NORGES HANDELSHØYSKOLE
Bergen, Spring 2009

Estimation and selection of time-varying volatility models

Master Thesis within the main profile of Finance

By: Øystein Skregelid

Thesis Advisor: Maytinee Wasumadee

Abstract

This paper describes methods that can be applied to select the best conditional volatility model for an individual asset. Three exchange traded funds (ETFs) for the financials, energy and utilities sectors of the Dow Jones Total Market Index are evaluated to illustrate the complexity of model selection. For the univariate series, the symmetric GARCH model and three asymmetric models (EGARCH, GJR-GARCH and TGARCH) with a variety of lag structures are parameterized under the assumption of both normal and t-distributed errors. The ranking of these models are based on how well the parameters of each model fit to the underlying data set (the likelihood), on selection criterions (AIC and BIC) and on their forecasting ability (through statistic and economic loss functions). The results show that different volatility models with different lag structures are selected for each of the three sectors. For the financial sector a t-distributed EGARCH(1,2,1) model gives the most satisfying results. The energy sector is best described by a t-distributed GJR-GARCH(1,1,2) model, while a normal distributed GJR-GARCH(1,1,1) model is recommended for the utilities sector.

In addition to the selection of univariate models, multivariate models are described and tested. The main focus in this part is on the Dynamic Conditional Correlation model that builds on univariate parameterizations of the volatility. A DCC model based on three univariate normal distributed GJR-GARCH(1,1,1) models is compared to the BEKK model and to a multivariate EWMA model. This comparison shows that while the DCC model performs best when it comes to minimizing the risk of a portfolio, the BEKK model is superior when evaluated on the reward-to-variability ratio (Sharpe). This is mainly due to the fact that the DCC model is unable to catch the time-varying correlation between the three chosen assets.

Foreword

Writing this master thesis has been a challenging, but highly interesting process. It has given me valuable insight and knowledge around topics that are not devoted much time in the obligatory courses within the Finance profile at the Norwegian School of Economics and Business Administration (NHH).

Another valuable experience is the knowledge and ability I have obtained by using Matlab. I wish to thank Kevin Sheppard for making his Oxford MFE Toolbox[1] for Matlab available at his homepage. This is a toolbox I would recommend to anyone who wants to study time-varying volatility.

My interest around the topic of conditional volatility modeling is due to one lecture around this theme in the course Applied Finance. Still, I was unaware of the complexity of time-varying volatility, the richness of alternative models to choose from, and the high amount of research being done around it. Due to this, a large part of this thesis is built on theory and literature review. Although the initial working title was "Testing volatility forecasting models in a portfolio optimization framework", it was the process of selecting the "best" univariate volatility models that fascinated me most, something that is reflected in the final result of this thesis.

I am grateful to Professor Richard D.F. Harris for suggesting this fascinating topic to me. I also want to thank my thesis advisor Maytinee Wasumadee for her help and inputs while writing this thesis. Finally I wish to express my gratitude to Hella for her invaluable support and for motivating me throughout the whole writing process.

Øystein Skregelid

Bergen, June 2009

[1] Can be found at http://www.kevinsheppard.com/wiki/MFE_Toolbox (Accessed February 12, 2009)

Contents

List of figures

List of tables

1. Introduction

Traditional technical analysis, which involves predicting the future by looking at historical variables, is based on the assumption of a long-term mean, constant variances, and thus constant covariances among assets. These constants calculated on e.g. five years of historical data, and become the information set used to determine the allocation of asset weights in a portfolio for the next period. The main idea behind such assumptions is that asset prices are mean-reverting. The volatility of financial assets are, however, time-varying, meaning that there will be periods with high volatility and periods of low volatility. Portfolios that need frequent updating regarding the allocation of assets therefore need to reflect the current level of volatility. If the current volatility is low, while it was high three years ago, then the past high volatility should not be given as much weight as the current low. Conditional time-varying volatility models and historical volatility models where the recent past observations are given higher weights than distant past observations, e.g. the EWMA model, will therefore give more realistic variance-covariance-matrices than those obtained from assuming that the volatility is constant.

The main focus in this paper is on time-varying volatility. Three exchange traded funds (ETFs) on the Dow Jones Total Market Index are evaluated at the univariate level. The three assets are arbitrarily chosen as the aim is to highlight the difficulty and sophistication of conditional volatility modeling. For each asset, a variety of models will be estimated under different assumptions of the distribution of the data set. Every financial asset will have their unique features, so it is impossible to say that one model is superior to another, as each model catches different stylized facts associated to financial time series. The goal is therefore to find the one model that fits best to each asset. Univariate processes will be estimated by the symmetric GARCH model and the asymmetric EGARCH, GJR-GARCH and TARCH models. The final selection of what is the best model will not only be based on the in-sample fit, but also on their out-of-sample performances.

Multivariate time-varying correlation models are also considered. The main focus here will be on the Dynamic Conditional Correlation model, which is a multivariate model that builds on the already estimated univariate processes of volatility. This will lead to the construction of a time-varying portfolio. Portfolios based on the DCC model will be compared to the more parsimonious multivariate EWMA portfolios and the more complex BEKK portfolios.

The paper is organized as follows. Chapter 2 gives a definition and introduction to the motivation of using conditional volatilities, together with stylized facts of financial assets that needs to be captured. Historical volatility models are given a more thorough introduction, and models for the estimation of conditional mean will be presented briefly. In Chapter 3 the conditional volatility models that will be estimated in this paper are described. Tests that should be applied in the pre- and post-estimation to control the appropriateness and how well specified the estimations are respectively are also presented. The evaluation of the forecasting ability of each model is also well-described. Finally, the focus of this chapter moves to alternative multivariate correlation models. Chapter 4 describes the data used in this paper, and Chapter 5 gives the results of the parameterized univariate and multivariate models. Conclusions are given in Chapter 6.

2. Background

It is a well known fact that for financial data, the errors made when predicting markets are not of a constant magnitude. The market fluctuations will be large for some periods and smaller for others (Engle et al., 2008). The rate of new information connected to financial assets is time-varying, and thus the variances of returns and the covariances between the assets are time-varying as well. This time-varying behaviour is referred to as heteroskedasticity, meaning that the volatility of an asset or market tends to cluster in periods of high volatility and in periods of low volatility. Time-varying mean, variance and covariances are based on the information currently available and are referred to as conditional mean, variance and covariances. The time-varying conditions that they are based on are the values of variables determining the level of parameters that describes the time-varying process (Bodie et al., 2008). If the mean, variance and covariances are treated as time-invariant, i.e. as constants, they are said to be unconditional estimates. The usual estimate of return variance is then given by the average of squared deviations over the sample period.

Robert F. Engle (1982) was the first to introduce the concept of conditional heteroskedasticity. He proposed a model where the conditional time series is a function of past shocks. The model, called the autoregressive conditional heteroskedasticity (ARCH) model, led to a breakthrough in financial econometrics. The impact this model has had on the research around time-varying volatility gave him the Nobel Prize in Economic Sciences in 2003. Although the initial ARCH model was designed to capture persistence in inflation, the model fits to a number of other financial time series. The model has had an enormous influence on theoretical and applied econometrics and was influential in the establishment of Financial Econometrics as a discipline (Franses and McAleer, 2002). This discipline can be defined as the application of econometric tools to financial data (Engle, 2001a). The introduction of ARCH class models extended traditional time-series tools such as autoregressive moving average (ARMA) models concerning the mean to equivalently essential models for the variance (Bauwens et al., 2003).

Engle (2001a) points to the fact that the least-squares method for many years served as a satisfactory tool in applied econometrics for the implementation of stock market forecasts, tests of efficient markets and tests of portfolio models such as the CAPM. The basic version of the least square model assumes that the expected value of all squared error terms is the same at any given point, or in other words that the variance of the error terms is constant (Engle, 2001b). This assumption is referred to as homoskedasticity. When the variances of error terms are not equal, but are expected to be larger for some periods and smaller for others, the data are heteroskedastic. When heteroskedasticity is present in the time series, the regression coefficients for an ordinary least squares regression are still unbiased, but the standard errors and confidence intervals estimated will be too narrow, giving a false sense of precision (Engle, 2001b).

In the ARCH model, the conditional variance is allowed to change over time as a function of past errors, while the unconditional variance is left constant (Bollerslev, 1986). A generalization of the ARCH model was proposed by Bollerslev (1986). His generalized autoregressive conditional heteroskedasticity (GARCH) model allows for past conditional variances in the current conditional variance equation. This generalization leads to models that are parsimonious and easy to estimate. Even in its simplest form it has proven successful in predicting conditional variance (Engle, 2001b).

The advantage of ARCH and GARCH models, and other models built on these, is that heteroskedasticity is treated as a variance to be modeled. They thus correct for the deficiencies of the least squares model and computes a prediction for the variance of each error term (Engle, 2001b). ARCH and GARCH models have become popular tools for dealing with time series heteroskedastic models. The aim of models like these is to provide a volatility measure that can be used for e.g. derivates pricing, optimal portfolio selection, and risk management (Fleming et al., 2000). These models do not only give an estimate of the conditional variance of the time series, they also enable forecasts of future conditional variance to be computed (Harris and Sollis, 2003). Modeling and estimation of time-varying return variances and covariances can lead to a better understanding of the expected returns as well (Bodie et al., 2008).

The vast research on time-varying volatilities over the last decades has had a primarily focus on univariate volatilities and not on correlations (Cappiello et al., 2006). Kroner and Ng (1998) argue that changes in asset returns due to the occurrence of time-varying conditional volatility should imply that this asset also has a time-varying correlation with other assets displaying time-varying volatility. Time-varying covariances are often estimated using the constant conditional correlation (CCC) model of Bollerslev (1990) to simplify the computational task of estimation. There is, however, no theoretical justification for this assumption (Cappiello et al., 2006). Other multivariate correlation models, like the BEKK model, require the estimation of a large number of parameters, and thus needs a lot of processing time for systems containing many assets. The Dynamic Conditional Correlation (DCC) model of Engle (2002) is a generalization of the CCC model where the correlation between the assets can be time-varying. The main advantage of the CCC and DCC models over the BEKK model is that they are based on univariate GARCH processes. This enables the conditional correlations to be calculated between assets based on the standardized residuals of the estimated univariate volatility models.

2.1 Defining volatility

Unlike prices, volatility and correlations cannot be directly observed in the market, so models are needed to generate estimates for them (Alexander, 2001). Alexander divides the procedure of volatility forecasting between implied volatility and statistical forecasting. The former gives the volatility forecast over the life of an option, a topic that won't be described in this paper. The latter usually refers to time-series models, such as a moving average model or a GARCH model. When applied to historical data, such models will give a statistical estimate of the volatility in the past, in addition to the generation of forecasts until some future point in time. A stochastic process governing price movements can be referred to as a volatility process. The realizations of this process are called the realized volatility, measured using historical price data. If the price process turns out to have a constant volatility, the realized volatility will simply be the sample standard deviation of observed returns (Alexander, 2001). Realized volatility is the ex-post estimate of the process volatility. It is difficult to forecast the realized volatility ex-ante, because it is likely to be affected by market movements during the forecasting horizon (Alexander, 2001).

Volatility is a measure of the dispersion in a probability density. The variance is a measure of the dispersion of the density function around its mean. The standard deviation, σ, which is the square root of the variance, is the most common measure of dispersion for a random variable (Alexander, 2001), as it is measured in the same units as the original data (Sheppard, 2009a). The sample standard deviation σ is given by:

$$\hat{\sigma} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(r_t - \mu)^2} \qquad (2.1)$$

where r_t is the return on day t and μ is the average return from the T-day period.

The standard deviation is more stable and more desirable for computational estimation and volatility forecast evaluation than variance (Poon and Granger, 2003). In this paper volatility therefore refers to the standard deviation. Equation 2.1 gives the constant volatility, also known as the unconditional volatility, of the return process. It can only be defined when asset returns are assumed to be generated by a stationary stochastic process where the variance is finite (Alexander, 2001).

Since volatility for financial assets is time-varying, a more suitable information at time t is the conditional volatility $\sigma_{t,\tau}$. The conditional volatility is the expected volatility at a future point in time, $t+\tau$, based on all information available up to time t (Sheppard, 2009b). Daily returns used to proxy daily volatility will give a very noisy estimate of the volatility. The ARCH, and subsequent conditional volatility models, is a less noisy approach for this task (Poon, 2005). In such models the conditional mean is often assumed to be constant, even though it is actually time-varying, when the purpose is to estimate and forecast conditional volatility (Alexander, 2001).

Even though volatility is related to risk, it is not strictly the same. The volatility is a measure of uncertainty involving both positive and negative outcomes of a return, while the risk is associated with undesirable outcomes only (Poon, 2005).

As it is not possible to compare a *n*-day variance with a *m*-day variance on the same scale, the volatility is often expressed as the standard deviation in annual terms, so that the annual volatility is defined as $100\sigma\sqrt{A}$, where *A* is the annualizing factor representing the number of returns per year (Alexander, 2001). An *A* equal to 252 days is normal, and will be used in this paper.

In most of the ARCH class volatility models the measure of volatility is based on squared returns. Poon (2005) investigates 93 studies related to volatility models. She reports several studies suggesting that volatility should be measured from absolute returns. Davidian and Carroll (1987) show that absolute returns are more robust against asymmetry and non-normality. McKenzie (1999) among others, proves that absolute return based models produce better forecasts than models based on square returns.

That volatility change over time has a number of explanations, but individually they are not completely satisfactory. Phenomenon's like illiquidity and news announcements are examples of such explanations (Sheppard, 2009a). Illiquidity refers to situations where shocks have a large impact on prices due to few participants being willing to trade an asset. This normally only lasts up until a couple of days, so it cannot explain the long cycles in present volatility (Sheppard, 2009a). News announcements make investors update their beliefs, leading to portfolio rebalancing and thus higher volatility. But the periods of higher volatility are generally short also for this phenomenon (Sheppard, 2009a).

2.2 Stationarity

Economic time-series are considered as realizations of stochastic processes, meaning that each observation is a random variable (Engle et al., 2008). The simplest stochastic process would be one where the random variables are independent and identically distributed (i.i.d.) for some distribution, for example a normal distribution (Sheppard, 2009a). A sequence of variables in a stochastic process is characterized by joint-probability distributions for every finite step at different time periods (Engle et al., 2008). A stochastic process that has a finite mean and variance is covariance stationary, or weakly stationary, if for all *t* and *t-s* the mean μ, variance σ_y^2 and autocorrelation γ_s are constant through time (Enders, 2004):

$$
\begin{aligned}
&E(y_t) = E(y_{t-s}) = \mu \\
&var(y_t) = var(y_{t-s}) = \sigma_y^2 \\
&cov(y_t, y_{t-s}) = cov(y_{t-j}, y_{t-j-s}) = \gamma_s \qquad (2.2)
\end{aligned}
$$

For a covariance stationary time series, the autocorrelation between y_t and y_{t-s} can be defined as $\rho_s \equiv \gamma_s/\gamma_0$ (Enders, 2004). The autocorrelation is time-independent since the autocorrelation between y_t and y_{t-1} must be identical to that between y_{t-s} and y_{t-s-1}. Covariance stationarity applies only to the unconditional moments, so it might still have a varying conditional mean (Sheppard, 2009a).

If none of the process' finite distributions depends on time so that the only factor to have an influence between two observations is the gap between them, the process is said to be strictly stationary (Sheppard, 2009a). A strictly stationary series is weaker than i.i.d. as the process might be dependent (Sheppard, 2009a).

The properties of one part of a stationary series are in other words similar to the properties of another part of the series. The stationary property is only defined for a model, so a real stationary time-series is not likely to exist. The time-series can, however, exhibit the characteristics of a stationary process (Chatfield, 2003).

2.3 White noise

White noise is a basic building block of discrete stochastic time series (Enders, 2004). Imagine a time-series $x_t = \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$, where ε_t is the uncontrollable portion of the series. When the sequence ε_t has a mean equal to zero, a constant variance and uncorrelated realizations, the sequence is said to be white noise (Enders, 2004). A white noise process is therefore also covariance stationary as all three conditions are met. If one or more of these conditions are not met then x_t is not a white noise process.

2.4 Stylized facts

When modeling volatility, time series properties and stylized facts have to be exploited. Each individual financial time series will have their unique features. This makes volatility modeling difficult and sophisticated (Poon, 2005). To be able to choose the right volatility forecasting model, some insight into the stylized facts associated with financial time series is needed as the models should be able to pick up these stylized facts (González-Rivera et al., 2004). Stylized facts include fat tails, volatility clustering and leverage effects. Engle's ARCH model was made to catch clusters and fat-tail behavior of the data. Subsequent models account for more complex issues, for example asymmetric responses to volatility news and the persistence of the volatility process (González-Rivera et al., 2004). In the end there is no straight-forward answer to the question which volatility model to use; it all comes down to the objectives of the study (González-Rivera et al., 2004). The most important stylized facts will be presented in the subsequent chapters. Bollerslev et al. (1993) give a thorough examination of stylized facts associated with financial time series.

2.4.1 Thick tails

Mandelbrot (1963) and Fama (1965) both document the fact that asset returns tend to be leptokurtic, i.e. the time series of returns exhibit fatter tails than a normal (Gaussian) distribution. A normal distribution has a skewness equal to zero and a kurtosis equal to three. Mandelbrot (1963, p.394) finds that *"… the empirical distributions of price changes are usually too 'peaked' to be relative to samples from Gaussian populations".* The kurtosis of a time-series measures the tail thickness. Excess kurtosis, that is kurtosis above 3, implies that the distribution has a sharper peak and fatter tails than a normal distribution. A low kurtosis, on the other hand, implies that the distribution has a rounder peak and shorter, thinner tails. A negative skewness, for example, tells us that the distribution will have a longer left tail than a right tail. In other words, a negative skewness indicates extreme losses, while a positive skewness indicates extreme gains.

The kurtosis and skewness are very sensitive to outliers in the time-series. By removing or 'dummying out' extreme outliers, both the kurtosis and the skewness will drop significantly (Poon, 2005). The "black Monday" on the 19th of October 1987 is an example of an occurrence that might be better left out for volatility forecasting purposes. Removing outliers does not remove volatility persistence, but has a great impact on the variance and thus increases the autocorrelation coefficient (Poon, 2005).

2.4.2 Volatility clustering

It is a well known fact that financial market volatility tends to cluster. This means that volatile periods tend to persist for some time before the market returns to normality (Poon, 2005). Mandelbrot (1963, p.418) for example points out that *"… large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes,…".* This effect can visually be seen when plotting a series of returns through time, which will be shown later in the paper (see Figure 5.2). A plot of the returns, together with statistical tests, shows that financial returns are not i.i.d. through time (Bollerslev et al., 1993). The positive and negative disturbances given by the day-to-day changes become a part of the information set used to construct variance forecasts for the coming period. This means that large shocks of either sign can have an influence on the forecasts for several periods to come. When the clustering is significant, the time series is said to display autoregressive conditional heteroskedasticity (Alexander, 2001). The effect becomes more pronounced the higher the frequency of the sample data is. Daily data is often sufficient to see the clustering, but it becomes clearer from intra-day data. The consequence of volatility clustering is that future volatility can be predicted by past and current volatility.

Rob Engle's (1982) ARCH model, which will be described in Chapter 3.2.1, captures this kind of volatility persistence. There is a close relationship between clustering and thick tails. The volatility clustering is a type of heteroskedasticity and accounts for some of the excess kurtosis typically observed in the distribution of a financial time series. Another part of the excess kurtosis can be due to the presence of a non-normal asset distribution, e.g. the Student's T, which happens to have fat tails.

2.4.3 Leverage effects

The leverage effect refers to the tendency of volatility to increase if the previous days returns are negative, i.e. changes in stock prices are negatively correlated with changes in stock volatility (Bollerslev et al., 1993). A fall in stock price causes leverage and financial risk of a firm with outstanding debt and equity to increase. A leverage effect results in volatility asymmetry, a phenomenon most marked during large falls (Poon, 2005). For time series exhibiting leverage effects, asymmetric GARCH models should be applied because the asymmetry cannot be captured by symmetric GARCH models. Asymmetric GARCH models will be presented in Chapter 3.3.

2.5 Conditional mean models

Autoregressive moving average (ARMA) processes are often considered as the core of time-series analysis. ARMA models can be divided into two smaller classes; the autoregressive (AR) processes and the moving average (MA) processes. For full derivation of the formulas for unconditional and conditional means and variances in the following sections, see Enders (2004) or Sheppard (2009a).

A first-order moving average, MA(1), process can be given by:

$$y_t = \varphi_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t \tag{2.3}$$

where φ_0 and θ_1 are parameters, while ε_t is a series of white noise. The current value of y_t thus depends on both a new shock and the previous shock (Sheppard, 2009a). The unconditional mean from this process is simply $E[y_t] = \varphi_0$ while the conditional mean is given as $E_{t-1}[y_t] = \varphi_0 + \theta_1 \varepsilon_{t-1}$, where the difference reflects the persistence of previous shocks in the current period (Sheppard, 2009a). The unconditional variance is $V[y_t] = \sigma^2(1 + \theta_1^2)$ while the conditional variance is $V_{t-1}[y_t] = \sigma_t^2$ so that the unconditional variance is larger than the conditional variance, reflecting the extra variability given by the moving average term (Sheppard, 2009a). The autocovariance for an MA(1) process will be $\theta_1 \sigma^2$ between y_t and y_{t-s} when $s=1$ and zero when $s>1$. Adding additional lagged errors gives an MA(Q) process.

A first order autoregressive process, AR(1), can be given as:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t \tag{2.4}$$

where the unconditional mean is $E[y_t] = \varphi_0/(1-\varphi_1)$ and the conditional mean is $E_{t-1}[y_t] = \varphi_0 + \varphi_1 y_{t-1}$. The unconditional variance is $V[y_t] = \sigma^2/(1-\varphi_1^2)$ while the conditional variance is still $V_{t-1}[y_t] = \sigma_t^2$. The unconditional variance is still larger than the conditional, and will explode as $|\varphi_t|$ approaches 1 or -1 (Sheppard, 2009a). The autocovariance of an AR process is given as:

$$E[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])] = \varphi_1^s \frac{\sigma^2}{1-\varphi_1^2} \tag{2.5}$$

Additional lags of y_t gives an AR(P) process.

An ARMA(P,Q) model can be given as:

$$y_t = \varphi_0 + \sum_{p=1}^{P} \varphi_p y_{t-p} + \sum_{q=1}^{Q} \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2.6}$$

where the unconditional mean of an ARMA(1,1) is $E[y_t] = \varphi_0/(1-\varphi_1)$. This is the same as for the AR(1) process since the moving average terms are mean zero. The conditional mean is $E_{t-1}[y_t] = \varphi_0 + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1}$. Unconditional variance is $V[y_t] == \sigma^2(\frac{1+2\varphi_1\theta_1+\theta_1^2}{1-\varphi_1^2})$ while conditional variance is still $V_{t-1}[y_t] = \sigma_t^2$.

2.6 Historical volatility models

Historical volatility models (HIS) are easy to manipulate and construct and have showed good forecasting performance compared to other time series volatility models such as ARCH and stochastic volatility (SV) (Poon, 2005). In the historical volatility models the conditional volatility is modeled separately from the returns, making them less restrictive and more able to respond to changes in the volatility dynamic. Poon's (2005) extensive research on the large amount of papers studying forecasting performance of various volatility models lists seven papers concluding that historical volatility models give better forecasts than ARCH and/or SV models. Among these are Andersen et al. (2001) using a realized volatility variant of the historical volatility models. Realized volatility will be briefly discussed in Chapter 3.5.

In the ARCH family of models the conditional volatility σ^2 is modeled as a byproduct of a return equation of the form $r_t = \mu + \varepsilon_t$ through maximizing the likelihood of observing ε_t using a normal, or another chosen, density. These models are explained in depth in Chapter 3. The HIS models, on the other hand, are built directly on conditional volatility, for example an AR(1) model given as $\sigma_t = \gamma + \beta_1 \sigma_{t-1} + \upsilon_t$ (Poon, 2005). The estimation of the parameters γ and β_1 is done through minimizing the in-sample forecast errors υ_t and the forecaster can choose between reducing the mean square errors, the mean absolute errors, etc. (see Chapter 3.4).

Poon (2005) divides the HIS models in two: the single-state and the regime-switching models. The single-state models include e.g. the random walk, the historical average method and the exponentially weighted moving average (EWMA) method. These are given a short introduction below. Examples of regime switching models referred to by Poon (2005) include the threshold autoregressive model of Cao and Tsay (1992) and the smooth transition exponential smoothing model of Taylor (2004). These models are not described in this paper.

The random walk model is the simplest of the HIS models. In the random walk model the difference between consecutive period volatility is modeled as a random noise where the best forecast for tomorrow's volatility, $\hat{\sigma}_{t+1}$, is today's volatility σ_t.

The historical average method makes a forecast based on the entire history of the time series. It assumes that the distribution of volatility has a stationary mean so that all variation in estimated volatility can be attributed to measurement error. The historical average is therefore computed as the unweighted average of volatility observed in-sample:

$$h_{t+1} = \bar{\sigma}^2 = \frac{1}{T}(\sigma_t + \sigma_{t-1} + \cdots + \sigma_1) \quad (2.7)$$

The forecasts based on the mean can provide a benchmark for comparative evaluation of the alternative forecasting models (McMillan et al., 2000).

A simple moving average method is similar to the historical mean method, but it discards older information. The lag length, τ, to past information can be subjectively chosen or based on minimizing the in-sample forecast error $\varsigma_{t+1} = \sigma_{t+1} - \hat{\sigma}_{t+1}$ (Poon, 2005). The moving average method is given as:

$$\hat{\sigma}_{t+1} = \frac{1}{\tau}(\sigma_t + \sigma_{t-1} + \cdots + \sigma_{t-\tau-1}) \qquad (2.8)$$

The multi-period forecasts $\hat{\sigma}_{t+\tau}$ for τ > 1 will be the same as the one-step ahead forecast.

The exponentially weighted moving average (EWMA) model gives more weight to the recent past and less to the distant past observations by letting their importance decline smoothly. The decay factor λ, for a value between 0 and 1, determines how rapidly the weights on past observations decline. The variance of the EWMA model is given as a weighted average of yesterday's variance and yesterday's squared return:

$$\hat{\sigma}_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)r_t^2 \qquad (2.9)$$

λ can be estimated by minimizing the in-sample forecast errors ξ_t (Poon, 2005). A common value for λ is 0.94, as used by RiskMetrics™ (J.P.Morgan, 1997), because it has been found as the average value that minimizes the one-step-ahead error variance for a number of financial assets.

2.7 Time-varying conditional volatility models

Analyses of time series are often treated in terms of the long-run moments of the series. That is the mean, variance and covariance as time approaches infinity. The ARCH model, developed by Engle (1982), allows for time-varying conditional variance, while the unconditional variance is constant. In other words, it is a model with conditional heteroskedasticity, but unconditional homoskedasticity (Harris and Sollis, 2003). When the mean, variance or covariance of a time series are time-varying, the series is non-stationary, so one might assume that a series with conditional heteroskedasticity is non-stationary. When defining 'non-stationarity', however, it is referred to the long-run or unconditional moments of the series. Therefore a time series with conditional heteroskedasticity can be stationary as long as the unconditional moments are constant (Harris and Sollis, 2003).

In the ARCH model, 'autoregressive' refers to the fact that high or low volatility tends to persist, 'conditional' means time-varying or with respect to a certain point in time and 'heteroskedasticity' is a technical expression for non-constant volatility (Poon, 2005).

The conditional and unconditional properties of a time series can be distinguished through its probability distribution. When using a maximum likelihood to estimate econometric models, it is typically assumed that the series has a conditional normal distribution (Harris and Sollis, 2003). But it also has an unconditional probability distribution which may not take the same form as the conditional distribution. When a normal distribution is assumed for an ARCH model, the unconditional distribution turns out to be non-normal. It will more specifically be leptokurtic, meaning it has fat tails (Harris and Sollis, 2003).

Time-varying volatility was not an unknown property before Engle's introduction of the ARCH model. Earlier informal procedures like recursive estimates of variance over time or moving variance were typically used (Bera and Higgins, 1993). The ARCH model, however, was the first formal model designed to capture volatility persistence. Since Engle's initial model there has been a large number of generalizations of the model, each capturing various stylized facts, to accommodate for real world features (Bera and Higgins, 1993). The initial ARCH model alone cannot capture stylized facts like the leverage effect, excess kurtosis and the high degree of nonlinearity (Bera and Higgins, 1993).

After Engle's introduction of the ARCH model, the focus shifted over from returns themselves to return volatility. Daily and monthly returns are approximately unpredictable, but it is widely agreed that return volatility is highly predictable (Andersen et al., 2001).

There are now very many models building on the ARCH model and only a handful of them will be presented in this paper. Bollerslev et.al (1993), among others, presents a large number of models in the ARCH family. They regard the richness of the ARCH family as both a blessing and a curse since the flexibility of ARCH class models enables us to formulate the appropriate model for a given analysis, but complicates the search for a "true" model.

3. Methods

In this chapter the various ARCH class models that will be used in this paper are presented. Firstly, univariate conditional volatility models are presented. These can be divided into two main groups; symmetric and asymmetric models. The main difference between these two classes is that symmetric models, including ARCH and GARCH do not capture leverage effects in the time-series, as opposed to the asymmetric models. Section 3.6 describes tests that can be applied to the time series to ensure that conditional volatility modeling is appropriate, and to check that an estimated model is well specified. Finally in section 3.8 multivariate models used to construct portfolios of assets will be presented.

3.1 Maximum Likelihood and parameter estimation

Estimation of ARCH class models mostly involves maximizing a likelihood function since it is known to produce consistent, asymptotically normal and efficient estimates (Alexander, 2001). Under an assumption about the shape and distribution of the data generation process, the set of parameters, ϑ, should maximize the likelihood of the data. This involves optimization of a function of several variables. Algorithms to solve this problem are often iterative, involving the parameter estimates to update using a scheme. For a normal distributed GARCH(1,1) model (described in Chapter 3.2.2) where the variance is time-varying given by the parameters *ϑ=(ω,α,β),* the log likelihood function of each observation is:

$$l_t = -\frac{1}{2}\left[\ln(2\pi) + ln\ \sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2}\right] \tag{3.1}$$

This procedure can for example be solved for an initial conditional variance on day 1, σ_1^2, equal to the unconditional variance of the whole sample, or even zero. The conditional variances for the remainder of the sample are then updated using an updating rule. The maximum log likelihood is the sum of the individual log likelihoods, l_t, so by adding the necessary constraints to obtain positive conditional variances, the maximum log likelihood is found by iteratively changing the values for the three parameters. The goal is to obtain the largest maximum likelihood possible, so the procedure should be repeated with different starting values for the parameters to ensure that the global optimum of the likelihood function is found (Alexander, 2001).

The easiest way to estimate the parameters so that the maximum likelihood can be obtained is to use statistical software such as Matlab, EViews, PcGive or Gauss, just to mention a few, where the most common GARCH models are incorporated. Brooks et al. (2001) give a review to determine the accuracy of coefficient and standard error estimates for a number of softwares. It is a good starting point when selecting the right software, but it might be a bit out of date. For this thesis Matlab is used, but instead of the built in Econometric Toolbox, the Oxford MFE Toolbox by Kevin Sheppard[2] is used. In the MFE Toolbox the starting values are computed using a grid of experience driven reasonable values. If the optimizer based on these starting values fails to converge, then other starting values are tried (Sheppard, 2009b). Therefore, in this work, the parameter estimates given by the MFE estimation are accepted without specifying the starting values manually.

3.2 Symmetric GARCH models

The most common univariate symmetric model for conditional volatility is Bollerslev's (1986) GARCH model. The GARCH model is presented in Chapter 3.2.2. The GARCH model is a generalization of Engle's (1982) ARCH model. Even though the ARCH model is not much used in this thesis, it is an important model for the understanding of conditional volatility models, and it will therefore be given a thorough introduction in the following section.

3.2.1 ARCH

Before Engle's introduction of the ARCH model in 1982, there was much effort on forecasting future returns, but virtually no methods were available to forecast future variance. The most popular tool until then was the rolling standard deviation, calculated using a fixed number of the most recent *n* observations (Engle, 2001b). This model assumes that the variance of tomorrow's return is an equally weighted average of the squared residuals from the last *n* days. Since the model gives zero weight to observations more than *n* days old, and because more recent events will probably be of higher relevance than the first day of the estimation window, it can be argued that a specification like this is not sufficient. In the ARCH model these weights are parameters to be estimated. This is done by

[2]The toolbox can be obtained from http://www.kevinsheppard.com/wiki/MFE_Toolbox (Accessed February 12, 2009).

allowing the data to determine the best weights to be used to forecast the variance (Engle, 2001b). The ARCH process explicitly recognizes the difference between the unconditional and conditional variance, and allows the latter to change over time as a function of past errors (Bollerslev, 1986).

Under the assumptions of normality, the ARCH process, based on the current information set ψ, and it's likelihood, *l*, are given by:

$$y_t|\psi_{t-1} \sim N(0, h_t),$$
$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 \qquad (3.2)$$
$$\varepsilon_t = y_t - x_t\beta,$$
$$l = \frac{1}{T}\sum_{t=1}^{T} l_t, \qquad (3.3)$$
$$l_t = -\frac{1}{2}log h_t - \frac{1}{2}\varepsilon_t^2 / h_t$$

In this specification, *p* is the order of the ARCH process and h_t is the variance function. $x_t\beta$ is the mean of the return series y_t and may include lagged dependent and exogenous variables. The use of square residuals ε_t^2 and the constraints $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1,\ldots,p$ ensures that the conditional variance is positive. Since the residuals are squared, leverage effects cannot be captured by the ARCH model.

The estimation of the unknown parameters α and β can be done by maximizing the likelihood function (Engle, 1982). The estimation of α and β can be considered separately without loss of asymptotic efficiency, and both can be estimated with full efficiency based only on a consistent estimate of the other (Engle, 1982). In his paper, Engle (1982) recommends to initially estimate β by ordinary least squares, and obtain the residuals. The residuals are then used to estimate α, and based on these $\hat{\alpha}$ estimates, efficient estimates of β can be found.

The simplest form of the ARCH(q) model is the first-order linear model ARCH(1) given under the assumptions of normality as:

$$y_t|\psi_{t-1} \sim N(0, h_t),$$
$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 \quad (3.4)$$

For a one-lagged ARCH model, a large observation of *y* will lead to a large variance for the disturbance in the next period. If $\alpha_1 = 0$, *y* will be Gaussian white noise.

The conditional variance function, h_t, is formulated to resemble the phenomena of clustering of large shocks to the dependent variable. Large shocks in the regression model will be represented by a large deviation of y_t from its conditional mean. In the ARCH model the variance of the current error ε_t conditional on realized values of the lagged errors is an increasing function of the magnitude of the lagged errors. The sign of the error terms does not matter, so large errors of either sign tend to follow large errors of either sign, while small errors tend to follow small errors of either sign. The lag-order *p* determines how long a shock persists in conditioning the variance of subsequent errors (Bera and Higgins, 1993) and is typically of high order due to the phenomenon of volatility persistence (Poon, 2005).

The one-step ahead forecast of an ARCH model is h_t since we know h_{t-1}, while the multi-step ahead forecasts can be formulated through the assumption that $[\varepsilon_{t+\tau}^2] = h_{t+\tau}$ (Poon, 2005).

The unconditional variance of the time-series y_t is given by $\sigma^2 = \frac{\omega}{1-\sum_{j=1}^{q}\alpha_j}$.

The ARCH process is covariance stationary only if the sum of the autoregressive parameters is less than one (Poon, 2005), i.e. $\sum_{j=1}^{q}\alpha_j < 1$.

3.2.2 GARCH

Bollerslev’s (1986) GARCH model is, as Engle’s ARCH, a weighted average of past squared residuals, but with declining weights that never go completely to zero. While the ARCH model has a rather random, but often long, linear declining lag structure for the conditional variance equation, the GARCH model allows for a much more flexible lag structure (Bollerslev, 1986). The GARCH model includes lags of the conditional variance (h_{t-1}, h_{t-2},… h_{t-p})

as regressors in the model for the conditional variance in addition to lags of the squared error term ($\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots, \varepsilon_{t-q}^2$) (Harris and Sollis, 2003).

A GARCH process with the orders *p* and *q* is denoted as GARCH(p,q) where *p* refers to the number of autoregressive lags (ARCH terms) in the equation and *q* refers to the number of moving average lags (GARCH terms) specified (Engle, 2001b). When *p=0,* the process is a ARCH(q) process, and for *p=q=0,* ε_t is white noise. Usually a GARCH(p,q) model with low values for *p* and *q* provides a better fit than an ARCH(q) model with a high value of *q* (Harris and Sollis, 2003). The generalization of ARCH to GARCH is similar to the generalization of a MA to ARMA process. The intention of the generalization is that GARCH parsimoniously can represent a high-order ARCH process (Bera and Higgins, 1993). The analogy to the ARMA class of models means that the time-series techniques used to identify the ARMA models can be used to identify the orders of *p* and *q* in a GARCH model (Bollerslev et al., 1993).

The most common GARCH model, the GARCH(1,1), states that the best estimate of the variance for next period is given as a weighted average of the long-run average variance, the current predicted variance, and the new information captured by the last squared residual (Engle, 2001b). An updating rule like this is according to Engle (2001b) a simple description of learning behavior that can be thought of as Bayesian updating. Sometimes, models with more than 1 lag are needed to ensure that the variance forecasts are good (Engle, 2001b).

By letting ε_t denote a real-valued discrete-time stochastic process, ψ_t denote the information set of all information through time *t,* and h_t denote the variance of the residuals of a regression $r_t = m_t + \sqrt{h_t}\varepsilon_t$, the GARCH(p,q) process is given by:

$$\varepsilon_t|\psi_t \sim N(0, h_t),$$
$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \tag{3.5}$$

To ensure nonnegativity, the lag orders and parameters are constrained so that $p \geq 0$, $q > 0$, $\omega > 0$, $\alpha_i \geq 0$ for *i= 1,…q* and $\beta_i \geq 0$ for *i= 1,…p*. The parameters *ω*, *α* and *β* are estimated through the same log likelihood function as for the ARCH(p) model (Equation 3.3). Updating then only requires knowing the previous forecast h and the residual.

Nelson and Cao (1992) shows that less severe inequality constraints for higher order GARCH models than those suggested by Bollerslev (1986) are sufficient to keep the conditional variance positive.

The GARCH(p,q) process is covariance stationary only if $\sum_{i=1}^{q}\beta_i+\sum_{j=1}^{p}\alpha_j<1$ (Poon, 2005). The short-run dynamics of the volatility series is determined by the GARCH error coefficient, α, and the GARCH lag coefficient, β. A large β indicates that shocks which have entered the system take long time to die out, making the volatility persistent. A large α implies that the volatility reacts intensely to market movements. For a large α combined with a low β, the volatility tends to be more 'spiky' (Alexander, 2001). According to Alexander (2001), common estimates for the parameters of financial markets based on daily data are usually above 0.8 for β and below 0.2 for α. All three parameters in the GARCH(1,1) model are sensitive to the data used. This means that the choice of historic data will affect the current volatility forecast (Alexander, 2001). The GARCH model captures thick tailed returns and volatility clustering, and can also be modified to capture stylized facts such as non-trading periods and predictable events, but not leverage effects (Bollerslev et al., 1993).

The GARCH(p,q) model gives the forecast for the next period directly. Based on the one-period forecast, a two-period forecast can be made. By repeating this step many times, the long-horizon forecasts can be constructed. For each step, the forecast will be a little closer to the long-run average variance, and ultimately, the distant-horizon forecast is the same for all time periods as long as the covariance stationary requirement is met. The long-run forecast is just the unconditional variance given by:

$$\sigma^2=\frac{\omega}{1-\sum_{i=1}^{q}\beta_i-\sum_{j=1}^{p}\alpha_j} \qquad (3.6)$$

This means that the GARCH models are mean reverting and conditionally heteroskedastic, but they have a constant unconditional variance (Engle, 2001b). The one-step-ahead forecast of the conditional variance at time t is given by $\hat{h}_{t+1}=\omega+\alpha_1\varepsilon_t^2+\beta_1 h_t$ and the multi-step-ahead forecast is:

$$\hat{h}_{t+\tau} = \frac{\omega}{1-(\alpha_1+\beta_1)} + (\alpha_1+\beta_1)^{\tau}[\alpha_1\varepsilon_t^2 + \beta_1 h_t] \quad (3.7)$$

As long as $\alpha_1 + \beta_1 < 1$ the second term of this equation eventually dies out and thus $\hat{h}_{t+\tau}$ converges to the unconditional variance (Poon, 2005).

The GARCH(1,1) can be extended and modified in a variety of ways (Engle, 2001b). According to Bera and Higgins (1993) it has frequently been demonstrated that a GARCH(1,1) process is able to represent the majority of financial time series, and data sets requiring models of higher order than GARCH(1,2) or GARCH(2,1) are rare.

Engle's (1982) ARCH model was initially applied to economic data. According to Alexander (2001), Bollerslev's (1986) GARCH model is more appropriate for financial data. Figure 3.1 illustrates how a GARCH(1,1) process models an infinite ARCH process for IYF, which is one of the assets I will study in Chapter 5. This is done with more sensible constraints on coefficients and with fewer parameters. An ARCH model with a few number of lags, such as the ARCH(5) in the figure, will be too variable because the lag is too short. The more lags applied, the more similar is the ARCH(p) process to a GARCH(1,1) process, which can be seen in Figure 3.1 for ARCH(20). The difference is the amount of noise around the estimates. The problem with increasing the lag in an ARCH model is of course that more parameters must be estimated. This is difficult because more parameter estimates make the likelihood function very flat (Alexander, 2001). A GARCH(1,1) model, on the other hand, requires only the estimation of three parameters.

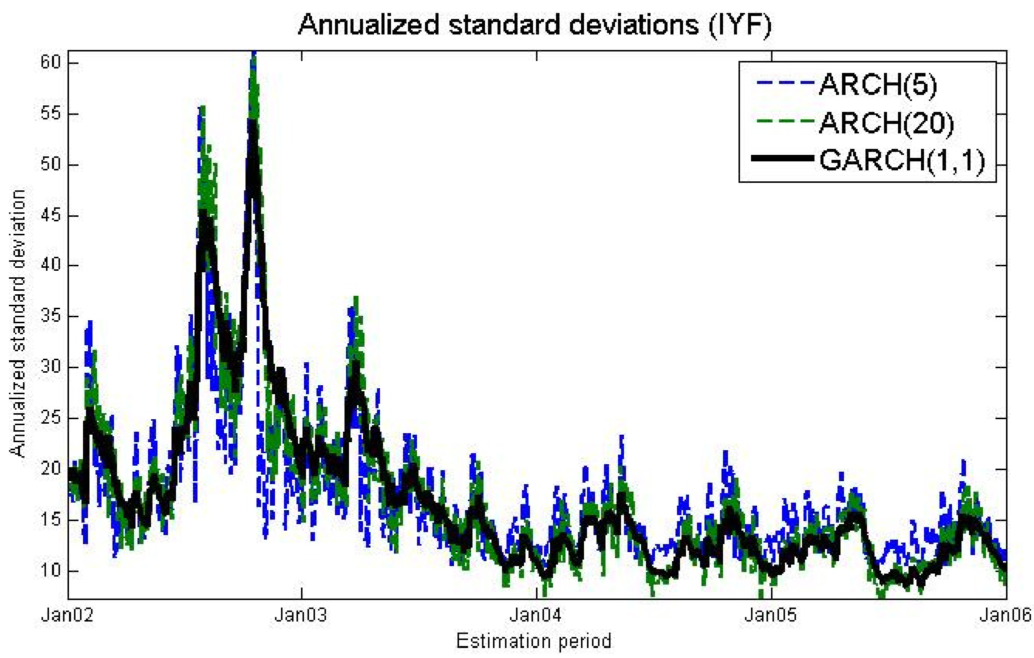


Figure 3.1 Comparing ARCH(p) models with a GARCH(1,1) model for IYF.

3.2.3 T-distributed GARCH

According to Bollerslev (1987), a conclusion drawn from many studies is that speculative price changes and returns are approximately uncorrelated (but not independent) over time, and therefore well described by a distribution with fatter tails than a normal. Bollerslev (1987) therefore suggests an extension to the GARCH model to allow for conditionally t-distributed errors, giving a distinction between conditional heteroskedasticity and a conditional leptokurtic distribution. Both can account for the observed unconditional kurtosis in the data. The conditional variance of a return series estimated under the t-distribution will have an extra parameter v, which is the degrees of freedom parameter. The t-distribution is symmetric around 0. When $1/v > 0$, the t-distribution has fatter tails than the corresponding normal distribution, while when $1/v$ approaches zero, the t-distribution approaches a normal distribution. According to Bollerslev (1987), a fat-tailed conditional distribution might be superior to a conditional normal distribution. A model with t-distributed errors allows for this possibility.

According to Sheppard (2009a), the motivation behind using another distribution than the normal is that a better approximation to the conditional distribution of standardized returns may improve the precision of the estimated volatility process parameters, something that can be important when using GARCH models for purposes like Value-at-Risk and option pricing. Sheppard (2009a) lists Bollerslev's standardized Student's t, Nelson's (1991) generalized error distribution (GED) and Hansen's (1994) skewed t distribution as alternatives to the normal assumption, and concludes that all three may be better approximations as they allow for kurtosis greater than that of a normal. In this paper only normal and t-distributions will be studied.

The log-likelihood of a t-distribution is given as:

$$L_T = \ln\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - 0.5\ln[\pi(\nu-2)] - 0.5\sum_{t=1}^{T}\left[\ln\sigma_t^2 + (1+\nu)\ln\left(1+\frac{z_t^2}{\nu-2}\right)\right] \tag{3.8}$$

where ν is the degrees of freedom in the range $2 < \nu \leq \infty$, Γ(.) is the gamma function and $z_t = \varepsilon_t/\sigma_t$, which is the standardized residual (Peters, 2001).

3.2.4 Integrated GARCH (IGARCH)

When $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j = 1$ the unconditional variance σ^2 is no longer definite for a GARCH(p,q) process, meaning that the series r_t is no longer covariance stationary. It is, however, still strictly stationary (Poon, 2005). For a case like this, the conditional variance is an integrated GARCH process. The EWMA model is a non-stationary version of GARCH(1,1) where the persistence parameters sum up to one: $\alpha_1 + \beta_1 = 1$ (Poon, 2005). Therefore the EWMA specification can be interpreted as an integrated GARCH model where the parameters are $\omega=0$, $\alpha=\lambda$ and $\beta=1-\lambda$ (González-Rivera et al., 2004). Alexander (2001) recommends the use of IGARCH or EWMA when the sum of $\alpha+\beta$ is close to or equal to 1. The EWMA model should in this case have a decay factor λ equal to the estimated GARCH β.

Although the EWMA model has an infinite variance, it is regarded to be a powerful model for volatility forecasting as it is not constrained by a mean level of volatility but adjusts to changes in unconditional volatility (Poon, 2005). The GARCH volatility forecasts of most financial markets tend to mean-revert, meaning that there is a convergence towards a long-term average volatility level. But for some markets, for example the currency market, the volatilities might not mean-revert at all (Alexander, 2001). Generally, for many applications using high-frequency financial data, $\alpha+\beta$ turns out to be close to unity (Bollerslev et al., 1993).

The IGARCH is similar to a symmetric GARCH(p,q) model. For $\alpha+\beta=1$ (or very close to 1), IGARCH is given by (Alexander, 2001):

$$\sigma_t^2 = \omega + (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2 \tag{3.9}$$

where $0 \leq \lambda \leq 1$. The unconditional variance is no longer defined and term structure forecasts do not converge. The variance is now non-stationary. When $\omega=0$, the model is simply an EWMA model (Alexander, 2001).

In the IGARCH(p,q) model, introduced by Bollerslev and Engle (1993), a shock to the conditional variance is persistent, meaning that it remains important for future forecasts of all horizons (Bollerslev et al., 1993).

3.3 Assymetric GARCH models

The ARCH and GARCH models described so far ignore information on the direction of returns since they specify a symmetric response to market news. A symmetric volatility refers to the fact that the unexpected returns are squared so that only the magnitude matters. There is, however, convincing evidence that direction affects volatility (Engle, 2001b). Symmetric volatility models like the GARCH model often give a too low estimate of the conditional volatility after a price drop, while it's too large after a price increase. This can lead to asset mispricing and poor in- and out-of-sample forecasts (Cappiello et al., 2006). The leverage effect has become quite noticeable during the last years (Alexander, 2001). Many asymmetric GARCH models are available. These include the EGARCH model of Nelson (1991), the GJR-GARCH of Glosten, Jagannathan and Runkle (1993) and the TARCH (threshold ARCH) model of Zakoian (1994). The problem with applying symmetric GARCH models to time-series showing leverage effect is that the conditional volatilities are likely to be very spiky, meaning they will show a large reaction (large α) and low persistence (low β) (Alexander, 2001). The parameter estimates from a symmetric GARCH will in other words be quite different from time-series not exhibiting leverage effects.

3.3.1 Exponential GARCH (EGARCH)

The EGARCH(p,o,q) model, where o refers to the number of asymmetric lags, proposed by Nelson (1991), models the natural logarithm of the variance rather than modeling the variance directly. It is then no need to impose estimation constraints to obtain nonnegative conditional variance (Poon, 2005). The EGARCH process is defined as (Sheppard, 2009a):

$$r_t = \mu_t + \epsilon_t$$
$$\mu_t = \varphi_0 + \varphi_1 r_{t-1} + \cdots + \varphi_s r_{t-s}$$
$$\ln(\sigma_t^2) = \omega + \sum_{p=1}^{P} \alpha_p \left(\left| \frac{\epsilon_{t-p}}{\sigma_{t-p}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{o=1}^{O} \gamma_0 \frac{\epsilon_{t-o}}{\sigma_{t-o}} + \sum_{q=1}^{Q} \beta_q \ln\left(\sigma_{t-q}^2\right) \quad (3.10)$$
$$\epsilon_t = \sigma_t e_t$$
$$e_t \sim i.i.d. N(0,1)$$

For the EGARCH model, σ_t^2 depends on both the sign and the magnitude of ϵ_t. It thus captures the leverage effect where negative shocks lead to higher conditional variance in subsequent periods than positive shocks (Poon, 2005). For an EGARCH(1,1,1) model the log variance will thus be a constant, ω, plus three terms. The α-term is the absolute value of a normal random variable minus its expectation. It is thus a mean zero shock. The γ-term is also a mean zero shock. The difference between these shocks is that the former produce a symmetric rise in the log variance while the latter creates an asymmetric effect (Sheppard, 2009a). A negative and significant estimated γ indicates the presence of asymmetry effects (Harris and Sollis, 2003) so that volatility rises more subsequent to negative shocks than to positive shocks. The β-term is the lagged log variance. The EGARCH process is covariance stationary only if $\sum_{j=1}^{q} \beta_j < 1$.

EGARCH models are difficult to use for volatility forecasting because no analytic form is defined for the volatility term structure (Alexander, 2001). The multi-step-ahead forecast for the EGARCH model is given by (Poon, 2005):

$$\hat{h}_{t+\tau} = h_t^{2\alpha_1}(\tau - 1)\exp(\omega)\, exp\{[0.5(\theta+\gamma)^2]\Phi(\theta+\gamma) + exp[0.5(\theta-\gamma)^2]\Phi(\theta-\gamma)\} \tag{3.11}$$

where $\omega = (1-\alpha_1)\alpha_0 - \gamma\sqrt{2/\pi}$.

3.3.2 GJR-GARCH

GJR refers to Glosten, Jagannathan and Runkle (1993). The GJR-GARCH(p,o,q) model which also allows for asymmetrical dependencies is given by:

$$h_t = \omega + \sum_{q=1}^{Q}\beta_q\, h_{t-q} + \sum_{p=1}^{P}\alpha_p\varepsilon_{t-p}^2 + \sum_{o=1}^{O}\gamma_o D_{o,t-1}\varepsilon_{t-o}^2 \tag{3.12}$$

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

When $\alpha_0 > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ and $\beta_j \geq 0$, for *i = 1, …, p* and *j = 1, …, q*, the conditional volatility is positive (Poon, 2005). An estimated *γ > 0* is evidence that asymmetry is present, indicating that negative shocks increase the volatility of returns by more than positive shocks of the same magnitude (Harris and Sollis, 2003).

The GJR-GARCH process is only stationary when $\sum_{q=1}^{Q}\beta_q + \sum_{p=1}^{P}\left(\alpha_p + \frac{1}{2}\gamma_p\right) < 1$ (Poon, 2005).

For a GJR-GARCH(1,1) model the one-step-ahead forecast is given by (Poon, 2005):

$$\hat{h}_{t+1} = \omega + \beta_1 h_t + \alpha_1\varepsilon_t^2 + \gamma_1\varepsilon_t^2 D_t \tag{3.13}$$

while the multi-step-ahead forecast for time τ is given by (Poon, 2005):

$$\hat{h}_{t+\tau} = \omega + \left(\frac{1}{2}(\alpha_1+\gamma_1) + \beta_1\right) h_{t+\tau-1} \tag{3.14}$$

Bollerslev et al. (1993, p.11) summarize the GJR-GARCH model as a model that *"allows a quadratic response of volatility to news with different coefficients for good and bad news, but maintains the assertion that the minimum volatility will result when there is no news"*.

3.3.3 Threshold GARCH (TARCH)

The TARCH, also known as ZARCH, model by Zakoian (1994) is similar to the GJR-GARCH. The difference is that it parameterizes the conditional standard deviation as a function of the lagged absolute value of shocks instead of squared residuals (Sheppard, 2009a). TARCH(p,o,q) conditional standard deviation is specified as:

$$\sigma_t = \omega + \sum_{p=1}^{P} \alpha_p \left|\varepsilon_{t-p}\right| + \sum_{o=1}^{O} \gamma_o \, D_{o,t-o} |\varepsilon_{t-o}| + \sum_{q=1}^{Q} \beta_q \sigma_{t-q} \qquad (3.15)$$

The conditional volatility from the TARCH model is positive when the parameters satisfy $\omega > 0$, $\alpha_p \geq 0$, $\alpha_p + \gamma_o \geq 0$ and $\beta_q \geq 0$. The process is covariance stationary for the case p=q when (Poon, 2005):

$$\beta_1^2 + \frac{1}{2}[\alpha_1^2 + (\alpha_1 + \gamma_1)^2] + \frac{2}{\sqrt{2\pi}} \beta_1 (\alpha_1 + \gamma_1) < 1 \qquad (3.16)$$

According to Sheppard (2009a) models of conditional standard deviations often outperform models where the conditional variance is directly parameterized since absolute shocks are less responsive than squared shocks.

The one-step-ahead forecast of the TARCH(1,1,1) is given by:

$$\hat{\sigma}_{t+1} = \omega + \beta_1 \sigma_t + \alpha_1 |\varepsilon_t| + \gamma_1 |\varepsilon_t| D_t \qquad (3.17)$$

3.3.4 Asymmetric Power ARCH (APARCH)

The APARCH model by Ding, Granger and Engle (1993) is a model that nests several other popular univariate parameterizations and therefore allows the data to determine the true form of asymmetry (Harris and Sollis, 2003). It extends TARCH and GJR-GARCH models in the sense that non-linearity in the conditional variance is directly parameterized through a parameter δ. It thus gives a greater flexibility when modeling the memory of volatility, while remaining parsimonious (Sheppard, 2009a). The APARCH(p,q) model is given by:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^\delta \qquad (3.18)$$

To satisfy non-negative conditional variances, it is necessary that $\omega > 0$, $\alpha_k \geq 0$ and $-1 \leq \gamma_i \leq 1$ (Sheppard, 2009a).

The model for example nests a standard linear GARCH(p,q) model when $\delta = 2$ and $\gamma_i = 0$, a GJR-GARCH model when $\delta = 2$ and a TARCH model when $\delta = 1$ (Laurent, 2004). To test if an APARCH specification fits better than for example a GARCH model the null hypothesis that $\delta = 1$ and $\gamma = 0$ must be tested (Laurent, 2004).

3.4 Forecasting performance of the various GARCH models

Poon's (2005) survey of volatility forecasting models reveals that there is no clear consistency between different researcher's findings when it comes to comparing the ability of the numerous models available. She finds, however, that models allowing asymmetry do well due to the strong negative relationship between volatility and shocks. Cao and Tsay (1992) and Lee (1991) are among those who support the EGARCH model for stock indices volatility. Furthermore, Brailsford and Faff (1996) and Taylor (2004) find that GJR-GARCH is superior to GARCH for stock indices.

When it comes to GARCH versus historical models such as the EWMA and rolling window, Poon (2005) refers to Akigray (1989) who finds that GARCH outperforms EWMA and rolling window models in all sub periods and under all evaluation measures.

Poon (2005) also refers to plenty of studies preferring exponential smoothing methods over GARCH for volatility forecasting. This is often the case when the GARCH family of models leads to convergence problems, possibly due to the fact that parameter estimation becomes unstable when the data period is short or when there is a change in the volatility level.

Finally, Poon (2005) shows to a lot of studies where the results are not so clear. She suggests that the reasons for this could be that

- The studies test a large number of very similar models.

- They use a large number of different forecasting error statistics with different loss functions.
- The forecasts and error statistics are calculated for variance and not standard deviations.
- Squared returns are used to proxy 'actual' volatility, making the results extremely noisy.

3.5 Realized volatility

With so many competing parametric volatility models, all with different properties, it is obvious that misspecifications do exist in the volatility forecasting methodology. At least one of the models could be correct, but none is strictly correct (Andersen et al., 2001). A relatively new way to measure volatility, supported by the ever-improving technology to handle large amounts of data, and databases making intraday data available, is termed realized volatility. Andersen et al. (2001) compute daily realized volatility by summing intraday squared returns. The theory behind this is that realized volatility from very high frequency data (10 years of continuously recorded 5-min returns is used in the paper) will be close to the underlying integrated volatility, which is the integral of instantaneous volatility over the interval studied. This is according to the authors a natural volatility measure. So by treating volatility as observed, they are able to examine the properties of the volatility directly, using much simpler techniques than those required from complicated econometric models. Poon (2005) expects to see a big increase in the research on realized volatility variant of historical volatility models in the coming years. In this paper, however, daily returns will be used.

3.6 Testing appropriateness of GARCH class modeling

Before taking the step of applying conditional volatility models to a time-series, it should be checked if such a procedure is appropriate. Tests that should be undertaken in the pre-estimation analysis are presented in Chapter 3.6.1. In a similar fashion, the results from a conditional volatility specification should be checked to test whether the model is well specified. These post-estimation tools are presented in Chapter 3.6.2.

3.6.1 Pre estimation analysis

The main tests before actually estimating the conditional volatility are Engle's ARCH test and Box-Pierce-Ljung's Q-test. The former tests whether the data is heteroskedastic, while the latter tests whether volatility clustering is present. Tests for normality in the raw returns will also be presented, although they will almost certainly reject normality. These normality tests are more useful in the post-estimation phase, where they are applied to standardized residuals of the estimated volatility models. Finally, a test for leverage effects in the returns series is presented. This test will give a useful insight to whether a symmetric or an asymmetric univariate volatility model should be applied.

Engle's ARCH test

Since the ARCH class of models requires iterative procedures, it is desirable to test whether it is appropriate before the estimation. An ideal tool for this is the Lagrange multiplier test (Engle, 1982). The null hypothesis assumes that the model is a standard dynamic regression model written as $y_t = x_t\beta + \varepsilon_t$, where x_t is a set of weakly exogenous and lagged dependent variables and ε_t is a Gaussian white noise process, meaning that the data are homoskedastic and that variance cannot be predicted (Bollerslev et al., 1993). The alternative is that the errors are ARCH(q), or in other words that there are ARCH effects in the sample as large values of ${\varepsilon^2}_t$ will be predicted by large values of the past squared residuals (Bollerslev et al., 1993). The reason for using a test like this is that if data turn out to be homoskedastic, then the variance cannot be predicted and variations in the squared residuals are purely random. If the test shows that ARCH effects are present, then large values of the squared residuals will be predicted by large values of past squared residuals (Bollerslev et al., 1993). This test is often referred to as Engle's ARCH test.

Autocorrelation functions

Autocorrelations and partial autocorrelation functions can be useful tools for identifying and checking GARCH behavior in the conditional variance equation (Bollerslev, 1986). A standard test for autoregressive conditional heteroskedasticity is to study the autocorrelation of squared returns, since returns themselves may not be autocorrelated. This is often the case for financial time-series. Volatility clustering implies that the autocorrelation of squared returns is strong. The significance of the autocorrelation in the squared returns can be tested by the Box-Pierce LM test (Alexander, 2001). It tests for autocorrelation up to order *p*. It is a form of Lagrange multiplier test meaning it is asymptotically distributed as chi-squared with *p* degrees of freedom. The test statistic is:

$$Q = T\sum_{n=1}^{p}\varphi(n)^2 \qquad (3.19)$$

where *T* is the sample size and $\varphi(n)$ is the *n*-th order sample autocorrelation given by:

$$\varphi(n) = \frac{\sum_{t=n+1}^{T} r_t^2 r_{t-n}^2}{\sum_{t=1}^{T} r_t^4} \qquad (3.20)$$

Even though volatility clustering is present, this test might show insignificant autocorrelation in the squared returns. This could be caused by one or more extreme negative returns in the sample. A negative skewness and extreme excess kurtosis (significantly above 3) will tell if this is the fact, and the solution then is to remove or 'dummy out' these outliers (Alexander, 2001).

Testing for normality

The skewness, τ, which is the standardized third moment of the time-series distribution, is given as (Alexander, 2001):

$$\tau = E[(X-\mu)^3]/\sigma^3 \qquad (3.21)$$

The kurtosis, κ, is the standardized fourth moment, given as:

$$\kappa = E[(X-\mu)^4]/\sigma^4 \qquad (3.22)$$

A Jarque-Bera normality test is defined in terms of the sample estimates of τ and κ. For a sample with the size n the test is:

$$JB = n[(\hat{\tau}^2/6) + (\hat{\kappa}^2/24)] \tag{3.23}$$

This test statistic is asymptotically chi-squared with 2 degrees of freedom (Alexander, 2001).

Quantile-Quantile (QQ) plots can be used to observe departures from normality and thus give a nonparametric view to assess whether a distribution is skewed or heavy tailed. A density plot shows the relative frequency distribution of the time-series compared to e.g. the normal density of the same mean and standard deviation. The QQ-plot is a scatter plot of empirical quantiles of a given distribution on the vertical axis against theoretical quantiles on the horizontal axis. Points along a 45° line indicate a good distributional fit. Returns that have excess kurtosis have a greater probability of large negative or large positive values than a corresponding normal density function. The lower quantiles are then less than the normal, and the upper quantiles are greater. Fat tails are displayed in the QQ plots as deviations below the line for the lower quantiles and above the line for upper quantiles (Alexander, 2001). A better understanding of the QQ-plot can be obtained through Figure 5.3, where the raw returns of the IYF are plotted against a normal distribution.

Testing for asymmetry

To investigate the leverage effect of a time series, an asymmetric GARCH test can be applied. This test is calculated by computing the first-order autocorrelation coefficient between lagged returns and current squared returns (Alexander, 2001):

$$\text{AGARCH Autocorrelation} = \frac{\sum_{t=2}^{T} r_t^2 r_{t-1}}{\sqrt{\sum_{t=2}^{T} r_t^4 \sum_{t=2}^{T} r_{t-1}^2}} \tag{3.24}$$

When the asymmetric autocorrelation is negative and the corresponding Box-Pierce test is significantly different from zero, there is asymmetry in volatility clustering. When asymmetry is present the variances after a negative shock will be greater than after a positive shock (Cappiello et al., 2006). This asymmetry cannot be captured by symmetric GARCH models.

3.6.2 Post estimation analysis

The true variance process could be different from the one specified by the conditional volatility models. Many diagnostic tests are available in order to test this. The simplest test is to construct the series of residuals, ε_t. This series is supposed to have constant mean and variance if the model is specified correctly. Tests for autocorrelation in the squares are also able to detect model failures. A Ljung and Box (1978) portmanteau test with 15 lagged autocorrelations for ε^2 is often used (Engle, 2001b). The Ljung-Box Q statistic tests the null hypothesis that the first *s* (for example 15) autocorrelations are all zero. The alternative is that at least one is non-zero (Sheppard, 2009a).

Remaining ARCH effects

If a GARCH model has captured volatility clustering, the residuals standardized by their conditional volatility $\left(\varepsilon_t/\sqrt{h_t}\right)$ should have no significant ARCH effects left. Standardized returns are then nearly normally distributed (Alexander, 2001). To test whether there are remaining ARCH effects, Engle's ARCH test is therefore applied to the standardized residuals.

Remaining autocorrelation

Just as in the pre-estimation analysis, the autocorrelation function is useful in the post-estimation analysis. The standardized returns squared $(r_t^{*2} = r_t^2/\hat{\sigma}_t^2)$, should have no remaining autocorrelation if the GARCH model is well specified (Alexander, 2001).

When applying different ARCH class models on the same time-series, and more than one show no autocorrelation in squared standardized returns, a simple procedure is to choose the model giving the highest maximum likelihood for the sample, implying that this model is more likely under the density generated by the volatility forecasts (Alexander, 2001).

AIC and BIC

The Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (BIC or SBC) are useful for model selection among specifications with different numbers of parameters. Adding lags for *p* and *q* requires estimation of additional coefficients and are associated with a loss of degrees of freedom. The use of a maximum likelihood function in the model parameterization often leads to adding additional parameters to increase the likelihood. This could result in overfitting of the model. The AIC and BIC models trade off a reduction in the sum of squares of the residuals for a more parsimonious model, i.e. it introduces a penalty term for the number of parameters in the model. The penalty for additional parameters is higher for the BIC than for AIC. For *n* parameters estimated and T usable observations (to deal with the fact that observations are lost when increasing the lags), these formulas are given by (Enders, 2004):

$$AIC = T\ln(sum\ of\ squared\ residuals) +\ 2 \tag{3.25}$$

$$BIC = T\ln(sum\ of\ squared\ residuals) + n\ln(T) \tag{3.26}$$

The AIC and BIC should be as low as possible, and the methods will thus select the model with the lowest value for AIC and SBC respectively. *T* must therefore be fixed between each model, since fewer observations reduce the AIC and the BIC. BIC will always select a more parsimonious model than AIC since ln(T) will be greater than 2 (Enders, 2004). If both tests select the same model, the choice of model is easy. When the two tests do not select the same model, certain diagnostics can be done to end up with the right selection. For the model selected by BIC it should be determined whether the residuals appear to be white noise since BIC selects the most parsimonious model, while the significance of all coefficients should be tested for the model selected by AIC (Enders, 2004).

3.7 Evaluation of volatility forecasts

The objective of applied econometrics is often to find the superior forecasting model. Traditionally this is done by direct comparison of the mean squared error (MSE) of the forecasts, while more popular tests in recent literature evaluate the statistical significance of differences in MSE and compare the informational content of forecasts (Harris and Sollis, 2003).

According to Gonzáles-Rivera et al. (2004) the task of comparing the relative performance of different volatility models is built on either a statistical loss function or an economic loss function. Statistical loss functions are based on moments of forecast errors, and include statistics such as the mean error (ME), the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percent error (MAPE):

$$\text{Mean Error (ME)} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t = \frac{1}{N}\sum_{t=1}^{N}(\hat{\sigma}_t - \sigma_t) \qquad (3.27)$$

$$\text{Mean Square Error (MSE)} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2 = \frac{1}{N}\sum_{t=1}^{N}(\hat{\sigma}_t - \sigma_t)^2 \qquad (3.28)$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{\sigma}_t - \sigma_t)^2} \qquad (3.29)$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{N}\sum_{t=1}^{N}|\varepsilon_t| = \frac{1}{N}\sum_{t=1}^{N}|\hat{\sigma}_t - \sigma_t| \qquad (3.30)$$

$$\text{Mean Absolute Percent Error (MAPE)} = \frac{1}{N}\sum_{t=1}^{N}\frac{|\varepsilon_t|}{\sigma_t} = \frac{1}{N}\sum_{t=1}^{N}\frac{|\hat{\sigma}_t - \sigma_t|}{\sigma_t} \qquad (3.31)$$

The best model would be the one that minimizes such a function of the forecast errors. While the evaluation of such models are straight forward for conditional mean models, forecasts of the conditional variance are more complex to evaluate as observed values of the conditional variance are not available even when holding back a part of the sample for this purpose (Harris and Sollis, 2003). Squared returns, r_t^2, are traditionally used as a proxy, enabling forecasts to be evaluated in the same way as the forecasts of the return series. The squared values are, however, often a poor proxy of the conditional variance leading to many studies that apply such a proxy concluding that GARCH models produce poor forecasts of the conditional variance (Harris and Sollis, 2003). According to Andersen and Bollerslev (1998, p.3) this is *"an inevitable consequence of the inherent noise in the return generating process"*. They suggest a method involving high-frequency data for the construction of volatility measurements. According to Sheppard (2009a) squared returns are, however, reasonable for short-horizon problems using daily data when the squared conditional mean is small relative to the variance. For samples containing monthly data, Sheppard (2009a) suggests the use of squared residuals, ε_t^2, produced from a conditional mean model instead.

A generalized Mincer-Zarnowitz (GMZ) R^2 regression can be applied to evaluate the optimality of the forecasts (Patton and Sheppard, 2008, Sheppard, 2009a). The GMZ regression is given by:

$$r_{t+h}^2 - \hat{\sigma}_{t+h|t}^2 = \gamma_0 + \gamma_1 \hat{\sigma}_{t+h|t}^2 + \gamma_2 z_{1t} + \cdots + \gamma_{K+1} z_{Kt} + \eta_t \quad (3.32)$$

where z_{jt} e.g. can be r_t^2, $|r_t|$ or r_t. If the forecast is correct, the regression of the proxy (for realized volatility) on its forecast should give (approximately) γ_0 = γ2 = γ_{K+1} = 0 and γ_1 = 1 (Sheppard, 2009a). This can be tested e.g. by a Wald test under the null that the coefficients are 0.

A common measure of risk is Value-at-Risk (VaR). This is an example of an economic loss function. VaR is the maximum loss (stated at a 1-α confidence level) that can be expected from normal market movements for a specified time period under the assumption that the portfolio is unmanaged during this time (Alexander, 2001). It gives a more sensible measure of risk than variance because the focus is on losses (Sheppard, 2009a). The probability that a portfolio will lose more than its VaR for a particular time horizon is equal to α, which is a prespecified number (Engle and Manganelli, 1999). A 1% VaR, for example, is the number of dollars (or other currency) one can be 99% sure exceeds any losses for the next day. This is a 1% quantile, as 1% of the outcomes are worse and 99% are better (Engle, 2001b). VaR involves forecasting a value at each time period that will be exceeded by a confidence level of 1-α.

Three approaches for the calculation of VaR will be examined in this paper. These are the fully parametric conditional VaR, the semi-parametric conditional VaR and the historical simulation models.

A fully parametric conditional VaR is estimated using maximum likelihood. At time *t* the conditional VaR is given by:

$$VaR_{t+1} = -\hat{\mu} - \hat{\sigma}_{t+1} F_\alpha^{-1} \quad (3.33)$$

where F_α^{-1} is the α-quantile for the distribution of residuals (Sheppard, 2009a). The VaR thus depends on the mean and standard deviation. The mean μ is assumed to be constant, while the standard deviation σ is the conditional volatility from an estimated model, like for example a GARCH(1,1) model. F_{α}^{-1} is an inverse cumulative density function (CDF) with probability α, mean 0 and variance equal to 1. This is approximately equal to 2.33, 1.65 and 1.28 for a normal distribution with 99%, 95% and 90% confidence level respectively. So a one-day ahead 1% VaR using the fully parametric method equals:

$$VaR_{t+1} = -\hat{\mu} - \hat{\sigma}_{t+1} \cdot 2.33 \qquad (3.34)$$

F can also be a Student's t distribution with ν degrees of freedom. One limitation of this model is that knowledge of the density is required; if the distribution is misspecified then the quantile used is wrong (Sheppard, 2009a). An alternative when the distribution is unknown is a semi-parametric estimation. In this case a standard maximum likelihood estimation (MLE) is not available, but the model can be estimated through a quasi-MLE (QMLE) where $\hat{F}_{\alpha}^{-1}$ is now the empirical α-quantile of a series of ordered standardized residuals (Sheppard, 2009a). The use of standardized residuals has an advantage over the use of residuals since the density does not have to be assumed.

A third method is the unconditional, or historical, VaR. This is the VaR computed directly from unmodified returns in order to compute the appropriate quantile (Sheppard, 2009a). At a time t the historical VaR is simply the α% quantile of the empirical distribution of the data set in the range 1 to t-1.

One way to rank the forecasting performance of each model based on VaR is to find the total "hits": $N = \sum_{t=1}^{T} HIT_{t+1}$, where HIT_{t+1} is equal to 1 if the return on day t+1 is less than VaR at day t+1, and 0 otherwise. A Mincer-Zarnowitz regression of $HIT_{t+1|t}$ on HIT_t in the out-of-sample period after demeaning the series (e.g. subtract 0.01 from each value of HIT_t at the 99% confidence level) can be calculated to control the quality of the model (see Sheppard, 2009a). For a correctly specified model, all coefficients should be zero. If the model is rejected, another confidence level can be tested.

3.8 Multivariate models

The motivation behind the use of multivariate modeling is that financial volatilities move together across markets and assets. In multivariate modeling we need to capture movements in the conditional variances.

If the joint distribution of two stationary return processes, r_1 and r_2, has stable properties over time, they are said to be jointly covariance stationary (Alexander, 2001). Stable properties means e.g. that the covariance between the series is a constant at all times, so that at every point in time $cov(r_{1t}, r_{2t}) = \sigma_{12}$. Only then can the constant correlation ρ be defined as $\rho_{12} = \sigma_{12}/\sigma_1\sigma_2$ where σ_i is the standard deviation of series *i*.

In financial markets the correlations between assets change from day to day, so that the unconditional correlation does not exist. A conditional correlation model allows correlations in a conditional joint distribution to change over time (Alexander, 2001). This is done by dividing the conditional covariance by the product of the conditional standard deviations of each return so that $\rho_{12,t} = \sigma_{12,t}/\sigma_{1,t}\sigma_{2,t}$. Multi period forecasts can be defined similarly.

Conditional correlation models are typically very unstable over time, in contrast to unconditional correlation models. The reason is that the past plays a weaker role in the conditional models, while events happening many periods ago still affect the unconditional models. The only variation seen in correlation estimates of an unconditional model is due to sampling errors, and the past can affect the sampling error just as much as the present. For time-varying correlation models the variations in the estimates are also due to changes in the true value of the parameters (Alexander, 2001).

Many models are available for the estimation of conditional correlations. The estimates of the various models can be quite different. In practice, simple methods like a rolling historical correlation or exponential smoothing are often preferred, but more advanced models like multivariate GARCH models can have many advantages. In the following, some different approaches to correlation modeling will be presented.

EWMA correlation

The EWMA correlation estimator uses declining weights given by the decay factor λ so that the latest data is given more weight, while past data never become uninformative. The conditional correlation is defined as:

$$\hat{\rho}_{12,t} = \frac{\sum_{s=1}^{t-1} \lambda^{t-j-1} r_{1,s} r_{2,s}}{\sqrt{\left(\sum_{s=1}^{t-1} \lambda^{t-s-1} r_{1,s}^2\right)\left(\sum_{s=1}^{t-1} \lambda^{t-s-1} r_{2,s}^2\right)}} \tag{3.35}$$

For the multivariate EWMA, λ should be the same for all assets to ensure a positive definite correlation matrix (Engle, 2002). RiskMetrics™ (J.P.Morgan, 1997) uses λ=0.94, and this will also be used in this paper. The value of the decay factor could otherwise be estimated in a number of ways, e.g. to a value that optimizes an economic or statistical criterion (such as the value that minimizes the one-step ahead forecast error variance). The estimator of the EWMA correlation H_t is given as:

$$H_t = \lambda\left(r_{t-j} r_{t-j}'\right) + (1-\lambda) H_{t-1} \tag{3.36}$$

Many practitioners prefer EWMA to GARCH correlations. EWMA correlations are easier to estimate, but also have their limitations including the fact that, in contrast to univariate GARCH, the term structure forecasts does not mean revert (Alexander, 2001).

Vech and BEKK

For a multivariate time series y_t the conditional variances of individual series and the conditional covariances between each series are estimated simultaneously by maximum likelihood (Harris and Sollis, 2003). The conditional mean of y_t is a $n \times 1$ vector denoted μ_t, while the conditional variance is a $n \times n$ matrix denoted H_t. The H_t matrix has the variances on the diagonal and covariances on the off-diagonal elements. Many different multivariate GARCH models are available. Two of the most common are the vech and the BEKK specifications (Harris and Sollis, 2003).

A *n*-dimensional multivariate GARCH model has *n* conditional mean equations. The diagonal vech model has a separate GARCH equation for each asset. It parameterizes the vector of variances and covariances. A *n*-dimensional vech model is defined as (Alexander, 2001):

$$vech(H_t) = A + Bvech(r_{t-1}r'_{t-1}) + Cvech(H_{t-1}) \quad (3.37)$$

where H_t is the conditional covariance matrix at time *t,* so that vech(H_t) is a vector containing all elements of the covariance matrix. For the bivariate case, $r_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, $A = (\omega_1, \omega_2, \omega_3)'$, $B = diag(\alpha_1, \alpha_2, \alpha_3)$ and $C = diag(\beta_1, \beta_2, \beta_3)$.

The BEKK model, named after Baba, Engle, Kraft and Kroner, first introduced by Engle and Kroner (1995), is a general parameterization that involves a minimum number of parameters. This model has no cross equation restrictions and it ensures positive definiteness for any parameter (Alexander, 2001). The BEKK model is specified as:

$$H_t = A'A + B'(r_{t-1}r'_{t-1})B + C'H_{t-1}C \quad (3.38)$$

A, B and C are *n x n* matrices and A is triangular. This method involves eleven parameters for a bivariate GARCH(1,1) model, which is two more than for a bivariate diagonal vech. The more assets involved, the higher becomes the number of parameters, making the estimation procedure difficult. This feature is shared by the diagonal vech procedure. For a ten-dimensional system for example, at least 175 parameters have to be estimated using one of these multivariate GARCH methods. All these parameters have to be estimated simultaneously, which can lead to extreme convergence problems (Alexander, 2001).

For special cases, the B and C matrices can be for example scalar or diagonal, refered to as SCALAR BEKK and DIAGONAL BEKK respectively (Engle, 2002).

The multivariate GARCH models can be estimated by the “variance targeting” constraint that the long run covariance matrix is the sample covariance matrix (Engle, 2002).

For portfolio problems where possibly thousands of assets are considered, a large number of correlations are required. The construction of an optimal portfolio requires forecast of the covariance matrix of returns, and the calculation of the standard deviation of today's portfolio requires a covariance matrix of all portfolio assets (Engle, 2002). Models such as vech and BEKK are difficult to use in those cases; when there are many assets the number of parameters is too big to optimize the portfolio (Engle, 2002). There is therefore a lot of research on new classes of multivariate models to handle such tasks. The CCC and DCC models are approximations to multivariate GARCH covariance matrices H_t, generated by univariate GARCH models. These will be described next. Another alternative is an orthogonal GARCH model. For more on this, see Alexander (2001).

Constant conditional correlation (CCC)

Bollerslev's (1990) constant GARCH correlation is given by $H_t = D_t R D_t$, where $D_t = \sqrt{h_{it}}$ is a *k x k* diagonal matrix of conditional volatilities from univariate GARCH models and R is the constant correlation matrix (Alexander, 2001). It is thus a product of time-varying volatilities and a time-invariable correlation matrix. Using the individual return series for each asset, GARCH volatilities can be estimated using univariate GARCH models. The correlation matrix can then be estimated using equally weighted moving averages over the whole data period (Alexander, 2001). By assuming that the correlation is constant, the model becomes feasible and ensures that the estimator is positive definite under the requirement that each conditional variance is non-zero (Engle and Sheppard, 2001).

Dynamic conditional correlation multivariate GARCH (DCC-MVGARCH)

The DCC model by Engle and Sheppard (2001) is a generalization of Bollerslev's constant conditional correlation model where the correlation matrix can be time-varying (Alexander, 2001). It has the flexibility of univariate GARCH models combined with parsimonious parametric models for the correlations. It parameterizes the conditional correlations directly in two steps. Firstly the volatilities and standardized residuals are calculated for each asset through univariate GARCH models. The standardized residuals of each asset are then used to calculate conditional covariances between them, using a maximum likelihood criterion and one of many models for the correlations. The covariance and correlation matrices are guaranteed to be positive definite (Engle and Colacito, 2004). Two examples of such correlation models are Engle's (2002) DCC model, allowing an asymmetric response in variances, and the asymmetric DCC (A-DCC) model by Cappiello et al. (2006), which also allows for asymmetric dynamics in the correlation. The difference between these two specifications is thus that the latter allows correlation to rise more when both returns are falling than when both are rising (Engle and Colacito, 2004).

The standard errors from the univariate step remain consistent, while the standard errors for the correlation parameters need to be modified (Engle and Sheppard, 2001). This gives a computational advantage over multivariate GARCH models because the number of parameters to be estimated for the correlation step is independent from the number of series to be correlated (Engle, 2002). DCC is thus very convenient for big systems (Engle and Colacito, 2004).

The DCC models have a covariance matrix given as $H_t = D_t R_t D_t$, where D_t is a k x k diagonal matrix of time-varying standard deviations from univariate GARCH models and R_t is the (possibly) time-varying correlation matrix (Cappiello et al., 2006). The parameterizations of R follow the same requirements as a parameterization of H, except that the conditional variances must be unity (Engle, 2002).

An asymmetric DCC for two assets is formulated by Engle and Colacito (2004) as:

$$y_{i,t} = h_{i,t}^{-\frac{1}{2}} \xi_{i,t}$$
$$H_t = \begin{bmatrix} h_{1,t} & \rho_t \sqrt{h_{1,t} h_{2,t}} \\ \rho_t \sqrt{h_{1,t} h_{2,t}} & h_{2,t} \end{bmatrix} \quad (3.39)$$

where $y_{i,t}$ is the standardized residuals of each asset, defined as the residuals ξ divided by the conditional standard deviations of each series. The D_t elements in the covariance matrix ($h_{i,t}$ for $i=1,2$) are given by any univariate GARCH process (i.e. symmetric or asymmetric) with normally distributed errors satisfying stationary conditions and non-negativity constraints (Engle and Sheppard, 2001) so that:

$$h_{i,t} = \omega_i + \sum_{q=1}^{Q} \beta_{iq}\, h_{i,t-q} + \sum_{p=1}^{P} \alpha_{ip} y_{i,t-p}^2 + \sum_{o=1}^{O} \gamma_{io} d_{io,t-1} y_{i,t-o}^2 \quad (3.40)$$

The subscripts on P and Q indicate that the lag lengths for the two assets can be different. The correlation $\rho_t = \frac{h_{12,t}}{\sqrt{h_{1,t}^* h_{2,t}^*}}$ (the R_t elements) is given by:

$$\begin{aligned} h_{i,t}^* &= \left(1 - \theta_1 - \theta_2 - \frac{\theta_3}{2}\right) + \theta_1 \varepsilon_{i,t-1}^2 + \theta_2 h_{i,t-1}^* + \theta_3 d_{i,t-1}^* \varepsilon_{i,t-1}^2 \\ h_{12,t} &= \varphi_{12} \cdot (1 - \theta_1 - \theta_2) - \varphi_3 \theta_3 + \theta_1 \varepsilon_{1,t-1} \varepsilon_{2,t-1} + \theta_2 h_{12,t-1} \\ &\quad + \theta_3 \left(d_{1,t-1}^* \varepsilon_{1,t-1}\right)\left(d_{2,t-1}^* \varepsilon_{2,t-1}\right) \end{aligned} \quad (3.41)$$

The variables $d_{i,t}$ and $d_{i,t}^*$ are dummies for $y_{i,t}$ and $\varepsilon_{i,t}$ respectively, where $\varepsilon_{i,t}$ are the standardized residuals of the two univariate asymmetric volatility processes. The dummies are equal to one when these variables are negative and zero otherwise. The coefficient $\theta_3/2$ relies on the assumption that the standardized residuals have a symmetric distribution. φ_{12} is the average correlation of returns, while φ_3 is the average asymmetric component $\left(d_{1,t-1}^* \varepsilon_{1,t-1}\right)\left(d_{2,t-1}^* \varepsilon_{2,t-1}\right)$ (Engle and Colacito, 2004).

The log-likelihood of the estimator H_t is given in Engle and Sheppard (2001) as:

$$L = -\frac{1}{2} \sum_{t=1}^{T} \left(k \log(2\pi) + 2\log(|D_t|) + \log(|R_t|) + \epsilon_t' R_t^{-1} \epsilon_t\right) \quad (3.42)$$

3.9 Portfolio optimization

An optimal portfolio needs to match two conditions: it must be feasible and efficient. If the weights of the assets in the portfolio sum to one, it is a feasible portfolio. This gives a set of all portfolio returns and standard deviations that are feasible. The portfolio with the minimum variance is on the envelope of the feasible set. An efficient portfolio is one that for a given level of risk maximizes the return, and the set of all possible efficient portfolios is called the efficient frontier (Benninga, 2008).

The return of a linear portfolio of *n* assets at day *t* is the sum of individual asset returns multiplied with the corresponding weight *w* invested in this asset, so that:

$$E(r_{p,t}) = w_{1,t}E(r_{1,t}) + \cdots + w_{k,t}E(r_{n,t}) \qquad (3.43)$$
$$\text{where } \sum w_i = 1$$

The variance of such a portfolio is:

$$\sigma_{p,t}^2 = w_{1,t}^2\sigma_{1,t}^2 + \cdots + w_{k,t}^2\sigma_{n,t}^2 + 2w_{1,t}w_{2,t}cov(w_{1,t}, w_{2,t}) + \cdots$$
$$+2w_{n-1,t}w_{n,t}cov(w_{n-1,t}, w_{n,t}) \qquad (3.44)$$

In matrix form the variance of the portfolio is simply $\sigma_{p,t}^2 = w'H_tw$, where w is a vector of weights, w' is the inverse vector of weights and H_t is the covariance matrix of asset returns.

A global minimum variance portfolio (GMVP) is a portfolio strategy that minimizes the expected variance of the portfolio. It is thus done on the basis of risk characteristics (Alexander, 2001). The asset allocation of a GMVP is given by:

$$\min_{\omega_t} \omega_t' H_t \omega_t \quad \textit{where} \sum w_i = 1 \qquad (3.45)$$

When short-selling is allowed, meaning that asset weights can be negative or zero, the optimal weights of GMVP is given by (Alexander, 2001):

$$w_i^* = \psi_i / \sum \psi_i \qquad (3.46)$$

where ψ_i is the sum of elements in the *i*th column of H_t^{-1}, and $\sum \psi_i$ is the sum of all elements of H_t^{-1}. The variance of the GMVP is then $1/\sum \psi_i$. When short-selling is not allowed the solution is more complex.

The GMVP approach ignores the portfolio return characteristics (Alexander, 2001). An alternative optimization could be an allocation that allows more risk, accompanied by higher returns. Another option is to find the weights of the portfolio that gives the best return for the risk undertaken. This reward-to-variability can be measured by the Sharpe ratio θ by dividing the risk premium of the portfolio (expected return of portfolio minus the risk free rate) on the risk of the portfolio (Bodie et al., 2008):

$$\theta_t = \frac{E(r_{p,t}) - r_{f,t}}{\sigma_{p,t}} \tag{3.47}$$

The slope of θ_t, which is connecting the risk free rate with the optimal portfolio, is known as the capital market line (CML). The optimal portfolio, given by the asset weights that maximize the Sharpe ratio, is the tangent portfolio between the CML and the efficient frontier.

4. Data

For GARCH models daily or intra-day returns are commonly used since the GARCH effects at lower frequencies are less apparent. The inclusion of major market events in the estimation period will raise long-term volatility forecasts by several percent, and this needs to be considered when choosing the time period (Alexander, 2001). All parameter estimates in a GARCH model are sensitive to the choice of historic data used, so it has to be decided whether events that occured a long time ago should influence current forecasts. To ensure that the likelihood function is well defined, and that the models properly converge, a few years of data are needed, but not so many years that current market conditions are not reflected. A too short period may give parameter estimates that lack robustness (Alexander, 2001). In this paper four years of daily data, leading up to December 31, 2005, are used for the estimation of various volatility models. The reason for this choice of time span is that the volatility has been relatively stable over this period, apart from the first year where it was slightly higher, and the models perform best during stable times. It also gives room for comparing the forecasts with the "actual" volatility. For this comparison, data from 2006 are used as the out-of-sample period. This means that a total of 1258 daily observations for each asset are collected. Of course, when constructing portfolios in real life the estimation has to be done up to the current date.

Three arbitrarily chosen ETFs from the Dow Jones Total Market Index are used in this paper as a proxy for the investment in three different industries. These industries are the financials (IYF), energy (IYE) and utilities (IDU) sectors of the Dow Jones index. The data are collected from Yahoo! Finance[3]. The returns of each asset, calculated as log differences using the dividend adjusted closing prices, represent the total return in each sector.

[3] http://finance.yahoo.com

5. Results

This chapter is organized as follows: firstly, some descriptive statistics for the three individual time series are outlined. This gives a useful input to the stylized facts of each time-series. Secondly, the results of the three univariate time series are presented one-by-one in Chapters 5.2 to 5.4. Several models, estimated under the assumption of both normal and t-distributed errors, will be considered for each asset. The choice of the right model is based on how well the estimated parameters fit the underlying data set and on how well each model predicts future volatilities. Lastly, Chapter 5.5 presents the results of multivariate modeling, where portfolios of the three assets are created.

5.1 Descriptive statistics

Figure 5.1 describes the price development of the three sectors from 2002 to 2007; the starting prices are set equal to 100. All series have had an upward trend in the analysis period and especially the energy sector has performed well in this time-span.

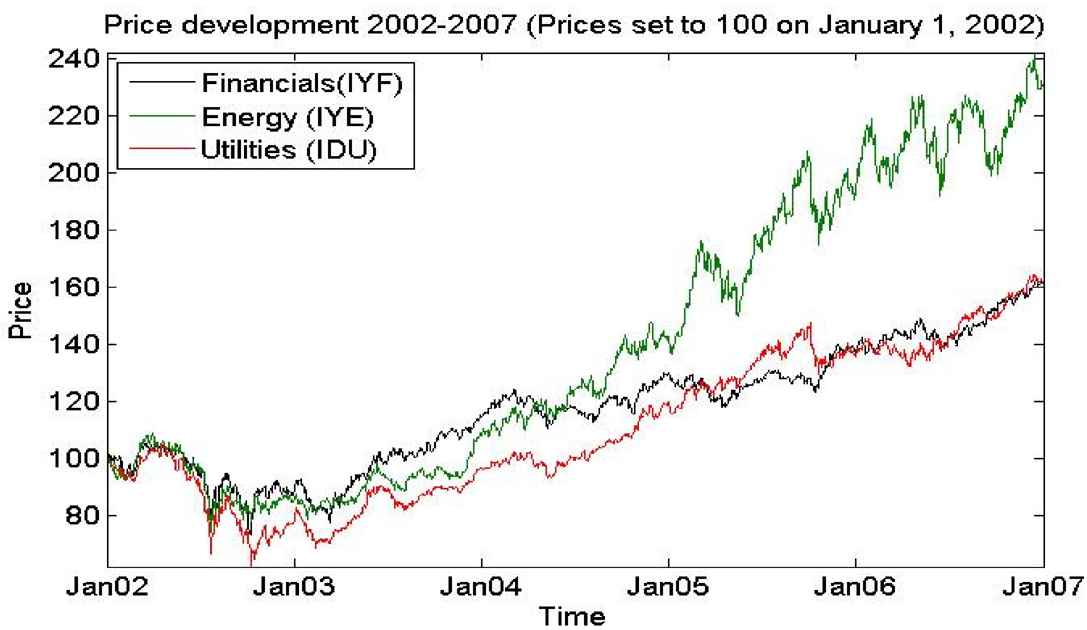


Figure 5.1 Price development of the three assets. Initial prices set to 100.

Table 5.1 provides the relevant information about the three time-series. The table shows that all three series possess standard properties of financial returns: they have a higher kurtosis than that of a normal distribution, and a skewness deviating from zero. The leptokurtic is most pronounced for IDU with an excess kurtosis of about 10. IYF is positively skewed, while IYE and IDU are negatively skewed. A formal test to check whether a distribution is normal, is the Jarque-Bera test (Equation 3.23). This test strongly rejects that the raw returns of all three series are normally distributed at all significance levels. Even though the raw returns are far from normal, it is well known that returns standardized by corresponding conditional standard deviations from a conditional volatility model can be normal or close to normal (Cappiello et al., 2006). Conditional volatility models under the assumption of normal distributions will therefore be estimated and tested for each asset.

Descriptive statistic	**IYF**	**IYE**	**IDU**
Observations	1258	1258	1258
Daily mean	0.038%	0.070%	0.038%
Kurtosis	7.1358	4.3926	13.0100
Skewness	0.2238	-0.2860	-0.3687
Unconditional variance	0.0001	0.0002	0.0001
Unconditional st.dev.	1.12%	1.43%	1.12%
Minimum return	-5.33%	-6.97%	-8.20%
Maximum return	6.45%	6.88%	7.73%
GARCH Autocorrelation	0.3074	0.1373	0.4012
GARCH LM	118.99	23.75	202.68
A-GARCH Autocorrelation	-0.1042	-0.0748	-0.2119
A-GARCH LM	13.72	7.10	56.58
Jarque-Bera	905.65	118.61	5272.22
P value	0.0000	0.0000	0.0000

Table 5.1 Descriptive statistics for the three returns series.

The three sectors all have a positive average return over the five year period, with the energy sector producing the highest average of 17.6% in annualized terms. The daily return plots for the three indices, multiplied by 100 for estimation purposes, are shown in the left pane of Figure 5.2. The absolute returns are plotted in the right pane to better illustrate the volatility clusters that are evident from all three indices.

A formal test to check whether the time-series display volatility clustering is the Box-Pierce LM test. Volatility clustering implies that there should be a strong autocorrelation in the squared returns. Table 5.1 reports GARCH Autocorrelation for the first-order autocorrelation of squared returns (Equation 3.20). The corresponding Box-Pierce statistic is also given as GARCH LM (Equation 3.19). This test confirms that the three indices all have significant autocorrelation (the 0.1% chi-squared critical value is 10.83) in the squared returns, implying that volatility clustering is present.

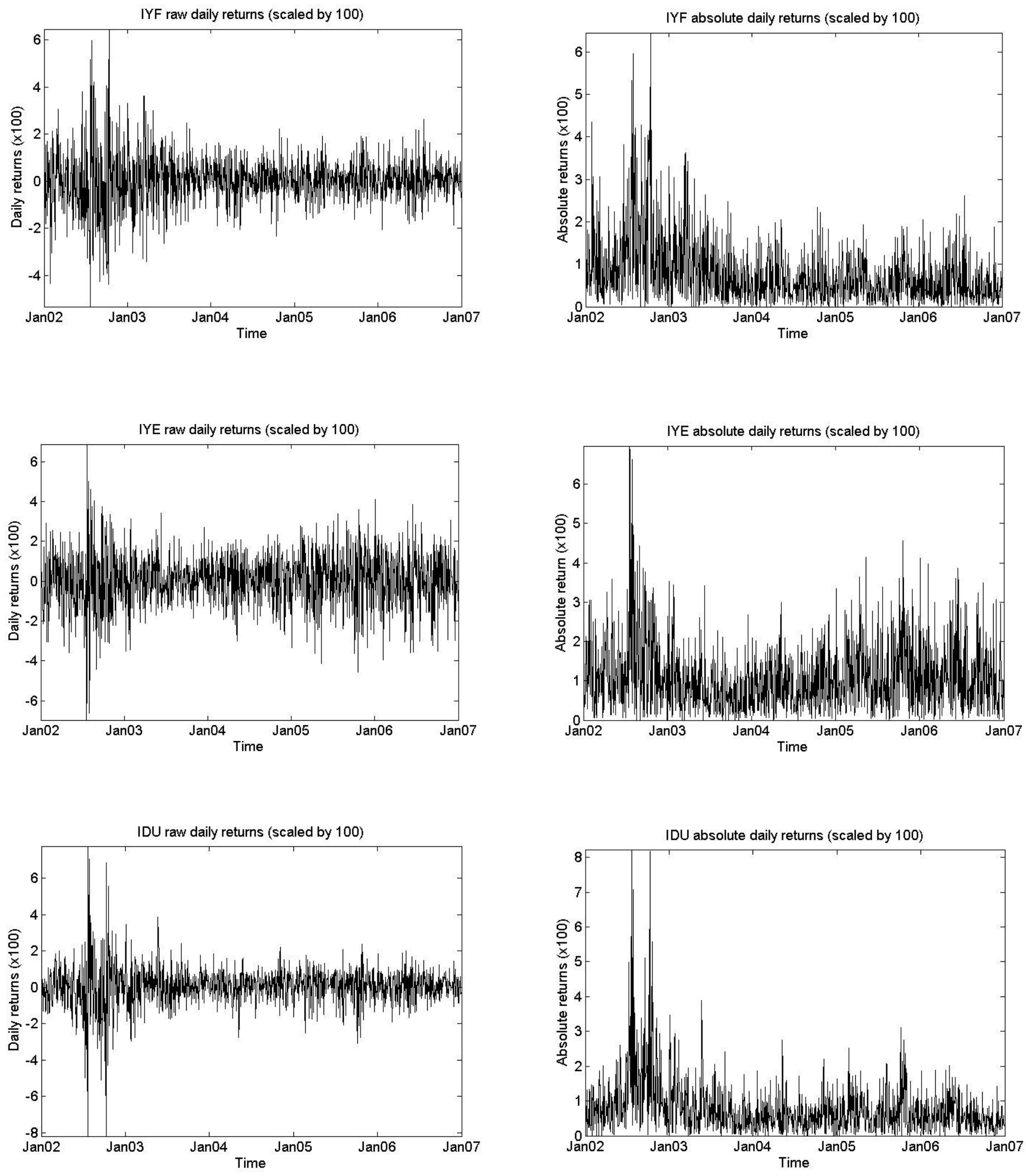


Figure 5.2 Plots of daily raw returns scaled by 100 (left) and corresponding absolute returns (right) for IYF, IYE and IDU respectively.

The asymmetric GARCH test in the last two rows of Table 5.1 investigates the leverage effect of the three indices (Equation 3.24). IYF and IDU have very significant leverage effects, implying that asymmetric volatility models should be used rather than symmetric models. For IYE the evidence is a bit weaker, but it is still significant against the 1% chi-squared critical value of 6.63. Estimating one of the asymmetric volatility models for the time series will eventually confirm whether the asymmetry is significant.

In the following the results for the three ETF's are given one by one, starting with the financial sector.

5.2 iShares Dow Jones US Financial sector (IYF)

The IYF fund tracks the financial and economic sectors of the U.S. equity market. As mentioned, it is arbitrarily chosen as the focus is on showing how to select the right time-varying volatility model. This chapter is divided into six parts. The steps that should be analyzed before estimating volatility models are presented first to control whether the time-series meets the requirements of conditional volatility estimation. Chapter 5.2.2 presents the parameterizations of a variety of conditional volatility models under the assumption of normally distributed errors for IYF, and tries to rank the models based on how well they fit the data in the estimation-period. The normality of the standardized residuals from each model is evaluated in the post-estimation analysis in Chapter 5.2.3. Chapter 5.2.4 shows the parameterization of the same conditional volatility models when assuming t-distributed errors, while Chapter 5.2.5 gives a visual comparison of the volatility processes estimated via the different models and distributions. Finally, Chapter 5.2.6 evaluates the forecasting performance of the significant estimated volatility models.

5.2.1 Pre-estimation analysis

The Kernel density plot in Figure 5.3 can be used to get a non-parametric view to assess whether the raw returns distribution is skewed or heavy tailed. This is done by plotting the individual distribution against a normal. The IYF density clearly has a higher peak than a normal distribution, implying that there is excess kurtosis. The skewness is close to normal. A QQ-plot for IYF is also shown in Figure 5.3 in order to visualize how far from normal the data set is. This plot shows that the data are heavier tailed than a normal.

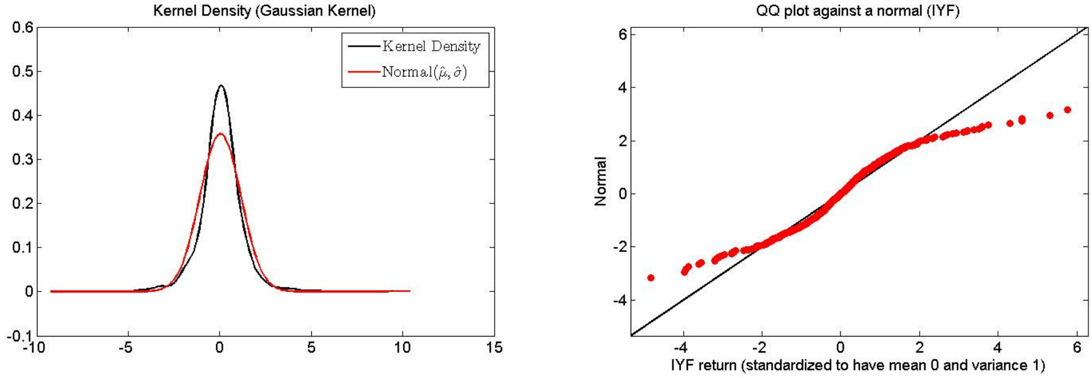


Figure 5.3 Kernel density plot (left) and QQ-plot (right) for the IYF raw returns against a normal distribution.

The Ljung-Box-Pierce Q-test for departure from randomness is a hypothesis test that can be used to quantify the correlations. This test is often used as a post-estimate lack-of-fit test on the fitted innovations of the time-series process, but it is also convenient in the pre-fit analysis because the default model assumes that returns are defined as a constant plus an innovation process. The null hypothesis of the test is that there is no serial correlation. The Q-test is asymptotically χ-distributed. The alternative is that there is significant serial correlation. At the 0.05 significance level, this test rejects the null at 10, 15 and 20 lags for both the raw returns and the squared returns, as illustrated in Table 5.2.

Lags	Raw returns		Squared returns		Critical value
	Q-stat	*P value*	*Q-stat*	*P value*	
10	20.46	0.0252	1303.14	0.0000	18.31
15	27.26	0.0267	1522.47	0.0000	25.00
20	40.98	0.0037	1658.46	0.0000	31.41

Table 5.2 Box-Pierce-Ljung's Q-test of no serial correlation at the 0.05 significance level.

The autocorrelations of squared returns up to 20 lags are shown on the left hand side of Figure 5.4. The broken lines in the figure give the confidence interval using heteroskedastic robust errors, and as the autocorrelations exceeds the errors there is significant autocorrelation in the squared returns, implying that there is volatility clustering in the returns series.

As there are signs that volatility clustering is present, GARCH parameterization seems to be ideal. A final test is to use Engle's ARCH test on the residual returns to test for the presence of ARCH-effects. The null hypothesis of the ARCH test is that the series is a random sequence of Gaussian disturbances, or in other words that the series is homoskedastic, and thus that there is no ARCH-effect. Since the Q-statistic of the raw returns in Table 5.2 indicates that there is serial correlation in the raw returns, ARMA modeling should perhaps be used when modeling the conditional mean. A common practice when modeling conditional volatility, however, is to treat the mean as constant μ. The residuals are therefore calculated as $\varepsilon_t = r_t + \mu$. The result of the ARCH test is given on the right hand side of Figure 5.4. It shows that homoskedasticity can be rejected as the p values are very small. The alternative hypothesis that heteroskedasticity is present is therefore accepted.

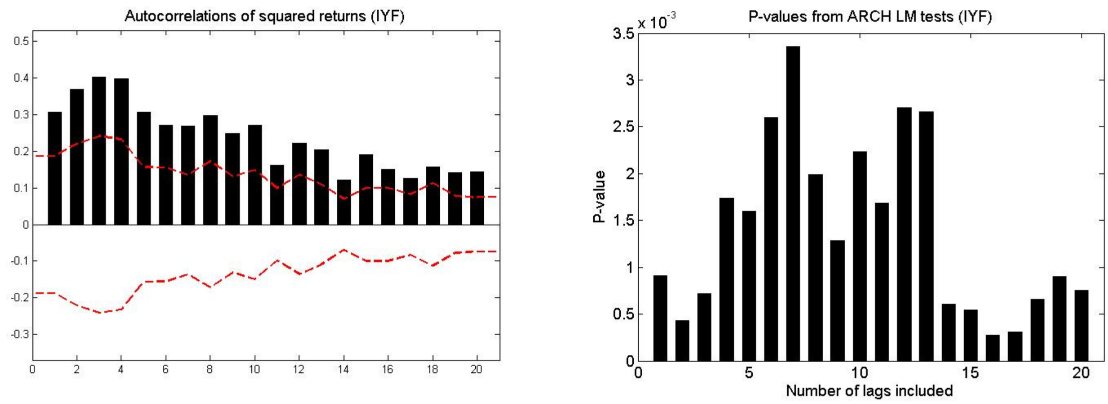


Figure 5.4 Autocorrelations of squared returns with heteroskedastic robust errors (left) and Engle's ARCH test (right).

5.2.2 Model parameterization and selection

Table 5.3 reports the parameterizations of the various ARCH class models applied under the assumption of normal distribution for IYF. The process leading towards a choice of model follows the steps of Sheppard (2009a).

A natural starting point is to examine a symmetric GARCH(1,1) model. This specification gives alpha and beta coefficients that are very significant. It is therefore checked whether any more lags of the square residual (symmetric lags) or variance are needed. Estimating the GARCH(2,1) shows that the second symmetric lag is significant, but that the first lag becomes highly insignificant. The GARCH(1,2) shows that a second lag of the variance is not needed either. A GARCH(1,1) parameterization so far seems to be the most reasonable specification for the capture of symmetric dynamics. This means that the model describing the volatility process so far is best given by $h_t = 0.000009 + 0.072\varepsilon_{t-1}^2 + 0.919h_{t-1}$, which corresponds to Equation 3.5.

	ω (x 10-3)	α1	α2	γ1	γ2	β1	β2	LL
GARCH(1,1)	0.009	0.072				0.919		-1418.4
	(.067)	*(.000)*				*(.000)*		
GARCH(2,1)	0.013	0.000	0.097			0.892		-1414.5
	(.047)	*(.999)*	*(.027)*			*(.000)*		
GARCH(1,2)	0.009	0.072				0.919	0.000	-1418.4
	(.056)	*(.000)*				*(.000)*	*(.999)*	
GJR(1,1,1)	0.011	0.000		0.106		0.936		-1406.2
	(.020)	*(.980)*		*(.000)*		*(.000)*		
GJR(2,1,1)	0.011	0.000	0.021	0.096		0.920		-1405.3
	(.768)	*(.999)*	*(.990)*	*(.848)*		*(.000)*		
GJR(1,2,1)	0.011	0.000		0.106	0.000	0.936		-1406.2
	(.016)	*(.974)*		*(.008)*	*(.999)*	*(.000)*		
GJR(1,1,2)	0.011	0.000		0.106		0.936	0.000	-1406.2
	(.011)	*(.999)*		*(.000)*		*(.000)*	*(.999)*	
TARCH(1,1,1)	0.010	0.009		0.077		0.952		-1406.6
	(.032)	*(.454)*		*(.000)*		*(.000)*		
TARCH(2,1,1)	0.014	0.000	0.030	0.074		0.933		-1404.8
	(.513)	*(.999)*	*(.932)*	*(.367)*		*(.000)*		
TARCH(1,2,1)	0.010	0.009		0.077	0.000	0.952		-1406.6
	(.703)	*(.802)*		*(.901)*	*(.999)*	*(.000)*		
TARCH(1,1,2)	0.010	0.009		0.077		0.952	0.000	-1406.6
	('.031)	*(.512)*		*(.000)*		*(.000)*	*(.999)*	
EGARCH(1,1,1)	-0.001	0.091		-0.073		0.991		-1406.4
	(.755)	*(.000)*		*(.000)*		*(.000)*		
EGARCH(2,1,1)	-0.001	-0.127	0.245	-0.083		0.987		-1400.7
	(.696)	*(.034)*	*(.000)*	*(.000)*		*(.000)*		
EGARCH(1,2,1)	-0.001	0.083		-0.229	0.159	0.993		-1401.5
	(.665)	*(.004)*		*(.000)*	*(0.005)*	*(.000)*		
EGARCH(1,1,2)	-0.001	0.093		-0.075		0.966	0.025	-1406.4
	(.750)	*(.000)*		*(.000)*		*(.000)*	*(0.035)*	
EGARCH(2,2,1)	-0.001	-0.176	0.284	-0.242	0.162	0.990		-1394.5
	(.607)	*(.019)*	*(.000)*	*(.000)*	*(.002)*	*(.000)*		
EGARCH(2,2,2)	-0.001	-0.174	0.279	-0.243	0.166	1.015	-0.024	-1394.4
	(.610)	*(.021)*	*(.000)*	*(.000)*	*(.002)*	*(.000)*	*(.428)*	

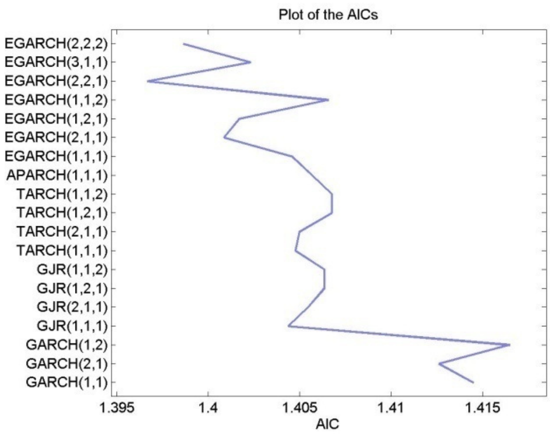
Table 5.3 Parameterization of conditional volatility models for IYF under the assumption of a normal distribution. P-values are reported in brackets, while LL reports the maximum log likelihood for the different models.

As noted in section 5.1, there is evidence for asymmetry in the time series of IYF. The GJR-GARCH(1,1,1) model confirms this as the asymmetric lag *γ* is significant. The symmetric lag, however, is insignifcant, impliying that it could be dropped in the parameterization. The symmetric lag will be kept as it can be assumed that there should be a rise in volatility after large positive shocks. This rise will not be large as the estimated *α* is zero to at least the third decimal. The GJR-GARCH(1,1,1) is superior to the GARCH(1,1) specification as it has a higher likelihood. Again additional lags are examined. The estimated parameters of the GJR-GARCH(2,1,1), GJR-GARCH(1,2,1) and GJR-GARCH(1,1,2) models respectively show that an additional symmetric lag, asymmetric lag and variance lag are all insignificant.

The TARCH parameterization is examined next to check whether a nonlinearity is needed. The TARCH(1,1,1) model also has an insignificant symmetric lag, and the specification adds nothing over the GJR-GARCH(1,1,1) model as the log likelihood is lower. Adding more lags to the TARCH does not change anything. The GJR-GARCH(1,1,1) model still fits the time-series best.

Finally, it is tested whether an EGARCH model may fit better. All parameters in the EGARCH(1,1,1) model are significant at any level, but the log likelihood is slightly lower than for GJR-GARCH. A second lag of the symmetric term is significant, but the first symmetric lag is negative and thus insignificant. The likelihood is, however, improved over the GJR-GARCH(1,1,1). Adding an extra asymmetric lag shows that all parameters are significant at the 0.05 level (although the second asymmetric lag is positive, which is a slightly contradictiv) and the EGARCH(1,2,1) has a much higher likelihood than GJR(1,1,1). In terms of the likelihood, an EGARCH(1,1,2) adds nothing. An EGARCH(2,2,1) model improves the log likelihood further, but α_1 is negative and thus insignificant. This is also the case for EGARCH(2,2,2), where β_2 is insignificant as well.

From this analyzis EGARCH(1,2,1) seems to give the best fit, followed by GJR-GARCH(1,1,1) and EGARCH(1,1,1). Finally the AICs and BICs are examined. These are plotted in Figure 5.5. These criterions select the model with the lowest AIC and BIC respectively, meaning that plots to the left of are preferred.



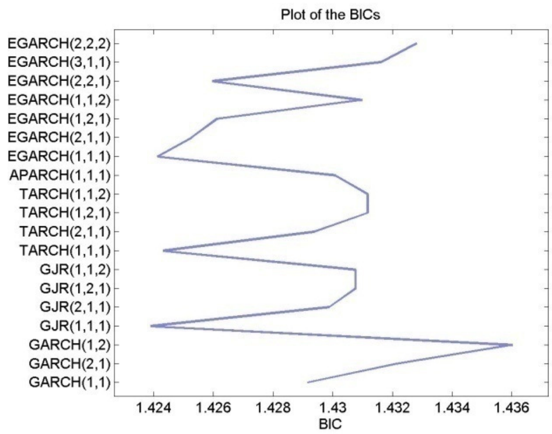


Figure 5.5 Plots of AIC (left) and BIC (right) for IYF.

The AIC selects EGARCH(2,2,1). Generally, the EGARCH models seem to do well according to the AIC selection criterion. The BIC criterion will always select more parsimonious models, i.e. models with fewer parameters, and is often exclusively used as a criterion to select the best model (see e.g. Cappiello et al. (2006)). GJR(1,1,1) is selected, but closely followed by EGARCH(1,1,1) and TARCH(1,1,1). The GARCH models perform badly as expected, since they do not include an asymmetric term. Both criterions show that adding an extra variance lag can be ruled out for IYF. Since the EGARCH(1,2,1) seems highly significant, has a high likelihood and does reasonably well according to both the AIC and BIC selection criterion, EGARCH(1,2,1) is selected as the best model for IYF.

5.2.3 Post-estimation analysis

The parameterizations in Table 5.3 are estimated under the assumption of a normal distribution. The standardized residuals from the models therefore need to be examined for normality. Table 5.4 shows the skewness, kurtosis and Jarque-Bera statistics for the raw returns in the estimation period and the standardized residuals from various GARCH class models.

	Skew	**Kurt**	**JB**	**P value**
Raw returns	0.229	6.624	558.840	0.0000
Standardized residuals				
GARCH(1,1)	-0.090	3.390	7.742	0.0208
GJR(1,1,1)	-0.124	3.376	8.519	0.0141
TARCH(1,1,1)	-0.102	3.426	9.340	0.0094
EGARCH(1,1,1)	-0.093	3.394	7.932	0.0190
EGARCH(1,2,1)	-0.061	3.360	6.033	0.0490

Table 5.4 Normality test of raw returns and standardized residuals from a set of GARCH specifications for IYF.

The raw returns, with significant positive skewness and high excess kurtosis, reject normality strongly, as can be seen by the Jarque-Bera statistic with a null hypothesis that the skewness is 0 and the kurtosis is 3. The standardized residuals come closer to being normally distributed. The kurtosis is closer to 3, while the skewness is only slightly negative, indicating that the actual return distribution is asymmetric. The p-value of the Jarque-Bera test confirms this, especially for EGARCH(1,2,1), which is almost significant at the 0.05 level. As the standardized residuals are so close to normal, the assumption of normal errors can explain most of the leptokurtis in the returns. If this was not the case, mistakes would be made when predicting returns. The density- and the QQ-plots for the standardized residuals from the EGARCH(1,2,1) model in Figure 5.6 confirm that the standardized residuals look more normal than the raw returns in Figure 5.3, but the QQ-plot shows signs of deviation from normality, especially in the lower tail.

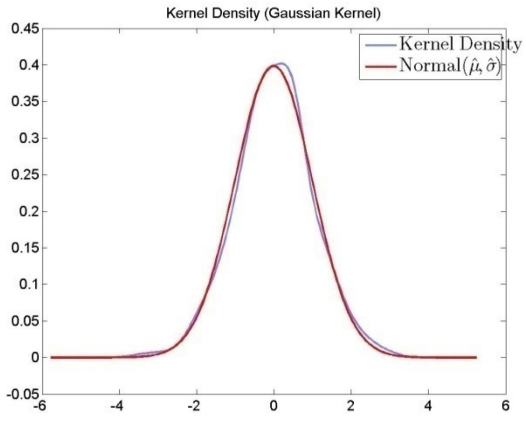


Figure 5.6 Kernel density plot (left) and QQ-plot (right) of the standardized residuals from an EGARCH(1,2,1) model against a normal distribution.

The fact that the standardized residuals might not be completely normal can be neglected as they are close. Figure 5.7 reports the result of an ARCH test of the standardized residuals from the EGARCH(1,2,1) model. The high p-values reject the null of remaining significant ARCH effects, meaning that the volatility clusters have been captured. The only worry is that there is still significant autocorrelation in the first lag of the squared standardized residuals. This is also the case for the other evaluated models.

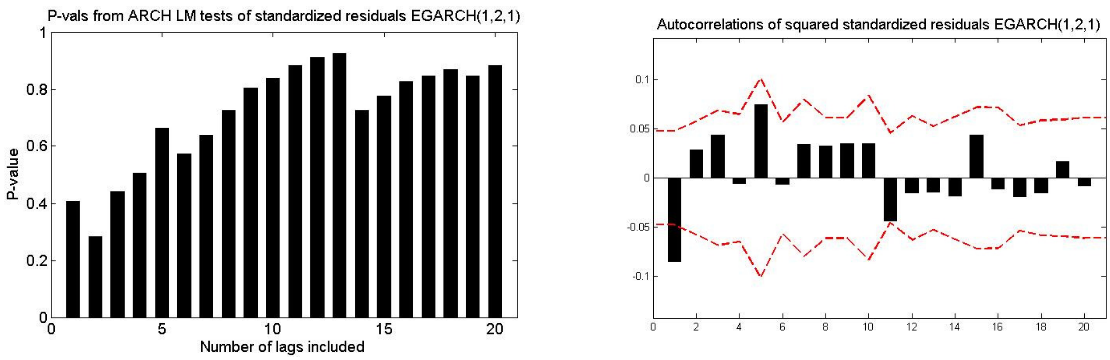


Figure 5.7 ARCH LM test of the standardized residuals (left) and autocorrelations of squared standardized residuals (right) for EGARCH(1,2,1).

5.2.4 Student's T distribution

Since there is still some excess kurtosis in the standardized residuals, the assumption of t-distributed errors is used to parameterize the GARCH class models again. The student t's distribution parameterizations are given in Table 5.5.

	ω (x 10-3)	α1	α2	γ1	γ2	β1	β2	ν	LL
t-GARCH(1,1)	0.010	0.076				0.913		15.7422	-1415.6
	(.060)	*(.000)*				*(.000)*		*(.032)*	
t-GARCH(2,1)	0.011	0.000	0.118			0.930		16.451	-1411.9
	(.019)	*(.999)*	*(.000)*			*(.000)*		*(.040)*	
t-GARCH(1,2)	0.010	0.076				0.913	0.000	15.742	-1415.6
	(.057)	*(.000)*				*(.000)*	*(.999)*	*(.029)*	
t-GJR(1,1,1)	0.011	0.000		0.118		0.930		16.451	-1403.6
	(.019)	*(.999)*		*(.000)*		*(.000)*		*(.040)*	
t-GJR(2,1,1)	0.012	0.000	0.022	0.107		0.913		17.012	-1402.8
	(.040)	*(.999)*	*(.875)*	*(.036)*		*(.000)*		*(.100)*	
t-GJR(1,2,1)	0.011	0.000		0.118	0.000	0.930		16.455	-1403.6
	(.019)	*(.999)*		*(.085)*	*(.999)*	*(.000)*		*(.032)*	
t-GJR(1,1,2)	0.011	0.000		0.118		0.930	0.000	16.456	-1403.6
	(.017)	*(.999)*		*(.000)*		*(.000)*	*(.999)*	*(.036)*	
t-TARCH(1,1,1)	0.010	0.004		0.091		0.949		14.510	-1403.3
	(.026)	*(.735)*		*(.000)*		*(.000)*		*(.016)*	
t-TARCH(2,1,1)	0.014	0.000	0.025	0.085		0.931		16.122	-1402.0
	(.075)	*(.999)*	*(.577)*	*(.000)*		*(.000)*		*(.034)*	
t-TARCH(1,2,1)	0.010	0.004		0.091	0.000	0.949		14.512	-1403.3
	(.090)	*(.737)*		*(.271)*	*(.999)*	*(.000)*		*(.014)*	
t-TARCH(1,1,2)	0.011	0.003		0.097		0.886	0.061	14.411	-1403.3
	(.030)	*(.830)*		*(.003)*		*(.000)*	*(.765)*	*(.018)*	
t-EGARCH(1,1,1)	-0.001	0.097		-0.087		0.990		14.712	-1403.4
	(.797)	*(.000)*		*(.000)*		*(.000)*		*(.018)*	
t-EGARCH(2,1,1)	-0.001	-0.128	0.249	-0.096		0.987		16.421	-1397.9
	(.696)	*(.033)*	*(.000)*	*(.000)*		*(.000)*		*(.030)*	
t-EGARCH(1,2,1)	-0.001	0.085		-0.245	0.165	0.992		15.214	-1398.8
	(.691)	*(.001)*		*(.000)*	*(.004)*	*(.000)*		*(.015)*	
t-EGARCH(1,1,2)	-0.001	0.101		-0.093		0.921	0.068	14.617	-1403.4
	(.790)	*(.000)*		*(.003)*		*(.000)*	*(.757)*	*(.018)*	
t-EGARCH(2,2,1)	-0.002	-0.193	0.300	-0.266	0.177	0.990		16.249	-1391.7
	(.578)	*(.007)*	*(.000)*	*(.000)*	*(.001)*	*(.000)*		*(.018)*	
t-EGARCH(2,2,2)	-0.001	-0.188	0.289	-0.269	0.185	1.045	-0.055	16.089	-1391.6
	(.580)	*(.013)*	*(.000)*	*(.000)*	*(.001)*	*(.000)*	*(.698)*	*(.019)*	

Table 5.5 Parameterization of conditional volatility models for IYF under the assumption of a t-distribution.

For a t-distribution an extra parameter, ν, is estimated. This parameter, which is the shape parameter, is significant at the 0.05 level for all models appart from GJR(2,1,1). Testing the covariance stationarity restriction shows that GARCH(2,1) is the only model not to meet this as $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j > 1$. When an estimated model is no longer covariance stationary, the unconditional variance does not exist so that the asymptotic properties of the maximum likelihood estimators are unclear.

Under the assumption of a t-distribution, TARCH(1,1,1) fits better than the GJR-GARCH(1,1,1). The symmetric lag of both models is still insignificant, but close to zero. The choice is now between TARCH(1,1,1) and one of the EGARCH models. EGARCH(1,1,1) has a slightly lower likelihood, but all parameters are significant at the 0.05 level. As for the normal distribution, the EGARCH(1,2,1) has the highest likelihood of the significant models. It is therefore ranked ahead of TARCH(1,1,1), EGARCH(1,1,1) and GJR(1,1,1).

The AIC selection criterion in the left pane of Figure 5.8 selects EGARCH(2,2,1). This model, however, has a negative first symmetric lag which is undesirable. The EGARCH models are generally preferred by AIC. The BIC criterion ranks TARCH(1,1,1) ahead of EGARCH(1,1,1) and GJR(1,1,1), but these are extremely close. EGARCH(1,2,1) is also ranked high for both models, in fact it is the highest ranked significant model according to AIC.

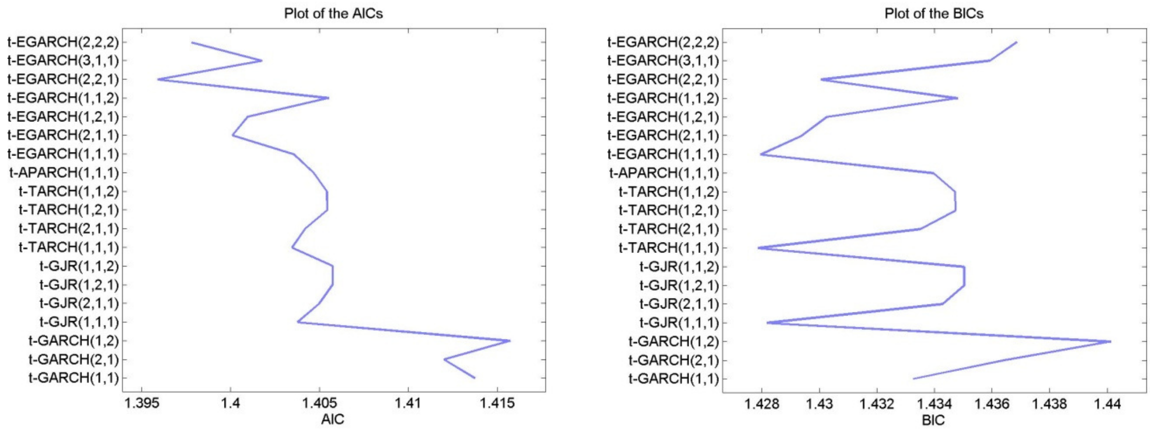
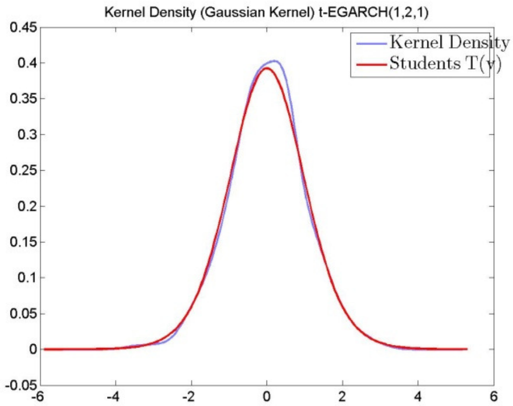


Figure 5.8 Plots of AIC (left) and BIC (right) for IYF for t-distributed errors.

Figure 5.9 examines whether the assumption of a t-distribution seems reasonable. The standardized residuals from the EGARCH(1,2,1) model are drawn in the density plot together with a student's t density with 15.2 degrees of freedom, corresponding to the estimated ν for EGARCH(1,2,1). The series seems to be just as well explained by t-distributed as by normal distributed errors, altough the QQ-plot shows deviations from the t(15.2) distribution in both tails.



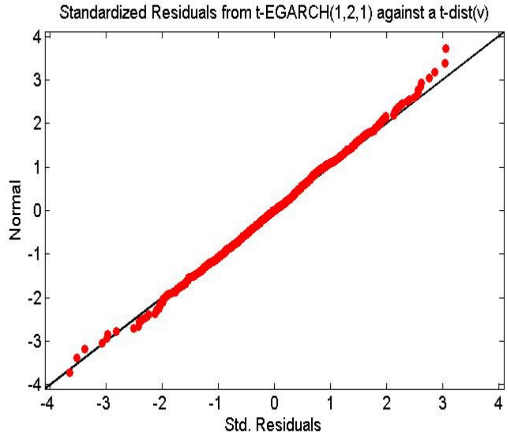


Figure 5.9 Kernel density plot (left) and QQ-plot (right) of the standardized residuals from an t-distributed EGARCH(1,2,1) model against a t(15.2).

All models were tested for remaining ARCH effects in the standardized residuals and autocorrelations in the squared standardized residuals, with the same result as for the normal distribution.

The most interesting comparision that can be made between the normal distributed and t-distributed parameterizations is to study the log likelihood. Table 5.6 shows that the likelihood is improved for all models when using t-distribution. The t-distributed EGARCH(1,2,1) model is therefore the model that gives the best fit for IYF, ahead of the normal distributed EGARCH(1,2,1) .

	Skew	Kurt	LL	AIC	BIC
Normal					
GARCH(1,1)	-0.090	3.390	-1418.4	1.4145	1.4292
GJR(1,1,1)	-0.124	3.376	-1406.2	1.4044	1.4239
TARCH(1,1,1)	-0.102	3.426	-1406.6	1.4048	1.4243
EGARCH(1,1,1)	-0.093	3.394	-1406.4	1.4046	1.4241
EGARCH(1,2,1)	-0.061	3.360	-1401.5	1.4017	1.4261
Students t					
GARCH(1,1)	-0.096	3.403	-1415.6	1.4137	1.4333
GJR(1,1,1)	-0.135	3.414	-1403.6	1.4038	1.4282
TARCH(1,1,1)	-0.117	3.500	-1403.3	1.4035	1.4279
EGARCH(1,1,1)	-0.111	3.480	-1403.4	1.4036	1.4280
EGARCH(1,2,1)	-0.071	3.401	-1398.8	1.4010	1.4303

Table 5.6 Descriptive statistics for normal and student's t distributed models.

5.2.5 Visual comparision of the time-varying volatilities

Even though a lot of models have been tested, each giving different estimates of the parameters, there is not a big visual difference in how the time-varying volatility evolves over time. Figure 5.10 shows that for a TARCH(1,1,1) model, the volatility process is approximately identical for normal and t-distributed errors, something that is a common feature when estimating under alternative distributional assumptions (Sheppard, 2009a). According to Sheppard (2009a), the alternative distributions are more meaningful when looking at the Value-at-Risk (see Chapter 5.2.6). Figure 5.11 shows the difference between the symmetric GARCH(1,1) process and the asymmetric EGARCH(1,1,1) process. The two volatility series are similar because the asymmetric coefficient of the EGARCH model is of a very small magnitude. The forecasts, however, will be different for a symmetric model compared to that of an asymmetric model. This is studied in section 5.2.6.

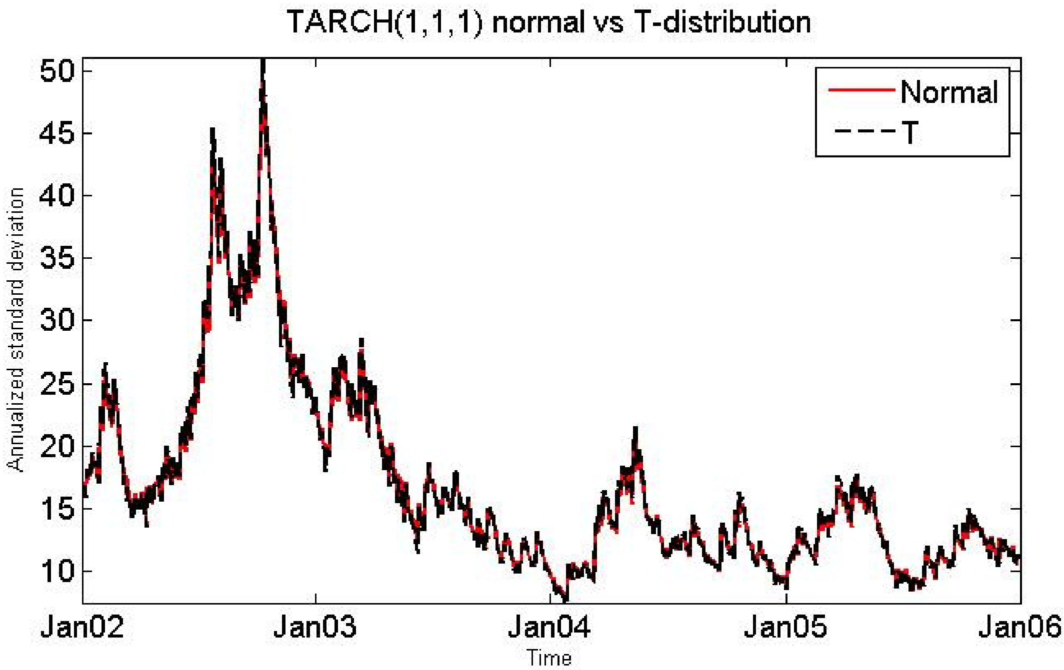


Figure 5.10 Estimated time-varying volatility using a TARCH(1,1,1) specification for both normal and t-distributed errors.

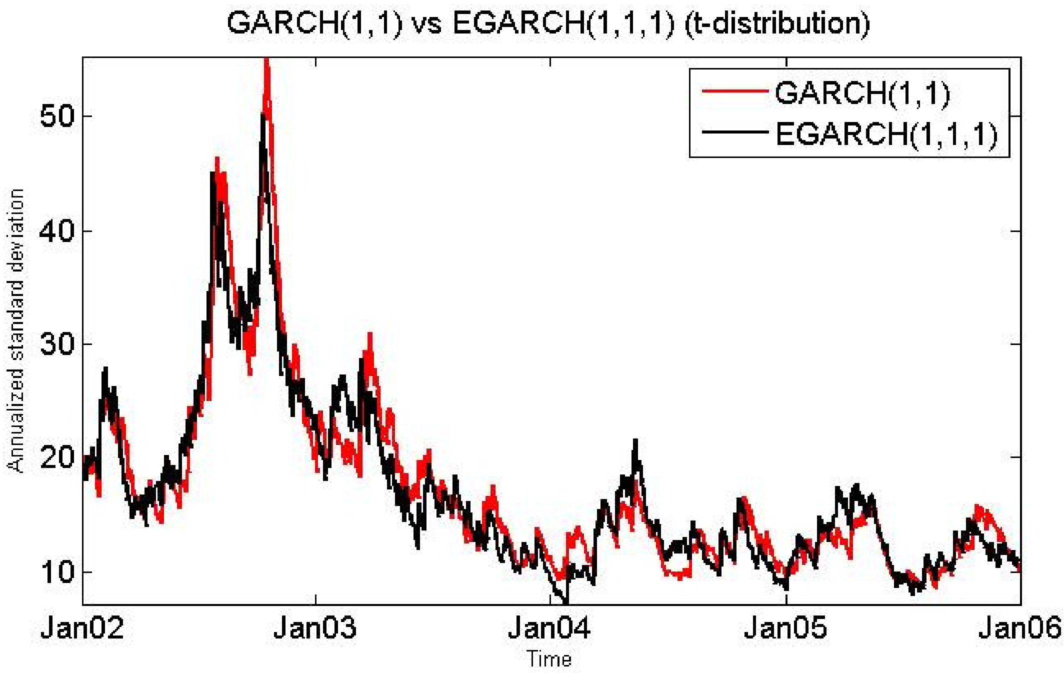


Figure 5.11 Estimated time-varying volatility using a symmetric GARCH(1,1) and an asymmetric EGARCH(1,1,1), both t-distributed.

Finally, Figure 5.12 illustrates the difference between the RiskMetrics™ (J.P.Morgan, 1997) EWMA model with a decay factor of 0.94 and a normal distributed GJR(1,1,1). The figure also shows the unconditional volatility recalculated every sixth month.

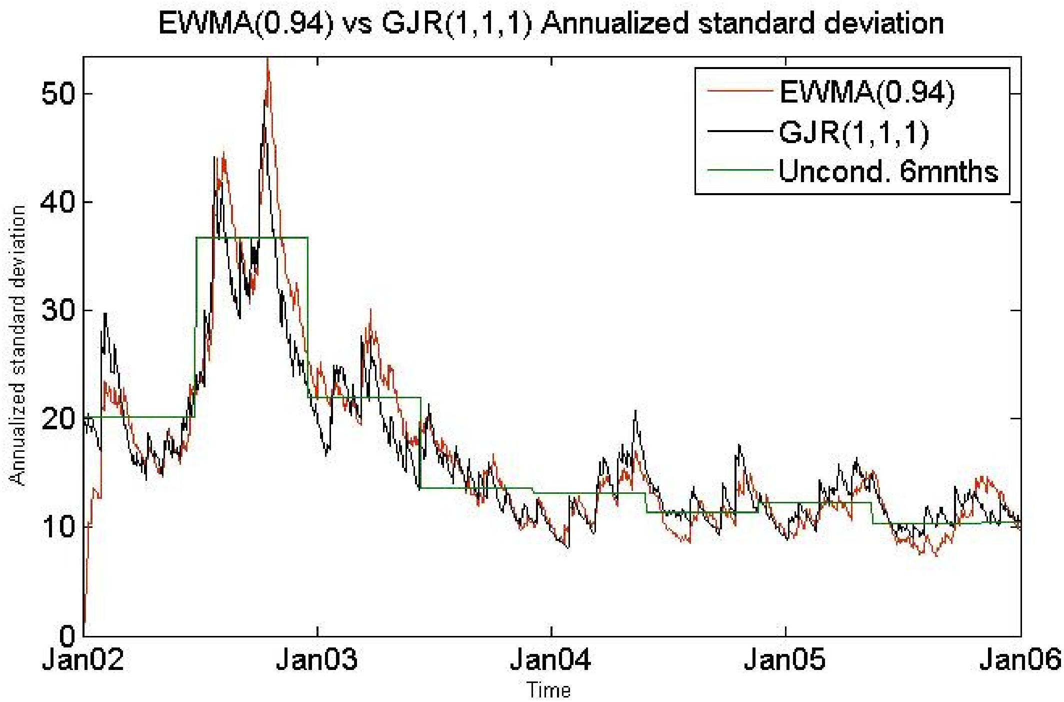


Figure 5.12 Estimated time-varying volatility from RiskMetrics™ versus the normal distributed GJR(1,1,1). These are compared to the unconditional volatility calculated at 6 month intervals.

5.2.6 Forecast evaluation

Figure 5.13 illustrates what a conditional volatility forecasting model actually does. Four years of data have been used to estimate the parameters of an EGARCH(1,1,1) model under the assumption of a normal distribution. The estimated parameters are then used to forecast the volatility for the next year. This forecast must then be compared to the "actual" volatility, which is unobservable. Therefore squared residuals for the last year of the time-series, i.e. 2006, are calculated to provide a proxy for the actual volatility. The residuals are calculated as daily returns minus the mean of the first four years of returns. The estimated conditional volatility models should therefore be compared on how well they predict the future.

As with the time-varying volatility in the estimation period, there is no big difference in the various model's forecasts for the upcoming year. This is illustrated in Figure 5.14 for a t-distributed EGARCH(1,2,1) process and a normal-distributed GJR(1,1,1) process. This figure also shows the six-month unconditional volatilities for the five years of data.

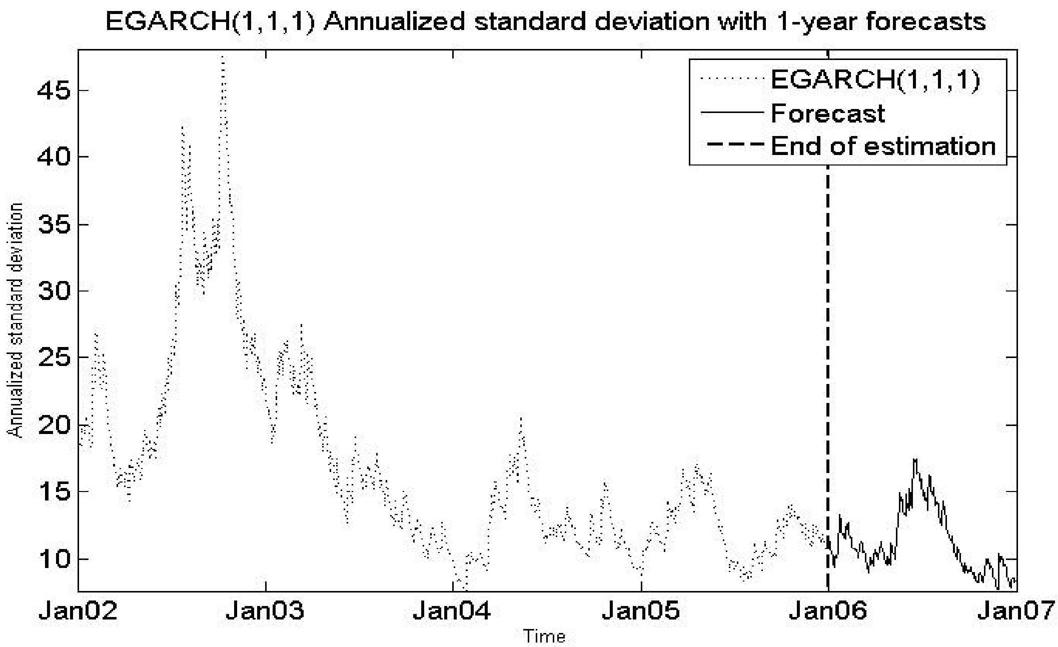


Figure 5.13 Time-varying volatility and volatility forecasts for a normal distributed EGARCH(1,1,1) process for IYF.

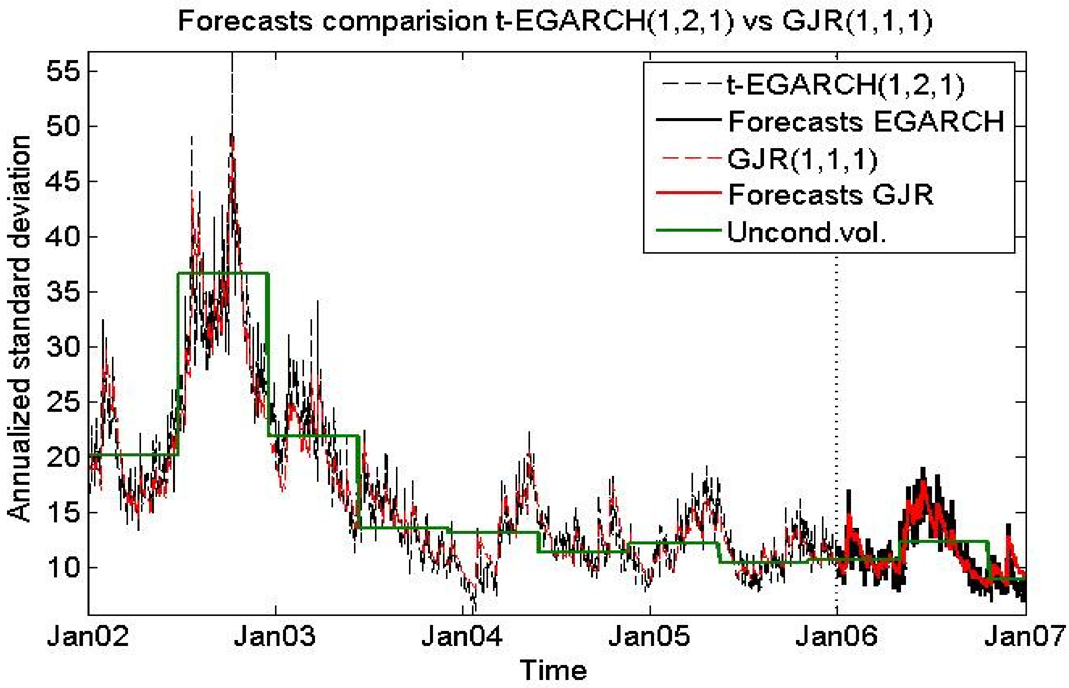


Figure 5.14 Time-varying volatility and volatility forecasts for a t-distributed EGARCH(1,2,1) and a normal distributed GJR(1,1,1) process for IYF. The unconditional volatility is updated every sixth month for the five years of collected data.

The first step that is applied when evaluating the volatility forecasts of the various models is the Mincer-Zarnowitz regression (R^2), using squared residuals as a proxy for volatility. The null hypothesis using a Wald test is that the forecasts are not optimal. Optimal forecasts should have a MZ statistic close to one. The MZ procedure does not reject optimality for any of the asymmetric models as the p-values exceed any realistic significance level. The GARCH(1,1) models are also optimal at the 0.05 level, but not at the 0.1 level for the t-distribution.

	Normal		**T-distribution**	
	MZ	*P value*	*MZ*	*P value*
GARCH(1,1)	4.209	0.1219	4.909	0.0859
EGARCH(1,1,1)	0.422	0.8098	0.673	0.7145
EGARCH(1,2,1)	0.893	0.6400	1.267	0.5308
TARCH(1,1,1)	0.408	0.8154	0.586	0.7462
GJR(1,1,1)	0.653	0.7216	0.873	0.6462

Table 5.7 Mincer-Zarnowitz R^2 for normal and t-distributed squared residuals as proxy for volatility.

Table 5.8 ranks the estimated forecasts by four statistical loss functions; ME, MSE, RMSE and MAE. These four statistics are also used by among others Akgiray (1989) and Brailsford and Faff (1996). The ranking is based on how well the model minimizes the statistical loss function of the forecast errors, where squared residuals are still used as a proxy for volatility. In addition, the R^2 ranking is based on how close to one the Mincer-Zarnowitz regression is.

	ME	MSE	MAE	MAPE	R^2	Total	Total rank
GARCH(1,1)	2	9	9	9	9	38	**9**
EGARCH(1,1,1)	2	3	3	6	7	21	**3**
EGARCH(1,2,1)	1	6	1	2	1	11	**1**
TARCH(1,1,1)	7	2	4	8	8	29	**6**
GJR(1,1,1)	9	5	7	4	5	30	**7**
t-GARCH(1,1)	5	10	10	10	10	45	**10**
t-EGARCH(1,1,1)	5	4	4	5	4	22	**4**
t-EGARCH(1,2,1)	4	7	2	1	3	17	**2**
t-TARCH(1,1,1)	8	1	6	7	6	28	**5**
t-GJR(1,1,1)	10	8	8	3	2	31	**8**

Table 5.8 The forecasting performance ranked by statistical loss functions and the Mincer-Zarnowitz R^2.

The EGARCH models seem to minimize the errors mostly. Interestingly, it is the normal distributed, rather than the t-distributed, EGARCH(1,2,1) model which provides the best forecasts. The ranking is also quite consistent with previous conclusions regarding which models are best.

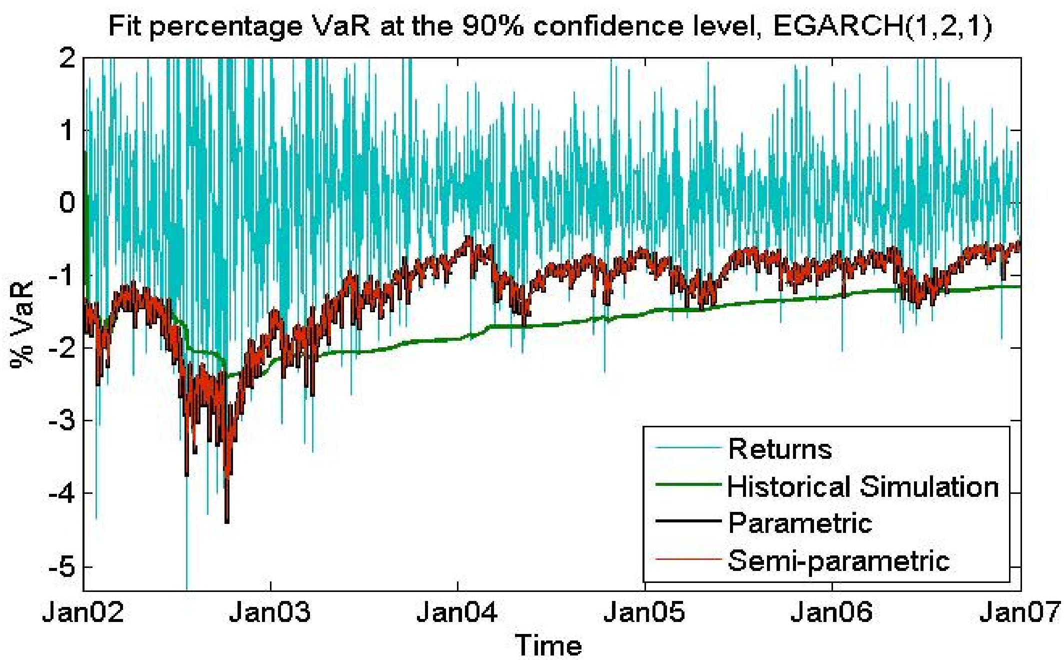


Figure 5.15 Estimated % VaR using historical simulation, parametric and semi-parametric VaR models for IYF at the 90% confidence level. The period Jan06 – Jan07 is the out-of-sample period.

Figure 5.15 shows the development in the one-day ahead VaR at the 90% confidence level for a normal distributed EGARCH(1,2,1) model using the historical simulation, fully parametric and semi-parametric VaR models. A 10% VaR is the return value of an investment we are 90% confident we will not lose more than. In other words, a 10% 1-day VaR represents the loss level that would be exceeded under normal market cirumstances one day in every ten days, when the portfolio is left unmanaged (Alexander, 2001). For a 10%-VaR this should occur approximately for 10% of the returns. When the model predicting the VaR is too low, more than 10% exceptional losses (i.e. hits) will be observed in the out-of-sample period. The parametric and semi-parametric models are very similar for this case.

		Hits	% Hits	MZ	P-value	Ranking
Historical VaR Simulation		11	4.38%	20.18	0.0000	
GARCH(1,1)	Parametric	19	7.57%	2.42	0.2979	
	Semi parametric	18	7.17%	3.12	0.2104	
GJR(1,1,1)	Parametric	21	8.37%	0.88	0.6445	4
	Semi parametric	21	8.37%	0.88	0.6445	4
TARCH(1,1,1)	Parametric	21	8.37%	1.55	0.4617	10
	Semi parametric	21	8.37%	1.55	0.4617	10
EGARCH(1,1,1)	Parametric	21	8.37%	1.55	0.4617	10
	Semi parametric	21	8.37%	1.55	0.4617	10
EGARCH(1,2,1)	Parametric	23	9.16%	0.19	0.9085	2
	Semi parametric	24	9.56%	0.07	0.9637	1
t-GARCH(1,1)	Parametric	17	6.77%	4.10	0.1289	
	Semi parametric	18	7.17%	3.12	0.2104	
t-GJR(1,1,1)	Parametric	21	8.37%	0.88	0.6445	4
	Semi parametric	21	8.37%	0.88	0.6445	4
t-TARCH(1,1,1)	Parametric	20	7.97%	1.97	0.3733	
	Semi parametric	21	8.37%	1.55	0.4617	10
t-EGARCH(1,1,1)	Parametric	21	8.37%	1.55	0.4617	10
	Semi parametric	22	8.76%	1.46	0.4820	8
t-EGARCH(1,2,1)	Parametric	22	8.76%	1.46	0.4820	8
	Semi parametric	23	9.16%	0.19	0.9085	2

Table 5.9 Out-of-sample hits and percentage hits for various VaR models at the 90% confidence level. There is a total of 251 days in the out-of-sample period. The Mincer-Zarnowitz regression asserts the quality of each specification, while the ranking is based on how close to zero the regression coefficients are, i.e. how close to 10% the out-of-sample hits percentages are.

The parametric VaR is calculated from the volatility estimated by each GARCH class model leading up to December 29, 2005 and the corresponding volatility forecasts for the next year. The last year is thus the out-of-sample period. At the 90% confidence level there will be 10% hits in the in-sample period. The question is how many hits occure in the out-of-sample period.

VaR models at the 95% and 99% confidence levels were tested, using parametric, semi-parametric and historical simulation. All these VaR models were by Mincer-Zarnowitz (MZ) in the out-of-sample period when comparing the forecasts of the VaR model with actual returns. Table 5.9 therefore reports the hits and percentage hits in the out-of-sample period at the 90% confidence level, together with the MZ regression and its p value. It shows that the MZ regression rejects that the historical VaR simulation is a good model for IYF, while the quality of the other models are sufficient enough not to be rejected. The closer the MZ statistic is to zero, which depends on how close the percentage hits in the out-of-sample period are to 10%, the better is the quality of the VaR model. The problem associated with using a model that understates the true VaR, is that the firm will keep less capital in reserve to cover unexpected losses. None of the forecasted VaR models have more than 10% hits in the out-of-sample period. A semi-parametric VaR model for the standardized residuals of an EGARCH(1,2,1) model gives the most precise out-of-sample forecasts of the value at risk. If IYF had been a firm concerned with risk management this model would thus be the best choice concerning the calculation of the amount of capital to hold back.

Finally, the 1-day 10% VaR for the first forecast in the out-of-sample period for a normal distributed and a t-distributed EGARCH(1,2,1) model are studied. A parametric VaR model is calculated as $VaR = \sigma\phi^{-1}(\alpha)$, where $\phi^{-1}(\alpha)$ is -1.28 for all the normal distributed models, while the value is different for the t-distributed models depending on the degrees of freedom parameter[4]. The t-distributed EGARCH(1,2,1) model for example, where the estimated ν=15.2, has a $\phi^{-1}(\alpha)$ approximately equal to -1.34. The one-day parametric VaR is equal to -0.9608% for the normal model. This can be interpreted as a 10% chance that the investment loses 0.96% or more over the next day. The corresponding VaR for a t-distributed model is -0.9661%. On a $1,000,000 portfolio invested in IYF on the last day of the estimation period, the 10% 1-day VaR is therefore $9608 and $9661 for the normal and t-distributed EGARCH(1,2,1) models respectively. The VaR is thus a bit higher for the t-

[4] $\phi^{-1}(\alpha)$ for a normal distribution is an inverse CDF with probability α, mean 0 and variance equal to 1. For a t-distribution it is an inverse CDF with probability α and ν degrees of freedom (Sheppard, 2009a).

distribution, which is not uncommon as it accounts for fatter tails in the distribution of returns.

For the semi-parametric VaR models the one-day 10% VaR is calculated as the 10% quantile of standardized residuals from the EGARCH(1,2,1) model. For the normal distributed model the 10% quantile of standardized residuals is equal to -1.2626, while it is -1.2648 for the t-distributed standardized residuals. This results in a 10% 1-day VaR of $9466 for the normal distributed EGARCH(1,2,1) and $9535 for the t-distributed EGARCH(1,2,1)[5]. This shows that the semi-parametric VaR model is better than the parametric VaR model, and that normal distribution is better than t-distribution for VaR purposes.

5.3 iShares Dow Jones US Utilities sector (IDU)

The IDU fund tracks the utilities stocks in the U.S. equity market. The same steps as for IYF are in the following applied on IDU.

5.3.1 Pre estimation analysis

The raw returns are not normal, as can be seen from the skewness, kurtosis and Jarque-Bera in Table 5.1. There is an excess kurtosis of about 10 and the returns are negatively skewed. The Box-Pierce-Ljung Q-test statistics for IDU are shown in Table 5.10. For the raw returns there are no significant serial correlations at the 0.05 level up to 20 lags. The squared returns, however, have as expected significant serial correlation up to at least 20 lags, implying that volatility clustering is also present in the IDU returns series.

Lags	Raw returns		Squared returns		Critical value
	Q-stat	*P value*	*Q-stat*	*P value*	
10	14.67	0.1446	767.78	0.0000	18.31
15	17.24	0.3047	864.72	0.0000	25.00
20	29.07	0.0864	946.60	0.0000	31.41

Table 5.10 Box-Pierce-Ljung's Q-test of no serial correlation for IDU.

Engle's ARCH test on the residual returns for IDU shows that there is homoskedasticity at the 0.05 level at the first lag of the squared innovations. This could imply a problem as the

[5] $VaR_{t+1,\alpha} = h_{t+1} * \alpha\%$ quantile (Sheppard, 2009a).

variance cannot be predicted from homoskedastic data and the variations in the squared residuals will be purely random. This is shown in the left pane of Figure 5.16. Removing the most extreme outliers solves this issue. The returns plot for IDU in Figure 5.2 shows some extreme outliers in the first year of the estimation period that should not influence the current and future level of volatility. All outliers with an absolute value above 7.5% were therefore simply dummied out. These are given in Table 5.11.

Date	**Return**
Jul 23, 2002	-8.20%
Jul 24, 2002	7.73%
Oct 9, 2002	-8.17%

Table 5.11 Removed outliers (IDU).

The p-values from the ARCH test on the returns series, after dummying out the outliers, are given in the right pane of Figure 5.16. Homoskedasticity in the residuals is now strongly rejected.

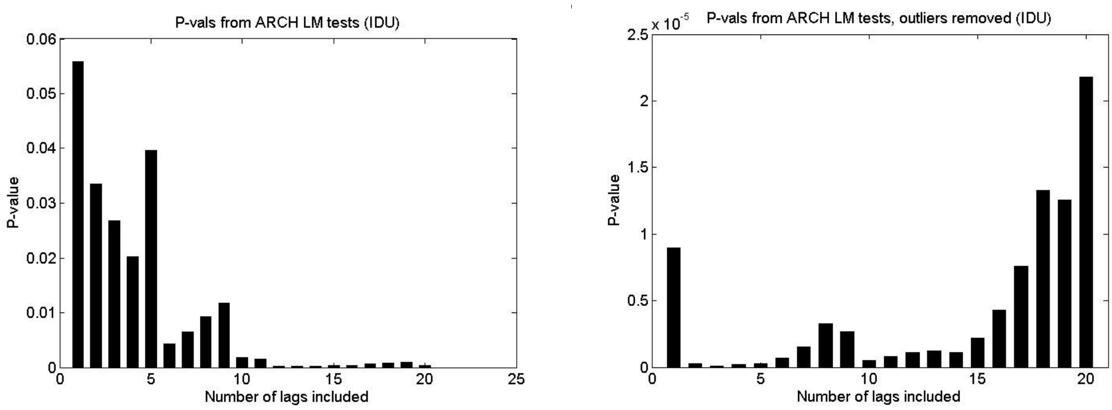


Figure 5.16 P-values from the ARCH LM tests before (left) and after (right) removing outliers.

The Q-test on the altered return series still shows significant correlation in the squared returns.

5.3.2 Model parameterization and selection

A similar analysis as for the IYF index was done for IDU. The estimated parameters after the three outliers have been removed, is given in Table 5.12. The initial analysis shows that all models are covariance stationary. Looking at the log likelihoods and the significance of the parameters at the 0.10 significance level, GJR(1,1,2) turns out to meet the requirements best, followed by GJR(1,1,1), TARCH(1,1,2) and EGARCH(1,1,2). In contrast to the parameterization for IYF, the symmetric term of IDU is significant at the 0.1 level for all these models. At the 0.05 significance level TARCH(1,1,1) is the best model for IDU, followed by GARCH(1,1) and EGARCH(1,1,1).

For comparision, Table 5.13 shows the estimated parameters for the best models including the three extreme outliers. GJR(1,1,2) and TARCH(1,1,2) are no longer significant at the 0.10 level, so the GJR(1,1,1) seems to give the best fit, followed by EGARCH(1,1,2), TARCH(1,1,1) and EGARCH(1,1,1). The sum of the likelihoods are higher for all models when the three outliers are dummied out, which is natural as the likelihoods on these dates become higher.

	ω (x 10-3)	**α1**	**α2**	**γ1**	**γ2**	**β1**	**β2**	**LL**
GARCH(1,1)	0.021	0.100				0.882		-1381.5
	(.043)	*(.000)*				*(.000)*		
GARCH(2,1)	0.021	0.100	0.000			0.882		-1381.51
	(.166)	*(.010)*	*(.999)*			*(.000)*		
GARCH(1,2)	0.026	0.137				0.360	0.481	-1380.05
	(.049)	*(.000)*				*(.180)*	*(.060)*	
GJR(1,1,1)	0.026	0.057		0.074		0.881		-1377.44
	(.035)	*(.063)*		*(.030)*		*(.000)*		
GJR(2,1,1)	0.026	0.057	0.000	0.074		0.881		-1377.44
	(.393)	*(.623)*	*(.999)*	*(.089)*		*(.000)*		
GJR(1,2,1)	0.026	0.057		0.074	0.000	0.881		-1377.44
	(.030)	*(.061)*		*(.091)*	*(.999)*	*(.000)*		
GJR(1,1,2)	0.031	0.075		0.104		0.372	0.471	-1375.7
	(.032)	*(.052)*		*(.029)*		*(.072)*	*(.018)*	
TARCH(1,1,1)	0.022	0.058		0.057		0.910		-1380.87
	(.035)	*(.041)*		*(.000)*		*(.000)*		
TARCH(2,1,1)	0.022	0.058	0.000	0.057		0.910		-1380.87
	(.432)	*(.757)*	*(.999)*	*(.030)*		*(.000)*		
TARCH(1,2,1)	0.022	0.058		0.057	0.000	0.910		-1380.87
	(.043)	*(.048)*		*(.133)*	*(.999)*	*(.000)*		
TARCH(1,1,2)	0.028	0.085		0.079		0.377	0.497	-1379.12
	(.029)	*(.044)*		*(.019)*		*(.079)*	*(.014)*	
EGARCH(1,1,1)	0.001	0.166		-0.046		0.981		-1382.01
	(.740)	*(.001)*		*(.025)*		*(.000)*		
EGARCH(2,1,1)	0.001	0.314	-0.183	-0.049		0.986		-1378.96
	(.711)	*(.000)*	*(.038)*	*(.010)*		*(.000)*		
EGARCH(1,2,1)	0.001	0.164		-0.076	0.033	0.982		-1381.71
	(.738)	*(.002)*		*(.147)*	*(.516)*	*(.000)*		
EGARCH(1,1,2)	0.002	0.244		-0.067		0.451	0.523	-1379.75
	(.745)	*(.000)*		*(.029)*		*(.009)*	*(0.002)*	
EGARCH(2,2,1)	0.001	0.309	-0.178	-0.065	0.017	0.987		-1378.88
	(.709)	*(.001)*	*(.045)*	*(.262)*	*(.755)*	*(.000)*		
EGARCH(2,2,2)	0.001	0.309	-0.146	-0.055	-0.004	0.760	0.223	-1378.73
	(.738)	*(.002)*	*(.533)*	*(.702)*	*(.985)*	*(.492)*	*(.838)*	

Table 5.12 Parameterization of conditional volatility models for IDU under the assumption of a normal distribution. Three extreme outliers have been removed from the time series.

	ω (x 10-3)	α1	α2	γ1	γ2	β1	β2	LL
GARCH(1,1)	0.031	0.129				0.845		-1392.90
	(.028)	*(.000)*				*(.000)*		
GJR(1,1,1)	0.039	0.064		0.107		0.844		-1386.76
	(.008)	*(.058)*		*(.006)*		*(.000)*		
GJR(1,1,2)	0.049	0.083		0.146		0.304	0.493	-1384.37
	(.009)	*(.056)*		*(.009)*		*(.184)*	*(.030)*	
TARCH(1,1,1)	0.034	0.066		0.081		0.883		-1389.87
	(.012)	*(.044)*		*(.000)*		*(.000)*		
TARCH(1,1,2)	0.044	0.097		0.111		0.273	0.564	-1386.84
	(.006)	*(.030)*		*(.003)*		*(.126)*	*(.001)*	
EGARCH(1,1,1)	0.002	0.205		-0.066		0.971		-1391.18
	(.737)	*(.001)*		*(.004)*		*(.000)*		
EGARCH(1,1,2)	0.002	0.296		-0.093		0.377	0.584	-1387.27
	(.800)	*(.000)*		*(.006)*		*(.009)*	*(.000)*	

Table 5.13 Parameterization of conditional volatility models for IDU without removing outliers.

The AIC and BIC criterias (see Figure 5.17) support the choice of GJR(p,o,q) models. GJR(1,1,1) is selected by AIC and does well for the BIC. It is also interesting to see that the BIC criteria selects GARCH(1,1) because it is the most parsimonious model. In comparision to IYF, the EGARCH models does not perform well for IDU according to these selection criterias. An APARCH(1,1,1) model has also been estimated, although it is not presented in the tables above as it due to it's complexity will not be used in the forecasting evaluation (see Chapter 5.3.5). It has a higher likelihood than all the listed models, with significant parameters at the 0.05 significance level. As can be seen from the AIC selection criteria, it is also ranked as the third best model. The estimated delta of the APARCH(1,1,1) is however not significantly different from two, which means it is actually just nesting a GJR(1,1,1).

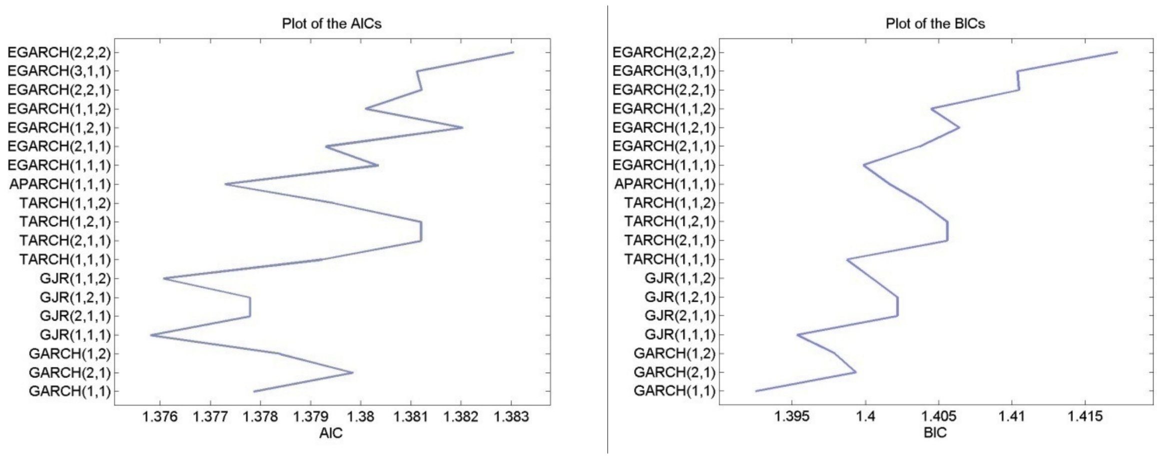


Figure 5.17 Plots of AIC (left) and BIC (right) for IDU for normal distributed errors.

5.3.3 Post estimation analysis

Table 5.14 reports the skewness, kurtosis and Jarque-Bera statistic for the raw returns (after the removal of the extreme outliers) in the estimation period and the standardized residuals from various GARCH class models. The Jarque-Bera statistic rejects normality for both the raw returns and for all the volatility models. It is, however, no doubt that the standardized residuals are closer to normal than the raw returns. There is a noticable left skewness for all the residuals, while the excess kurtosis exceeds 0.5.

The density- and the QQ-plots for the standardized residuals from the GJR(1,1,1) model in Figure 5.18 also indicate that an alternative distribution like Bollerslev's Student's T or Hansen's Skewed T might be desirable. The first will be examined in the following section.

	Skew	Kurt	JB	P value
Raw returns	-0.004	8.135	1104.36	0.0000
Standardized residuals				
GARCH(1,1)	-0.182	3.669	24.281	0.0000
GJR(1,1,1)	-0.121	3.663	20.861	0.0000
GJR(1,1,2)	-0.104	3.627	18.297	0.0001
TARCH(1,1,1)	-0.115	3.762	26.549	0.0000
TARCH(1,1,2)	-0.109	3.725	23.989	0.0000
EGARCH(1,1,1)	-0.120	3.771	27.308	0.0000
EGARCH(1,1,2)	-0.109	3.723	23.850	0.0000
EGARCH(1,2,1)	-0.106	3.783	27.521	0.0000

Table 5.14 Testing for normality in standardized residuals (IDU).

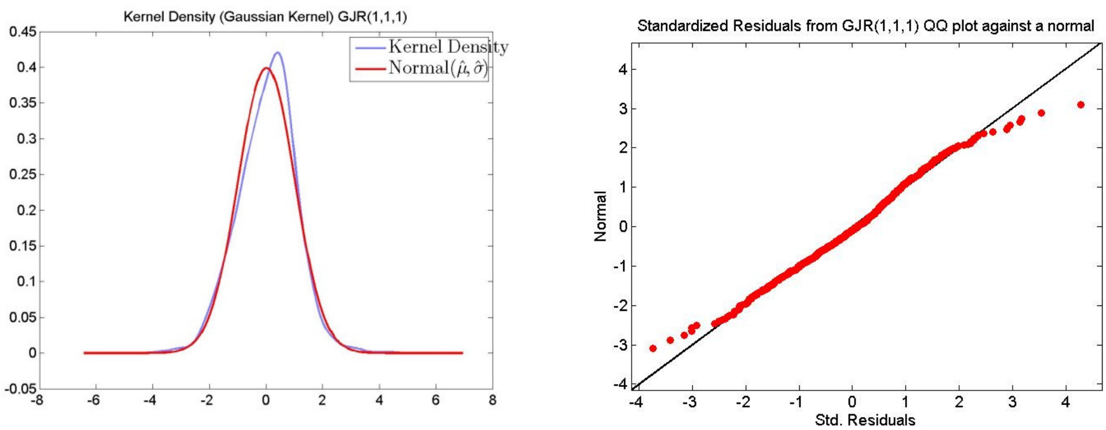


Figure 5.18 Kernel density plot (left) and QQ-plot (right) of the standardized residuals from an GJR(1,1,1) model for IDU against a normal distribution.

Engle's ARCH test on the standardized residuals rejects remaining ARCH effects for all models analyzed, and there is also no remaining significant autocorrelation for any of the models.

5.3.4 Student's T distribution

The student's t distribution parameterizations for IDU are given in Table 5.15. The shape parameter, ν, is significant for all models. The ranking of the various models is not much different than for the normal distribution. GJR(1,1,2) is the best model at the 0.1 significance level, ahead of GJR(1,1,1). The latter is now also significant at the 0.05 level. Specifications with two variance lags perform well.

	ω (x 10-3)	α1	α2	γ1	γ2	β1	β2	ν	LL
t-GARCH(1,1)	0.020	0.096				0.885		12.4974	-1375.8
	(.033)	*(.000)*				*(.000)*		*(.002)*	
t-GARCH(2,1)	0.024	0.050	0.076			0.886		13.398	-1375.8
	(.019)	*(.019)*	*(.013)*			*(.000)*		*(.004)*	
t-GARCH(1,2)	0.026	0.132				0.391	0.453	13.038	-1374.9
	(.035)	*(.000)*				*(.173)*	*(.091)*	*(.003)*	
t-GJR(1,1,1)	0.024	0.050		0.076		0.886		13.398	-1372.2
	(.019)	*(.019)*		*(.013)*		*(.000)*		*(.004)*	
t-GJR(2,1,1)	0.024	0.050	0.000	0.076		0.886		13.398	-1372.2
	(.031)	*(.057)*	*(.999)*	*(.014)*		*(.000)*		*(.005)*	
t-GJR(1,2,1)	0.024	0.050		0.076	0.000	0.886		13.398	-1372.2
	(.023)	*(.022)*		*(.036)*	*(.999)*	*(.000)*		*(.004)*	
t-GJR(1,1,2)	0.031	0.068		0.109		0.383	0.461	14.116	-1371.0
	(.018)	*(.012)*		*(.013)*		*(.084)*	*(.027)*	*(.007)*	
t-TARCH(1,1,1)	0.023	0.056		0.061		0.907		12.596	-1374.6
	(.017)	*(.011)*		*(.006)*		*(.000)*		*(.002)*	
t-TARCH(2,1,1)	0.023	0.056	0.000	0.061		0.907		12.596	-1374.6
	(.039)	*(.337)*	*(.999)*	*(.006)*		*(.000)*		*(.003)*	
t-TARCH(1,2,1)	0.023	0.056		0.061	0.000	0.907		12.596	-1374.6
	(.137)	*(.057)*		*(.572)*	*(.999)*	*(.000)*		*(.003)*	
t-TARCH(1,1,2)	0.031	0.082		0.086		0.387	0.482	13.161	-1373.4
	(.015)	*(.017)*		*(.005)*		*(.105)*	*(.032)*	*(.004)*	
t-EGARCH(1,1,1)	0.001	0.168		-0.051		0.979		12.281	-1375.5
	(.769)	*(.000)*		*(.009)*		*(.000)*		*(.002)*	
t-EGARCH(2,1,1)	0.001	0.290	-0.144	-0.051		0.983		13.411	-1373.9
	(.767)	*(.000)*	*(.071)*	*(.008)*		*(.000)*		*(.003)*	
t-EGARCH(1,2,1)	0.001	0.166		-0.087	0.040	0.981		12.300	-1375.1
	(.769)	*(.000)*		*(.063)*	*(.382)*	*(.000)*		*(.002)*	
t-EGARCH(1,1,2)	0.001	0.245		-0.073		0.469	0.503	13.022	-1374.0
	(.785)	*(.000)*		*(.009)*		*(.015)*	*(.009)*	*(.004)*	
t-EGARCH(2,2,1)	0.001	0.282	-0.137	-0.081	0.033	0.984		13.327	-1373.7
	(.765)	*(.000)*	*(.085)*	*(.105)*	*(.506)*	*(.000)*		*(.004)*	
t-EGARCH(2,2,2)	0.001	0.287	-0.099	-0.069	0.004	0.719	0.259	13.446	-1373.6
	(.783)	*(.000)*	*(.554)*	*(.271)*	*(.964)*	*(.396)*	*(.755)*	*(.004)*	

Table 5.15 Parameterization of conditional volatility models for IDU under the assumption of a t-distribution.

The plots for the AIC and BIC selection criterions can be seen in Figure 5.19. The curves are almost identical to the normal distribution. GJR-GARCH models seem to give the best fit of the returns series for IDU.

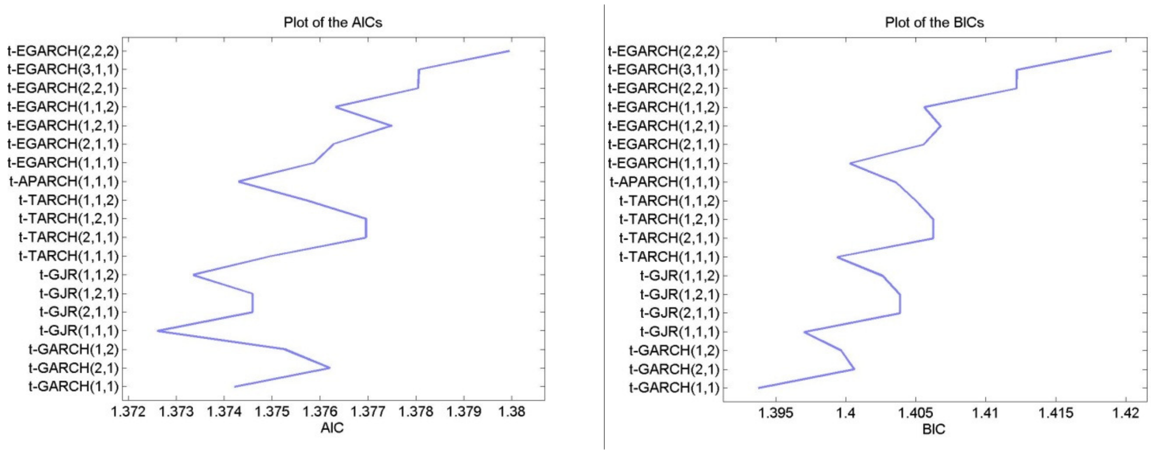


Figure 5.19 Plots of AIC (left) and BIC (right) for IDU for t-distributed errors.

The kernel density and QQ plots for the GJR(1,1,1) standardized residuals against a t(13.4) distribution, where 13.4 corresponds to the degrees of freedom parameter estimated, are given in Figure 5.18. It is obvious that the GARCH-filtered residuals have a higher kurtosis and a left skew compared to a t(13.4) distribution.

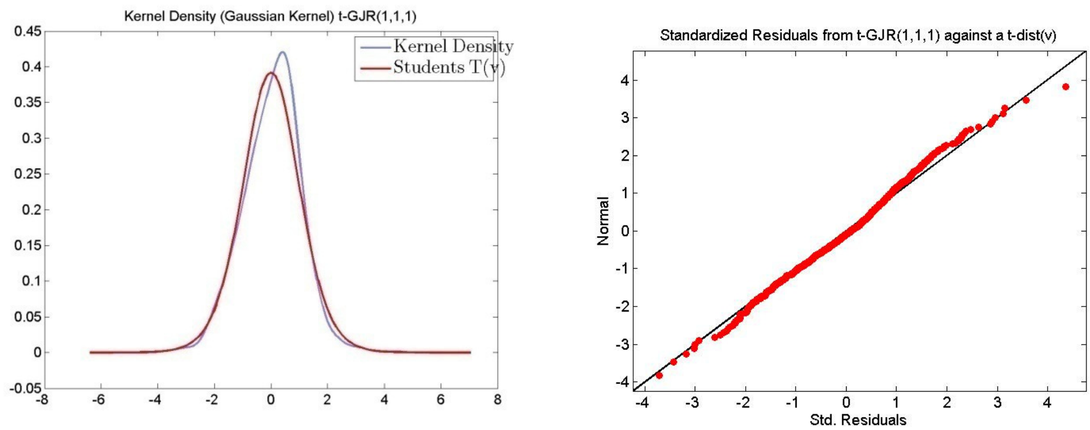


Figure 5.20 Kernel density plot (left) and QQ-plot (right) of the standardized residuals from a t-distributed GJR(1,1,1) model for IDU against a t(13.4) distribution.

Since neither the t-distribution nor the normal distribution is completely satisfying, it is difficult to conclude which assumption is best. The best way to analyze this is the direct comparison of the fits of corresponding models between the two distributions. As for IYF,

Table 5.16 shows that the likelihoods are improved under the assumption of t-distribution for all models.

	Skew	Kurt	LL	AIC	BIC
Normal					
GARCH(1,1)	-0.182	3.669	-1381.5	1.3779	1.3925
GJR(1,1,1)	-0.121	3.663	-1377.4	1.3758	1.3953
GJR(1,1,2)	-0.104	3.627	-1375.7	1.3761	1.4005
TARCH(1,1,1)	-0.115	3.762	-1380.9	1.3792	1.3987
TARCH(1,1,2)	-0.109	3.725	-1379.1	1.3795	1.4039
EGARCH(1,1,1)	-0.120	3.771	-1382.0	1.3804	1.3999
EGARCH(1,1,2)	-0.109	3.723	-1379.8	1.3801	1.4045
EGARCH(1,2,1)	-0.106	3.783	-1381.7	1.3820	1.4064
Students t					
GARCH(1,1)	-0.183	3.671	-1375.8	1.3742	1.3937
GJR(1,1,1)	-0.118	3.676	-1372.2	1.3726	1.3970
GJR(1,1,2)	-0.099	3.638	-1371.0	1.3734	1.4026
TARCH(1,1,1)	-0.109	3.770	-1374.6	1.3750	1.3994
TARCH(1,1,2)	-0.101	3.738	-1373.4	1.3757	1.4050
EGARCH(1,1,1)	-0.114	3.779	-1375.5	1.3759	1.4003
EGARCH(1,1,2)	-0.100	3.735	-1374.0	1.3763	1.4056
EGARCH(1,2,1)	-0.096	3.796	-1375.1	1.3775	1.4068

Table 5.16 Descriptive statistics for normal and student's t distributed models (IDU).

The t-distributed GJR (1,1,2) is the best model based on the log likelihood, but t-GJR(1,1,1) is a strong alternative since it has more significant parameters and is chosen ahead of t-GJR(1,1,2) according to AIC and BIC.

5.3.5 Forecast evaluation

The Mincer-Zarnowitz regression rejects optimality for all estimated models for the IDU time-series at the 0.05 significance level (see Table 5.17). The GJR-models are closest to being optimal. The fact that optimality is rejected does not mean that the forecasts have no value, rather that it does not satisfy the condition of optimality. It could be caused by an unstable data generation process, and might be improved if the first year of data was omitted, as this is the most volatile period.

	Normal		T-distribution	
	MZ	*P value*	*MZ*	*P value*
GARCH(1,1)	8.181	0.0167	7.989	0.0184
GJR(1,1,1)	8.042	0.0179	7.770	0.0206
GJR(1,1,2)	7.775	0.0205	7.746	0.0208
TARCH(1,1,1)	12.598	0.0018	12.979	0.0015
TARCH(1,1,2)	12.611	0.0018	13.222	0.0013
EGARCH(1,1,1)	11.434	0.0033	11.919	0.0026
EGARCH(1,1,2)	12.659	0.0018	13.152	0.0014

Table 5.17 Mincer-Zarnowitz R^2 for normal and t-distributed squared residuals as proxy for volatility for IDU.

Table 5.18 ranks the models based on statistical loss functions and on how far from optimal the R^2 is. T-distributed models dominate normal distributed models, as the top three ranked models regarding forecasting performance are t-distributed. The GJR models perform best for the IDU time series. The GARCH(1,1) models also perform well for the volatility forecasts both according to the R^2 and the statistical loss functions.

	ME	MSE	MAE	MAPE	R^2	Total	Total rank
GARCH(1,1)	2	4	2	14	6	28	**4**
GJR(1,1,1)	5	2	6	12	5	30	**6**
GJR(1,1,2)	5	6	3	11	3	28	**4**
TARCH(1,1,1)	13	9	11	6	9	48	**10**
TARCH(1,1,2)	10	11	12	5	10	48	**10**
EGARCH(1,1,1)	8	7	7	1	7	30	**6**
EGARCH(1,1,2)	7	12	9	3	11	42	**9**
t-GARCH(1,1)	3	3	1	13	4	24	**3**
t-GJR(1,1,1)	4	1	4	9	2	20	**1**
t-GJR(1,1,2)	1	5	5	10	1	22	**2**
t-TARCH(1,1,1)	14	10	13	8	12	57	**13**
t-TARCH(1,1,2)	12	14	14	7	14	61	**14**
t-EGARCH(1,1,1)	11	8	8	2	8	37	**8**
t-EGARCH(1,1,2)	9	13	10	4	13	49	**12**

Table 5.18 The forecasting performance ranked by statistical loss functions and the Mincer-Zarnowitz R^2.

Since the Mincer-Zarnowitz R^2 rejects optimality, it comes as no surprise that the hits-percentages at the 90% confidence level for VaR for all models are far from 10%. A parametric VaR for the normal distributed GARCH(1,1) model has a hit-percentage in the out-of-sample period of 7.2% versus 6.77% for almost all the other specifications. It is therefore difficult to draw conclusions from the VaR analysis of IDU.

5.4 iShares Dow Jones US Energy sector (IYE)

The IYE fund tracks U.S. energy stocks as represented by the Dow Jones U.S. Oil & Gas index. The same steps as for IYF and IDU are applied in the following sections.

5.4.1 Pre estimation analysis

The squared returns of IYE have significant serial correlation up to at least 20 lags, implying that volatility clustering is also present here. But as for IDU, Engle's ARCH test on the residual returns cannot reject homoskedasticity for the first lag of squared innovations at the 0.05 significance level. The most extreme outliers are removed to work around this issue. The returns plot for IYE in Figure 5.2 shows three outliers in excess of 6% (absolute value) in the first year of the estimation period. Although these returns are by no means extreme, they should not influence the main task of forecasting volatility after 2006. The three outliers that were dummied out for IYE are given in Table 5.19.

Date	Return
July 19, 2002	-6.97%
July 24, 2002	6.88%
August 1, 2002	-6.64%

Table 5.19 Removed outliers (IYE).

The p-values from the ARCH LM test after removing these outliers are now low enough to reject homoskedasticity. The Q-test on the altered return series still shows significant correlation in the squared returns.

5.4.2 Model parameterization and selection

Table 5.20 shows the parameterization of conditional volatility models for IYE under the assumption of a normal distribution. Only significant models at the 0.1 level are presented here. As can be seen from Table 5.1, the evidence of asymmetry was weaker for IYE than for the two other assets. This fact is also somehow confirmed by the parameterizations below, as the GARCH(1,1) model has the second highest likelihood. The estimated asymmetry coefficients are however significant for all the presented asymmetry models. The GJR(1,1,1) model gives the best fit of the data.

	ω (x 10-3)	α1	γ1	β1	LL
GARCH(1,1)	0.019	0.052		0.938	-1693.17
	(.131)	*(.000)*		*(.000)*	
GJR(1,1,1)	0.029	0.028	0.042	0.934	-1691.268
	(.103)	*(.054)*	*(.095)*	*(.000)*	
TARCH(1,1,1)	0.023	0.035	0.037	0.940	-1693.926
	(.064)	*(.006)*	*(.000)*	*(.000)*	
EGARCH(1,1,1)	0.008	0.102	-0.031	0.984	-1693.487
	(.070)	*(.000)*	*(.075)*	*(.000)*	

Table 5.20 Parameterization of conditional volatility models for IYE under the assumption of a normal distribution.

Figure 5.21 plots the AICs and BICs for IYE. The AIC selection criteria selects GARCH(1,1) as the best model, narrowly ahead of GJR(1,1,1). This is also the ranking of the BIC criteria, where the margin between them is larger.

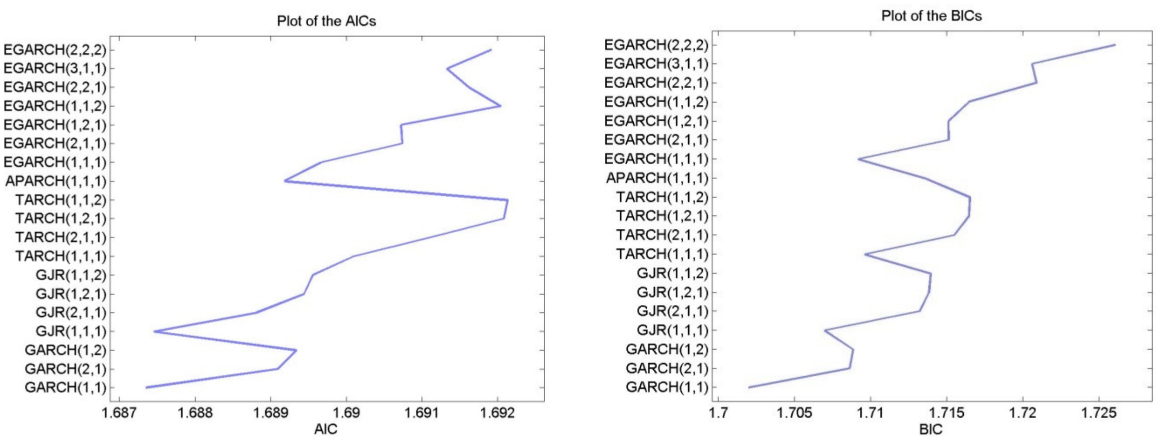


Figure 5.21 Plots of AIC (left) and BIC (right) for IYE under the assumption of normal distributed errors.

Table 5.21 shows the parameterization for t-distributed errors. GJR(1,1,1) is still the strongest, now followed by a GARCH(2,1) model. AIC and BIC, however, maintain the same ranking as for the normal distributed parameterizations, including that GARCH(1,1) is ranked ahead of GJR(1,1,1).

Model	ω (x 10-3)	α1	α2	γ1	γ2	β1	β2	ν	LL
t-GARCH(1,1)	0.010	0.050				0.930		101.4178	-1693.1
	(.131)	*(.000)*				*(.000)*		*(.000)*	
t-GARCH(2,1)	0.020	0.020	0.040			0.930		194.826	-1692.9
	(.098)	*(.046)*	*(.094)*			*(.000)*		*(.000)*	
t-GJR(1,1,1)	0.020	0.020		0.040		0.930		194.826	-1691.3
	(.098)	*(.046)*		*(.094)*		*(.000)*		*(.000)*	
t-TARCH(1,1,1)	0.020	0.030		0.030		0.930		133.701	-1693.9
	(.063)	*(.006)*		*(.072)*		*(.000)*		*(.001)*	
t-EGARCH(1,1,1)	0.000	0.100		-0.030		0.980		141.260	-1693.5
	(.077)	*(.000)*		*(.078)*		*(.000)*		*(.005)*	

Table 5.21 Parameterization of conditional volatility models for IYE under the assumption of a t-distribution.

5.4.3 Post estimation analysis

Table 5.22 shows that the raw returns in the estimation period are far from normal (but closer than the raw returns of IYF and IDU). The Jarque-Bera statistic also rejects normality for the standardized residuals from GARCH(1,1) and GJR(1,1,1) under the assumption of normal distribution, although the latter is very close to being significant at the 0.05 level. The standardized residuals of TARCH(1,1,1) and EGARCH(1,1,1) are normal distributed.

	Skew	Kurt	LL	AIC	BIC	JB	P value
Raw returns	-0.2277	3.7229				30.5624	0
Normal							
GARCH(1,1)	-0.2052	3.0594	-1693.17	1.6874	1.702	7.2021	0.0273
GJR(1,1,1)	-0.1889	3.0254	-1691.27	1.6875	1.707	6.0011	0.0498
TARCH(1,1,1)	-0.1724	3.0408	-1693.93	1.6901	1.7096	5.0489	0.0801
EGARCH(1,1,1)	-0.1693	3.0374	-1693.49	1.6897	1.7092	4.8595	0.0881
Student's T							
t-GARCH(1,1)	-0.2054	3.0595	-1693.1	1.6893	1.7088		
t-GARCH(2,1)	-0.2059	3.0483	-1692.87	1.691	1.7154		
t-GJR(1,1,1)	-0.1889	3.0256	-1691.25	1.6894	1.7138		
t-TARCH(1,1,1)	-0.1725	3.0412	-1693.89	1.692	1.7164		
t-EGARCH(1,1,1)	-0.1632	3.0522	-1693.83	1.694	1.7233		

Table 5.22 Postestimation statistics for IYE.

In contrast to the IYF and IDU parameterization there is no evidence that a t-distribution gives a better fit to the data than a normal distribution. The likelihoods are similar, and in most cases slighly lower for the t-distributed models. A normal and a t-distributed GJR(1,1,1) have about the same likelihood, so the conclusion is that a GJR(1,1,1) model under any of the two distributions is recommended for IYE.

5.4.4 Forecast evaluation

The Mincer-Zarnowitz regression fails to reject optimality for all estimated models for the IYE time-series at the 0.05 significance level. This is shown in Table 5.23.

	Normal		T-distribution	
	MZ	*P value*	*MZ*	*P value*
GARCH(1,1)	2.2479	0.3250	2.2521	0.3243
GJR(1,1,1)	1.5987	0.4496	1.6088	0.4473
TARCH(1,1,1)	0.8871	0.6418	0.9014	0.6372
EGARCH(1,1,1)	0.7669	0.6815	0.7786	0.6775

Table 5.23 Mincer-Zarnowitz R^2 for normal and t-distributed squared residuals as proxy for volatility for IYE.

Table 5.24 ranks the models based on statistical loss functions and on how far from optimal the R^2 is. Suddenly the TARCH(1,1,1) models are to prefer, followed by the normal distributed GJR(1,1,1). The GARCH(1,1) models, which gives a very good fit in the estimation period, fails to keep up the performance in the out-of-sample period.

	ME	MSE	MAE	MAPE	R^2	Total	Total rank
GARCH(1,1)	8	8	8	8	7	39	**8**
GJR(1,1,1)	2	5	3	5	5	20	**3**
TARCH(1,1,1)	4	1	1	2	2	10	**2**
EGARCH(1,1,1)	6	3	5	3	4	21	**4**
t-GARCH(1,1)	7	7	7	7	8	36	**7**
t-GJR(1,1,1)	1	6	3	6	6	22	**6**
t-TARCH(1,1,1)	3	2	1	1	1	8	**1**
t-EGARCH(1,1,1)	5	4	5	4	3	21	**4**

Table 5.24 The forecasting performance ranked by statistical loss functions and the R^2 for IYE.

VaR modeling using parametric and semi-parametric VaR models at the 90% confidence level for the normal distributed GJR(1,1,1) model is shown in Figure 5.22, together with a historical simulation of the returns data. In the out-of-sample period the historical simulation has a hit-percentage of 9.6%, compared to 10.36% for the semi-parametric and 10.76% for the parametric models for GJR(1,1,1). The last two thus overhit at the 90% confidence level. According to the hits-percentages the semi-parametric GJR(1,1,1) models (normal and t) and the GARCH(1,1) models (normal and t) perform equally well, ahead of all other specifications. This confirms the presumption that the GJR(1,1,1) specification is the best alternative for IYE for the chosen time horizon. Since it is better than the t-distributed GJR(1,1,1) for forecasting purposes, the normal distributed GJR(1,1,1) model is the recommended model for IYE.

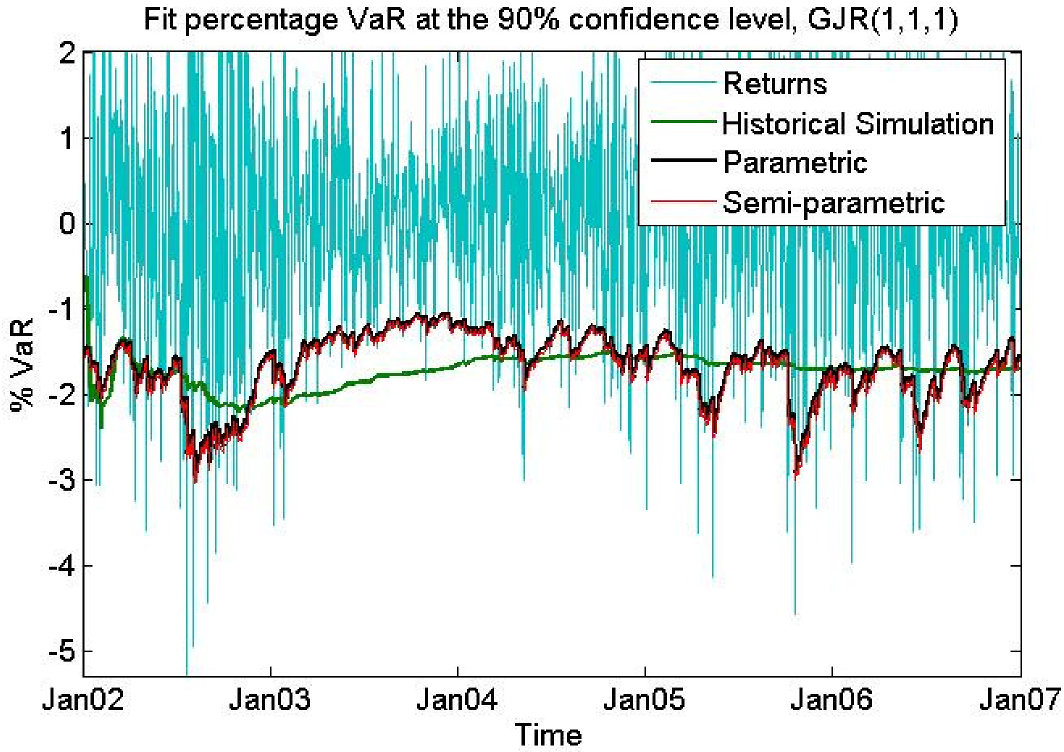


Figure 5.22 Estimated % VaR based on a normal distributed GJR(1,1,1) model, using historical simulation, parametric and semi-parametric VaR models for α=10%.

5.5 Multivariate models

The previous sections show evidence for time-varying volatility for individual time-series. The probability of time-varying correlations (and thus covariances) between each pair of assets is therefore large. This should lead to re-allocations of the total portfolio assets based on the time-varying correlations. Cappiello et al. (2006) use an asymmetric version of the DCC model to explore the dynamics and changes in correlation of asset markets to investigate whether the correlation demonstrate evidence of asymmetric response to negative returns. For national equity index returns series they find asymmetry both in the conditional volatility and in conditional correlations. An asymmetric DCC model will also be used in this paper to construct a time-varying portfolio of the three ETF's studied. Although different volatility models have been recommended for each individual ETF, the asymmetric models have performed well. For IYF the suggestion a t-distributed EGARCH(1,2,1) model, for IDU a t-distributed GJR(1,1,2) model and for IYE a normal distributed GJR(1,1,1) model. The multivariate DCC parameterization for the construction of a portfolio of the three assets will be based on normal distributed GJR(1,1,1) models for all three assets. The reason for this choice is not only that GJR(1,1,1) models have performed well for the three assets (ranked highest of the asymmetric models by the BIC criterion for all normally distributed models), but also the fact that the current version of the MFE Oxford Toolbox for Matlab builds on GJR(p,o,q) specifications for the univariate time-series. DCC models are based on the assumption of normal distributed standardized residuals.

Looking at the unconditional correlation matrix and the unconditional variance-covariance-matrix between the three ETF's in Table 5.25, shows that the assets are reasonably correlated. The correlation is highest between financials and utilities, while the correlation is, as expected, a bit lower (but positive) between energy and utilities.

	IYF	**IYE**	**IDU**
IYF	1.0000	0.4951	0.5709
IYE		1.0000	0.4844
IDU			1.0000

Table 5.25 Unconditional correlation of returns of the three ETF's.

	IYF	**IYE**	**IDU**
IYF	1.4301	0.8141	0.7659
IYE		1.8908	0.7472
IDU			1.2584

Table 5.26 Unconditional Variance-Covariance (VCV) matrix (scaled by 10,000) of the returns.

Three time-varying global minimum variance portfolios are created for the three-asset problem: a DCC(1,1) model, a BEKK model and an EWMA model with a decay factor of 0.94, i.e. RiskMetrics' decay factor. The three portfolios will be ranked on how well they minimize the variance.

A DCC(M,N) model represents a dynamic conditional correlation model with M (positive scalar) innovation lags and N (non-negative scalar) variance lags in the correlation model. A DCC(1,1) model for three univariate GJR-GARCH(1,1,1) processes will thus require the estimation of 17 parameters (ω_i, α_i, γ_i and β_i for i=1,2,3 parameters from the three univariate processes and $\theta_1,...,\theta_5$ correlation parameters). The first step of the DCC(1,1) model is to parameterize the univariate series. Table 5.27 shows the DCC parameterizations, which are equal to the estimated univariate GJR(1,1,1) models of each asset in previous sections of Chapter 5.

	ω (x 10^{-3})	α	γ	β
IYF	0.0107	0.0004	0.1055	0.9358
IYE	0.0288	0.0284	0.0423	0.9338
IDU	0.0256	0.0566	0.0743	0.8814

Table 5.27 DCC parameterizations of the univariate GJR(1,1,1) models for normal distributed errors.

The parameterization of the second step of the DCC process is given in Table 5.28. The alpha and beta parameters correspond to θ_1 and θ_2 in Equation 3.41 respectively, while the correlation parameters γ_{12} (IYF-IYE), γ_{13} (IYF-IDU) and γ_{23} (IYE-IDU) represent θ_3 .

	γ_{12}	γ_{13}	γ_{23}	α	β	LL
DCC(1,1)	0.4341	0.5799	0.4888	0.0103	0.9895	-4092.7

Table 5.28 Parameterization of the DCC(1,1) model, step 2. LL is the log likelihood of the DCC(1,1) model.

The BEKK[6] parameterization is done in one step, and for the three-asset model this involves the estimation of 24 parameters. It needs much more processor time than the estimation of a DCC model, which would make it unsuitable for higher systems.

Figure 5.23 shows the time-varying annualized standard deviations given by the three multivariate models over the estimation period. As expected, there is little deviation between the models when it comes to volatility. By looking at the time-varying correlations in Figure 5.24, however, it can be seen that while the correlation changes much over time for both the RiskMetrics and BEKK models, there are no meaningful changes in the correlation for the DCC model, partly due to the use of standardized residuals.

[6] BEKK modeling in Matlab is done with Kevin Sheppard's UCSD GARCH toolbox as it is not implemented in the MFE toolbox. The UCSD toolbox can be found at http://www.kevinsheppard.com/wiki/UCSD_GARCH (Accessed February 12, 2009).

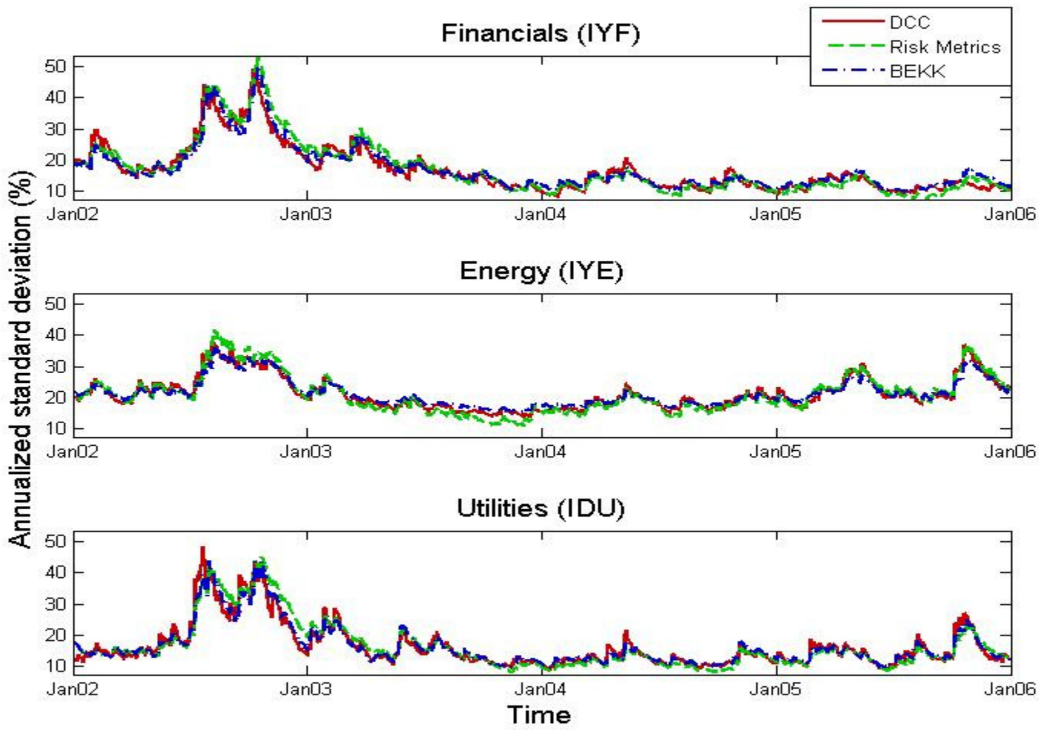


Figure 5.23 Development in annualized standard deviations for the three assets given by the DCC(1,1), RiskMetrics and BEKK estimations.

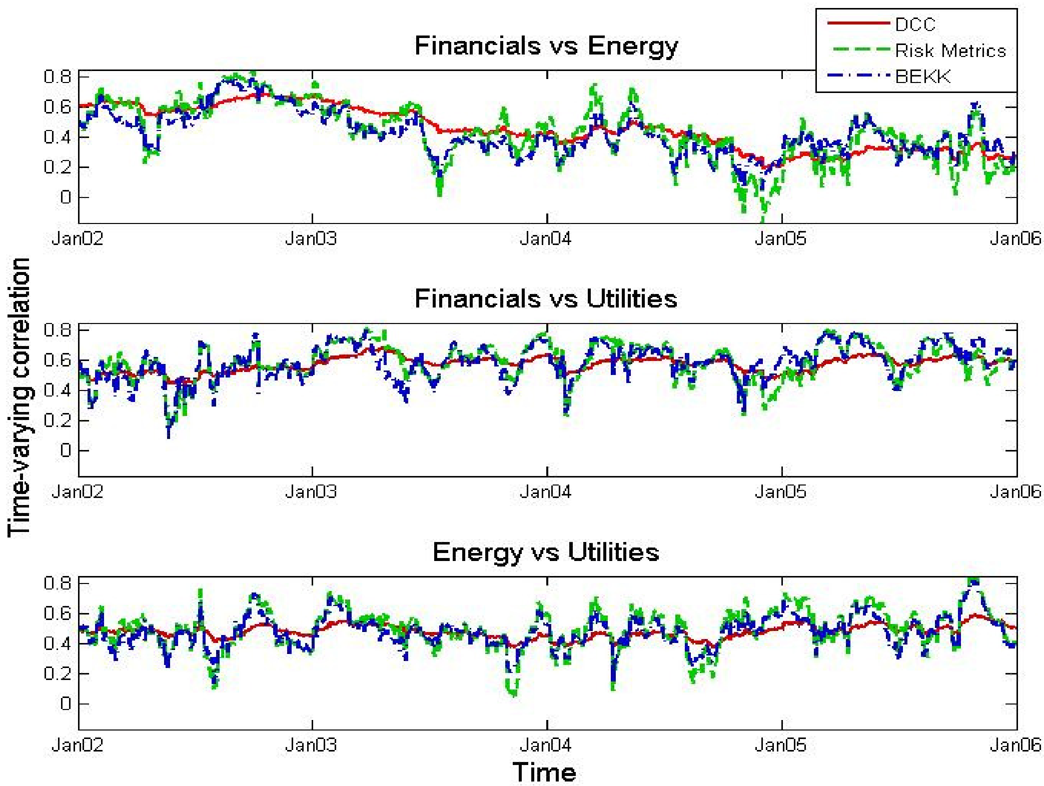


Figure 5.24 Time-varying correlations between each pair of assets given by the DCC(1,1), BEKK and RiskMetrics models.

Based on the daily estimates of the correlation matrix H_t from the three multivariate models, daily optimal weights on each asset are created in order to minimize the variance. In other words, three time-varying global minimum variance portfolios are constructed. Figure 5.25 shows the development in the annualized standard deviation of the DCC model compared to the annualized standard deviation of the three individual assets. The figure shows that while the minimum variance portfolio from the DCC model has volatility in excess of 30% in the third quarter of 2002, it is possible to construct a portfolio of these three assets at around 10% volatility in December 2005.

The time-varying weights of the three portfolios are shown in Figure 5.26. The weights changes considerably during the estimation period. There is a strong relationship between the volatility of the individual assets and the weights. An example of this can be seen by looking at the weights on IYF at the end of the estimation period. The weight is high compared to the other assets because the volatility in IYF was much lower than for IYE and IDU at the end of 2005.

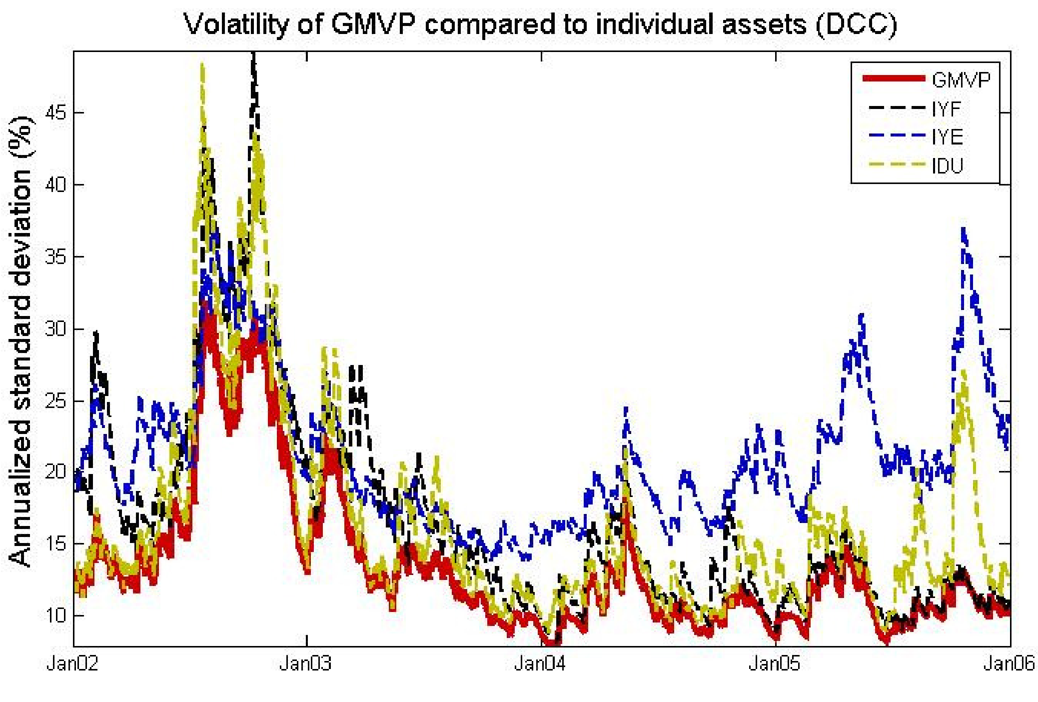


Figure 5.25 Volatility of the global minimum variance portfolio from DCC(1,1) compared to the three individual assets given by univariate GJR(1,1,1) processes.

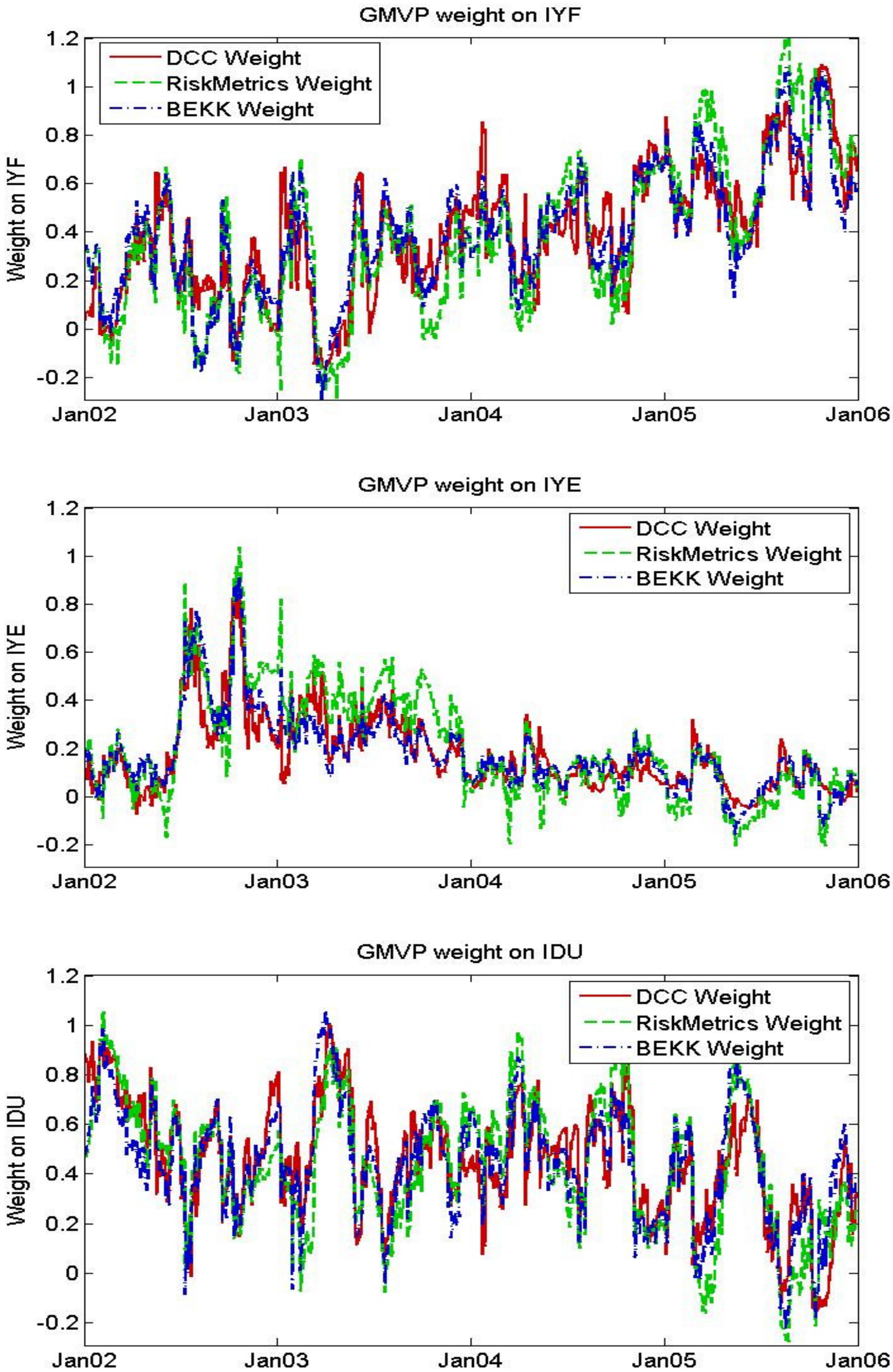


Figure 5.26 Daily optimal GMVP weights for IYF (top), IYE (middle) and IDU (bottom) given by the three multivariate models.

The annualized standard deviations for the three GMVP-portfolios are given in Table 5.29. It is calculated as the standard deviation of the daily returns of the portfolio, given by the time-varying weights in each asset multiplied by the actual returns on that day. The annualized standard deviations show that the DCC method minimizes the volatility better than BEKK and RiskMetrics, although the difference is not huge. The Sharpe ratio (calculated for an assumed risk-free rate of zero), however, representing the reward-to-variability, is much higher for the BEKK model. The reason for this is probably that the BEKK model captures the time-varying correlation better than the DCC model for these three assets. The transaction costs of a daily rebalancing scheme as illustrated in Figure 5.26 will of course be high, but since the weights of the three models evolves in a similar way, transaction costs are neglected in this paper.

	RiskMetrics	DCC	BEKK
Annualized returns	5.64%	5.77%	7.78%
Annualized std.dev	15.76%	15.64%	15.68%
Sharpe ratio	0.36	0.37	0.50

Table 5.29 Annualized returns and standard deviations for the minimum variance portfolios.

The result of an optimization procedure of the three portfolios so that the Sharpe ratio each day is maximized, is given in Table 5.30. The BEKK model again has the highest Sharpe ratio, while the DCC is narrowly beaten by the RiskMetrics optimization. The daily weights of these portfolios are calculated under the assumption that the mean of the estimation period represent the long-term average for each asset. The risk-free return in the Sharpe formula (Equation 3.47) is once again set to zero, as it would be the same for all assets. The portfolio weights are also constrained to lay within the interval of -1.2 to 1.2, to correspond with the weights of the minimum variance portfolios.

	RiskMetrics	DCC	BEKK
Annualized returns	13.05%	12.50%	14.53%
Annualized std.dev	17.12%	16.76%	17.30%
Sharpe ratio	0.76	0.75	0.84

Table 5.30 Optimization based on maximizing daily Sharpe ratio.

Although BEKK is the clearly the most adequate model with regards to maximizing the Sharpe ratio, it is less convenient for higher systems. For ten assets the parameters of the

BEKK model may take several hours to estimate. The DCC model will also probably be more useful than the multivariate EWMA model when it comes to correlation forecasting. This paper only presents an ex-post evaluation of the three correlation models.

6. Conclusions

The goal of this thesis was to present time-varying volatility in theory and practice. Even though a variety of models dealing with this issue has been presented and parameterized, the list of conditional volatility models that have not been mentioned is large. The paper gives, however, a comprehensive overview of symmetric versus asymmetric models. The conclusion regarding these two classes of models is that the asymmetric models generally perform better than symmetric models due to the fact that the asymmetric effects were evident for all three individual series. Alexander's (2001) statement that the leverage effect has become quite noticeable during the last years can therefore be supported, but this paper is unable to confirm it as analysis of older time-series is neglected.

Different models with different lag structures were recommended for each of the three time-series. These recommendations are subjective choices based on the likelihood and significance of estimated parameters, selection criterions (AIC and BIC), and forecasting ability through both economic and statistical loss functions. It is no surprise that no single model is rated above all other models for all assets, as each asset has its own unique features. The finding is consistent with Poon's (2005) survey over different researcher's findings regarding the ability of different models.

The final recommendation of the best model for each asset in this paper does not necessarily reflect the choice that would be undertaken by other researchers. If, based on these three U.S. sector indices, one single model should be chosen as the best model for all U.S. sectors, then the choice would be a GJR-GARCH(1,1,1) model. On average, this model gives the best fit of the three time-series according to both the significance of the parameters and the likelihood. According to the BIC selection criteria, which in the literature often is used as the main basis for the final selection, the GJR-GARCH(1,1,1) model is also superior to other models. This choice is in agreement with Brailsford and Faff (1996) and Taylor (2004) who find that GJR-GARCH is superior to GARCH for stock indices.

This thesis has also compared parameterizations under the assumption of alternative distributions of the errors, namely the normal distribution versus the student's T distribution. The fit and forecasting ability of two of the three assets were markedly improved by using the student's T distribution. These two assets, IYF and IDU, have a higher excess kurtosis in their raw returns over a normal than IDU. It therefore makes sense that the t-distribution gives a better fit, as it has fatter tails than a corresponding normal distribution. Ideally, other alternative distributions, like Nelson's (1991) generalized error distribution (GED) and Hansen's (1994) skewed t distribution could be tested as well to find out if such parameterizations would improve the fit of the time-series even more. This is therefore a suggestion for further investigation of the subject.

Finally, multivariate modeling shows that portfolio optimization using the Dynamic Conditional Correlation model, which builds on the already parameterized univariate GJR-GARCH(1,1,1) models for each asset, perform well when the target is to minimize the risk of the portfolio. For the three asset portfolio, however, BEKK gives a higher return on the risk undertaken than the DCC model. Analysing the time-varying correlation from the two models shows that the BEKK model is superior to the DCC model when it comes to catching this feature for the chosen assets. That the time-varying correlation between assets is this flat over time is not a feature that is shared by Cappiello et al. (2006) who study correlations in international equity indices and bonds. The DCC model performs equally well to the multivariate EWMA model, but this is for the in-sample data only. A more realistic portfolio problem would be to look at the out-of-sample performances of each time-varying correlation model, and for more assets than the three analysed. As the main focus in this paper was on univariate time-series, correlation forecasting was neglected. Further investigation on this issue therefore also goes down as a suggestion for further investigation.

Literature

Akgiray, V. (1989) Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. The Journal of Business 62**,** 55-80.

Alexander, C. (2001) Market models: a guide to financial data analysis*,* Chichester, Wiley.

Andersen, T. G. & T. Bollerslev (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39**,** 885-905.

Andersen, T. G., T. Bollerslev, F. X. Diebold & P. Labys (2001) The Distribution of Realized Exchange Rate Volatility. Journal of the American Statistical Association 96**,** 42-55.

Bauwens, L., S. Laurent & J. V. K. Rombouts (2003) Multivariate GARCH Models: A survey [online]. CORE Discussion Paper No. 2003/31. Available from http://ssrn.com/abstract=411062. [Accessed May 8, 2009].

Benninga, S. (2008) Financial Modeling*,* Cambridge, Massachusetts, The MIT Press.

Bera, A. K. & M. L. Higgins (1993) ARCH models: properties, estimation and testing. Journal of Economic Surveys 7**,** 305-366.

Bodie, Z., A. Kane & A. J. Marcus (2008) Investments*,* Boston, Mass., McGraw-Hill.

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

Bollerslev, T. (1987) A conditional heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics 69**,** 542-547.

Bollerslev, T. (1990) Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model. The Review of Economics and Statistics 72**,** 498-505.

Bollerslev, T. & R. F. Engle (1993) Common persistence in conditional variances. Econometrica 61**,** 167-186.

Bollerslev, T. P., R. F. Engle & D. B. Nelson (1993) ARCH models. IN: Bollerslev, T. P., R. F. Engle & D. B. Nelson (Eds.) The Handbook of Econometrics*.* Department of Economics, University of California. Available from http://finance.martinsewell.com/stylized-facts/BollersvlevEngleNelson1994.pdf.

Brailsford, T. J. & R. W. Faff (1996) An evaluation of volatility forecasting techniques. Journal of Banking & Finance 20**,** 419-438.

Brooks, C., S. P. Burke & G. Persand (2001) Benchmarks and the accuracy of GARCH model estimation. International Journal of Forecasting 17**,** 45-56.

Cao, C. Q. & R. S. Tsay (1992) Nonlinear Time-Series Analysis of Stock Volatilities. Journal of Applied Econometrics 7**,** 165-185.

Cappiello, L., R. F. Engle & K. Sheppard (2006) Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. Journal of Financial Econometrics 4**,** 537-572.

Chatfield, C. (2003) The analysis of time series: an introduction, CRC Press.

Davidian, M. & R. J. Carroll (1987) Variance Function Estimation. Journal of American Statistical Association 82**,** 1079-1091.

Ding, Z., C. W. J. Granger & R. F. Engle (1993) A long memory property of stock market returns and a new model. Journal of Empirical Finance 1**,** 83-106

Enders, W. (2004) Applied econometric time series*,* Hoboken, N.J., Wiley.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50**,** 987-1007.

Engle, R. F. (2001a) Financial econometrics - A new discipline with new methods. Journal of Econometrics 100**,** 53-56.

Engle, R. F. (2001b) GARCH 101: The use of ARCH/GARCH models in Applied Econometrics. Journal of Economic Perspectives 15**,** 157-168.

Engle, R. F. (2002) Dynamic conditional correlation. Journal of Business and Economic Statistics 20**,** 339-350.

Engle, R. F. & R. Colacito (2004) Dynamic Conditional Correlation - A simple class of multivariate GARCH models [online]. Draft. Available from http://accounting.uwaterloo.ca/finance/documents/RColacito_001.pdf. [Accessed June 9, 2009].

Engle, R. F., S. M. Focardi & F. J. Fabozzi (2008) ARCH/GARCH Models in Applied Financial Econometrics. IN: Fabozzi, F. J. (Ed.) Handbook of Finance*.* John Wiley and Sons. Available from http://pages.stern.nyu.edu/~rengle/ARCHGARCH.pdf.

Engle, R. F. & K. F. Kroner (1995) Multivariate Simultaneous Generalized ARCH. Econometic Theory 11**,** 122-150.

Engle, R. F. & S. Manganelli (1999) CAViaR: Conditional Autoregressive Value at Risk by regression quantiles [online]. NBER Working Paper No. W7341. Available from http://ssrn.com/abstract=195948. [Accessed June 6, 2009].

Engle, R. F. & K. Sheppard (2001) Theoretical and empirical properties of Dynamic Conditional Correlation multivariate GARCH. USCD Discussion Paper 2001-15, NYU Discussion Paper.

Fama, E. F. (1965) The behavior of stock-market prices. The Journal of Business 38**,** 34-105.

Fleming, J., C. Kirby & B. Ostdiek (2000) The economic value of volatility timing. Journal of Finance 56**,** 329-352.

Franses, P. H. & M. McAleer (2002) Financial Volatility: An Introduction. Journal of Applied Econometrics 17**,** 419-424.

Glosten, L. R., R. Jagannathan & D. E. Runkle (1993) On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance 48**,** 1779-1801.

González-Rivera, G., T.-H. Lee & S. Mishra (2004) Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of Forecasting 20**,** 629-645.

Hansen, B. E. (1994) Autoregressive conditional density estimation. International Economic Review 35**,** 705-730.

Harris, R. I. D. & R. Sollis (2003) Applied time series modelling and forecasting*,* Chichester; Hoboken, N.J., J. Wiley.

J.P.Morgan (1997) RiskMetrics Technical Documents, 4th edition*,* New York.

Kroner, K. F. & V. K. Ng (1998) Modeling Asymmetric Comovements of Asset Returns. The Review of Financial Studies 11**,** 817-844.

Laurent, S. (2004) Analytical derivates of the APARCH model. Computational Economics 24**,** 51-57.

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

Ljung, G. M. & G. E. P. Box (1978) On a Measure of Lack of Fit in Time Series Models. Biometrika 65**,** 297-303.

Mandelbrot, B. (1963) The variation of certain speculative prices. The Journal of Business 36**,** 394-419.

McKenzie, M. D. (1999) Power transformation and forecasting the magnitude of exchange rate changes. International Journal of Forecasting 15**,** 49-55.

McMillan, D., A. Speight & O. Apgwilym (2000) Forecasting UK stock market volatility. Applied Financial Economics 10**,** 435-448.

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

Nelson, D. B. & C. Q. Cao (1992) Inequality Constraints in the Univariate GARCH Model. Journal of Business & Economic Statistics 10**,** 229-235.

Patton, A. J. & K. Sheppard (2008) Evaluating volatility and correlation forecasts. IN: Andersen, T. G., R. A. Davis, J.-P. Kreiss & T. Mikosh (Eds.) Handbook of financial time series*.* New York, Springer. Available from http://www.kevinsheppard.net/images/c/c6/Patton_Sheppard.pdf.

Peters, J.-P. (2001) Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities [online]. Preprint. Ecole d'Administration des Affaires, University of Liége, Belgium. Available from http://www.unalmed.edu.co/~ndgirald/Archivos%20Lectura/Archivos%20curso%20Series%20II/jppeters.pdf. [Accessed June 1, 2009].

Poon, S.-H. (2005) A practical guide for forecasting financial market volatility*,* Chichester ; Hoboken, N.J., Wiley.

Poon, S.-H. & C. W. J. Granger (2003) Forecasting volatility in financial markets: a review. Journal of Economic Litterature 41**,** 478-539.

Sheppard, K. (2009a) Financial Econometric Notes [online]. University of Oxford. Available from http://www.kevinsheppard.com/images/a/af/Financial_Econometrics_2008-2009.pdf. [Accessed April 17, 2009].

Sheppard, K. (2009b) MFE Matlab Function Reference Financial Econometrics [online]. Available from http://www.kevinsheppard.com/images/9/95/MFE_Toolbox_Documentation.pdf. [Accessed May 11, 2009].

Taylor, J. W. (2004) Volatility forecasting with smooth transition exponential smoothing. International Journal of Forecasting 20**,** 273-286.

Zakoian, J.-M. (1994) Threshold heteroskedastic models. Journal of Economic Dynamics and Control 18**,** 931-955.