School of Economics and Management

Lund University

Department of Economics

Modeling and Forecasting Volatility in Copper Price Returns with GARCH Models

Bachelor Thesis: NEKH01 (15 credits ECTS)

Supervisor: Karl Larsson   Author: Dominice Goodwin

August 2012

Abstract.

   This thesis offers a study on how well the standard GARCH(1,1) model, the GJR-GARCH(1,1) model and the QGARCH(1,1) model, were able to model (in-sample) and forecast (out-of-sample) the volatility of copper spot price returns in four equally large subsamples within the period July 21, 1993 to 22 Mars, 2012. The results shows that the GARCH models' ability to model the conditional variance (in-sample) was highly satisfactory for the three subsamples in which there was found significant presence of ARCH effects. In out-of-sample forecasting the GARCH models dominated a Random Walk model across all four subsample (17 to 29 % lower Mean of Absolute Errors for the GARCH models). In three out of the four subsamples the GARCH models also dominated the unconditional variance (8 to 19 % lower Mean of Absolute Errors for the GARCH models). In one subsample the GARCH models produced about 8 to 10 % higher MAE than the unconditional variance. The results did also indicate that no one of the three GARCH models used in this thesis could generally be recommended over the others for the purpose of forecasting the one-step ahead one day variance in copper spot price returns. This conclusion is based on the facts that the differences between the GARCH models in the out-of-sample forecasting was small, and that it was difficult to predict which model that would have the slightly better forecasting performance. This conclusion may be interpreted as that the standard GARCH(1,1) model was sufficient.

Key words: GARCH(1,1), GJR-GARCH(1,1), QGARCH(1,1), forecasting volatility, Copper

Contents

1. Introduction

The recent years' dramatic price increases in many commodities are one of the many noteworthy events which have occurred in the world commodity markets in the last decades. Alongside of this development there has followed a rapidly increasing activity in the commodity markets by market participants other than the 'traditional' ones whom has industrially related exposure to the physical commodity, namely by institutional investors, investment banks, hedge funds, mutual funds, and other speculators. In contrast to the traditional market participants, who mainly uses the commodity markets for spot transitions and hedging purposes, these new market participants have been drawn to the commodity markets by the prospect of further strong price movements, and by the risk reduction effect (through diversification by uncorrelated returns) and inflation protection, which adding commodities to a traditional portfolio asset mix could bring (for more in-depth discussions on this topic see Geman, H. (2005) page 2-7; Russel-Walling, E. (2005); and Watkins, C. and McAleer, M. (2008)).

However, since sound investment decisions in general are based on the trade-off between expected risk and return, these increased investment orientated activities in the commodity markets calls for the same careful analysis and estimation of the expected future return, volatility and other key variables, as is standard when it comes to more traditional types of investment assets.

With that in mind, and when considering the growing number of studies suggesting that ARCH class models, especially some types of GARCH models[1], are able to deliver more accurate forecasts of future variance compared to the simple historical variance (unconditional variance) and a Random Walk model, an important area for further academic research should therefore be this kind of studies with focus on the volatility of commodity returns. Although the volatility of some commodities are more frequently studied, such as crude oil, some researchers (e.g. Bracker, K. and Smith, K.L. 1999, page 80) have pointed out that other commodities, such as copper, has been given surprisingly little attention in this regard, given that copper (along side of aluminum) has the highest trading volume (spot and future) of the industrially used non-ferrous metals on the London Metal Exchange (LME).

This Bachelor Thesis therefore aims to extend the literature in this area by investigating the extent to which ARCH effects has been present in copper price returns since mid 1993 to

---

[1] These studies have mainly been focused at the volatility in returns of equities, fixed income securities and exchange rates. A few examples of such studies are listed the Appendix.

the present, and to evaluate how well GARCH models have been able to model (in-sample) and forecast (out-of-sample) the copper price volatility within this period.

The value of such a study lies, first and foremost, in the importance of accurate estimates of the future volatility in financial decision making and many financial applications, such as portfolio management, derivatives pricing, and value at risk (VaR) estimation. Secondly, GARCH models have proven to be able to model conditional volatility and improve the forecasting accuracy of the future volatility of many financial time series. To fully appreciate the value of this fact, it should be remembered that it has proven difficult to forecast volatility using external regressors, thus meaning that GARCH models have been able to improve the forecasting accuracy of the future volatility were other methods empirically have shown to be unsatisfactory[2]. The results presented in this thesis should in other word be relevant because; knowledge about GARCH models' ability to model and forecast the future volatility in copper price returns are (or at least, should be) important to both traditional market participants (who doesn't fully hedge their physical exposure) and for the growing number of investment-orientated market participants in the commodity markets who faces price risk and portfolio decisions involving copper.

Furthermore, another aspect of the changes and the dramatic events which have taken place in the commodity markets and the macroeconomic and financial environment as a whole over the last two decade, is that it is possible or even likely that there could have been some significant changes in the dynamics of the volatility in copper price returns over this period. This further motivates updating the literature since the study by Bracker, K. and Smith, K.L. (1999 and 2003) who, for the studied period December 31, 1974 to June 28, 1996, found the GARCH, EGARCH and GJR-GARCH model outperformed the random walk model in out-of-sample forecasting of the one-step ahead future variance of copper price returns. And, the study by Watkins, C. and McAleer, M. (2008), who found the volatility of

---

[2] It is also worth to point out that GARCH models can be combined with external variance regressors, if such can be found to have a significant explanatory power of the future variance. See for example Kogan, L., Livdan, D. and Yaron, A. (2009) who presented a model for modeling the future volatility of oil futures conditional on a GARCH term and the slope of the term structure.

For a study on the drivers of volatility in the non-ferrous metals market see Brunetti, C. and Gilbert, C.L. (1995).

copper price returns to be non-increasing and time-varying (when analyzed over a long horizon) for the period January 5, 1976 to July 24, 2006.

The research method used in this thesis could be divided into four main parts. First the possibility of modeling the mean with an autoregressive model is investigated by studying the presence of autocorrelation in the returns. This is done for all four subsamples which the studied period in this thesis is divided into. Based on this analysis an appropriate model for the mean is then selected for each of the subsamples. The reason for dividing the studied period into subsamples is to be able to analyze how the GARCH models' ability to model the volatility of copper price returns has change over time.

In the next step the presence of time-varying volatility and 'volatility clustering' (ARCH effects), and asymmetry in the distribution of returns are investigated, both visually and by various econometric tests. This is done to derive a prior expectation regarding the appropriateness of modeling the volatility of the copper price returns with symmetric and asymmetric GARCH models in each of the subsamples.

In the third part the GARCH models are estimated in each of the subsamples. The GARCH models are estimated on the residuals of the model for the mean which was selected for that subsample. The regression outputs are then analyzed to conclude whether it was appropriate to model the variance (in-sample) with the GARCH models, and to derive prior expectations of the forecasting performances of the GARCH models.

The GARCH models analyzed in this thesis are the standard symmetrical GARCH model developed by Bollerslev, T. (1986), and two asymmetrical GARCH models, namely the GJR-GARCH model proposed by Glosten, L., Jagannathan, R. and Runkel, D. (1992), and the QGARCH model proposed by Engel, R. F. and Ng, V. (1993) and further developed by Sentana, E. (1995). The asymmetric GARCH models are included in this study since the distributions of copper price returns were skewed throughout all four subsamples. Furthermore, including these asymmetric GARCH models in the study also opens up the possibility to analyze if the standard GARCH model is sufficient to model the volatility in copper price returns even when the distribution of returns is skewed.

In this thesis the QGARCH was chosen (in favor of the EGARCH) since this model wasn't covered in the previously mentioned studies of the copper return volatility. The QGARCH was, furthermore, the best performing GARCH model regarding out-of-sample forecasting in the study by Franses, P.H. and van Dijk, D. (1996), who analyzed the volatility of weekly returns of several major stock market indices.

In the fourth and last part the forecasting performances of the estimated GARCH models are being evaluated. This is done by comparing the variance forecasts from the GARCH models, a Random Walk model and the unconditional variance, to a proxy for the 'true' realized one day variance. Mean Squared Error (MSE), the Mean Absolute Error (MAE) and the Median Squared Error (MedSE) are then calculated based on the deviations between the variance forecasts and the proxy for the 'true' realized one day variance. The GARCH models' forecasting performances are then evaluated by comparing the MSE, MAE and MedSE of the GARCH models' variance forecast to the MSE, MAE and MedSE generated by the Random Walk model's and the unconditional variance's variance forecasts. Conclusions are however primarily based on the MAE values for reasons discussed in section 4.6.

There are a few things that should be noted regarding the limitations of this study. To begin with, the estimated GARCH models were only of a (1,1)-order. However, judging by the diagnostics tests after the model estimations, this specification seemed to be sufficient to fully capture the ARCH effects in the data.

Secondly, only the volatility in the close-to-close daily log returns is analyzed. It was in other words not investigated whether or not the GARCH models were suitable for modeling and forecasting the volatility of open-to-close (intraday) returns, close-to-open returns, or returns of higher frequencies.

Thirdly, there was in this thesis no concern taken to a possible 'information accumulation effect' as a consequence of the observations in the data being irregularly spaced in time.

And finally, the data sample was arbitrarily divided into four equally large subsamples. It is therefore possible that better results could have been derived by instead dividing the sample into subsamples with a more methodical approach. An example would be the iterated cumulative sum of squares (ICSS) algorithm, which Bracker, K. and Smith, K.L. (1999) used to detect breakpoints in the volatility series.

The results presented in this thesis shows that there was strong presence of ARCH effects in the volatility of the copper spot price returns in the studied period. This presence did however vary across the subsamples and was found to be highly significant in three out of the four subsamples.

The GARCH models' ability to model the conditional variance (in-sample) was highly satisfactory for the three subsamples in which there was significant presence of ARCH effects. The results also suggested that the GARCH models' ability to model the conditional

variance might have increased over time. The GARCH models removed the largest portion of excess kurtosis in the distribution of the residuals from the mean equation in the last subsample (between 75 to 85 % of the excess kurtosis were removed by the GARCH models here).

The economic and statistical significance of the asymmetry parameters of the GJR-GARCH and the QGARCH did also increase over time. In the last two subsamples the asymmetry parameters provided significant explanatory power. The skewness in the distribution of demeaned returns or the distribution of the residuals from the mean equation were, however, not a good indicator of whether the two asymmetric GARCH models would be better than the standard symmetrical GARCH model in modeling the conditional variance (in-sample or out-of-sample).

The (1,1) order of the GARCH models did furthermore show to be sufficient to fully capture the ARCH effects in all of the subsample periods.

In out-of-sample forecasting the GARCH models, in terms of MAE, dominated the Random Walk model across all subsamples. The improvements over the Random Walk model varied between 17 to 29 % lower MAE.

However, since the unconditional variance produced lower MAE values than the Random Walk model, which was expected due to the mean reversion in the volatility, the unconditional variance proved to be a more legitimate benchmark for the forecasting performance of the GARCH models than the Random Walk model in this case.

The GARCH models had about 8 to 19 % lower MAE than the unconditional variance for three out of four subsamples. However, in one subsample the GARCH models had about 8 to 10 % higher MAE than the unconditional variance. The three subsamples where the GARCH models outperformed the unconditional variance were not the same as the three subsamples which had significant presence of ARCH effects. This suggests that it's hard to predict the forecasting performances of the GARCH models based on the in-sample estimations. However, the results also suggested that if more thoughts were to be given into the subsample selection, more accurate expectations of the GARCH models out-of-sample forecasting performance could probably be derived.

The results from the out-of-sample forecasting also showed that the difference between the GARCH models (in MAE) was small, and that it was difficult to determine which one of the GARCH models that would perform best in out-of-sample forecasting based on the results from the in-sample estimations. Therefore, no one of the three GARCH models used in this thesis can generally be recommended over the others.

2. Theoretical framework

This section of the thesis offers a dense discussion on the topic of modeling the conditional variance with GARCH models, and an incomplete overview of related studies. For readers who are already familiar with the purposes and strengths of using GARCH, this section could be redundant and therefore skipped.

2.1 Modeling non-constant conditional variance

The necessity of modeling the conditional variance is consequence of the well known tendency for the variance in many time series of financial returns to be non-constant over time. When estimating coefficients of a standard regression model (e.g. CLRM and ARMA models) using the ordinary least squares (OLS) on such data, the error variance of the model could be non-constant (heteroscedastic), which violates the assumption of homoscedasticity in the classical linear regression model (CLRM)[3]. Although the OLS estimator of the coefficients is still unbiased in such a case, the problematic consequence of this is that the standard formula of calculating the standard error isn't, thus making inference (e.g. confidence intervals and p-values) unreliable.

One commonly used solution to this problem when estimating a model on cross-sectional data is to use a robust standard error, such as White's Robust S.E., which delivers consistent estimates of the standard errors [4]. When it comes to time series models on the other hand, the forecasting accuracy of the model is often important or the main concern. Simply adjusting the standard error and keeping the assumption of constant error variance is therefore not a satisfying solution if the heteroscedastic error variance instead can be explained by the model.

The reason for why modeling the heteroscedastic error variance would be a preferred method, is based on the fact that the error variance can be shown to be the same as variance of the dependent variable y_t in a regression model. This is so since it's also assumed that the values of the explanatory variables are given, so that the deviations of the dependent variable

---

[3] The assumption of homoscedasticity: the expected value of all squared errors is the same at any given point in time and doesn't depend on the values taken by any other variables. Formally; $Var[\varepsilon] = \sigma^2$. Also, remember that it has been shown (Gauss-Markov theorem) that the OLS estimators of the coefficients have minimum variance of all linear and unbiased estimators (BLUE) only when the assumptions of the classical linear regression model (CLRM) holds.

[4] In this case the error variance varies with some explanatory variable(s) rather than over time.

y_t from its expected value (its mean) are fully explained by the error term ε_t. The conditional variance of y_t and ε_t are therefore the same and assumed to be constant, which formally could be express as:

$$var(y_t|\Omega_{t-1}) = var(\varepsilon_t|\Omega_{t-1}) = \sigma^2 \qquad (2.1.1)$$

, where Ω_{t-1} being some information set available at time t-1. In other words, the assumption of constant error variance (homoscedasticity) also means that conditional distribution of each y_t population is assumed to be spread around its conditional mean with the same (or constant) variance (see Gujarati, D. N. and Porter, D.C. (2010) p.55).

Now, since the dependent variable y_t in many financial time series are some asset returns, and since financial decisions are generally bases upon the tradeoff between expected risk and return, an accurate forecast of future variance of y_t should be important[5]. Therefore, if the assumption of constant error variance is known to not hold, it should for financial time series models be a preferred approach (compared to using robust standard error) to try modeling the non-constant error variance since this could both solve the problems with OLS estimation as described above and lead to higher forecasting accuracy of the future variance of y_t[6].

In order to model non-constant conditional variance (or conditional heteroscedasticity), changes in the structure of the standard type of regression model first has to be made. One way to do this is by multiplying the error term with the conditional standard deviation. For such model, the general form for a regression of a dependent variable y_t on a number of explanatory $(x_{1,t}, \dots, x_{p,t})$ can be expressed as:

$$y_t = f(x_{1,t}, \dots, x_{p,t}) + \sigma(x_{1,t}, \dots, x_{p,t})v_t \qquad (2.1.2)$$

, where v_t is a stochastic error term with a zero mean and unit variance. Both f and σ are conditional upon the values taken by some explanatory variables. Therefore such a model has

---

[5] Without claiming to deliver complete coverage of situations where the accuracy of the forecast of future variance is important in finance, some areas where it widely accepted to be so is in; portfolio management, derivatives pricing, value at risk (VaR) estimation, and for other investment related decisions where accurate measures of expected risk and return are key inputs.

[6] Remember that if homoscedasticity is assumed, the one-step ahead forecast variance is assumed to be the constant unconditional variance.

a conditional mean (or expectation) of $f(x_{1,t}, \ldots, x_{p,t})$, and a conditional variance of $\sigma^2(x_{1,t}, \ldots, x_{p,t})$, both of which being nonconstant (see Ruppert, D. (2006) p.364-365)

Adding the assumption of normality, and if f is assumed to be linear and $\sigma^2(x_{1,t}, \ldots, x_{p,t})$ is denoted as σ_t^2, the model can be expressed in the following structure used by Engel, R. F. (1982), and Bollerslev, T. (2011).

$$y_t = x_t\beta + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_t^2) \tag{2.1.3}$$

$$\varepsilon_t = v_t\sigma_t \qquad v_t \sim N(0,1) \; i.i.d \tag{2.1.4}$$

, where β being a vector of unknown parameters, and where $x_t\beta$ (the conditional mean of y_t) could consist of both endogenous and exogenous variables. In other words, non-constant conditional variance can be modeled in both structural models and univariate time series models (e.g. the Classical Linear Regression Model (CLRM) and the AutoRegressive Moving Average (ARMA) model). $x_t\beta$ and σ_t^2 can also be expressed in terms of the information set available at time t-1, so that y_t can be expressed using conditional densities.

$$y_t \mid \Omega_{t-1} \sim N(x_t\beta, \sigma_t^2) \tag{2.1.5}$$

$$E(y_t \mid \Omega_{t-1}) = x_t\beta \tag{2.1.6}$$

$$Var(y_t \mid \Omega_{t-1}) = E([y_t - E(y_t \mid \Omega_{t-1})]^2 \mid \Omega_{t-1}) = E(\varepsilon_t^2 \mid \Omega_{t-1}) =$$
$$= Var(\varepsilon_t \mid \Omega_{t-1}) = \sigma_t^2 \tag{2.1.7}$$

As we see in eq. 2.1.7, the conditional variance of y_t and ε_t are still the same. However, their conditional variance are now non-constant (σ_t^2) rather than constant as before (σ^2).

Although the model with this modification allow for non-constant conditional variance, an important part remains to be considered, namely the way in which the conditional variance should be modeled for the best fit to the data so that better forecasts of future variance can be acquired. This task is however, as forecasting the future intrinsically is, a difficult one. Finding external variables which can forecast the changes in variance has shown to be difficult, and to use external variables that are found to correlate with the changes in the variance has to be forecasted as well (which could be just as, or even more, difficult) in order

to forecast the future variance. These difficulties of modeling the conditional variance bring us to the next section where the ARCH class models for the conditional variance is discussed.

2.2 ARCH

The ARCH model, which stands for AutoRegressive Conditional Heteroscedastic model, was introduced by Engel, R. F. (1982). In short, the ARCH model models the conditional error variance as a function of the values taken by past squared errors, thus exploiting the commonly found presence of autocorrelations in these. Engel stated by the example of an autoregressive model of the first-order, that the use of a conditional mean in time series model clearly could bring vast improvements in forecasting the dependent variable, and by that logic argued that; *"For real processes one might expect better forecast intervals if additional information from the past were allowed to affect the forecast variance"* (Engel, R. F. (1982) p.988).

What Engel proposed was to model the conditional variance as a weighted average of the values taken by a number of the most recent lags of squared errors, and to parameterize their weights so that they could be estimated for the best fit to the sample data using maximum likelihood[7]. Engel did in other words propose that the conditional error variance should be conditional upon the values taken by previous errors[8]. Formally this could be express as:

$$\sigma_t^2 = Var(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = E(\varepsilon_t^2 \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \qquad (2.2.1)$$

The possibility to model the conditional variance this way is due to another empirical characteristic (besides the heteroscedasticity) in time series of financial returns, which is commonly referred to as volatility clustering. What this means is that the volatility is clustered together so that there are periods of persistent high volatility followed by periods of persistent low volatility, and so on. There is in other words commonly observed to be positive and statistically significant autocorrelation in the squared errors.

To understand why modeling the conditional variance with ARCH could lead to more accurate forecasts for the future variance in the presence of non-constant error variance and positive autocorrelation in the squared errors, remember first that the historical (or, empirical-,

---

[7] Engel, R. F. (1982) showed that the use of ML rather than OLS brought efficiency gains.

[8] It should also be mentioned that it's possible to include relevant exogenous variance regressors (if such could be found), so that the conditional variance is not only conditional upon the values taken by previous errors.

simple-) variance of a variable is the equally weighted average of all the squared deviations from the sample mean in that sample. This estimate of the future variance would be optimal if the errors were to be strict white noise (zero mean, independent variables with constant variance) since the conditional error variance then would be constant and equal to the unconditional[9].

However, being aware of the presence of non-constant error variance and positive autocorrelation in the squared errors, the historical variance as an estimate for the future variance appears unattractive. An improvement could then be to use an equally weighted moving average variance, of most recent X-days' squared errors. However, if it's also known that the size of the most recent squared errors have a larger impact on the future variance, this approach doesn't appear completely satisfactory either. Another improvement under these conditions could then be to allow each of the X most recent squared errors to have unequal weights. This is one of the essential features of the ARCH model. And, as explained above, the weights are estimated for the best fit to the sample data by using maximum likelihood estimation. The conditional variance equation of the ARCH(q) model can in a general from be expressed as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \qquad (2.2.2)$$

, where σ_t^2 is the one-step ahead forecast of the variance (the conditional variance), q is the number the most resent squared errors included in the calculation of the conditional variance, and α_i are the weights given to each corresponding lag of squared error[10]. To ensure that the conditional variance would not take a negative value, all weights and the constant term are required to be non-negative (Non-negativity constrains: $\alpha_i \geq 0 \; \forall \; i, \; and \; \omega \geq 0$)[11].

This formula does, however, raise the question of what the optimal size of q are given a certain data sample of a time series. Two problems with the ARCH model is that there isn't

---

[9] It's here desirable to have as big sample as possible since the sample variance approaches the population variance as the number of observations increases.

[10] Recognize that this formula also could represent an equally weighted moving average if $\omega = 0$ and $\alpha_1 = \ldots = \alpha_q$.

[11] σ_t^2 could take a negative value given at least one negative weight and a sufficiently large lagged error attached to that negative weight.

any clear best approach to determine this, and that q might need to be quite large in order to fully capture the dynamics in the variance. And, having a large q introduces the problem of having many parameters to estimate in the regression and an increased risk of breaching the non-negativity constrains.

To counter these problems some ad hoc solutions has been proposed. One example of such ad hoc solution (eq. 2.2.3) is the one Engel, R. F. (1982) used to model the variance of UK inflation in the same study in which he first introduced the ARCH model.

$$\sigma_t^2 = \omega + \alpha_1(0.4\varepsilon_{t-1}^2 + 0.3\varepsilon_{t-2}^2 + 0.2\varepsilon_{t-3}^2 + 0.1\varepsilon_{t-4}^2) \qquad (2.2.3)$$

Another such example (eq. 2.2.4) is the model Engel, R. F. and Kraft, D. (1983) used.

$$\sigma_t^2 = \omega + \alpha_1 \sum_{i=1}^{8} \frac{(9-i)}{36\varepsilon_{t-i}^2} \qquad (2.2.4)$$

As we see, both these models solve the problem of large q since only two parameters needs to be estimated.

However, this ad hoc approach still leaves the researcher with the inconvenience of having to pre-determine how many lags of the squared errors the conditional variance should be conditional upon, and what weights they should have relative to one and another. Finding the most suitable model for the conditional variance could therefore be a time consuming task. Furthermore, and possibly more serious than the inconvenience, there also exists the possibility that the pre-determine specification (like those above) isn't optimized to the sample data since an estimation of α_1 leaves the relative size of the weights and the size of the q unchanged.

Due to these weaknesses, the ARCH(q) model are rarely used in practice today. However, the ARCH model has provided a framework for modeling the conditional variance upon lags of squared errors, a framework which the other still widely used models has been developed on. One of these models is the widely used GARCH model which is covered in section 2.3 below.

However, before proceeding to this section, it's worth pointing out one additional aspect about ARCH models in general. In Engel's own words:

"...the ARCH regression model is an approximation to a more complex regression which has non-ARCH disturbances. The ARCH specification might then be picking up the effect of variables omitted from the estimated model. The existence of an ARCH effect would be

interpreted as evidence of misspecification, either by omitted variables or structural change. If this is the case, ARCH may be a better approximation to reality than making standard assumption about the disturbances, but trying to find the omitted variable or determine the nature of the structural change would be even better." Engel, R. F. (1982), p.990.

"The goal of volatility analysis must ultimately be to explain the cause of volatility. While time series structure is valuable for forecasting, it does not satisfy out need to explain volatility. ... Thus far, attempts to find the ultimate cause of volatility are not very satisfactory." Engel, R. F. (2001) p.166.

In other words, it's important to not forget that no matter how accurate forecasts of the future variance that ARCH class models produces; they do not explain the true cause of volatility. And, the popularity of models based on ARCH today, should arguably not only be contributed to the ability of these models to model the non-constant variance, but also due to the difficulty in finding external variables that in a satisfactory manner can explain the true process.

2.3 GARCH

The GARCH model, which stands for Generalized AutoRegressive Conditional Heteroscedastic model, was developed by Bollerslev, T. (1986)[12], and builds on the ARCH framework of modeling the conditional error variance. The difference between the GARCH model and the ARCH model is that it 'generalizes' the selection of q and the scheme in which the weights decline. There is in other words no need to pre-specify an ad hoc solution for the number of lags (q) to include and their relative weights before estimating a GARCH model, in order to avoid the problems of the ARCH model. To be more specific, the weights in a GARCH model decline exponentially, thus never reaching completely zero. Therefore, the GARCH model (in its most simple form) can be thought of as an ARCH(∞) model with exponentially declining weights.

Although the benefits from this generalization might not be obvious at first, it does bring of some very important and desirable properties. To begin with, having the weights decline exponentially solves the problem of having too many parameters to estimate. This property of the GARCH model can be shown by studying the properties of an Exponentially Weighted Moving Average (EWMA) model for the conditional variance.

---

[12] Some sources do however state that the GARCH model also was developed independently by Taylor S.J.

A EWMA is simply the historical average variance but with exponentially declining weights, and can formally be express as:

$$\sigma_t^2 = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j}^2 \qquad (2.3.1)$$

At first glance, introducing an infinite number of weights and lagged squared errors doesn't seem to solve the problem of having a too large number of parameters to estimate as described above, rather the opposite[13]. However, the infinite series of exponentially declining weights can through algebra be reduced to the following recursive formula for the conditional variance:

$$\sigma_t^2 = (1 - \lambda) \varepsilon_{t-1}^2 + \lambda \sigma_{t-1}^2 \qquad (2.3.2)$$

, where σ_{t-1}^2 is the most resent one-step ahead forecast of the variance, and λ is the so called decay factor, which through regression can be estimated to the best fit for a particular data set, but is for EWMAs typically set to some standard value as RiskMetrics' 0,94. Since λ by specification can take values between zero and one ($0 < \lambda < 1$), the conditional variance is a weighted average of the most resent one-step ahead forecast of the variance and the most resent lag of squared error[14]. From this formula it's evident that letting the weights decline exponentially solves the problem of having too many parameters to estimate since there are only two weights to be estimated.

The GARCH model uses this feature of a EWMA so that the conditional variance equation for a GARCH(1,1) model can be written:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \qquad (2.3.3)$$

As we see, there are some minor differences between a GARCH(1,1) model and a EWMA. In the GARCH(1,1) the $(1 - \lambda)$ and λ has been changed to α and β, which together with the

---

[13] Since it's impossible to weight an infinite number of the most resent errors, the series has to be truncated at some lag.

[14] For large λ new shocks (errors) affects the conditional variance relatively little, and the last one-step ahead forecast of the variance predominates the value taken by the new one-step ahead forecast of the variance. λ is therefore also commonly called the "volatility persistence parameter".

constant term ω are estimated to the best fit for a given data sample using maximum likelihood estimation. The conditional variance in the GARCH model is therefore a weighted average of the long term variance, the most resent one-step ahead forecast of the variance and the most resent lagged error squared. The variables ε^2_{t-1} and σ^2_{t-1} are commonly referred to as the ARCH term and the GARCH term.

The addition of the constant term in the GARCH(1,1) model lets the conditional variance depend upon a long run average variance (or, the constant unconditional variance) when the model is stationary (that is when: $\alpha + \beta < 1$), which it is in most cases. This property separates the GARCH(1,1) mode from a EWMA model, and is a desirable feature since it's well known that volatility series are mean-reverting, that is, the variance tend to move towards long run average when being relatively high or relatively low (see Brooks, C. (2008) p.385).

To highlight the benefits of the GARCH(1,1) model compared to the described ad hoc solutions for the ARCH model in the previous section, consider figure 2.3.1.

Figure 2.3.1 - Weight declining schemes, GARCH vs. ad hoc ARCH
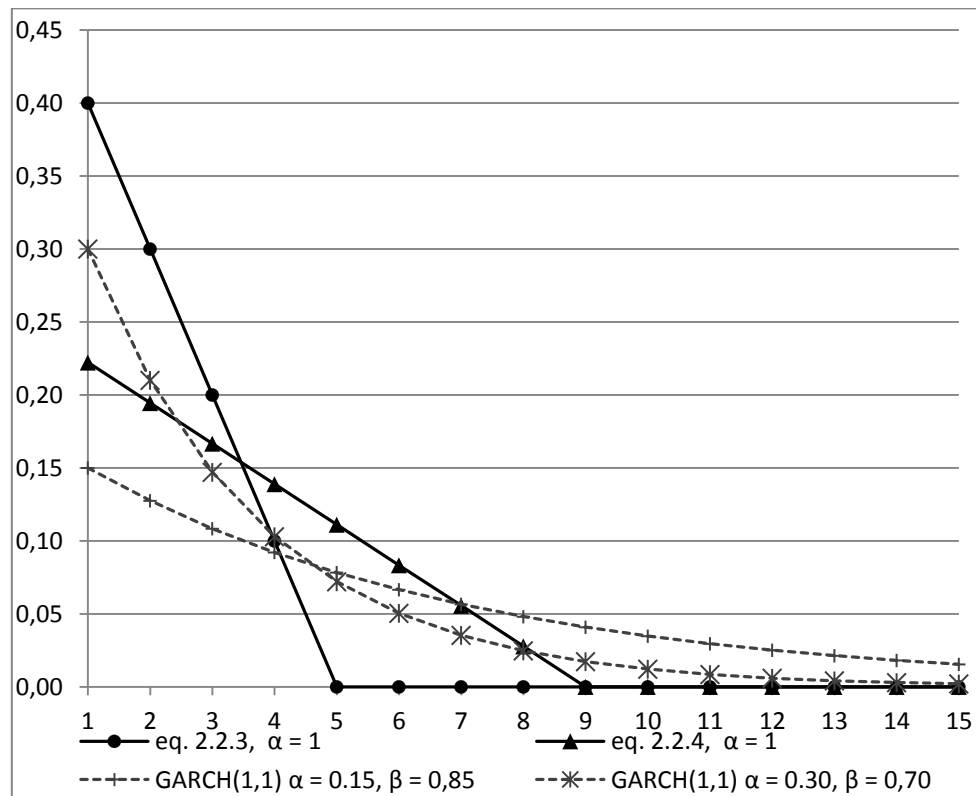
Notes: $\omega = 0$ in the GARCH models here

Here we see that the more weight given to the most resent squared error in a GARCH(1,1) model, the faster will the rest of the weighs decline (although, still exponential). By this property, the estimation of α and β in a GARCH model will automatically adjust the relative weights of the lagged squared errors. And, although the weights in the GARCH model never completely reaches zero, the estimation of α and β dose in one sense also automatically adjusts how many of the most resent lags of squared error that the conditional variance should be conditional upon. For example, the GARCH model with $\beta = 70$ could almost be considered an ARCH(15) model with exponentially declining weights. This feature of the GARCH model makes it a very flexible model compared to ad hoc solutions in eq. 2.2.3 and eq. 2.2.4, where the estimation of α_1 merely shifts the curve downwards (if $\alpha_1 < 1$) since all weights are equally affected by the value of this parameter.

Although the GARCH(1,1) in general is sufficient to capture the volatility clustering, the GARCH model can be expanded to hold q number of ARCH terms and p number of GARCH terms. The conditional variance equation for such GARCH(p, q) model can formally be written as[15]:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{2.3.3}$$

2.4 Modifications of the standard GARCH model

Since the introduction of the standard GARCH model there has been developed a large number of extensions or modifications to this model. One of the more commonly used groups of such modifications has been developed with the aim of dealing asymmetries in the distribution of returns. The usefulness of this group of GARCH models arises from yet another commonly found characteristic of financial returns, namely the presence of negative skewness in the distribution of returns.

A distribution which has negative skew has a longer left tail and a shorter right tail compared to the symmetrical normal distribution. This means that negative returns are less concentrated close to the mean compared to positive returns. Negative returns are in other words more likely to take on extreme absolute values than positive returns. It has therefore

---

[15] If all $\beta_j = 0$ this formula is an ARCH (q) model.

been argued that negative shocks would have larger impact on future volatility than positive shocks in financial time series[16].

Although the standard GARCH model has proven the be successful in modeling the volatility of financial returns in the presence of time-varying volatility and 'volatility clustering', it isn't able to account for asymmetry in the distribution of returns since only the size of lagged errors are assumed to explain future volatility[17]. The Basic linear GARCH model can in other word not take into account that volatility might be higher or lower depending on the sign of previous errors. In the presence of asymmetry in the distribution of returns, the standard symmetric GARCH model might therefore not be the most suitable model for the conditional variance, and the use of a modified version of the standard GARCH model might be appropriate.

As was mentioned in the introduction in this thesis only two modified versions of the GARCH model will be used, the GJR-GARCH and the QGARCH. For an incomplete, but yet extensive list of more modified GARCH models, see Bollerslev, T. (2011).

---

[16] One well-known hypothesis explaining why such dynamic would present in time series of equity returns is known as *leverage effects*. Relatively higher volatility after negative shocks is in this case said to be a consequence of that future cashflow streams, from holding a particular stock, appears more risky after fall in the stock price since the firm's debt to equity ratio then rises (see Brooks, C. (2008) p.404).

[17] Remember that the explanatory variables in the standard GARCH model are lagged conditional variances (GARCH terms) and lagged squared errors (ARCH terms), and that sign of the errors are lost when squaring them.

3. Model specifications

This section of this thesis reviews the model specifications and statistical properties of the standard GARCH(1,1) model, the GJR-GARCH(1,1) model, the QGARCH(1,1) model, and the AutoRegressive model for the mean.

3.1 Mean equation

A GARCH model could be estimated on the residuals from a model (for the returns) with ether a constant mean or a non-constant conditional mean. Since there is sometimes small but still significant autocorrelation in financial returns, GARCH models are frequently estimated on the residuals from an autoregressive model (AR) or an autoregressive moving average model (ARMA). In this thesis the model for the returns (the mean equation) will be either a AR(k) model or a model with constant mean (unconditional mean), as specified respectively in eq. 3.1.1 and eq. 3.1.2 below.

$$y_t = \emptyset_0 + \emptyset_1 y_{t-1} + \cdots + \emptyset_k y_{t-k} + \varepsilon_t \qquad (3.1.1)$$

$$y_t = \emptyset_0 + \varepsilon_t \qquad (3.1.2)$$

$$\varepsilon_t = v_t \sigma_t \qquad (3.1.3)$$

, where y_t being the returns, v_t a strong white noise with zero mean and unit variance, σ_t the conditional standard deviation, and $\emptyset_0$ a constant term in the regression[18].

The reason for having this flexible specification is because each of the GARCH models will be estimated and tested in four separate subsamples. Which one of these equations that will be used for the mean, and what k to use if the AR(k) model will be used, are determined by the presence of autocorrelation (or lack of such) in the returns of each individual subsample.

3.2 GARCH(1,1)

The conditional variance equation of the GARCH(1,1) model is given in eq. 3.2.1 below.

---

[18] It's also common to add the assumption of v_t being normally distributed, $v_t \sim iid\ N(0,1)$. In that case it can be shown that $\varepsilon_t \sim iid\ N(0, \sigma_t)$ which can be compared with $\varepsilon_t \sim iid\ N(0, \sigma)$ under the assumption of CLRM.

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.2.1}$$

, where σ_t^2 being the one-step ahead forecast for the future variance (or the conditional variance), ε_{t-1}^2 the most resent error squared, and σ_{t-1}^2 the last estimate of the conditional variance produced by the model. The model imposes the negativity constrains: $\omega > 0$, α *and* $\beta \geq 0$, to ensure that σ_t^2 is positive.

Since the conditional variance in the GARCH model is only conditional upon the values taken by previous lags of the errors, the conditional variance could also be express as[19]:

$$Var(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sigma_t^2 \tag{3.2.2}$$

, where it's assumed that $E(\varepsilon_t) = 0$ and $Cov(\varepsilon_t\varepsilon_j) = 0, t \neq j$.

The GARCH model is stationary if $\alpha + \beta < 1$. Given stationary the model has a mean-reverting property, meaning that the models conditional variance will be depend upon the long run average variance (unconditional variance) to some extent. This is the case since the constant term is the product of $(1 - \alpha - \beta)$ times the unconditional variance ($\bar{\omega}$). If we denote $(1 - \alpha - \beta)$ as γ, the constant term can be expressed as: $\omega = \gamma * \bar{\omega}$. Since the model requires that $\gamma + \alpha + \beta = 1$, γ and therefore also ω will only be positive when $\alpha + \beta < 1$. The degree to which the model will mean-revert depends on how close $\alpha + \beta$ is to one. When $\alpha + \beta \approx 1$, the models one-step ahead forecast for the variance depends relatively little on the unconditional variance and is therefore said to have slow mean reversion or high volatility persistence. If $\alpha + \beta = 1$ the model has a 'unit root in variance'. In this case there is no mean-reversion accounted for by the model, and the model then has the same properties as an Integrated GARCH model (IGARCH) or a EWMA. If $\alpha + \beta > 1$, the model is 'non-stationarity in variance'.

When the model is stationarity the unconditional variance is given by (see Brooks, C. (2008) p.394):

$$Var(y_t) = Var(\varepsilon_t) = \frac{\omega}{1 - \alpha - \beta} = \bar{\omega} \tag{3.2.3}$$

---

[19] Remember that a GARCH(1,1) model can be thought of as an ARCH(∞) model.

Comparing eq. 3.2.1 or eq. 3.2.2 with eq. 3.2.3, we see that the GARCH(1,1) model has constant unconditional variance $\bar{\omega}$ (unconditional homoscedastic), but non-constant conditional variance σ_t^2 (conditionally heteroscedastic).

Although, the model (eq. 3.2.1) only deliver a one-period forecast, longer prediction horizons can be created by using a transformed version of the formula (eq. 3.2.4) below, and repeatedly using the previous forecast for the variance as input for the next step-ahead forecast for the variance.

$$E(\sigma_{t+k|t}^2) = \bar{\omega} + (\alpha + \beta)^k(\sigma_t^2 - \bar{\omega}) \tag{3.2.4}$$

When the model is stationary the forecast of the k-step ahead future variance will converge with the historical average variance ($\bar{\omega}$) as the prediction horizon increases, which formally can be expressed as:

$$\lim_{k \to \infty} \sigma_{t+k|t}^2 = \frac{\omega}{1 - \alpha - \beta} = \bar{\omega} \tag{3.2.5}$$

When the model has a 'unit root in variance' the forecast variance is the same for all k-steps-ahead forecasts.

3.3 GJR-GARCH(1,1)

The conditional variance equation of the GJR-GARCH(1,1) model is given in eq. 3.3.1 below:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \delta_G\varepsilon_{t-1}^2 D_{t-1} + \beta\sigma_{t-1}^2 \tag{3.3.1}$$

$$D_t = \begin{cases} 1 & if\ \varepsilon_t < 0 \\ 0 & otherwise \end{cases}$$

, where the difference to the conditional variance equation of the standard GARCH(1,1) model is the addition of $\delta_G\varepsilon_{t-1}^2 D_{t-1}$, an asymmetry parameter times the most resent error squared and a dummy variable which takes the value 1 if $\varepsilon_{t-1} < 0$.

The non-negativity constrains are similar to those of the standard GARCH model. In the GJR-GARCH these are: $\omega > 0\ and\ \alpha,\ \alpha + \delta_G,\ \beta \geq 0$, which means that its acceptable that $\delta_G < 0$ as long as $\alpha + \delta_G \geq 0$.

The model is stationary when $\delta_G < 2(1 - \alpha - \beta)$. In this case the unconditional variance is given by eq. 3.3.2 below.

$$Var(y_t) = Var(\varepsilon_t) = \frac{\omega}{1 - \alpha - \beta - 0.5\delta_G} = \overline{\omega} \tag{3.3.2}$$

When $\delta_G \neq 0$ the conditional variance will not only be conditional upon the sizes of the previous lags of errors squared, but also on the sign of the most recent error. If $\delta_G > 0$ the models one-step ahead forecast for the variance will increase in response to the most resent error being negative, by the amount of $\delta_G \varepsilon_{t-1}^2$. If $\delta_G < 0$ on the other hand, it will decrease by the same amount.

3.4 QGARCH(1,1)

The conditional variance equation of the QGARCH(1,1) model is given in eq. 3.4.1 below:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \delta_Q \varepsilon_{t-1} \tag{3.4.1}$$

, where the difference to the conditional variance equation of the standard GARCH(1,1) model is the addition of $\delta_Q \varepsilon_{t-1}$, an asymmetry parameter times the most resent error.

In addition to the non-negativity constrains in the standard GARCH(1,1), $\delta_Q^2 < 4\alpha\overline{\omega}(1 - \alpha - \beta)$ is required to ensure positivity of σ_t^2.

The model is stationary if $\alpha + \beta < 1$. The stationarity of the model does in other words not depend on the asymmetry parameter. When the model is stationary the unconditional variance is given just as in the standard GARCH model (eq. 3.2.3). The mean-reverting property of this the QGARCH(1,1) model is also the same as in the standard GARCH(1,1) model. The closer $\alpha + \beta$ is to 1, the less will the one-step ahead forecast for the variance depend on the unconditional variance, thus having slower mean reversion or higher volatility persistence.

For some given values of ω, $\alpha\varepsilon_{t-1}^2$ and $\beta\sigma_{t-1}^2$, when $\delta_Q > 0$ the models one-step ahead forecast variance will increase when most resent error is positive, and decrease when most resent error is negative, by the amount of $|\delta_Q\varepsilon_{t-1}|$. If $\delta_Q < 0$ on the other hand, it will increase when most resent error is negative, and decrease when most resent error is positive, by the amount of $|\delta_Q\varepsilon_{t-1}|$. The case of $\delta_Q < 0$ can therefore be interpreted as the model predicts the one-step ahead future variance to be relatively higher after negative shocks compared to after positive shocks.

4. Data and Research Method

4.1 Variables and sample periods specifications

The data analyzed in this thesis are daily log-returns, calculated from daily close spot prices for copper (Grade A Cash U$/MT) noted at London Metal Exchange between 1993-07-20 and 2012-03-21 (Source of data: Datastream). The log-returns where calculated using the standard formula;

$$y_t = \ln\left(\frac{p_t}{p_{t-1}}\right) = \ln p_t - \ln p_{t-1} \tag{4.1.1}$$

where y_t being the log-returns and p_t being the close spot price for day t.

The total number of observations in the analyzed period is 4872, with the first observation at 1993-07-21 and the last at 2012-03-21[20]. The sample is further divided into four subsamples, each of which consists of one in-sample and one out-of-sample period. The GARCH models are estimated in the in-sample periods and their forecasting performance are tested by forecasting in the corresponding out-of-sample period. Each GARCH model will therefore be estimated and tested four times over the entire sample period.

The dates for which each subsamples, in-sample and out-of sample ranges between are specified in table 4.1.1. The in-sample period consists of 1155 observations each, making them approximately four and a half years. The out-of-sample-period consists of 252 observations each so that the estimated models will be forecasted over the subsequent year (assuming that one year consists of 252 trading days).

To make maximum use of the total sample, out-of-sample periods are overlapping the in-sample period, thus letting one in-sample period to start right after the previous one. This shouldn't cause any problems since the GARCH model estimation and prediction power testing are fully independent between subsamples. A graphical depicture of this structure can be seen in figure 4.1.1.

Table 4.1.1 - Specification of subsamples

Subsample	In-sample period	Out-of-sample period
Subsample 1	1993/07/21 - 1997/12/23	1997/12/24 - 1998/12/09
Subsample 2	1997/12/24 - 2002/05/28	2002/05/29- 2003/05/14
Subsample 3	2002/05/29 - 2006/10/31	2006/11/01 - 2007/10/17
Subsample 4	2006/11/01 - 2011/04/05	2011/04/06 - 2012/03/21

Notes: Each in-sample period and out-of sample period has 1155 and 252 observations of daily log-returns respectively.

---

[20] Note that the observation at 1993-07-21 is the return between 1993-07-20 and 1993-07-21.

Figure 4.1.1 - Graphical overview of the subsample layout
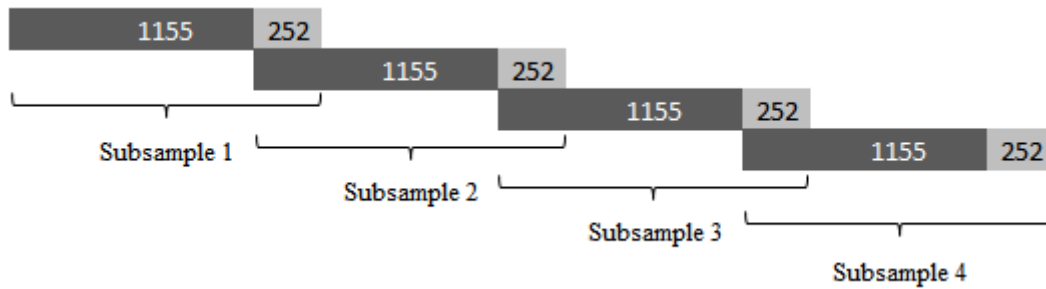
Notes: In-sample periods are marked will dark gray and the out-of-sample periods are marked with light gray. The number of observations for each period is written within each corresponding cell.

NOTE, however, that during the general data analysis in section 4.2, it was found that the LME data included observations of the spot close price at days for which the market had been closed, such as Christmas, New Year and Easter holidays. Since it makes no sense to estimate or forecast the variance for days when the return and variance is known to be zero, all observations with log-returns equal to zero was therefore removed (approximately 60 observations from each subsample) before proceeding with the analysis, model estimation and forecasting in section 4.3 to section 4.6.

4.2 Data exploration

The daily log-returns and spot price level for the entire sample period are graphed in figure 4.2.1 below. From this figure it, at least visually, seems to be some volatility-clustering present in the copper returns, suggesting that the volatility might be time-varying (heteroscedastic) with some positive autocorrelation in squared returns. The volatility-clustering dose however seem to be larger in in-sample 1, 3 and 4 than in in-sample 2, suggesting that the possibility of modeling the conditional variance with GARCH models might vary over time.

This suspicion is further strengthened by positive excess kurtosis shown in table 4.2.1 below[21]. The positive excess kurtosis differs in size between the in-samples, again suggesting that benefits of modeling the conditional variance with GARCH models might vary over time.

---

[21] A consequence of heteroscedastic unconditional variance is that the distribution of the errors (here the demeaned copper returns) around the unconditional mean will be leptokurtic (have a positive excess kurtosis). A leptokurtic distribution has fatter tails and is more peaked around the mean compared to a normal distribution.

The kurtosis is particularly high in in-sample 1 and 3. The kurtosis is about the same in in-sample 4 and in-sample 2.

In table 4.2.1 we also see that there is skewness in the distribution of returns in all of the in-sample period, suggesting that the asymmetric GARCH models might be better suitable for modeling the volatility than the symmetrical GARCH model. Interesting to note here is also that the skewness is negative in in-sample 1, 3 and 4 (which is the common case in returns of for example equities), but positive in in-sample 2. The skewness was, furthermore, significantly larger in in-sample 1 than in the other in-samples.

Figure 4.2.1 - Daily log-returns and spot price levels
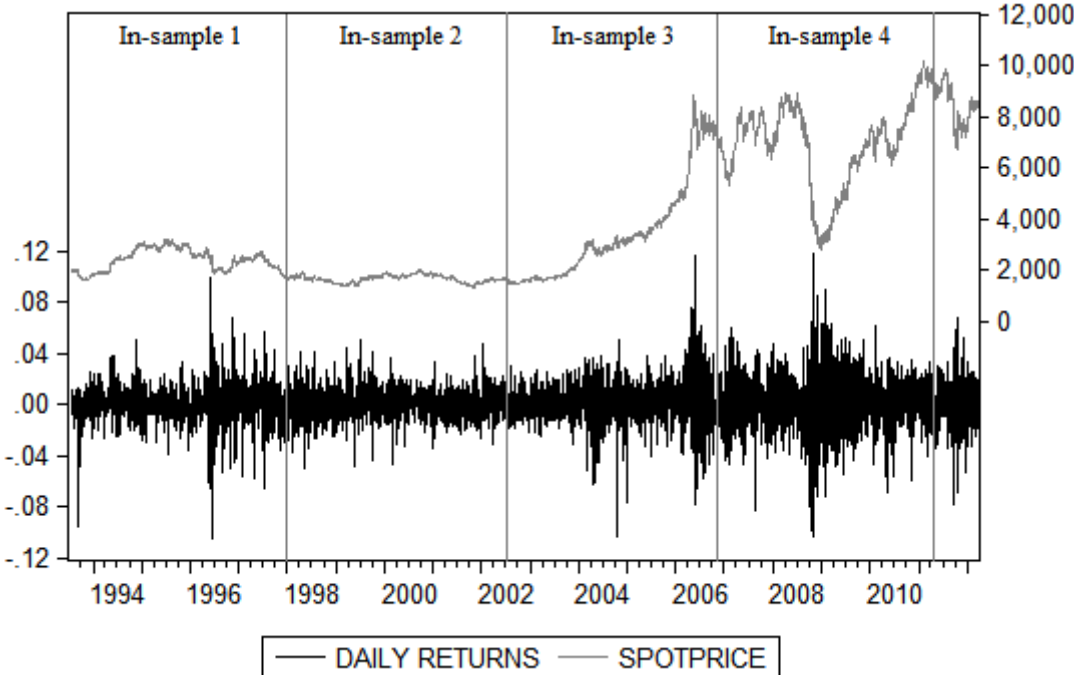
Notes: Daily log-returns are shown in black and their sizes are given by the left axis scale. The spot price is shown in light gray and its level is given by the right axis scale. For further illustration of the subsample periods, the in-sample periods are separated by horizontal lines.

Table 4.2.1 - Descriptive statistics of log-returns

Period	Mean	Std. Dev.	Skewness	Kurtosis
Entire sample	0,000303	0,017144	-0,225917	7,848979
In-sample 1	-0,0000795	0,014937	-0,702237	11,61832
In-sample 2	-0,0000519	0,011656	0,124801	4,827180
In-sample 3	0,001304	0,016791	-0,172543	8,686449
In-sample 4	0,000208	0,022739	-0,164349	5,042145

Notes: Descriptive statistics of the log-returns in each in-sample period and the entire sample.

To summarize, the copper price returns in all of the in-samples seem to have some volatility clustering and a distribution of returns that is characterized by skewness and excess kurtosis. The degree to which this is the case does however appear to be larger in in-sample 1, 3 and 4 than in in-sample 2. Some prior expectations are therefore that; it should be appropriate to try modeling the variance with the GARCH models, the asymmetric GARCH models might better explain the variance than the symmetrical GARCH model, and that the GARCH models might be more successful in modeling the variance in in-sample 1, 3 and 4 than in in-sample 2.

Before moving on the next section it should also be worth noticing the radical price level increase which occurred over in-sample 3, and the fact that the general level of the volatility has increased over the last decade compared to the previous one.

4.3 Selection of the model for the conditional mean

As was stated in section 3, each of the GARCH models will be estimated on the residuals of either mean equation 3.1.1 or 3.1.2. Therefore, before estimating the GARCH models, it will in this section be analyzed which of the two mean equations that is most suitable for each one of the subsamples.

In the case of significant autocorrelation (AC) in the returns, there might be a possibility to model the returns based on the values taken by precious lags of returns through an autoregressive model (eq. 3.1.1). Although the extent to which it is possible to model financial returns by an autoregressive model usually is rather small, there appears to be no good motivation for instead using an unconditional mean in the presence of significant AC in the returns and thus neglecting a small but still available improvement of the one-step ahead expectation for the returns[22]. Therefore, the AC of the returns and the significance of the estimated parameters in eq. 3.1.1 are investigated and presented in table 4.3.1 below[23].

---

[22] One reason for why an AR model would not be preferred for the mean equation, when the aim is to analyze the forecasting performance of GARCH models, even in the presence of AC, is however discussed in section 6 of this thesis.

[23] Only the standard GARCH model will be used for these tests since it's assumed that the standardized residuals from the GJR-GARCH or QGARCH will not differ enough to lead to any different conclusions about what mean equation that would be appropriate.

Model 4.3.1 - Regression outputs and diagnostics test of standardized residuals

In-sample	AR(1)-GARCH(1,1) $\emptyset_1$	AR(2)-GARCH(1,1) $\emptyset_1$	$\emptyset_2$	Corr_C	Corr_{AR1}	Corr_{AR2}
1	0,1049	0,1166	-0,1057	0,007	0,101	0,874
	(3,1205)*	(3,2759)*	(-2,6883)*	0,004	0,037	0,327
	(0,0018)^P	(0,0011)^P	(0,0072)^P	0,003	0,026	0,214
2	0,0213	0,0218	-0,0429	0,462	0,434	0,614
	(0,6596)*	(0,6770)*	(-1,4904)*	0,797	0,793	0,907
	(0,5095)^P	(0,4984)^P	(0,1361)^P	0,537	0,528	0,591
3	-0,0285	-0,0299	-0,0506	0,446	0,326	0,387
	(-0,8192)*	(-0,8593)*	(-1,6160)*	0,406	0,388	0,439
	(0,4127)^P	(0,3902)^P	(0,1061)^P	0,607	0,553	0,613
4	-0,0529	-0,0521	0,0149	0,501	0,690	0,540
	(-1,6128)*	(-1,5824)*	(0,4632)*	0,762	0,863	0,806
	(0,1068)^P	(0,1136)^P	(0,6432)^P	0,922	0,957	0,941

Notes: Regression output of the mean equation from estimation of AR(k)-GARCH(1,1) models on the log-returns in each corresponding in-sample. Estimation Method: ML - ARCH (Marquardt) - Normal distribution, with Bollerslev-Wooldridge robust standard errors & covariance. The first row of the $\emptyset_1$ and $\emptyset_2$ columns shows the estimated parameter values. The second and third row, marked with a * and a ^{P}, shows the corresponding z-statistics and the p-values (respectively). In the Corr columns are p-values based on a Ljung-Box joint test for autocorrelation in the standardized residuals. For the first row in these columns 5 lags where considered, in the second 10 lags and in the third row 15 lags. In Corr_C column the standardized residuals of a GARCH(1,1) model with constant mean was tested. In Corr_{AR1} column the standardized residuals of a AR(1)-GARCH(1,1) model was tested. In Corr_{AR2} column the standardized residuals of a AR(2)-GARCH(1,1) model was tested.

In the table we first see that only for in-sample 1 does the Ljung-Box joint test statistic reject the null hypothesis of no autocorrelation in the standardized residuals of the GARCH(1,1) model with a constant mean for the returns. The null hypothesis is rejected at 1 % significance level in in-sample 1, and can't be rejected even at a 10 % significance level for the other in-samples.

Estimating the AR(1)-GARCH(1,1) model, we see (as expected given the discussed observation above) that $\emptyset_1$ is significant at a 10 % level only for the model estimated in in-sample 1. For in-sample 1 we also see that, even after estimating the AR(1)-GARCH(1,1) model, the Ljung-Box joint test statistic still rejects the null hypothesis of no autocorrelation in the standardized residuals at 5 % significance level when 10 or 15 lags are considered. This suggests that a higher order of the autoregressive model in the mean might be appropriate. Naturally, we also see that the Ljung-Box joint test statistic cannot reject the null hypothesis

of no autocorrelation in the standardized residuals by an AR(1)-GARCH(1,1) model estimated in the other in-samples.

Finally, estimating the AR(2)-GARCH(1,1) model we see that $\emptyset_1$ as well as $\emptyset_2$ is significant at a 10 % level only for the model estimated in in-sample 1, where both $\emptyset_1$ and $\emptyset_2$ then are significant at a 1 % level. For in-sample 1 we now see that the Ljung-Box joint test statistic do not reject the null hypothesis of no autocorrelation in the standardized residuals at a 10 % significant level. Quickly estimating an AR(3)-GARCH(1,1) model for in-sample 1 (not covered in the table), it was, as expected, also observed that $\emptyset_3$ was not significant at a 10 % level.

Based on these results a constant unconditional mean (eq. 3.1.2) was selected for the GARCH models in subsamples 2 to 4, while an autoregressive model of the 2nd order was selected for the GARCH models in subsample 1.

The models that will be estimated and tested for their prediction power of the future variance in each respective subsample are listed in table 4.3.2 below. The random walk model is not included here but will, as mentioned, be used as a benchmark in all of the subsamples when evaluating the forecasting performances of the GARCH models.

Table 4.3.2 - The GARCH models that are estimated and tested in this thesis

Subsample 1	Subsample 2	Subsample 3	Subsample 4
AR(2)-GARCH(1,1)	GARCH(1,1)	GARCH(1,1)	GARCH(1,1)
AR(2)-GJR-GARCH(1,1)	GJR-GARCH(1,1)	GJR-GARCH(1,1)	GJR-GARCH(1,1)
AR(2)-QGARCH(1,1)	QGARCH(1,1)	QGARCH(1,1)	QGARCH(1,1)

Notes: The selection of the mean equation was based on the analysis in this section. Exclusion of an AR(k) model for the conditional mean equation (as in subsample 2, 3 and 4) means that only a constant is included in the mean equation as in eq. 3.1.2.

4.4 Motivation for the use of GARCH models

In section 4.2 there was found some indications of autocorrelation in the squared returns and heteroscedastic variance of the returns. However, before estimating the GARCH models listen in table 4.3.2, the appropriateness of doing so will be further examined in this section. This will be done by analyzing the squared residuals, produced by estimating each in-samples' corresponding mean equation by OLS regression, with Ljung-Box joint test for AC and Engel, R. F. (1982) test for ARCH effects.

Furthermore, in section 4.2 it was also found that the distributions of the returns in all of the in-samples were skewed, suggesting that the use of asymmetric GARCH models might be

appropriate. This will also be further analyzed in this section by conducting an Engel, R. F. and Ng, V. (1993) test for asymmetry effects in each of the in-samples.

Looking at table 4.4.1 below, we see that the Ljung-Box joint test statistic, and the F- and LM-statistic from an Engel, R. F. (1982) test for ARCH effects with five lags, are all highly statistically significant in in-samples 1, 3 and 4, thus rejects the null hypothesis of no ARCH effects or autocorrelation in the squared residuals at a 1 % significance level for these in-samples. For in-sample 2 on the other hand, the null hypotheses of neither the Ljung-Box joint test nor the Engel, R. F. (1982) test for ARCH effects could be rejected at a 10 % significance level.

Model 4.4.1 - Test for ARCH effects

In-sample	F-stat.	LM-stat.	Corr.
1	34,866	151,21	0,000
	$(0,0000)^P$	$(0,0000)^P$	0,000
			0,000
2	1,594	7,96	0,148
	$(0,1588)^P$	$(0,1586)^P$	0,465
			0,246
3	15,442	72,53	0,000
	$(0,0000)^P$	$(0,0000)^P$	0,000
			0,000
4	45,966	191,29	0,000
	$(0,0000)^P$	$(0,0000)^P$	0,000
			0,000

Notes: Tests based on the residuals from OLS regression of the mean equation in each in-sample. In the Corr. column are p-values based on a Ljung-Box joint test for autocorrelation in the squared residuals. For the first row in these columns 5 lags where considered, in the second 10 lags and in the third row 15 lags. F- and LM-statistic (Observations * R-squared) are from a Engel's (1982) test for ARCH effects with five lags. [P] denotes p-values based on corresponding F and LM-statistics.

These results suggest that trying to model the variance with GARCH models should be appropriate, and that doing so could be expected to be more suitable for subsample 1, 3 and 4, than for subsample 2. These results are consistent with the findings in section 4.2. Two prior expectations before estimating the GARCH models are therefore that α and β over all have a high probability of being statistically significant, and that this probability is higher for the models estimated at in-sample 1, 3 and 4 returns, than for the models estimated at in-sample 2 returns.

As previously mentioned, an Engel, R. F. and Ng, V. (1993) test for asymmetry effects will be conducted for each of the in-sample periods to analyze the appropriateness of using asymmetric GARCH models. The regression model of this test can formally be written as:

$$\hat{\varepsilon}_t^2 = \varphi_0 + \varphi_1 S_{t-1}^- + \varphi_2 S_{t-1}^- \hat{\varepsilon}_{t-1} + \varphi_3 S_{t-1}^+ \hat{\varepsilon}_{t-1} + v_t \qquad (4.4.1)$$

, where $\hat{\varepsilon}_t$ are the residuals from estimating the mean equation. S_{t-1}^- is an indicator dummy which takes the value 1 if $\varepsilon_{t-1} < 0$, and the value 0 otherwise. S_{t-1}^+ is another indicator dummy which takes the value 0 if $\varepsilon_{t-1} < 0$, and the value 1 otherwise. v_t is an i.i.d. error term.

Statistical significance of the parameter φ_1 indicates sign biases asymmetry, meaning that the sign of the most resent residual has an explanatory effect on the one step-ahead squared residual. Statistical significance of the parameter φ_2 and φ_3 indicates size biases asymmetry, meaning that not only the sign, but also the size of most resent residual has an explanatory effect on the one-step ahead squared residual.

The regression outputs from estimating this test model in each of the in-samples are presented table 4.4.2 below.

Model 4.4.2 - Engel and Ng (1993) asymmetry test

In-sample	φ_1	φ_2	φ_3
1	-1,12E-04	-0,0136	-0,0015
	(-1,9917)*	(-5,5716)*	(-0,5160)*
	(0,0466)[P]	(0,0000)[P]	(0,6060)[P]
2	2,29E-06	0,0008	0,0006
	(0,0911)*	(0,5166)*	(0,3832)*
	(0,9274)[P]	(0,6055)[P]	(0,7016)[P]
3	6,64E-05	-0,0066	0,0086
	(0,9807)*	(-2,4521)*	(2,9615)*
	(0,3269)[P]	(0,0144)[P]	(0,0031)[P]
4	1,22E-06	-0,0155	0,0181
	(0,0130)*	(-5,6977)*	(6,0610)*
	(0,9895)[P]	(0,0000)[P]	(0,0000)[P]

Notes: Engel, R. F. and Ng, V. (1993) test for asymmetry effects. Estimation method: OLS. The first row of the φ_1, φ_2 and φ_3 columns shows the estimated parameter values. The second and third row, marked with a * and a [P], shows corresponding t-statistics and the p-values (respectively).

As we here, φ_1 is not significant for in-sample 2, 3 and 4, and just barely significant at a 5 % significance level for in-sample 1. However, since the conditional variance equations of the

28

GJR-GARCH and the QGARCH are designed to be able to deal with size biased asymmetry, there shouldn't be a too big focus on the significance of φ_1 when evaluating the appropriateness of using these two asymmetric GARCH models.

In terms of statistic and economic significance of φ_2 and φ_3 we see that in-sample 4, followed by in-sample 3, appears to be the most suitable in-samples for the use of asymmetric GARCH models. It's however more uncertain whether or not in-sample 1 is suitable for the use of asymmetric GARCH models since φ_2, but not φ_3, is statistically significant. In-sample 2 does not appear suitable for asymmetric GARCH models since nether φ_2 nor φ_3 is statistically significant at a 10 % level.

Based on these results, the statistic and economic significance of the asymmetry terms of the estimated asymmetric GARCH models could therefore be expected to be largest in in-sample 4, second largest in in-sample 3, second lowest in in-sample 2, and lowest in in-sample 1.

4.5 Model estimation

The regression outputs from the GARCH model estimations, together with diagnostics test and statistics of the standardized residuals, are presented in tables 4.5.1 - 4.5.4 below.

The models were estimated using Maximum Likelihood - ARCH (Marquardt) - Normal distribution. Bollerslev-Wooldridge robust standard errors & covariance was used in all of the GARCH model estimations since the assumption of v_t being normally distributed, and ε_t being conditionally normally distributed, is clearly unlikely to hold when examining the skewness and kurtosis of the standardized residuals[24]. This was the reason for using the Bollerslev-Wooldridge robust standard errors & covariance when estimating the models in section 4.3 as well.

---

[24] v_t being normally distributed, and ε_t being conditionally normally distributed can formally be expressed as: $\varepsilon_t \sim N(0, \sigma_t^2)$, and $\varepsilon_t = v_t \sigma_t$, $v_t \sim N(0,1)$. Since v_t is unknown the sample counterpart, the standardized residuals ($\hat{v}_t$) was analyzed for normality. $\hat{v}_t$ was derived by dividing the models residual at time t by the conditional standard deviation at time t. Formally: $\hat{v}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$. Although not included in the tables, Jarque-Bera test statistic rejected the null-hypothesis of normality in the distribution of the standardized residuals from each of the twelve GARCH models at a 1 % significance level.

Model 4.5.1 - Regression outputs from estimating the AR(2)-GARCH models at in-sample 1 returns

Model				Regression output and parameter estimates						Standardized residuals stat. and diagnostics				
Model	ω	γ	α	β	δ_G	δ_Q	AIC	SBIC	Skew.	Kurt.	Corr.	Corr.[2]	ARCH
GARCH(1,1)	1,22E-05	0,0550	0,1073	0,8416			-5,718	-5,691	-0,677	8,871	0,874	0,787	0,960
	(2,2096)*		(3,2427)*	(18,567)*							0,327	0,946	0,960
	(0,0271)[P]		(0,0012)[P]	(0,0000)[P]							0,214	0,931	
GJR-GARCH(1,1)	1,23E-05	0,0591	0,1037	0,8407	0,0081		-5,716	-5,685	-0,671	8,833	0,875	0,792	0,962
	(2,1979)*		(2,7236)*	(18,616)*	(0,1555)*						0,325	0,945	0,962
	(0,0280)[P]		(0,0065)[P]	(0,0000)[P]	(0,8764)[P]						0,215	0,936	
QGARCH(1,1)	1,30E-05	0,0581	0,1131	0,8331		-0,0004	-5,717	-5,685	-0,629	8,646	0,877	0,793	0,963
	(2,2351)*		(3,4235)*	(18,367)*		(-0,5319)*					0,318	0,940	0,963
	(0,0254)[P]		(0,0006)[P]	(0,0000)[P]		(0,5948)[P]					0,214	0,936	

Notes: Method for the estimation: ML - ARCH (Marquardt) - Normal distribution, with Bollerslev-Wooldridge robust standard errors & covariance. The first row of the ω, α, β, δG and δQ columns shows the estimated parameter values. The second and third row, marked with a * and a [P], shows the corresponding z-statistics and the p-values (respectively). γ was calculated as $1-\alpha-\beta$. The skewness and kurtosis of the standardized residuals can be compared to those from the residuals when only the mean eq. was estimated (Skewness = -0,572 and Kurtosis = 10,535). In the Corr and Corr[2] column are p-values based on a Ljung-Box joint test for autocorrelation in the standardized residuals and squared standardized residuals respectively. For the first row in these columns 5 lags where considered, in the second 10 lags and in the third row 15 lags. The values in the ARCH column are p-values based on from F- and LM-statistic from a Engel's (1982) test for ARCH effects with five lags. First row's p-value corresponds to the F-statistic and the second row's p-value corresponds to the LM-statistic. AIC: Akaike's Information Criterion. SBIC: Schwarz's Bayesian Information Criterion.

Model 4.5.2 -Regression outputs from estimating the GARCH models at in-sample 2 returns

	Regression output and parameter estimates								Standardized residuals stat. and diagnostics				
Model	ω	γ	α	β	δ_G	δ_Q	AIC	SBIC	Skew.	Kurt.	Corr.	Corr.^{2}	ARCH
GARCH(1,1)	2,50E-06	0,0406	0,0204	0,9617			-6,028	-6,010	0,079	4,762	0,462	0,648	0,631
	(1,1060)*		(1,8730)*	(39,919)*							0,797	0,961	0,629
	(0,2687)[P]		(0,0611)[P]	(0,0000)[P]							0,537	0,864	
GJR-GARCH(1,1)	2,54E-06	0,0466	0,0184	0,9602	0,0066		-6,027	-6,004	0,088	4,475	0,468	0,686	0,672
	(1,0810)*		(1,4543)*	(38,322)*	(0,4458)*						0,804	0,968	0,670
	(0,2797)[P]		(0,1459)[P]	(0,0000)[P]	(0,6557)[P]						0,544	0,878	
QGARCH(1,1)	1,32E-06	0,0581	0,0169	0,9733		-0,0002	-6,029	-6,006	0,095	4,460	0,478	0,555	0,530
	(1,0361)*		(2,0963)*	(66,276)*		(-1,8675)*					0,813	0,922	0,529
	(0,3001)[P]		(0,0361)[P]	(0,0000)[P]		(0,0618)[P]					0,547	0,772	

Notes: Method for the estimation: ML - ARCH (Marquardt) - Normal distribution, with Bollerslev-Wooldridge robust standard errors & covariance. The first row of the ω, α, β, δG and δQ columns shows the estimated parameter values. The second and third row, marked with a * and a [P], shows the corresponding z-statistics and the p-values (respectively). γ was calculated as 1-α-β. The skewness and kurtosis of the standardized residuals can be compared to those from the residuals when only the mean eq. was estimated (Skewness = 0,125 and Kurtosis = 4,827). In the Corr and Corr^2 column are p-values based on a Ljung-Box joint test for autocorrelation in the standardized residuals and squared standardized residuals respectively. For the first row in these columns 5 lags where considered, in the second 10 lags and in the tired row 15 lags. The values in the ARCH column are p-values based on from F- and LM-statistic from a Engel's (1982) test for ARCH effects with five lags. First row's p-value corresponds to the F-statistic and the second row's p-value corresponds to the LM-statistic. AIC: Akaike's Information Criterion. SBIC: Schwarz's Bayesian Information Criterion.

Model 4.5.3 -Regression outputs from estimating the GARCH models at in-sample 3 returns

Model	Regression output and parameter estimates								Standardized residuals stat. and diagnostics				
	ω	γ	α	β	δ_G	δ_Q	AIC	SBIC	Skew.	Kurt.	Corr.	Corr.^{2}	ARCH
GARCH(1,1)	2,08E-06	0,0054	0,0522	0,9424			-5,533	-5,515	-0,356	5,830	0,446	0,239	0,264
	(1,8210)*		(2,6602)*	(46,912)*							0,460	0,586	0,263
	(0,0686)^P		(0,0078)^P	(0,0000)^P							0,607	0,721	
GJR-GARCH(1,1)	1,51E-06	-0,0259	0,0672	0,9587	-0,0583		-5,551	-5,528	-0,295	5,287	0,260	0,481	0,471
	(1,8433)*		(2,7852)*	(79,187)*	(-2,0464)*						0,297	0,674	0,469
	(0,0653)^P		(0,0053)^P	(0,0000)^P	(0,0407)^P						0,425	0,603	
QGARCH(1,1)	2,67E-06	0,0581	0,0410	0,9493		0,0006	-5,539	-5,517	-0,303	5,568	0,359	0,448	0,459
	(2,3128)*		(2,8288)*	(58,318)*		(1,6926)*					0,381	0,752	0,458
	(0,0207)^P		(0,0047)^P	(0,0000)^P		(0,0905)^P					0,484	0,775	

Notes: Method for the estimation: ML - ARCH (Marquardt) - Normal distribution, with Bollerslev-Wooldridge robust standard errors & covariance. The first row of the ω, α, β, δG and δQ columns shows the estimated parameter values. The second and third row, marked with a * and a ^{P}, shows the corresponding z-statistics and the p-values (respectively). γ was calculated as 1-α-β. The skewness and kurtosis of the standardized residuals can be compared to those from the residuals when only the mean eq. was estimated (Skewness = -0,173 and Kurtosis = 8,686). In the Corr and Corr^2 column are p-values based on a Ljung-Box joint test for autocorrelation in the standardized residuals and squared standardized residuals respectively. For the first row in these columns 5 lags where considered, in the second 10 lags and in the third row 15 lags. The values in the ARCH column are p-values based on from F- and LM-statistic from a Engel's (1982) test for ARCH effects with five lags. First row's p-value corresponds to the F-statistic and the second row's p-value corresponds to the LM-statistic. AIC: Akaike's Information Criterion. SBIC: Schwarz's Bayesian Information Criterion.

Model 4.5.4 -Regression outputs from estimating the GARCH models at in-sample 4 returns

Model	ω	γ	α	β	δ_G	δ_Q	AIC	SBIC	Skew.	Kurt.	Corr.	Corr.^{2}	ARCH
				Regression output and parameter estimates					Standardized residuals stat. and diagnostics				
GARCH(1,1)	9,83E-06	0,0207	0,0744	0,9049			-4,890	-4,872	-0,242	3,562	0,501	0,391	0,414
	(1,8298)*		(3,7365)*	(34,531)*							0,762	0,709	0,413
	(0,0673)^P		(0,0002)^P	(0,0000)^P							0,922	0,782	
GJR-GARCH(1,1)	7,10E-06	0,0545	0,0223	0,9232	0,0786		-4,902	-4,879	-0,196	3,393	0,442	0,301	0,292
	(1,9422)*		(1,4979)*	(52,364)*	(3,1085)*						0,742	0,690	0,292
	(0,0476)^P		(0,1342)^P	(0,0000)^P	(0,0019)^P						0,907	0,744	
QGARCH(1,1)	7,50E-06	0,0581	0,0548	0,9306		-0,0016	-4,910	-4,887	-0,098	3,283	0,143	0,675	0,676
	(1,9733)*		(3,8982)*	(52,391)*		(-3,4629)*					0,088	0,952	0,673
	(0,0485)^P		(0,0001)^P	(0,0000)^P		(0,0005)^P					0,252	0,995	

Notes: Method for the estimation: ML - ARCH (Marquardt) - Normal distribution, with Bollerslev-Wooldridge robust standard errors & covariance. The first row of the ω, α, β, δG and δQ columns shows the estimated parameter values. The second and third row, marked with a * and a ^{P}, shows the corresponding z-statistics and the p-values (respectively). γ was calculated as 1-α-β. The skewness and kurtosis of the standardized residuals can be compared to those from the residuals when only the mean eq. was estimated (Skewness = -0,164 and Kurtosis = 5,042). In the Corr and Corr^2 column are p-values based on a Ljung-Box joint test for autocorrelation in the standardized residuals and squared standardized residuals respectively. For the first row in these columns 5 lags where considered, in the second 10 lags and in the third row 15 lags. The values in the ARCH column are p-values based on from F- and LM-statistic from a Engel's (1982) test for ARCH effects with five lags. First row's p-value corresponds to the F-statistic and the second row's p-value corresponds to the LM-statistic. AIC: Akaike's Information Criterion. SBIC: Schwarz's Bayesian Information Criterion.

Before analyzing the model estimation outputs in more detail some general results are worth pointing out. To begin with, the null hypothesis of the Ljung-Box joint test for AC in the standardized residuals and squared standardized residuals, and the null hypothesis of Engel, R. F. (1982) test for ARCH effects, cannot be rejected at a 10 % significance level for any of the models (one not so worrisome exception is however the Ljung-Box joint test for AC in the standardized residuals from the QGARCH model in in-sample 4, which has a p-value of 0,088 when 10 lags are considered). This suggests that the models' specifications are sufficient to capture the AC in the ordinary residuals and squared ordinary residuals in all of the in-sample periods, and higher orders of AR-GARCH models shouldn't therefore be needed.

Secondly, all models were found to be stationary, and all models except for the QGARCH in in-sample 4 did fulfill the non-negativity constrains. In table 4.5.5 some calculations are presented to easier control that so is the case. The fact that the QGARCH in in-sample 4 didn't fulfill the non-negativity constrains shouldn't however be a problem for the end results since the model didn't produce any negative variance forecasts.

And finally, the weight distribution between γ, α and β has varied quite a bit between the in-sample periods. No clear trend in this variation can however be determined. The relatively low β:s and relatively high α:s in in-sample 1 do however stand out in comparison to the other in-samples. The α:s in in-sample 2 are also relatively low, which is probably explained by the lack of AC in the squared residuals from the mean equation. Also worth pointing out is that there does appear to be a mean reverting effect incorporated in most of the estimated models, suggesting that GARCH models are more suitable for modeling the conditional variance than an EWMA.

Model 4.5.5 - Non-negativity and stationarity of GJR-GARCH and QGARCH

In-sample	Non-neg. GJR $\alpha + \delta_G$	Stat. GJR $2(1-\alpha-\beta)$	Non-neg. QGARCH δ_Q^2	$4\alpha\omega$
1	0,1118	0,1182	1,60E-07	5,86E-06
2	0,0250	0,0932	5,99E-08	8,95E-08
3	0,0089	-0,0517	3,46E-07	4,38E-07
4	0,1009	0,1090	2,44E-06	1,64E-06

Notes: The table includes some results from calculations necessary to check the stationarity requirements and non-negativity constrains for the GJR-GARCH and the QGARCH.

Now, starting by examining the statistical significance of α and β for the standard symmetrical GARCH(1,1) model, we see that the β:s are highly statistically significant (p-values are zero down to four digits) in all four subsamples. The α:s, on the other hand, are

statistically significant at a 1 % level only for in-sample 1, 3 and 4 (p-value = 0,0611 for in-sample 2). Furthermore, looking at the kurtosis of the distribution of the standardized residuals from the standard symmetrical GARCH(1,1) model, we also see that the kurtosis has decreased substantially in comparison to the kurtosis of the distribution of the ordinary residuals in these three in-samples, but just marginal for in-sample 2[25]. These results are in line with our pre-estimation expectations discussed in the previous section, and thus demonstrate the relevance of such pre-estimation inspection and testing of the data.

Now, trying to determine which one of the GARCH models that are the most suitable one for each in-sample, we first see that in in-sample 1 there isn't much difference between the models, judging by the AIC and SBIC values and the level of kurtosis and skewness in the standardized residuals. This is not very surprising when considering the statistical insignificance and low economic significance of the asymmetry parameters in the GJR-GARCH and QGARCH model. One first conclusion would therefore be that the GJR-GARCH and QGARCH model was not able to model the conditional variance significantly better than the standard symmetric GARCH model in this in-sample. This is, however, rather surprising since the level of skewness in the distribution of the ordinary residuals was about four times higher for in-sample 1 than for the other in-samples[26].

For the models estimated in in-sample 2, we see that the asymmetric GARCH models was actually able to lower the kurtosis more than the standard GARCH model compared to in-sample 1 (when the overall level of kurtosis is considered). The asymmetry parameter in the QGARCH model was furthermore statistically significant at a 10 % level. However, for the GJR-GARCH model, α and the asymmetry parameter was not statistically significant at a 10 % level in this in-sample either. The AIC and SBIC values are, on the other hand, still not showing any significant differences between the models. The general conclusion would however, all things considered, be that; although the difference isn't very large in this in-sample either, the QGARCH model did possibly model the conditional variance slightly better than the standard symmetric GARCH model in this in-sample.

---

[25] When the conditional variance can be successfully (to some degree) modeled with GARCH models, some (but usually not all) of the leptokurtosis in the unconditional distribution of returns are captured by the GARCH model, thus making $\hat{v}_t$ less leptokurtic than $\hat{\varepsilon}_t$ (see Brooks, C. (2008) p.399).

[26] Remember that the skewness in the distribution of returns often is used as the main argument for the use of asymmetric GARCH models.

Moving on to in-sample 3, we see that the difference between the models, in terms of kurtosis, AIC and SBIC values, appears to have increased. The kurtosis, AIC and SBIC are all lower for the asymmetric GARCH models compared to the standard symmetrical GARCH model. The asymmetry parameters are in this in-sample statistically significant at a 5 % and a 10 % level for the GJR-GARCH and the QGARCH respectively. The GJR-GARCH model does overall appear to be the best model for the variance in in-sample 3. One noteworthy observation for this in-sample is however the sign of the asymmetry parameters. Only for in-sample 3 is the sign of the asymmetry parameter negative for the GJR-GARCH and positive for the QGARCH. The implication of this is that the models predict the one-step ahead variance to be higher when the sign of the most recent error is positive, compared to when the sign of the most recent error is negative. This is rather unusual and is probably caused by the fact that in-sample 3 mainly consists of the extraordinary price increase between 2004 and mid 2006.

Finally, taking a look at the model estimations for in-sample 4 we see that the in AIC and SBIC values again are significantly different between the asymmetric GARCH models and the standard GARCH model. We can also see that both the economic and statistical significance of the asymmetry terms of the GJR-GARCH and QGARCH model are higher here than for any other in-sample. The statistical significance of α and β is also highest in this in-sample, with the exception of α in the GJR-GARCH model (which could be seen as an even stronger evidence of the asymmetric models superiority in this in-sample)[27]. The QGARCH model did however appear to best explain the variance in this in-sample since it had the lowest kurtosis, AIC and SBIC of the three models (almost all excess kurtosis was removed).

---

[27] Remember that α is the corresponding parameter to the most resent squared residual, and that the asymmetry parameter is the corresponding parameter to the most resent squared residual if the most resent residual is negative. If α is statistically significant in the GARCH model but only β and the asymmetry parameter are statistically significant in the GJR-GARCH model, this could be interpreted as that the most resent squared residual actually only holds a significant explanatory power if the most resent residual is negative, but still makes α in the GARCH model statistically significant since a large portion (presumably about half in this in-sample) of the most resent residual is negative.

To summarize the results in this section, we first saw that the results from the Ljung-Box joint test for AC in the standardized residuals and squared standardized residuals, and Engel, R. F. (1982) test for ARCH effects in the standardized residuals, strongly suggests that the models' specifications are sufficient to capture the AC in the ordinary residuals and squared ordinary residuals in all of the in-sample periods. We also saw that the models was stationary and fulfilled the non-negativity constrains (or didn't produce any negative variance forecasts).

The GARCH models' ability to model the conditional variance (in-sample) was generally satisfying for in-sample 1, 3 and 4, but not for in-sample 2. This result was accurately predicted by Ljung-Box joint test for AC the squared residuals from the mean equation, and the Engel, R. F. (1982) test for ARCH effects in section 4.4. The results also suggested that the GARCH models' ability to model the conditional variance might have increased over time. In in-sample 4, followed by in-sample 3, the GARCH models could remove the largest portion of excess kurtosis in the distribution of the residuals from the mean equation. Between 75 to 85 % of the excess kurtosis were removed by the GARCH models in in-sample 4.

Furthermore, the economic and statistical significance of the asymmetry term appears to have increased over time. In in-sample 3 and 4 the asymmetry parameters of the QGARCH and GJR-GARCH model did provide significant explanatory power. The skewness in the distribution of demeaned returns or the distribution of the residuals from the mean equation were, however, not a good indicator of whether the two asymmetric GARCH models would be better than the standard symmetrical GARCH model in modeling the conditional variance. The Engel, R. F. and Ng, V. (1993) test for asymmetry effects did, however, provide useful indications of this.

In the next section the forecasting performance of the estimated models will be tested. Although it's generally hard to predict how well the models will do in out-of-sample forecasting based on the results from the in-sample estimations, two things could be expected.

Firstly, due to the high statistically significance of the parameters in the GARCH models and the success in lowering the kurtosis, the GARCH models should be able to deliver more accurate forecasts of the future variance than a Random Walk model and the unconditional variance, given that the out-of-sample data not differs drastically from the in-sample data.

Secondly, the differences in forecasting performances between the GARCH models should be rather small, but presumably slightly lager in subsample 3 and 4 compared to subsample 1 and 2. Also, if the asymmetric models would prove to perform significantly better than the symmetric GARCH model, this is likely to occur in subsample 3 or 4.

4.6 Out-of-sample forecasting

In this section of the thesis, the out-of-sample forecasting performances of the estimated GARCH models are tested. This is done by comparing the GARCH models' rolling one-step ahead forecast variance to a proxy for the 'real' realized one day variance over the entire out-of-sample period[28]. Mean Squared Error (MSE), the Mean Absolute Error (MAE) and the Median Squared Error (MedSE) are then calculated based on the deviations between the forecasts and the proxy.

As a method of evaluating the forecasting performances of the GARCH models, MSE, MAE and MedSE for a Random Walk (No Change) model will also be calculated to use as a benchmark. This is a common method of evaluating the out-of-sample forecasting performance of GARCH models, and is used in studies such as the one conducted by Franses, P.H. and van Dijk, D. (1996); Watkins, C. and McAleer, M. (2008); Bracker, K. and Smith, K.L. (1999 and 2003). Another well know method that doesn't involve a RW-model as a benchmark was introduced by Day, T. and Lewis, C. (1992). However, since the method of using a benchmark such as the RW-model appears to be more common, this approach will be used in this thesis.

In addition to the RW-model, the unconditional (average historical) error variance in the corresponding in-sample period will also be used as a forecasting benchmark. This approach was not used in the studies mentioned above, but should however make both theoretically and practical sense since the volatility of financial returns are widely known to be 'mean-reverting' (i.e. stationary), and since using the unconditional variance as a forecast always is an available option for the market participants. Furthermore, Due to the 'mean-reversion', a prior expectation is that the unconditional variance will produce better variance forecasts than the RW-model.

In this thesis the proxy for the 'real' realized one day variance is the squared realized errors (residuals) from the specified mean equation in each model, which formally can be expressed as:

$$\varepsilon_t^2 = [y_t - E(y_t)]^2 \tag{4.6.1}$$

---

[28] A proxy for the 'true' realized variance is needed since the 'true' variance can't be observed directly.

, where y_t being the realized return at time t, and $E(y_t)$ the one-step ahead forecast for the return at time t. This type of proxy for the 'true' realized variance is commonly used in studies on the forecasting performance of GARCH models (e.g. Franses, P.H. and van Dijk, D. (1996); Watkins, C. and McAleer, M. (2008); Bracker, K. and Smith, K.L. (1999 and 2003)).

These authors did however assume that the forecast for the future return was the unconditional mean for the returns in the in-samples, formally; $E(y_t) = \bar{y}$. This structure deserves a second thought since the variance forecasts from the GARCH models are forecasts for the error variance, which depends on the specified mean equation. Therefore, in this thesis, this structure is only used when the mean equation has a constant mean. In out-of-sample 1, however, the one-step ahead forecasts for the returns are based on the estimated AR(2) model in the mean equation for each GARCH model. For the random walk model a separate AR(2) model where estimated on the returns with OLS regression to produce forecasts for the one step ahead return. This should be fine since the statistical significance of an AR(2) model in the mean has already been confirmed. In the estimations of the GARCH models in subsamples 2 to 4 there is only a constant mean. The forecast for the one-step ahead returns are therefore the same throughout the entire out-of-sample period. The constant means are, for the GARCH models, the constant mean in the mean equation from the GARCH model estimations. For the random walk model the constant mean where estimated by OLS regression on the returns in each corresponding in-sample period.

The GARCH models one step ahead forecast for the one day error variance is given by the model specifications in section 3 and the parameter estimates from the model estimations. The parameters of the GARCH models are in other words not re-estimated during the forecasting procedure. For the random walk model the one step ahead forecast for the one day error variance is the last observed squared residual from the specified mean equation, which formally can be expressed as:

$$E(\varepsilon^2_{t+1\,|\,t}) = \varepsilon^2_t \qquad\qquad (4.6.2)$$

There is no drift in the random walk here since it's assumed that there is no long term trend in variance of copper price returns. This assumption is in accordance with the results presented in the study by Watkins, C. and McAleer, M. (2008).

The deviations (errors) by the variance forecasts from the 'true' variance proxy are then for the GARCH models and the random walk model given as below.

$$u_t = \varepsilon_t^2 - \sigma_{t\,|\,t-1}^2 \qquad or, \qquad u_t = \varepsilon_t^2 - E(\varepsilon_{t\,|\,t-1}^2) \qquad (4.6.3)$$

Based on u_t the MSE, MAE and MedSE are calculated as described below:

$$MSE = \frac{1}{N}\sum_{t=1}^{N} u_t^2 \qquad (4.6.4)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N} |u_t| \qquad (4.6.5)$$

The MedSE is simply the median value of u_t^2.

As mentioned above, comparing the MSE, MAE and MedSE (and sometimes also other measures) of the variance forecasts by the GARCH models and the Random Walk model is a common way to evaluate the forecasting performance of GARCH models. However, before drawing any conclusions bases on these measures it's important to consider what they actually measure and what economic implications this has.

To begin with, remember that the MSE is the average of the squared deviations by the variance forecasts from the 'true' variance proxy, which means that the deviations (absolute) are not weighted equally in the MSE. Large deviations are in other words, given a disproportionally large weight compared to small deviations. This means that a large error for a one day's variance forecast could lead to a lager increase in the MSE than several small errors which has a larger absolute value when summarized. This is much doubtfully a logical approach when it comes to evaluating variance forecast performance since market participants normally are exposed to price risks for periods longer than one day, and that larger than expected returns (negative or positive) are additive over time. It does therefore not make much sense to give disproportionally more weight to large deviations, since one large deviation isn't more problematic for the market participants than several small deviations of the same absolute size when summarized. This motivates for the use of MAE rather than MSE when evaluating the forecasting performance.

Furthermore, due to the nature of the MSE discussed above, outliers could have a severe impact on the result and are therefore frequently removed before OLS estimations and MSE calculations. When it comes to squared errors of a mean equation on returns, however, such interference in the data doesn't appear to be a rational approach. Firstly since it's hard to

motivate that some returns are too large (positive or negative) due to some abnormality, and thus should be inappropriate to be represented in the data. And secondly, since squaring the errors from the mean equation makes 'outliers' in the returns even more extreme.

With the aim of minimizing the impact of outlying observations, P.H. Franses and D. van Dijk (1996 p.231) approached this problem by using MedSE instead of MSE. However, it is here important to consider the value of a median error measurement to a market participant whose financial decision making are based on forecasts of risk and return. Basing financial decision on median squared errors of a forecast doesn't appear attractive when considering that all errors less than the median could be only slightly less while the errors larger than the median could be extremely large[29]. This also motivates for the use of MAE when evaluating the forecasting performance of the GARCH models.

Based on the reasoning above, conclusions regarding the forecasting performances of the GARCH models will therefore in this thesis primarily be based on the MAE values. MedSE and MSE values will however also be included for those who are of other opinion.

The results from the forecasting procedure are presented in table 4.6.1 below. Here we see that, in terms of MAE the GARCH models outperform the RW-model in all of the subsamples. The improvement varies between 17 to 29 % lower MAE. No clear trend in the improvement to the RW-model can be observed over the four subsamples.

Comparing the GARCH models to the unconditional variance, we see that in subsample 1 and 2 the GARCH models have about 8 to 12 % lower MAE. However, in subsample 3 the GARCH models have about 8 to 10 % higher MAE than the unconditional variance. This is the only case where ether the RW-model or the unconditional variance outperforms the GARCH models in terms of MAE. This is rather surprising considering that the GARCH models were relatively more successful in modeling the variance in in-sample 3 compared to in-sample 2. One possible explanation for the poor forecasting performance in subsample 3 could be that the GARCH models in in-sample 3 were estimated on data that are highly unrepresentative for the average dynamics in the copper price returns. This explanation is supported by the fact (as was discussed in section 4.5) that in-sample 3 mainly consists of the

---

[29] Remember that the median of the values 1, 2 and 3, and the median the values 1, 2 and 999, are both 2. If the numbers in the two series above were the amount of money you would risk to lose as a consequence of participating in two different investment alternatives over the next day, evaluating the risk of these two investments by median loss would clearly be irrational.

Model 4.6.1 - Forecasting results.

Subs.	Model	MAE	Percent diff.		MSE	Percent diff.		MedSE	Percent diff.	
1	AR(2)-GARCH(1,1)	**1,936E-04**	-(20,98%)*	-(8,25%)''	**9,270E-08**	-(49,09%)*	-(1,20%)''	2,156E-08	(82,88%)*	-(42,55%)''
	AR(2)-GJR-GARCH(1,1)	1,939E-04	-(20,86%)*	-(8,11%)''	9,283E-08	-(49,01%)*	-(1,06%)''	2,173E-08	(84,31%)*	-(42,10%)''
	AR(2)-QGARCH(1,1)	1,951E-04	-(20,34%)*	-(7,51%)''	9,328E-08	-(48,77%)*	-(0,58%)''	2,229E-08	(89,08%)*	-(40,60%)''
	AR(2)-RW	2,450E-04			1,821E-07			**1,179E-08**		
	Unconditional Var.	2,110E-04			9,382E-08			3,753E-08		
2	GARCH(1,1)	1,081E-04	-(26,26%)*	-(8,28%)''	2,086E-08	-(53,47%)*	-(4,58%)''	9,236E-09	(3,13%)*	-(29,75%)''
	GJR-GARCH(1,1)	1,078E-04	-(26,44%)*	-(8,51%)''	2,083E-08	-(53,52%)*	-(4,69%)''	8,978E-09	(0,25%)*	-(31,71%)''
	QGARCH(1,1)	**1,053E-04**	-(28,14%)*	-(10,62%)''	**2,075E-08**	-(53,70%)*	-(5,05%)''	**8,163E-09**	-(8,86%)*	-(37,92%)''
	RW	1,466E-04			4,482E-08			8,956E-09		
	Unconditional Var.	1,179E-04			2,186E-08			1,315E-08		
3	GARCH(1,1)	4,072E-04	-(27,33%)*	(10,48%)''	**5,158E-07**	-(48,15%)*	-(0,38%)''	8,207E-08	(4,43%)*	(45,08%)''
	GJR-GARCH(1,1)	4,051E-04	-(27,70%)*	(9,92%)''	5,281E-07	-(46,91%)*	(2,00%)''	7,370E-08	-(6,21%)*	(30,30%)''
	QGARCH(1,1)	3,975E-04	-(29,06%)*	(7,86%)''	5,227E-07	-(47,46%)*	(0,95%)''	7,027E-08	-(10,58%)*	(24,23%)''
	RW	5,603E-04			9,947E-07			7,858E-08		
	Unconditional Var.	**3,685E-04**			5,177E-07			**5,657E-08**		
4	GARCH(1,1)	4,156E-04	-(18,17%)*	-(18,24%)''	5,370E-07	-(40,67%)*	-(12,41%)''	6,777E-08	(24,28%)*	-(69,42%)''
	GJR-GARCH(1,1)	**4,110E-04**	-(19,07%)*	-(19,14%)''	**5,156E-07**	-(43,03%)*	-(15,91%)''	6,866E-08	(25,91%)*	-(69,02%)''
	QGARCH(1,1)	4,201E-04	-(17,29%)*	-(17,36%)''	5,211E-07	-(42,43%)*	-(15,02%)''	7,646E-08	(40,23%)*	-(65,50%)''
	RW	5,078E-04			9,051E-07			**5,453E-08**		
	Unconditional Var.	5,083E-04			6,131E-07			2,216E-07		

Notes: MAE: Mean of absolute Errors. MSE: Mean of squared Errors. MedSE: Median of absolute Errors. * Percent difference between the MAE, MSE or MedSE of the GARCH model and the RW-model. '' Percent difference between the MAE, MSE or MedSE of the GARCH model and the unconditional variance forecast.

extraordinary price increase between 2004 and mid 2006 and that the asymmetry terms was found to be of opposite signs compared to the normal. Also, considering that the mean of the daily log returns was about 0,13 % (per day!) in in-sample 3, such type of market could clearly not be expected to persist indefinitely.

Moving on to subsample 4, we here see that the GARCH models produce about 17 to 19 % lower MAE than the unconditional variance, which is the largest improvement over the unconditional variance by the GARCH models. This further strengthens the suspicion of that the poor performances of the GARCH models in in-sample 3 could have been the consequence of misfortunate subsample specifications.

Anyhow, this result show it's hard to predict whether or not the GARCH models will be able to deliver more accurate forecasts of the future variance than the unconditional variance base on the pre-estimation test for ARCH effects in section 4.4 and the models ability to model the variance in-sample. However, if more thoughts were to be given into the subsample selection, this might lead to more accurate expectations of the GARCH models out-of-sample forecasting performance.

In the table we also see that the differences between the GARCH models are small. The differences are however larger in subsamples 3 and 4 compared to subsamples 1 and 2. This result is however consistent with our pre-forecast expectations.

However, the QGARCH did actually perform worst of the GARCH models in subsample 4, but best in subsample 3. This suggest that it's hard to determine which one of the GARCH models that will perform best in out-of-sample forecasting, given the level of difference of the models parameter value and statistical significance which was the case here.

We also see that, in terms of both MAE and MSE, the unconditional variance produces much more accurate forecast for the future variance than the random walk. This result is also in line with our pre-forecast expectations discussed previously in this section. This indicates that the unconditional variance actually is a more legitimate benchmark for the forecasting performances of the GARCH models than the Random Walk model.

In the MSE section of the table, we see that the improvements compared to the RW-model are much larger here than they were for MAE, indicating that the RW-model generally produces a larger number of relatively small errors and a larger number of errors of relatively large errors compared to the GARCH models. The RW-model is significantly outperformed, with MSEs that are about 41 % to 53 % lower for the GARCH models in all of the samples.

On the other hand, the improvements by the GARCH models over to the unconditional variance in subsamples 1 and 2 are only about 1 % and 4,5 % respectively when measured as MSE. This indicates that the GARCH models produce a larger number of relatively small errors and a larger number of errors of relatively large errors compared to the unconditional variance in these two out-of-sample periods. The same is also the case for subsample 4, where improvement over the unconditional variance has decreased to about 12 to 16 %. The opposite is however true for subsample 3 where the MSE is only about 1 % higher for the GARCH compared to the unconditional variance.

5. Conclusion

Based on the results presented in section 4.4 to section 4.6 in this thesis, some conclusions could be made. These conclusions are listed below.

×   In the presence of ARCH effects in the volatility of copper spot price returns, the GARCH models were able to successfully model the conditional variance (in-sample) to a satisfying degree. The presence of ARCH effects in the volatility of copper spot price returns did however vary over the studied period in this thesis. There was only a significant presence of ARCH effects in three out of the four subsamples. It should therefore be important to keep updated on the current situation regarding this before modeling the conditional variance of copper spot price returns with GARCH models.

×   Forecasting the one-step ahead one day variance of copper spot price returns with the GARCH models can be recommended. This conclusion is based on the fact that the GARCH models provide more accurate forecasts (across all subsamples) compared to assuming that the one day error variance followed a Random Walk process (17 to 29 % lower MAE), and more accurate forecasts (in three out of the four subsamples) compared to assuming constant error variance and using the unconditional variance as a forecasts for the one-step ahead one day variance (8 to 19 % lower MAE).

A further motivation for this recommendation is due to the fact that the GARCH models' performance in modeling (in-sample) and forecasting (out-of-sample) the conditional variance was best in the last subsample. Here between 75 to 85 % of the excess kurtosis were removed by the GARCH models in in-sample estimation, and the GARCH models produced 17 to 19 % lower MAE than both the Random Walk model and the unconditional variance in out-of-sample forecasting.

×   Using the unconditional variance as a forecast for the one-step ahead one day variance produced lower MAE values than a Random Walk model, and could therefore arguably be consider a more legitimate benchmark for the forecasting performance of the GARCH models.

×   Before estimating the GARCH models with the purpose of later using the models to forecast the one-step ahead one day variance of copper spot price returns, it is probably

important to evaluate how well the in-sample data might represent the future. This conclusion is based on the 8 to 10 % higher MAE the GARCH models had compared to the unconditional variance in out-of-sample forecasting in subsample 3.

× No one of the three GARCH models used in this thesis can generally be recommended over the others due to the small differences between the GARCH models in the out-of-sample forecasting, and due to the difficulty in predicting which model that will have the slightly better performance.

× The skewness in the distribution of demeaned returns or the distribution of the residuals from the mean equation were not a good indicator of whether the two asymmetric GARCH models would be better than the standard symmetrical GARCH model in modeling the conditional variance (in-sample or out-of-sample).

× The (1,1) order of the GARCH models appears to be sufficient to fully capture the ARCH effects in the volatility of copper spot price returns.

6. Discussion

One of the more serious limitations of this thesis is that only two modified versions of the standard GARCH model was used. This limitation was added since testing all types of GARCH models would be a far too extensive study for this thesis. It is however possible that some other type of GARCH models would perform even better than those which have been used in this thesis. An important area of further research should therefore be to investigate the extent to which other types of GARCH models are able to improve the accuracy of the forecast of future copper return volatility.

Two examples of other modifications to the GARCH model that could be interesting to investigate closer are; GARCH models which are able to account for an 'information accumulation effect' since the observations in the sample data are irregularly spaced in time, and GARCH models which includes external variance regressors of fundamental volatility drivers in the copper price returns.

Another rather serious limitation to the study in this thesis is the structure of the subsample periods. In this thesis the data sample was arbitrarily divided into four equally large subsamples. It is however likely that different results would have been derived if instead the sample was dividing into subsamples of different sizes, or if a more methodical approach was used to divide the sample into subsamples. An example of such approach would be the iterated cumulative sum of squares (ICSS) algorithm, which Bracker, K. and Smith, K.L. (1999) used to detect breakpoints in the volatility series.

However, for a first investigation of whether or not modeling the copper return volatility with GARCH models is appropriate, the selection of equally large subsamples should be sufficient. But for further knowledge and more precise conclusions it could be worth investigating the GARCH models forecasting performances when the subsamples are selected differently.

Furthermore, only the volatility in the close-to-close daily log returns was analyzed in this thesis. For further researches about GARCH models' ability to model and forecast the volatility in copper price returns, it can also be interesting to investigate whether or not GARCH models are suitable for modeling and forecasting the volatility of open-to-close (intraday) returns, close-to-open returns, or the volatility of returns of higher frequencies (and maybe also for lower frequencies).

Finally, in this thesis a AR(2) model was selected for the conditional mean in subsample 1, and the proxy for the 'true' realized variance was therefore specified as eq. 4.6.1. Although the AR(2) lowers the sizes of the errors and increases the accuracy of the mean equation, it could however be the case that GARCH models are more suitable to model the error variance from a model with constant mean.

The reason for this would be that some of the underlying factors which create the volatility clustering might be due to some sort of psychological reaction of the market participants. Such kind of response is arguably unlikely to be based on the previous days' errors from a model for the returns with a conditional mean, since it's very unlikely that market participants use the same model for the mean. It's on the other hand arguably more likely that the market participants react to the mere sizes of previous returns or errors from a long term expected return.

However, since forecasting the returns are at least as important as forecasting the volatility, using only a constant mean where it possible to successfully model a conditional mean with an AR process, doesn't appear very attractive even if the accuracy of the variance forecasting are increased by doing so. Also, since the autocorrelations in the returns in this case are quite small, it's unlikely that the use of the AR(2) model in subsample 1 would have a significant impact on the end results.

7. Bibliography

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31 (1986), 307-327.

Bollerslev, T. (2011), "ARCH and GARCH models", http://public.econ.duke.edu/~boller/Econ.350/talk_garch_11.pdf.

Bracker, K. and Smith, K.L. (1999), "Detecting and Modeling Changing Volatility in the Copper Futures Market", *The Journal of Futures Markets*, 19 (1999), 79-100.

Bracker, K. and Smith, K.L. (2003), "Forecasting Changes in Copper Futures Volatility with GARCH Models Using an Iterated Algorithm", *Review of Quantitative Finance and Accounting*, 20 (2003), 245-265.

Brooks, C. (2008), "Introductory Econometrics for Finance", *Cambridge University Press*, Second Edition (2008).

Brunetti, C. and Gilbert, C.L. (1995), "Metals price volatility", *Resources Policy,* 21 (1995), 237–254.

Engel, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50(4) (1982), 987-1008.

Engel, R. F. (2001), "GARCH 101: The use of ARCH/GARCH models in applied econometrics", *Journal of Economic Perspectives*, 15 (2001), 157-168.

Engel, R. F. and Kraft, D. (1983), "Multiperiod forecast error variances of inflation estimated from ARCH models", *Bureau of the Census*, (1983), 293-302.

Engel, R. F. and Ng, V. (1993), "Measure and testing the impact of news on volatility", *Journal of Finance*, 48 (1993), 1749-1778.

Day, T. and Lewis, C. (1992), "Stock market volatility and the information content of stock index options", *Journal of Econometrics*, 50 (1992), 267-287.

Franses, P.H. and van Dijk, D. (1996), "Forecasting stock market volatility using (Non-linear) GARCH models", *Journal of Forecasting*, 15 (1996), 229-235.

Geman, H. (2005), " Commodities and commodity derivatives: modeling and pricing for agriculturals, metals, and energy ", *John Wiley & Sons Ltd*, (2005).

Glosten, L., Jagannathan, R. and Runkel, D. (1992), "On the relation between the expected value and the volatility of nominal excess return on stocks", *Journal of Finance*, 46 (1992), 1779-1801.

Gujarati, D. N. and Porter, D.C. (2010), "Essentials of Econometrics", *McGraw-Hill*, Fourth Edition (2010).

Kogan, L., Livdan, D. and Yaron, A. (2009), "Oil Futures Prices in a Production Economy with Investment Constraints", *Journal of Finance*, 64 (2009), 1345-1375.

Ruppert, D. (2006), "Statistics and Finance - An Introduction ", *Springer Texts in Statistics*, (2006).

Russel-Walling, E. (2005), "Commodities as a new institutional asset class", *The Ringsider: Metals 2005,* London Metal Exchange, (2005), 135–137.

Sentana, E. (1995), "Quadratic ARCH models", *Review of Economic Studies*, 62 (1995), 639-661.

Watkins, C. and McAleer, M. (2008), "How has volatility in metals markets changed?", *Mathematics and Computers in Simulation*, 78 (2008), 237–249.

8. Appendix

Below follows an incomplete list of studies in which the performance of ARCH type of models to model and forecast the conditional variance is investigated.

-Alberg, D., Shalit, H. and Yosef, R. (2008), "Estimating stock market volatility using asymmetric GARCH models", *Applied Financial Economics*, 18 (2008), 1201-1208.

-Akgiray, V. (1989), "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts", Journal of Business, 62(1) (1989), 55-80.

-Awartani, B.M.A. and Corradi, V. (2004), "Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries", *International Journal of Forecasting*, 21 (2004), 167-183.

-Balaban, E. (2004), "Comparative forecasting performance of symmetric and asymmetric conditional volatility models of an exchange rate", *Economics Letters*, 83 (2994), 99-105.

-Bracker, K. and Smith, K.L. (1999), "Detecting and Modeling Changing Volatility in the Copper Futures Market", *The Journal of Futures Markets*, 19 (1999), 79-100.

-Bracker, K. and Smith, K.L. (2003), "Forecasting Changes in Copper Futures Volatility with GARCH Models Using an Iterated Algorithm", *Review of Quantitative Finance and Accounting*, 20 (2003), 245-265.

-Brailsford, T.J. and Faff, R.W. (1996), "An Evaluation of Volatility Forecasting Techniques", *Journal of Banking and Finance*, 20 (1996), 419-438.

-Chen, C.W., Gerlach, R. and So, M. (2006), "Comparison of nonnested asymmetric heteroscedastic models", *Computational Statistics and Data Analysis,* 51 (2006), 2164-2178.

-Day, T. and Lewis, C. (1992), "Stock market volatility and the information content of stock index options", *Journal of Econometrics*, 50 (1992), 267-287.

-Engel, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50(4) (1982), 987-1008.

-Engel, R. F. and Kraft, D. (1983), "Multiperiod forecast error variances of inflation estimated from ARCH models", *Bureau of the Census*, (1983), 293-302.

-Franses, P.H. and van Dijk, D. (1996), "Forecasting stock market volatility using (Non-linear) GARCH models", *Journal of Forecasting*, 15 (1996), 229-235.

-Hansen, P.R., and Lunde, A. (2005), "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?", *Journal of Applied Econometrics*, 20 (2995) 873-889.

-Loudon, G.F., Watt, W.H. and Yadav, P.K. (2000), "An empirical analysis of alternative parametric ARCH models", *Journal of Applied Econometrics*, 15 (2000), 117-136.

-Pagan, A.R. and Schwert, G.W. (1990), "Alternative Models for Conditional Stock Volatilities", *Journal of Econometrics*, 45 (1990), 267-290.

West, K.D. and Cho, D. (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility, *Journal of Econometrics,* 69 (1995), 367-291.