On Return Innovation Distribution in GARCH Volatility Modelling

Empirical evidence from the Stockholm Stock Exchange

Kristofer Eriksson
krer@live.se
Abstract

Volatility is an important variable in many financial models and as such, it is of great importance that the estimations and predictions are accurate. This paper considers some of the volatility models in the GARCH family, namely; GARCH, EGARCH and a special case of TGARCH. These models are analyzed in the case of an assumed normal, Student-t and generalized error distribution for the return innovations. A nonparametric approach based on the algorithm by Bühlmann and McNeil (2002) is also included. The performance of these volatility models are evaluated both in-sample and out-of-sample by using the loss functions MSE and QLIKE on the Swedish stock index OMXS30. To avoid the risk of data snooping, this study uses the Superior predictive ability test. The main results and conclusions are; in the parametric volatility models, an assumed normal distribution is enough for OMXS30 since it can handle the leptokurtosis faced in this stock index equally good as the other distributions. This is a favourable result since the normal distribution needs less restriction for the EGARCH and TGARCH model to fulfil stationarity and is not as computationally demanding. A leverage effect is present in the OMXS30 which suggests the use of an asymmetric model. The nonparametric approach and TGARCH perform well both in-sample and out-of-sample but the nonparametric approach is extremely computationally demanding. This paper therefore suggests the use of the TGARCH model for OMXS30 with an assumed normal distribution for the return innovations.

Keywords: Volatility, GARCH, EGARCH, TGARCH, Nonparametric, Return innovation distribution.
## Table of Contents

1. Introduction .................................................................................................................. 1

2. Theory .......................................................................................................................... 4
   2.1 The volatility models ................................................................................................. 4
       2.1.1 GARCH ............................................................................................................. 4
       2.1.2 EGARCH .......................................................................................................... 5
       2.1.3 TGARCH ......................................................................................................... 6
       2.1.4 Comparison ....................................................................................................... 6
   2.2 Innovation distribution modelling ............................................................................. 6
       2.2.1 Parametric ........................................................................................................... 6
       2.2.2 Nonparametric .................................................................................................. 7
   2.3 Model evaluation ...................................................................................................... 8

3. Data .................................................................................................................................. 11

4. Results ........................................................................................................................... 12

5. Conclusions .................................................................................................................. 14

References ......................................................................................................................... 15
1. Introduction

Sound financial decisions are usually based on both the return and the risk involved since a trade-off between return and risk is almost always present. To understand risk, several measures are available; volatility, covariance, correlation and copulas to just name a few. In financial theory these measures are commonly used as inputs in models, for example in option pricing theory, portfolio allocation theory and risk management. It is therefore of great importance that the models used to estimate and forecast these measures are good and accurate. However, to determine what is a good model in a general way is not an easy task since the models are used for different applications. For example, an option trader that considers pricing an option with time to maturity of one year may need one kind of model while a risk manager that wants to calculate the Value-at-Risk for the next coming days is certainly in need of another.

This paper considers univariate models of the Generalized Autoregressive Conditional Heteroskedastic (GARCH) family, introduced by Engle (1982) and generalized by Bollerslev (1986). These models can be used to estimate and forecast the volatility in assets. Volatility is in fact the annual standard deviation of the asset price and as such it measures the variation in the price. When performing a least square regression, the error variance is assumed to be constant, i.e. homoskedasticity is assumed. When conditional heteroskedasticity is present we may want to turn to a model that can account for this. The GARCH family focus on this by modelling the conditional variance and therefore it can be used to model the volatility.

A commonly used assumption is that financial markets are efficient as described in the Efficient market hypothesis (EMH) developed by Fama (1970). However, many articles and research has criticized this framework. Anomalies as seasonal effects like the January effect and the day-of-week effect, the size effect and financial bubbles has among others been used as evidence against the EMH. There has been a stream of suggestions and evidence for market microstructure effects when considering high-frequency data, like intra-daily and to some extent daily data. Tsay (2010) summaries in his book several microstructure effects that the literature has proposed, for example, nonsynchronous trading i.e. assets trade with different intensity or are geographically placed in different time zones which imply that new information is incorporated in the prices at different points in time. This can lead to a lag-1 serial correlation in the return of portfolios. Another one is that quoted prices are discrete and usually in a bid and ask price structure which can give rise to an effect called bid-ask bounce. This is caused when the “true price” is different from the bid price which is different from the ask price and can give rise to a bouncing between the bid and ask which may introduce a negative lag-1 serial correlation in the return. Even if the EMH has been widely discussed in the literature, whether it is possible to predict future returns or not, the second moment of the returns, i.e. the variance, has been widely accepted to be predictable.

Some stylized facts has been presented in the modelling of volatility. Volatility clustering is present in most financial time series, i.e. the volatility changes with time, sometimes it is high and other times it is low and that there tends to be a persistence in volatility - a high volatility is often followed by high volatility. The returns have “fat tails”, i.e. leptokurtosis, and negative skewness. Black (1976) came up with what is called the leverage effect, i.e. volatility tends to increase more after a negative return than a positive one. To address these stylized facts, an abundance of different models has been proposed. This paper studies the GARCH model by Bollerslev (1986), the Exponential GARCH (EGARCH) model of Nelson (1991) and a special case of the Threshold GARCH (TGARCH) model by Zakoian (1994). In practice a parametric framework is usually adopted to estimate the parameters in these models. This is
done by maximizing a log-likelihood function, often from a standard normal, a Student-t or a standard generalized error distribution (GED), which implies that the distribution of the return innovation is assumed to come from either of these three distributions. Even the simple normal GARCH model takes to some extent into account the leptokurtosis in the return since it uses the sum of normal variables but with different variances. However, when extended to the Student-t distribution or GED it can model more severe leptokurtosis which is usually the case for financial data. The simple GARCH model that uses these parametric distributions cannot account for the leverage effect since it is a purely symmetric model. The EGARCH and TGARCH on the other hand are asymmetric models that were developed to address this and will therefore handle the negative skewness in the return innovations. All these parametric models are nevertheless subject to misspecification errors and researchers have therefore been developing semi- and nonparametric approaches to handle the distribution of the return innovations.

Semi- and nonparametric approaches has been studied for quite some time but with more intensity during recent years as the computational power has increased along with the invention of new and more sophisticated semi- and nonparametric techniques. However, in a recent article, Wang et al. (2012) wrote

“Non/semi-parametric methods enhance the flexibility of the volatility models that practitioners use. However, due to the limitations in either interpretability, computational complexity, or theoretical reliability, most of the nonparametric stochastic volatility models have not been widely used as general tools in volatility analysis.”

which summaries the problems faced in this framework. Pagan and Schwert (1990) studied nonparametric models by using a kernel estimator and a flexible Fourier form estimator and found that the nonparametric models adjusted more quickly to shocks in volatility than the parametric models. Engle and González-Rivera (1991) proposed a semiparametric GARCH model that uses a nonparametric estimator, namely, the discrete penalized likelihood estimation technique. They showed that under the normal distribution using the quasi-maximum likelihood estimator (QMLE), up to 86% of the efficiency could be lost due to misspecification and that their semiparametric approach could increase the efficiency with up to 50% more than the QMLE. This method was however only partly successful and many researchers set out to find a better estimator, for example Drost and Klaassen (1997). Bühlmann and McNeil (2002) developed a nonparametric approach to deal with the distribution of the return innovation in a GARCH model by using an iterative method including backfitting that they proved under weak assumptions to converge. A broad class of different semiparametric and nonparametric approaches has been developed in recent years, see for example Shin (2005), Yang (2006), Hudson and Gerlach (2008), Ausin, Galeano and Ghosh (2010) and Yang and Wu (2011), but from the citation above by Wang et al. (2012), we understand that there is a lot of research left to do in this area. This paper uses the model developed by Bühlmann and McNeil (2002) which is relatively easy to implement. It will be referred to as the nonparametric GARCH.

To sum up which volatility models this paper uses; GARCH, EGARCH, TGARCH in a normal, Student-t and GED approach, and the nonparametric GARCH. The main question in this paper is which one of these models are in statistical terms the best in estimating both in-sample and out-of-sample volatility for the daily close prices of OMXS30 (Swedish stock index containing the 30 greatest stocks in terms of market capitalization on the Stockholm Stock Exchange). As has already been pointed out, to determine which model is the best in a general way is not an easy task. It depends on the application it will be used in. Here the volatility is estimated on a daily basis and the out-of-sample forecast considers one day ahead.
It is tempting to develop models that take into account many stylized facts about how the market works and to use complex techniques that almost fits exactly to the data available. This usually leads to very precise in-sample estimates. However, in most applications we are more interested in predicting the volatility. One then has to accept that the market’s behaviour will always be to some extent unpredictable, which means that a robust but simple model may perform better out-of-sample than a more complex and over-fitted model. With this in mind, the out-of-sample analysis is of greater interest to focus on.

To evaluate whether a model is good or not in estimating and predicting volatility in statistical terms is also a bit difficult since volatility is not directly observable. Volatility is not something that is directly traded as the price and can only be determined by using models. Up until the late 90’s it was popular by both researchers and practitioners to perform the analysis by using the Mincer-Zarnowitz regression. This model uses the estimated volatility and the squared returns to obtain a $R^2$ value. When using daily data many papers reported that the GARCH family provided poor predictions by using this method. However, Andersen and Bollerslev (1998) showed that the daily squared return is a very noisy proxy. When estimating the realized squared return with intra-day data they could conclude that the GARCH models do actually provide good forecasts. Christoffersen (1998) came up with some new methods to evaluate the out-of-sample performance by predicting the tails of a distribution in what he calls conditional and unconditional coverage tests. Patton (2006) derived a family of loss functions that are robust to the most commonly used noisy volatility proxies; the daily squared return, the intra-daily range and the realized volatility estimator. The realized volatility estimator is the sum of intra-daily squared returns at some certain intra-daily frequency while the intra-daily range is an estimator that uses the high and low intra-day prices. As reported by Patton (2006), the intra-daily range in the framework of the MSE loss function, correspond to the realized volatility estimator when using 5 intra-daily observations. A smaller time grid in the realized volatility are shown in Andersen and Bollerslev (1998) to yield a more accurate volatility proxy but that a higher frequency may be subject to market microstructure effects. It can also be difficult to obtain intra-day data in certain markets. The high and low intra-day prices are easier to obtain. This paper therefore uses the intra-daily range proxy with the MSE and the QLIKE loss functions as proved by Patton (2006) to be robust to these proxies. When comparing different models in a single dataset a certain model could outperform a benchmark model by chance even if the benchmark model is better (this is called data snooping). To address this issue, Hansen (2005) developed a model called Superior Predictive Ability (SPA) which this paper uses.

Bühlmann and McNeil (2002) empirically showed that their model is good for stocks as they analyzed in-sample estimates of the BMW stock. Hou and Suardi used the Bühlmann and McNeil algorithm in a similar manner as in this paper for interest rate (2011) and crude oil prices (2012) with the evidence that the nonparametric approach is better than the parametric models. Malmsten (2004) tested the normal GARCH against EGARCH on the 30 most actively traded stocks on the Stockholm Stock Exchange and found that the in-sample estimates of the two models are almost equally good. To the author’s knowledge, no analysis has been done with the nonparametric GARCH in this paper on a stock index and that the Swedish stock market has not been fully explored when it comes to the model evaluation framework that this paper considers. Furthermore, an investigation of the assumed return innovation distribution for the parametric models along with the determination of which volatility model to use, is of interest for practitioners in volatility modelling of the OMXS30. This paper therefore contributes with empirical evidence from the Stockholm Stock Exchange in volatility modelling and forecasts.
2. Theory

In this section the theory is outlined and structured in the following way; first, the basic volatility models are presented, i.e. the GARCH, EGARCH and TGARCH and a comparison between them. Second, the choice for the return innovation distribution is outlined in the case of a parametric approach for a normal, Student-t and generalized error distribution, as well as in the case of a nonparametric approach. Third, the theory behind the in-sample and out-of-sample analysis is described.

2.1 The volatility models

The GARCH family models incorporates the possibility to estimate higher order models in terms of lag, in a general case GARCH\((p,q)\) would be such a model where it depends on the \(p\) first lags in the ARCH term and the \(q\) first lags in conditional variance. In practice a low order model is usually used and this is often as good as a higher order model which has been pointed out by several authors. For example, Malmsten (2004) in case of the EGARCH model on the 30 most actively traded stocks on the Stockholm Stock Exchange found some evidence to use a EGARCH\((1,2)\) model. Hansen and Lunde (2005) also suggest the use of a low order model since they found that \(p = q = 2\) models rarely performed better than the lower order models in the out-of-sample analysis. This paper considers only the first order lags in all volatility models, i.e. \(p = q = 1\).

Much research has been carried out to give the GARCH family models sufficient conditions such that the process is stationary and consistent. There are a great amount of tests to show whether the parameters in the models holds stationarity, constancy, restricted higher order moments and positivity, and how to restrict the models so that these are fulfilled. However, restricting the model may also restrict its ability to handle the dynamic in the volatility. To choose which restrictions to use is an entire science and art in itself. This paper uses the MFE toolbox for Matlab by Kevin Sheppard to estimate the parametric models. His algorithms impose the restrictions presented under each model subsection bellow.

All models use the daily log return, i.e. \(r_t = \ln(P_t) - \ln(P_{t-1})\). Now assume that this process can be described by

\[
y_t = \mu_t + \epsilon_t = \mu_t + \sqrt{h_t}z_t
\]

(1)

where \(\mu_t = E[y_t|I_{t-1}]\) is the conditional mean given the information set \(I_{t-1}\), \(\{\epsilon_t\}\) are the return innovations, \(h_t = \text{Var}[y_t|I_{t-1}] = E[(y_t - \mu_t)^2|I_{t-1}]\) is the conditional variance and \(\{z_t\}\) are i.i.d. random variables with mean 0, variance 1 and for which some distribution is assumed in the parametric case but not in the nonparametric GARCH.

2.1.1 GARCH

The GARCH\((1,1)\) model is given by

\[
h_t = \omega + \alpha \epsilon^2_{t-1} + \beta h_{t-1}
\]

(2)

where \(\omega, \alpha, \beta\) are parameters, \(\epsilon^2_{t-1}\) and \(h_{t-1}\) are the squared return innovation and conditional variance in the previous time step, respectively. Bollerslev (1986) showed restrictions on the parameters for positivity, \(\omega > 0, \alpha \geq 0, \beta \geq 0\), and the wide-sense stationarity condition, \(\alpha + \beta < 1\). Nelson (1990) proved that the GARCH\((1,1)\) process is uniquely stationary if \(E[\log(\beta + \alpha \epsilon^2_t)] < 0\), and Bougerol and Picard (1992a, 1992b) generalized this for any
GARCH\((p,q)\) order model, but neither the Nelson or the Bougerol and Picard restrictions are included in the algorithm. If the process is stationary, then higher order moments exist. Bollerslev (1986) proved that if the fourth order moment exists, then the model can handle leptokurtosis. By looking at the functional form in equation (2) we can see that the model incorporates the behaviour of volatility clustering, i.e. a high value in either \(\epsilon_{t-1}^2\) or \(h_{t-1}\) will imply a high volatility in the next time step \(h_t\).

The unconditional variance, also called the long term variance, is given in the case of a symmetric distribution by \(\omega/(1-(\alpha+\beta))\) and can be seen as the long term average value of the conditional variance. If the GARCH model is correctly specified it will converge to this long term variance as the forecast horizon is increased. For further reading about the GARCH model see for example, He (1997), Lundbergh and Teräsvirta (2002) and Berkes, Horváth and Kokoszka (2004).

2.1.2 EGARCH

To address the asymmetric behaviour of the leverage effect and to avoid the parameter restrictions for positivity as in the GARCH model, Nelson (1991) developed the EGARCH model. The EGARCH(1,1) model is given by

\[
\ln(h_t) = \omega + \theta z_{t-1} + \gamma(|z_{t-1}| - E[|z_{t-1}|]) + \beta \ln(h_{t-1})
\]

(3)

where \(\omega, \theta, \gamma, \beta\) are the parameters of the model and

\[
E[|z_{t-1}|] = \begin{cases} \sqrt{2/\pi}, & \text{for standard normal variable} \\ \sqrt{(v-2)/2} \cdot \Gamma((v-1)/2)/\Gamma(v/2), & \text{for Student-t} \\ \lambda 2^{1/v} \Gamma(2/v)/\Gamma(1/v), & \text{for GED} \end{cases}
\]

(4)

where \(\Gamma(\cdot)\) is the gamma distribution, \(v\) is the degrees of freedom and

\[
\lambda = [2^{-2/v} \Gamma(1/v)/\Gamma(3/v)]^{1/2}
\]

(5)

The algorithm uses only \(\sqrt{2/\pi}\) for all distributions. However, the values of \(v\) that we expect for the Student-t and GED gives \(E[|z_{t-1}|] \approx \sqrt{2/\pi}\). Nelson (1991) proved that if there exists a second order moment then the process is strictly stationary if and only if \(|\beta| < 1\) for a normal distribution or a GED with \(v > 1\). When \(v < 1\) for a GED and \(v > 2\) for the Student-t distribution the following conditions needs to be added for the process to be stationary, \(\gamma \leq -|\theta|\) and \(\gamma \beta \leq |\beta\theta|\). However, this case is only partly incorporated in the algorithm for which \(v\) is only allowed to be greater than 1 for the GED and greater than 2 for the Student-t. This means that the GED is always stationary if the first condition is fulfilled, and in the case of Student-t distribution the model is at least able to fulfil stationarity.

From equation (3) we see that the model incorporates the asymmetric effect by comparing when \(z_{t-1} > 0\), then we have \((\theta + \gamma)z_{t-1}\) and in the case of \(z_{t-1} < 0\), we have \((\theta - \gamma)z_{t-1}\). Hence, if \(\theta = \gamma\) the model will only respond to positive shocks and contrary if \(\theta = -\gamma\) it will only respond to negative shocks. The long term log variance is given by \(\omega/(1-\beta)\) which means that the unconditional variance for the EGARCH(1,1) has a different functional form than the GARCH(1,1). For further reading about the EGARCH model, see for example Malmsten (2004).
2.1.3 TGARCH

Another model developed to handle the leverage effect is the TGARCH by Zakoian (1994). This paper considers a special case of the original TGARCH(1,1) model, namely,

\[ \sqrt{h_t} = \omega + \alpha |\epsilon_{t-1}| + \gamma \cdot \mathbb{I}_{\{\epsilon_{t-1}<0\}} |\epsilon_{t-1}| + \beta \sqrt{h_{t-1}} \]  

where \( \omega, \alpha, \gamma, \beta \) are parameters and \( \mathbb{I}_{\{\epsilon_{t-1}<0\}} \) is the indicator function. The conditions for positivity are \( \omega > 0, \alpha \geq 0, \beta \geq 0, \alpha + \gamma > 0 \). The algorithm uses the conditions \( v > 1 \) for the GED, \( v > 2 \) for the Student-t distribution and \( \alpha + \gamma/2 + \beta < 1 \) to weakly fulfil stationarity.

It is easy to see from equation (6) that the model incorporates the leverage effect since it has a certain term for negative return innovations. The unconditional variance for the above model is given for a symmetric distribution by \( \left( \omega/(1 - (\alpha + \gamma/2)E[|\epsilon_{t-1}|] - \beta) \right)^2 \) where \( E[|\epsilon_{t-1}|] \) is given in equation (4). For further reading about TGARCH, see for example Rabemananjara and Zakoian (1993).

2.1.4 Comparison

The strength with the GARCH model is that it is simple and robust. It can handle volatility clustering and also the leptokurtosis, especially when a student-t distribution or GED are assumed for the return innovations. However, it may need to be constrained to fulfil the positivity condition and it is a completely symmetric model in the case of the distributions for the return innovations assumed in this paper. Since it is a symmetric model it cannot handle the leverage effect. It should also be noted that the GARCH model is a linear model and as such, it is assumed that it is possible to estimate the volatility in a linear manner. The EGARCH and TGARCH models are asymmetric models that can handle the leverage effect and allow to some degree the volatility to evolve in a non-linear manner. As such, the EGARCH and the TGARCH are quite similar. However, there are also some differences between them. The EGARCH is in terms of the logarithm which means that no positivity constraints are needed. On the other hand, the TGARCH model do more clearly separates the term for negative shocks and is more similar in form to the simple GARCH model. Common for both the EGARCH and TGARCH is that the stationarity conditions are highly dependent on the assumed distribution for the return innovations.

2.2 Innovation distribution modelling

The return innovations \( \{\epsilon_t\} \), defined in equation (1), has in the parametric models an assumed distribution. This paper uses the normal distribution, the Student-t distribution and the generalized error distribution for all three volatility models. For simplicity the distributional form will be included in the name of the model, for example, normal GARCH or Student-t TGARCH. When assuming a certain distribution for the return innovations we face misspecification errors which can damage the estimate and forecast of volatility. Therefore a nonparametric approach is also included.

2.2.1 Parametric

This subsection describes the log-likelihood functions used to estimate the parameters in the volatility models for all considered distributions. The problem faced when trying to maximize these log-likelihood functions are that their surfaces tends to be quite flat and they demands a minimum amount of data to yield good results and to be robust.
Normal distribution

In the case of a standard normal distribution for the i.i.d. random variables \( \{z_t\} \), the following log-likelihood function needs to be maximized

\[
\ln(L(\Phi)) = -\frac{1}{2} \sum_{t=1}^{n} \left( \ln(h_t) + \left( \frac{\varepsilon_t}{\sqrt{h_t}} \right)^2 \right)
\]

(7)

where \( \Phi \) is the vector of parameters that are going to be estimated and \( h_t \) is the chosen volatility model, i.e. GARCH as in equation (2), the exponential function of EGARCH see equation (3) and the square of TGARCH in equation (6).

Student-t distribution

The Student-t distribution can handle more severe leptokurtosis. The log-likelihood function is defined as

\[
\ln(L(\Phi)) = -\sum_{t=1}^{n} \left( \ln(\sqrt{h_t}) + \left(\frac{v+1}{2}\right) \ln \left( 1 + (v-2)^{-1} \left( \frac{\varepsilon_t}{\sqrt{h_t}} \right)^2 \right) \right) + T \psi_v
\]

(8)

where \( v \) is the degrees of freedom and

\[
\psi_v = \ln \left( (v-2)^{-1/2} \Gamma \left( \frac{v}{2} \right) \Gamma \left( \frac{v+1}{2} \right) \right)
\]

(9)

for which \( \Gamma(\cdot) \) is the gamma distribution. The Student-t distribution incorporates the standard normal distribution as a special case when \( v = \infty \) and the Cauchy distribution when \( v = 1 \). Hence, a lower parameter \( v \), yields a distribution with “fatter tails”.

Generalized error distribution

Nelson (1991) proposed the use of GED when estimating EGARCH since it is more appealing in terms of fulfilling stationarity compared to for example the Student-t distribution. In the case of a Student-t distribution the unconditional means and variances may not be finite in the EGARCH. The log-likelihood function for the standard GED is defined as

\[
\ln(L(\Phi)) = -\frac{1}{2} \sum_{t=1}^{n} \left( \ln(h_t) + \left( \frac{\varepsilon_t}{\lambda \sqrt{h_t}} \right)^v \right) - \phi_v
\]

(10)

where \( \lambda \) is defined in equation (5) and

\[
\phi_v = (1 + v^{-1}) \ln(2) + \ln\left( \Gamma(1/v) \right) + \ln(\lambda/v)
\]

(11)

GED incorporates both the normal \( (v = 2) \), the Laplace \( (v = 1) \) and the uniform \( (v = \infty) \) distribution as special cases, where “fatter tails” than the normal distribution is obtained for \( v < 2 \).

2.2.2 Nonparametric

The nonparametric GARCH assumes no distribution for the return innovations and are therefore a nonparametric method in that sense. It is also nonparametric in the way that no functional relationship between the estimated volatility and the lagged variables are assumed. The core idea in this model is to first get an initial estimate of the volatility by for example
using a parametric model, then using a nonparametric smoothing technique in an iterative way to get better and better estimates of the volatility. To outline the model more precisely, consider the process in equation (1) and

\[ h_t = f(\epsilon_{t-1}, h_{t-1}) \]  

(12)

where \( f \) is a strictly positive function. This is a first order model which will be estimated in a nonparametric way, in contrast to the previous described models. The algorithm developed by Bühlmann and McNeil (2002) is described in Algorithm 1.

**Algorithm 1 Nonparametric GARCH**

1. Estimate the conditional variance \( \{\hat{h}_{t,0}; 1 \leq t \leq n\} \), where \( n \) is the number of data points in the sample, by using a parametric GARCH(1,1).
2. for \( m = 1 \) to \( M \) do
3. Use a nonparametric smoothing technique in a weighted regression to estimate \( f_m \) by regress \( \{\epsilon_t^2; 2 \leq t \leq n\} \) against \( \epsilon_{t-1}; 2 \leq t \leq n \) and \( h_{t-1}; 2 \leq t \leq n \) with weights \( h_t, m-1; 2 \leq t \leq n \).
4. Calculate \( \{\hat{h}_{t,m} = f_m(\epsilon_{t-1}, \hat{h}_{t-1,m-1}); 2 \leq t \leq n\} \) and set \( \hat{h}_{1,m} = \hat{h}_{1,m-1} \)
5. end for
6. Averaging over the final \( K \) steps \( \hat{h}_{t,*} = \frac{1}{K} \sum_{m=M-K+1}^{M} \hat{h}_{t,m} \)
7. Use step 3 and 4 but with \( \hat{h}_{t,*} \) to obtain final estimate of the conditional variance \( \hat{h}_t \)

Bühlmann and McNeil (2002) found that the improvement mainly took place during the first 1-4 iteration steps and this paper uses their suggested values \( M = 8 \) and \( K = 5 \). The nonparametric smoothing technique is based on the LOESS method. If the model needs to be restricted to fulfil positivity, the kernel smoothing can be used with a positive kernel as pointed out by Bühlmann and McNeil (2002).

The algorithm has four assumptions; firstly, contraction of the function \( f \) with respect to the conditional variance is essential for the reasoning since without it, the model would not converge to the true conditional variance. Secondly, the fourth moment must be finite for both the initial process (the parametric GARCH in the algorithm above) and the final estimate of the conditional variance. The third and the fourth assumptions are dealing with the estimation error. See Bühlmann and McNeil (2002) for more precise definitions of the assumptions as well as the theorems that prove that the algorithm converges.

### 2.3 Model evaluation

The evaluation of the models is based on both an in-sample and out-of-sample analysis. To illustrate the difference between them, consider a dataset with historical data for time \( 1 \leq t \leq n \). The in-sample analysis sets the volatility model at \( t = n \) and estimates the volatility in every point in time before that for the entire sample. The out-of-sample analysis on the other hand starts by setting the volatility model at \( t = n - s \), where \( s < n \) is the time length we wish to perform the out-of-sample analysis on. At \( t = n - s \) the volatility model predicts the volatility for the next time step and saves this for the sample to be analyzed. Then repeating this procedure for every out-of-sample time step until \( t = n - 1 \). Hence, the out-of-sample analysis deals with the models ability to predict while the in-sample analysis is based on how good the model is to estimate past volatility.
Loss functions

Both the in-sample and out-of-sample analysis is performed by using the MSE and QLIKE loss functions. These loss functions are robust to noisy proxies as proved by Patton (2006). The proxy this paper uses is the adjusted intra-daily log range estimator. The intra-daily log range estimator is defined as

\[ RG_t = \max_{\tau} (\ln(P_t)) - \min_{\tau} (\ln(P_t)), \quad t - 1 < \tau \leq t \]  

(13)

where \( \tau \) is the intra-daily time frequency. However, Patton (2006) points out that this is a biased estimator and therefore needs to be adjusted to

\[ RG_t^* = \frac{RG_t}{2\sqrt{\ln(2)}} \]  

(14)

\( RG_t^* \) is the adjusted intra-daily log range estimator which is an unbiased proxy for the conditional variance when it is squared.

The idea behind loss functions is to get a measure on how well estimates corresponds to the true values. A lower value on the loss function implies a better estimate. This paper uses the adjusted intra-daily log range estimator as a proxy for the true conditional variance. The MSE and the QLIKE are defined respectively as

\[ MSE = \frac{1}{n} \sum_{t=1}^{n} L_{MSE} (\hat{h}_t, h_t) = \frac{1}{n} \sum_{t=1}^{n} (\hat{h}_t - h_t)^2 \]  

(15)

\[ QLIKE = \frac{1}{n} \sum_{t=1}^{n} L_{QLIKE} (\hat{h}_t, h_t) = \frac{1}{n} \sum_{t=1}^{n} \left( \ln(h_t) + \frac{\hat{h}_t}{h_t} \right) \]  

(16)

where \( L_{MSE}(\hat{h}_t, h_t) \), \( L_{QLIKE}(\hat{h}_t, h_t) \) are the MSE and QLIKE loss functions, respectively, \( \hat{h}_t \) is the estimated conditional variance and \( h_t \) is the true conditional variance.

Superior predictive ability test

Since several volatility models are compared in just a single set of data we faces the problem of data snooping, i.e. some model may by chance outperform a benchmark model even though the benchmark is actually better. To address this, the SPA test developed by Hansen (2005) can be used. It tests whether the benchmark model is as good as or better than any of the alternative models and therefore determines which model is actually the best. The test uses the loss functions in terms of a performance variable

\[ d_{t,k} = L(\hat{h}_{t,0}, h_t) - L(\hat{h}_{t,k}, h_t) \]  

(17)

where \( L(\hat{h}_{t,0}, h_t) \) is the loss value for the benchmark model and \( L(\hat{h}_{t,k}, h_t) \) is for the alternative model \( k \). Hence, if \( E[d_{t,k}] > 0 \) then the alternative model \( k \) is better than the benchmark. Now, we can construct the null hypothesis by first constructing the vector of all alternative models, \( l \), in terms of the performance variable \( d_t = (d_{t,1}, ..., d_{t,l}) \) and then taking the expected value of the vector, \( \mu_d = E[d_t] \). The null hypothesis is then

\[ H_0: \mu_d \leq 0 \]  

(18)
which is interpreted as, “the benchmark is not inferior to any of the alternatives”, as written by Hansen (2005). To test this, the following test statistic is used

\[
T^{SPA} = \max \left( \max_{k=1,\ldots,l} \left( \frac{n^{1/2} \bar{d}_k}{\hat{\eta}_k} \right), 0 \right)
\]  

(19)

where \( \bar{d}_k = n^{-1} \sum_{t=1}^{n} d_{t,k} \) and \( \hat{\eta}_k \) is when squared the estimated variance of \( n^{1/2} \bar{d}_k \). To estimate the distribution of the test statistics under the null hypothesis and \( \hat{\eta}_k \), a stationary bootstrap can be used in the following way as described by Hansen (2005). First of all, resample \( d_{t,k} \) to \( d_{t,b,k}^* \) by using a stationary bootstrap for \( b = 1, \ldots, B \) where \( B \) is the number of bootstrap samples. The stationary bootstrap can handle the dependency in the time series data and is basically a block bootstrap but with random block length. In this paper the average block length for the stationary bootstrap is 10. Calculate \( \hat{\eta}_k^2 = B^{-1} \sum_{b=1}^{B} \left( n^{1/2} d_{t,b,k}^* - n^{1/2} \bar{d}_k \right)^2 \)

(20)

Then under the null hypothesis and the estimate of \( \mu_d \) we have

\[
Z_{t,b,k}^* = d_{t,b,k}^* - \bar{d}_k \sqrt{n^{-1/2} \left( \hat{\eta}_k^2 + (\ln(n))^{2/3} \right)}
\]

(21)

From here, calculate \( \tilde{Z}_{b,k} = n^{-1} \sum_{t=1}^{n} Z_{t,b,k}^* \) and

\[
T_b^{SPA^*} = \max \left( \max_{k=1,\ldots,l} \left( \frac{n^{1/2} \tilde{Z}_{b,k}}{\hat{\eta}_k} \right), 0 \right)
\]

(22)

to obtain the p-value

\[
\hat{\pi}_{SPA} = \sum_{b=1}^{B} \frac{\mathbb{I}_{T_b^{SPA^*} > T^{SPA}}}{B}
\]

(23)

In this study \( B = 5000 \). For further reading about the SPA test, see Hansen (2005) and Hansen and Lunde (2005). When performing the test, all models are used as benchmark to compare if the null hypothesis is rejected for any of the models. When performing hypothesis tests there are always the risk of a Type I error, i.e. rejecting a true null hypothesis. This error is increased when comparing several hypothesis tests and therefore needs to be corrected so that the chosen significance level is still valid. This study uses Bonferroni correction which is a very simple and conservative way to correct for this. It is just to adjust the significance level to the number of test \( \alpha_{adj,sign.lvl} = \alpha_{sign.lvl} / \#Tests \). Hence, when comparing all ten volatility models as benchmarks against any of the others we need to perform ten tests and have to correct for this.

This paper uses a significance level of 5% for all tests.
3. Data

This paper uses data from the Stockholm Stock Exchange, more precisely, the OMXS30 stock index, and is obtained from Datastream. The OMXS30 contains the 30 most traded stocks on the Stockholm Stock Exchange and the data is based on historical constituents. This paper considers the sample from 1994-12-30 to 2012-12-28, even though the index was introduced in 1986. The sample therefore contains the dot-com bubble of 2000, the financial crisis of 2007-2008 and the ongoing European sovereign debt crisis. The sample kurtosis is 6.3640 and the skewness is 0.0818 for the entire sample. The Ljung-Box test for the 15 first lags cannot reject the null hypothesis that the data has no serial correlation for the first lag, but rejects the null hypothesis for all the other lags. The Engle’s ARCH test on the 15 first lags indicates that ARCH effects are present in all of those lags. Figure 1 and 2 shows the price, the return and the return distribution for the sample period.

![Price and Return](image1.png)

Figure 1: The price (top) and the return (bottom) for the OMXS30 on the entire sample period in this paper.

![Return Distribution](image2.png)

Figure 2: The histogram of the return for the OMXS30 on the entire sample period in this paper.
4. Results

The volatility models uses a starting value for the conditional variance to enable the estimation of the rest and therefore the model needs some time to fade away the starting value. For this reason, the in-sample analysis uses one year as burn in period and as a result considers only the period 1996-01-02 to 2012-12-28 which corresponds to 4261 data points. The out-of-sample analysis uses the sample period 2010-01-04 to 2012-12-28 which is 756 data points. Hence, the predictive performance of the volatility models is analyzed during the ongoing European sovereign debt crisis.

When estimating the parameters in the parametric volatility models for the entire sample period, the values in Table 1 are obtained. The table gives both the estimate and the standard error for the parameters along with the value of the log-likelihood function. An interpretation of the log-likelihood value is that a higher log-likelihood value yields a better fit.

Table 1: The parameter estimates of the parametric volatility models for the sample period 1995-01-02 to 2012-12-28 where the value inside the parenthesis is the standard error and $LL$ is the value of the log-likelihood. All parameter estimates are significant.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Student-t</th>
<th>GED</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\omega}$</td>
<td>2.2e-6 (6e-7)</td>
<td>2e-6 (1e-6)</td>
<td>2.0e-6 (5e-7)</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>0.08 (0.01)</td>
<td>0.08 (0.01)</td>
<td>0.08 (0.01)</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.91 (0.01)</td>
<td>0.91 (0.02)</td>
<td>0.91 (0.01)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>-</td>
<td>13 (2)</td>
<td>1.65 (0.06)</td>
</tr>
<tr>
<td>$LL$</td>
<td>12993</td>
<td>13019</td>
<td>13013</td>
</tr>
</tbody>
</table>

EGARCH

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Student-t</th>
<th>GED</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\omega}$</td>
<td>-0.17 (0.03)</td>
<td>-0.16 (0.03)</td>
<td>-0.16 (0.03)</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>-0.09 (0.01)</td>
<td>-0.09 (0.01)</td>
<td>-0.09 (0.01)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>0.15 (0.02)</td>
<td>0.15 (0.02)</td>
<td>0.15 (0.02)</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.979 (0.004)</td>
<td>0.980 (0.004)</td>
<td>0.980 (0.004)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>-</td>
<td>17 (4)</td>
<td>1.74 (0.06)</td>
</tr>
<tr>
<td>$LL$</td>
<td>13061</td>
<td>13077</td>
<td>13072</td>
</tr>
</tbody>
</table>

TGARCH

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Student-t</th>
<th>GED</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\omega}$</td>
<td>2.9e-4 (5e-5)</td>
<td>2.8e-4 (7e-5)</td>
<td>2.9e-4 (5e-5)</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>0.033 (0.008)</td>
<td>0.03 (0.01)</td>
<td>0.032 (0.008)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>0.10 (0.01)</td>
<td>0.11 (0.02)</td>
<td>0.10 (0.01)</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>0.914 (0.009)</td>
<td>0.92 (0.02)</td>
<td>0.915 (0.009)</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>-</td>
<td>17 (4)</td>
<td>1.74 (0.06)</td>
</tr>
<tr>
<td>$LL$</td>
<td>13066</td>
<td>13081</td>
<td>13076</td>
</tr>
</tbody>
</table>

In Table 1 we see that the assumed distribution for the i.i.d. random variables $\{z_t\}$ has basically no effect on the estimations of the parameters but that the EGARCH and TGARCH model seems to fit the data better than the GARCH model.
The loss values for both the in-sample estimates and the out-of-sample predictions are presented in Table 2. A lower loss value means that the model is better to estimate and forecast the conditional variance when compared to a proxy for the true conditional variance.

Table 2: The loss values for the in-sample estimates and out-of-sample predictions.

<table>
<thead>
<tr>
<th>Model</th>
<th>In-sample</th>
<th>Out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>QLIKE</td>
</tr>
<tr>
<td>Normal GARCH</td>
<td>9.0746e-8</td>
<td>-6.4154</td>
</tr>
<tr>
<td>Normal EGARCH</td>
<td>7.6839e-8</td>
<td>-6.5333</td>
</tr>
<tr>
<td>Normal TGARCH</td>
<td>7.9071e-8</td>
<td>-6.5448</td>
</tr>
<tr>
<td>Student-t GARCH</td>
<td>9.1015e-8</td>
<td>-6.4141</td>
</tr>
<tr>
<td>Student-t EGARCH</td>
<td>7.7125e-8</td>
<td>-6.5335</td>
</tr>
<tr>
<td>Student-t TGARCH</td>
<td>7.9314e-8</td>
<td>-6.5450</td>
</tr>
<tr>
<td>GED GARCH</td>
<td>9.0769e-8</td>
<td>-6.4167</td>
</tr>
<tr>
<td>GED EGARCH</td>
<td>7.6925e-8</td>
<td>-6.5337</td>
</tr>
<tr>
<td>GED TGARCH</td>
<td>7.9138e-8</td>
<td>-6.5451</td>
</tr>
<tr>
<td>NP GARCH</td>
<td>8.3927e-8</td>
<td>-6.4164</td>
</tr>
</tbody>
</table>

From Table 2 we see that the loss values also suggest that the EGARCH and TGARCH model is better than the GARCH model. The nonparametric model on the other hand seems to perform on average good in-sample but worse out-of-sample. The loss values are a bit smaller out-of-sample than in-sample for the MSE but about the same for the QLIKE.

Data snooping may be a problem in the results above since we are only considering one time series and analyzes many volatility models. In Table 3 the results from the Superior predictive ability test is presented. The null hypothesis is “the benchmark is not inferior to any of the alternatives”. The model column indicates the model that are used as benchmark model and if the null hypothesis is rejected it obtains a value 1.

Table 3: The Superior predictive ability test for the in-sample and out-of-sample in the case of MSE and QLIKE.
A 1 stands for the rejection of the null hypothesis and when the null hypothesis cannot be rejected a 0 is used.

<table>
<thead>
<tr>
<th>Model</th>
<th>In-sample</th>
<th>Out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>QLIKE</td>
</tr>
<tr>
<td>Normal GARCH</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Normal EGARCH</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Normal TGARCH</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Student-t GARCH</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Student-t EGARCH</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Student-t TGARCH</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GED GARCH</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GED EGARCH</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GED TGARCH</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NP GARCH</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

From Table 3 we see that the null hypothesis is rejected in-sample for the GARCH model and to some extent the EGARCH model. In out-of-sample on the other hand it seems like all models perform equally well or that the null hypothesis is rejected for the EGARCH, depending on which loss function we considers.
5. Conclusions

Risk is one of the most central concepts in finance since in financial decisions it is of great importance to understand the features of the risk involved. One of the measures that we usually use to understand risk is the volatility. Volatility measures the variation in the price of an asset and is used in many financial models, for example in option pricing, portfolio allocation and risk management. Volatility usually exhibit some stylized facts that has been presented in the literature. The most common ones are that financial time series usually are subject to volatility clustering, persistence, a leverage effect, leptokurtosis and negative skewness of the returns. This paper uses the GARCH, EGARCH and a special case of the TGARCH volatility models under the assumptions of a normal, Student-t and generalized error distribution to investigate how the models behaves on the Swedish stock index, OMXS30. A nonparametric approach is also included to further study the effect of the return distribution in the framework of GARCH models. The models are evaluated both in-sample and out-of-sample with the MSE and QLIKE loss functions. To avoid the risk of data snooping, the Superior predictive ability test is performed.

In Figure 2 we can see that the return distribution of the OMXS30 for the considered sample period is quite nicely shaped; it is symmetric, almost no skewness and has some “fat tails” even though they are not extremely severe. The results from the parametric models also suggest that the normal distribution can handle the leptokurtosis of OMXS30. This is consistent with the theoretical calculation of the kurtosis for the normal GARCH that Bollerslev (1986) derived to be $6\alpha^2(1 - \beta^2 - 2\alpha\beta - \alpha^2)^{-1}$. Inserting the values from Table 1 gives a kurtosis of 5.40 which is close to the observed sample kurtosis of 6.3640 for OMXS30. I suggest the use of the normal distribution in the parametric models since it is able to account for the leptokurtosis in OMXS30. This suggestion is mainly based on the fact that the stationarity conditions for EGARCH and TGARCH are easier to be fulfilled under this distribution and that maximizing the log-likelihood function for the normal distribution is not as demanding. The nonparametric model is more flexible since we are not facing the problem of distributional misspecification errors. But the distribution of the returns in OMXS30 is quite kind in the case of fulfilling the assumption of a parametric model. I therefore suspect that the nonparametric GARCH model would have performed better than the parametric models if the distributional form of the returns would not have been so nicely shaped.

The asymmetrical models indicates that there is a leverage effect present on the OMXS30 which is probably the reason why the EGARCH and TGARCH perform better in-sample than the GARCH model as can be seen in Table 1-2 (Table 3 also suggests this to some extent). When looking at Table 3 for the out-of-sample, the MSE suggests that no model is inferior to any of the others and the QLIKE indicate that EGARCH may be outperformed by the others. Hence, if we are only interested in the out-of-sample predictions, then all models but the EGARCH are suggested. If considering both in-sample estimates and out-of-sample predictions we see that the nonparametric GARCH model along with the TGARCH models are the only ones with zeros for all cases in Table 3. However, the nonparametric GARCH is very computational demanding so I therefore suggest the use of a normal TGARCH for OMXS30. This means that practitioners that need to estimate and predict the volatility of OMXS30 can use a rather easy parametric model and obtain relatively good results.

Further studies would be to consider a stock index with not as nicely shaped return distribution. The length of the sample and the forecasting horizon is also in need of further investigation. It would also be of interest to use the evaluation models developed by Christoffersen (1998) since this sets the framework closer to the application of Value-at-Risk.
References


