformed by the same models with Student-t errors. However, the measure values of the GARCH$(1,1)$, the EGARCH$(1,1)$ and the TGARCH$(1,1)$ with normal errors show better results compared to the ANNs.

## 6.3 Summary of Results and Analysis

The in-sample estimation results, presented in section 6.1, give an idea of the possible usefulness of the ARCH-type models. Estimated model parameters from the GARCH$(1,1)$, EGARCH$(1,1)$ and TGARCH$(1,1)$, for all the three indices, indicate that conditional volatilities are highly persistent. This means that shocks creating uncertainty in the Swedish, British and/or Australian markets show little tendency to dissipate. This can be considered as evidence for the stylized fact of volatility clustering.

For all indices, the estimated model parameter $\hat{\alpha}_1$ in the EGARCH$(1,1)$ models and the estimated model parameter $\hat{\delta}_1$ in the TGARCH$(1,1)$ models are significant different from zero at least at the 5% level. This indicates that the all data series are asymmetric. One explanation often used for this feature is that “bad” news often have a more pronounced effect on volatility for asset prices than “good” news. It is possible that this is the case for OMXS30, FTSE100 and S&P/ASX200 as well.

The model selection criteria (AIC and BIC) suggest that the EGARCH$(1,1)$ best fit the in-sample data for all the three indices. However, the measure values do not differ by much to the comparing measures values for the GARCH$(1,1)$ and TGARCH$(1,1)$. They are just slightly better.

The out-of-sample predictions from the ARCH-type models and from the ANNs, in section 6.2, can be summarized by stating that ANNs do not seem to outperform the ARCH-type models as weekly volatility-predictors for the OMXS30, the FTSE100 or the S&P/ASX200. In fact, the measure values from the RMSE, MAE, MAPE and out-of-sample $R^2$ actually show evidence of the opposite, namely that the GARCH$(1,1)$, EGARCH$(1,1)$ and TGARCH$(1,1)$, generally, are better weekly volatility-predictors compared to all different versions of ANNs evaluated.

The figures 9-11 illustrate the weekly volatility-predictions from the best predictors among the two families of models for the out-of-sample period, for each of the three indices. One can observe, for all indices, that the ANNs in a higher degree seem to “miss” the realized volatility spikes. That makes it look like the weekly predictions from the ANNs are lagging throughout the out-of-sample periods and are more sensitive to extreme volatility clusters, compared to the best ARCH-type models. Hu & Tsoukalas (1999) finds that an ANN model performs better as a volatility-predictor for the European Monetary System (EMS) exchange rates during extreme volatility subperiods, such as during the August 1993 crisis, compared to other models. According to the findings in my study, I would say that the ANNs performance as volatility-predictors for extreme volatility subperiods are the opposite. The less volatile the out-of-sample period is, the better the ANNs seem to work as volatility-predictors. Notice the difference by comparing figure 11 to figure 9 and 10.

I have not further analyzed the optimal number of hidden neurons in the hidden layer of the ANNs. Notice that for different indices, different numbers of hidden neurons seem to be optimal. I have neither further tried to determine the best weekly volatility-predictor among the ARCH-type models. These two issues go beyond the scoop of this study.

# 7 Conclusion

The purpose of this thesis is to investigate if ANNs outperform the more traditional ARCH-type models in predicting weekly stock index volatility. In this study, the back-propagation, dynamic, three-layer feed-forward neural network represents the ANN structure and the GARCH(1,1), the EGARCH(1,1) and the TGARCH(1,1) represent the ARCH-type models. The major stock indices in Sweden (OMXS30), in the UK (FTSE100) and in Australia (S&P/ASX200) are analyzed in order to give a broader perspective of potential model differences.

An out-of-sample testing methodology is applied to the most recent 20 percent of the data observations, which fully range from 8th February 2008 to 29th December 2017. The metrics used to evaluate the volatility-predicting performances of the different models are the RMSE, the MAE, the MAPE and the out-of-sample sample $R^2$.

Based on the results of this thesis, no conclusions regarding the possibility of accurately predicting volatility of the OMXS30, the FTSE100 or the S&P/ASX200 with ANNs can be drawn. The ANNs show no evidence of outperforming the GARCH(1,1), EGARCH(1,1) and TGARCH(1,1) in stock index volatility prediction. In fact, the results show the opposite, namely that the GARCH(1,1), the EGARCH(1,1) and the TGARCH(1,1), generally, are better weekly volatility-predictors compared to all different versions of ANNs evaluated.

It is hard to generalize beyond the proposed models and the time series of this study. There are several reasons to that. One reason is based on that constructing ANNs involve a vast amount of design choices. Potential design choices for the ANNs that could have an impact on the results, for better or for worse, include: class of ANNs, architecture, regularization technique, and more. There is a possibility that properly designed ANNs could improve the weekly volatility-predicting performance for the stock indices. If it was not for the fact that ANNs suffer from their inability to explain the steps by which they reach decisions and their inability to incorporate rules into their structure improvement actions would have been taken in order to find better ANNs.

Other factors that might as well have limited the results are based on the choices made by the author. To mention a few, I could have chosen to use daily data instead of weekly data. By involving more data observations, the results might have been different. I could have included more independent variables that possible would have an effect on stock index volatility, such as exchange rates, long-term interest rates or GDP growths, in the ANNs. The ANNs would then have more data patterns to learn from, which potentially could result in better volatility predictions. I could also have investigated different out-of-sample periods instead of a fixed period. The results might have been different if, for example, the out-of-sample periods were shorter in time. Finally, I could have used another proxy when specifying, estimating and evaluating volatility models. If I would have used, for example, the difference between the highest and lowest daily index prices as a proxy, it would potentially be less noisy which could have generated different results.

Despite the fact that the results in this thesis do not demonstrate the usefulness of ANNs for volatility-prediction, I believe that the technique has a bright future as a financial forecasting instrument. Gately (1996, pp. 126) states that neural network architecture and training algorithms are advancing at a remarkably fast pace. Sometime in the future, they will be the basis of a stand-alone trading system. However, for volatility prediction purposes, I say that it is best if they at this moment are used at most as confirming indicators with one or more other indicators.

The potential future research areas where ANNs can be tested and evaluated are endless for economists. To further explore ANNs as volatility-predicting tools, I would suggest the field of hybrids between ARCH-type models and ANNs, which needs further research. Another idea is that instead of trying to use ANNs as volatility-predicting tools, one could try to integrate this technique into fundamental or technical analysis in order to improve stock picking decisions. One could also ask banks or insurance companies and come up with an idea that potentially can improve their businesses. Kaastra & Boyd (1996) states that financial organizations have been the second largest sponsor of research in neural network applications.

# References

Akgiray, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts, *The Journal of Business*, vol.62, no. 1, pp. 55-80.

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. 1st edition. New York: Springer.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, vol.31, no. 3, pp. 307-327.

Brailsford, T.J., & Faff, R.W. (1996). An Evaluation of Volatility Forecasting Techniques, *Journal of Banking and Finance*, vol. 20, no. 3, pp. 419-438.

Brooks, C. (1998). Predicting Stock Index Volatility: Can Market Volume Help?, *Journal of Forecasting*, vol. 17, no. 1, pp. 59-80.

Brooks, C. (2008). *Introductory Econometrics for Finance*. 2nd edition. New York: Cambridge University Press.

Cao, C.Q., & Tsay, R.S. (1992). Nonlinear Time-Series Analysis of Stock Volatilities, *Journal of Applied Econometrics*, vol. 1, no. 1, pp. 165-185.

Enders, W. (2015). *Applied Econometric Time Series*. 4th edition. New York: Wiley.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation, *Econometrica*, vol.50, no. 4, pp. 987-1007.

Franses, P.H., & van Dijk, D. (1996). Forecasting Stock Market Volatility Using (Non-Linear) Garch Models, *Journal of Forecasting*, vol. 15, no. 3, pp. 229-235.

Gately, E. (1996). *Neural Networks for Financial Forecasting*. 1st edition. New York: John Wiley & Sons, Inc.

Ghatak, A. (2017). *Machine Learning with R*. 1st edition. Singapore: Springer.

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*, vol. 48, no. 5, pp. 1779-1801.

Hansen, P.R., & Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, *Journal of Applied Econometrics*, vol. 20, no. 7, pp. 873-889.

Heynen, R.C., & Kat, H.M. (1994). Volatility Prediction: A Comparison of Stochastic Volatility, GARCH(1,1) and EGARCH(1,1) Models, *Journal of Derivatives*, vol. 2, no. 1, pp. 50-65.

Hu, M.Y., & Tsoukalas, C. (1999). Combining Conditional Volatility Forecasts Using Neural Networks: An Application to the EMS Exchange Rates, *Journal of International Financial Markets, Institutions and Money*, vol. 9, no. 4, pp. 407-422.

Hull, J.C. (2015). *Risk Management and Financial Institutions*. 4th edition. Hoboken, NJ.: Wiley.

Indro, D.C., Jiang, C.X., Patuwo, B.E., & Zhang, G.P. (1999). Predicting Mutual Fund Performance Using Artificial Neural Networks, *Omega*, vol. 27, no. 3, pp. 373-380.

Kaastra, I., & Boyd, M. (1996). Designing a Neural Network for Forecasting Financial and Economic Time Series, *Neurocomputing*, vol. 10, no. 3, pp. 215-236.

Kayri, M. (2016). Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data, *Mathematical and Computational Applications*, vol. 21, no. 2, pp. 1-11.

Lee, K.Y. (1991). Are the GARCH Models Best in Out-of-Sample Performance?, *Economics Letters*, vol. 37, no. 3, pp. 305-308.

Lopez, J.A. (2001). Evaluating the Predictive Accuracy of Volatility Models, *Journal of Forecasting*, vol. 20, no. 2, pp. 87-109.

Maciel, L.S., & Ballini, R. (2008). Design a Neural Network for Time Series Financial Forecasting: Accuracy and Robustness Analysis. *Instituto de Economía, Universidade Estadual de Campiñas, São Paulo-Brasil*.

McCullock, W.S., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity, *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115-133.

Mitchell, T.M. (1997). *Machine Learning*. 1st edition. New York: McGraw-Hill.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, vol. 59, no. 2, pp. 347-370.

Poon, S.H. (2005). *A Practical Guide to Forecasting Financial Market Volatility*. 1st edition. Chichester: John Wiley.

Roh, T.H. (2007). Forecasting the Volatility of Stock Price Index, *Expert Systems of Applications*, vol. 33, no. 4, pp. 916-922.

Shaikh, A.H., & Iqbal, Z. (2004). Using Neural Networks for Forecasting Volatility of S&P500 Index Futures Prices, *Journal of Business Research*, vol. 57, no. 10, pp. 1116-1125.

Sharda, R. (1994). Neural Networks for the MS/OR Analyst: An Application Bibliography, *Interfaces*, vol. 24, no. 2, pp. 116-130.

Silva, T.C., & Zhao, L. (2016). *Machine Learning in Complex Networks*. 1st edition. Cham: Springer International Publishing.

Taylor, J.W. (2004). Volatility Forecasting with Smooth Transition Exponential Smoothing, *International Journal of Forecasting*, vol. 20, no. 2, pp. 273-286.

Tsay, R.S. (2002). *Analysis of Financial Time Series: Financial Econometrics*. 1st edition. New York: Wiley.

Verbeek, M. (2012). *A Guide to Modern Econometrics*. 4th edition. New York: Wiley.

West, K.D., & Cho, D. (1995). The Predictive Ability of Several Models of Exchange Rate Volatility, *Journal of Econometrics*, vol. 69, no. 2, pp. 367-391.

Zakoian, J.M. (1994). Threshold Heteroscedastic Models, *Journal of Economic Dynamics and Control*, vol. 18, no. 5, pp. 931-955.

# Appendix A

**Table 9:** Estimation results from GARCH(1,1) models with Normal errors, for the in-sample period from 15th February 2008 to 8th January 2016 (413 weeks). The relevant parameter estimates are reported together with the robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5%, and 1% levels, respectively. Next to the parameter estimates, the values of AIC and BIC are reported in order to compare models. The notation $(\times 10^{-4})$ means that the reported value should be multiplied with $10^{-4}$.

| *GARCH(1,1)* | **Parameter estimates** | | | **Diagnostics** | |
|---|---|---|---|---|---|
| **Index** | $\widehat{\boldsymbol{\omega}}$ $(\times \mathbf{10^{-4}})$ | $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}}$ | $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ | $\boldsymbol{AIC}$ | $\boldsymbol{BIC}$ |
| OMXS30 | 0,305** (0,141) | 0,140*** (0,028) | 0,843*** (0,036) | -4,28 | -4,24 |
| FTSE100 | 0,409** (0,201) | 0,205*** (0,048) | 0,770*** (0,066) | -4,54 | -4,50 |
| S&P/ASX200 | 0,308** (0,144) | 0,195*** (0,037) | 0,774*** (0,054) | -4,68 | -4,64 |

**Table 10:** Estimation results from EGARCH(1,1) models with Normal errors, for the in-sample period from 15th February 2008 to 8th January 2016 (413 weeks). The relevant parameter estimates are reported together with the robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5%, and 1% levels, respectively. Next to the parameter estimates, the values of AIC and BIC are reported in order to compare models.

| *EGARCH(1,1)* | **Parameter estimates** | | | | **Diagnostics** | |
|---|---|---|---|---|---|---|
| **Index** | $\widehat{\boldsymbol{\omega}}$ | $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}}$ | $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ | $\widehat{\boldsymbol{\gamma}}_{\mathbf{1}}$ | $\boldsymbol{AIC}$ | $\boldsymbol{BIC}$ |
| OMXS30 | -0,375** (0,121) | -0,184*** (0,023) | 0,958*** (0,013) | 0,097 (0,059) | -4,36 | -4,32 |
| FTSE100 | -0,620*** (0,131) | -0,276*** (0,026) | 0,933*** (0,017) | 0,159*** (0,037) | -4,64 | -4,60 |
| S&P/ASX200 | -0,319*** (0,103) | -0,134*** (0,027) | 0,970*** (0,012) | 0,117*** (0,041) | -4,73 | -4,68 |

**Table 11:** Estimation results from TGARCH(1,1) models with Normal errors, for the in-sample period from 15th February 2008 to 8th January 2016 (413 weeks). The relevant parameter estimates are reported together with the robust standard errors in brackets. *, ** and *** denote significance at the 10%, 5%, and 1% levels, respectively. Next to the parameter estimates, the values of AIC and BIC are reported in order to compare models. The notation $(\times 10^{-4})$ means that the reported value should be multiplied with $10^{-4}$.

| *TGARCH(1,1)* | **Parameter estimates** | | | | **Diagnostics** | |
|---|---|---|---|---|---|---|
| **Index** | $\widehat{\boldsymbol{\omega}}$ $(\times \mathbf{10^{-4}})$ | $\widehat{\boldsymbol{\alpha}}_{\mathbf{1}}$ | $\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ | $\widehat{\boldsymbol{\delta}}_{\mathbf{1}}$ | $\boldsymbol{AIC}$ | $\boldsymbol{BIC}$ |
| OMXS30 | 0,791*** (0,125) | -0,012 (0,038) | 0,718*** (0,048) | 0,473*** (0,069) | -4,35 | -4,30 |
| FTSE100 | 0,673*** (0,153) | -0,055 (0,035) | 0,731*** (0,036) | 0,515*** (0,066) | -4,63 | -4,58 |
| S&P/ASX200 | 0,338** (0,131) | -0,028 (0,056) | 0,784*** (0,055) | 0,281*** (0,060) | -4,71 | -4,66 |

# Appendix B

**Table 12:** Out-of-sample predicting performance of GARCH(1,1), EGARCH(1,1), TGARCH(1,1) and four ANN models with 1, 2, 5 and 10 hidden neurons, respectively, in the hidden layer, for the weekly volatility of OMXS30. The ARCH-type models have Normal errors. The out-of-sample period ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years/103 weeks). The four metrics used to compare the volatility-predicting performance are the Root Mean Squared Error (RMSE), the Mean Average Error (MAE), the Mean Average Percentage Error (MAPE) and the Out-of-sample $R^2$. The notations $(\times 10^{-3})$ and $(\times 10^{3})$ mean that the reported values should be multiplied with $10^{-3}$ and $10^{3}$, respectively.

| | ***Metrics*** | | | |
|---|---|---|---|---|
| **Model** | **RMSE** $(\times \mathbf{10^{-3}})$ | **MAE** $(\times \mathbf{10^{-3}})$ | **MAPE** $(\times \mathbf{10^{3}})$ | **Out-of-sample** $\boldsymbol{R^2}$ |
| GARCH(1,1) Normal errors | 0,547 | 0,440 | 13,6% | 0,587 |
| EGARCH(1,1) Normal errors | 0,588 | 0,481 | 13,5% | 0,524 |
| TGARCH(1,1) Normal errors | 0,611 | 0,479 | 13,3% | 0,485 |

**Table 13:** Out-of-sample predicting performance of GARCH(1,1), EGARCH(1,1), TGARCH(1,1) and four ANN models with 1, 2, 5 and 10 hidden neurons, respectively, in the hidden layer, for the weekly volatility of FTSE100. The ARCH-type models have Normal errors. The out-of-sample period ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years/103 weeks). The four metrics used to compare the volatility-predicting performance are the Root Mean Squared Error (RMSE), the Mean Average Error (MAE), the Mean Average Percentage Error (MAPE) and the Out-of-sample $R^2$. The notations $(\times 10^{-3})$ and $(\times 10^{3})$ mean that the reported values should be multiplied with $10^{-3}$ and $10^{3}$, respectively.

| | ***Metrics*** | | | |
|---|---|---|---|---|
| **Model** | **RMSE** $(\times \mathbf{10^{-3}})$ | **MAE** $(\times \mathbf{10^{-3}})$ | **MAPE** $(\times \mathbf{10^{3}})$ | **Out-of-sample** $\boldsymbol{R^2}$ |
| GARCH(1,1) Normal errors | 0,612 | 0,382 | 314% | 0,379 |
| EGARCH(1,1) Normal errors | 0,578 | 0,340 | 215% | 0,445 |
| TGARCH(1,1) Normal errors | 0,609 | 0,364 | 308% | 0,385 |

**Table 14:** Out-of-sample predicting performance of GARCH(1,1), EGARCH(1,1), TGARCH(1,1) and four ANN models with 1, 2, 5 and 10 hidden neurons, respectively in the hidden layer, for the weekly volatility of S&P/ASX200. The ARCH-type models have Normal errors. The out-of-sample period ranges from 15th January 2016 to 29th of December 2017 (approx. 2 years/103 weeks). The four metrics used to compare the volatility-predicting performance are the Root Mean Squared Error (RMSE), the Mean Average Error (MAE), the Mean Average Percentage Error (MAPE) and the Out-of-sample $R^2$. The notations $(\times 10^{-3})$ and $(\times 10^{3})$ mean that the reported values should be multiplied with $10^{-3}$ and $10^{3}$, respectively.

| | ***Metrics*** | | | |
|---|---|---|---|---|
| **Model** | **RMSE** $(\times \mathbf{10^{-3}})$ | **MAE** $(\times \mathbf{10^{-3}})$ | **MAPE** $(\times \mathbf{10^{3}})$ | **Out-of-sample $R^2$** |
| GARCH(1,1) Normal errors | 0,413 | 0,319 | 119% | 0,485 |
| EGARCH(1,1) Normal errors | 0,376 | 0,293 | 119% | 0,573 |
| TGARCH(1,1) Normal errors | 0,394 | 0,290 | 124% | 0,532 |