Stockholms universitet

The LBMA Gold Price: The one-day-ahead conditional variance predictability using the GARCH and EGARCH models

Hilding Köhler

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2020:4
Matematisk statistik
Juni 2020

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2020:4**
http://www.math.su.se

The LBMA Gold Price: The one-day-ahead conditional variance predictability using the GARCH and EGARCH models

Hilding Köhler*

June 2020

Abstract

The main goal of this study is to examine whether or not a set of ARMA-GARCH and ARMA-EGARCH models can be used to predict the one-day-ahead conditional variance of the logarithmic return of the LBMA gold price. A secondary goal is to examine the traits of the models that are able to forecast. The data used in the study is the LBMA gold price set at 10:30 AM on bank days between the dates 2000-01-04 to 2020-01-04. To answer the question, ARMA-GARCH type models are implemented where the GARCH type models are restricted to GARCH(1,1) and EGARCH(1,1). The forecasting is done via a rolling window forecast with three different window sizes: 365, 912 and 1825 days. In total are eight models tested in each window which entails there is a total of 24 models to evaluate. The forecasting ability of the models' is evaluated by backtesting the predictive distribution. The backtesting involves the Kolmogorov-Smirnov test as well as analysis of sample histograms. It is found that the best forecasting model over-all is the MA(1)-EGARCH(1,1) with t-distributed white noise terms. Furthermore, the results indicate that using the standard normal distribution to model the white noise will cause an inadequate forecasting ability. Based on the results, it is not possible to determine whether the models, with and without mean model, using the GARCH(1,1) performed better than the ones with the EGARCH(1,1).

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: hildingkohler@gmail.com. Supervisor: Mathias Millberg Lindholm, Kristoffer Lindensjö and Felix Wahl.

Acknowledgement

I would like to extend my deepest gratitude to my supervisors Mathias Millberg Lindholm, Kristoffer Lindensjö and Felix Wahl for sharing their expertise and input for countless of hours.

Furthermore, I would also like to thank my good friend and colleague Mathias Carlsson for the endless and great discussions as well as constantly inspiring me to work harder.

Lastly, I would like to thank my family for the support and advise.

Introduction

Financial investing occupies a lot of people be it an ordinary individual seeking to earn a few extra dollars on the side or a financial analyst at a corporation. Whomever the person might be, none of them can escape the uncertainty of not knowing if their investments will bring forth a loss or a gain. A financial asset that stands out is the precious metal, gold. Gold is often regarded as a safe haven in portfolio management, in other words it is on average negatively or uncorrelated with the rest of the assets in turbulent times [1]. In the paper *"Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold"* the authors found evidence to support this claim for stocks, indicating that the claims might be true [1]. Since it might be a lucrative strategy to invest in gold to protect against major losses in turbulent times, it is important to know when to buy and when to sell, given their level of risk aversion. Well, not a single person can be one hundred percent sure what the price of an asset will be one day ahead, there will always be some degree of uncertainty. One measurement of risk is the variance of the return or price of an asset [2]. With all this being said, it would be interesting to examine whether the one-day-ahead conditional variance of the LBMA gold price is forecastable.

This thesis will focus on examining whether the conditional variance of the logarithmic return of the LBMA gold price (set at 10:30 AM on bank days) is forecastable one day ahead. The London Bullion Market Association (LBMA) gold price is one of many and is set twice each day. The data used in this study stretches between the dates 2000-01-04 to 2020-01-04 (YYYY-MM-DD). Models implemented in the study are ARMA-GARCH type models, where the GARCH models are limited to the GARCH(1,1) and EGARCH(1,1). These types of models are widely implemented in studies concerning financial markets. Before the forecasting it is studied whether the data as a whole is compatible with ARMA-GARCH models. The forecasting method makes use of a rolling window forecast with three different window sizes: 365, 912 and 1825 days. A total of eight models are tested in each window which implies that there are 24 models to evaluate in terms of forecasting ability. To evaluate the forecasting ability of each model a backtesting of the predictive distribution is implemented, for further reading see [3]. The main goal of the study is to find one or several models that can be used to forecast the conditional variance of the log return of the LBMA gold price.

Contents

Theory

The theory presented in the section below can be assumed to be from [4] unless anything else is stated. As an effect of this most of the modelling in this study can be reproduced using this source. For readability let ⧫ mark the end of a definition. All sources can be found in the references section of this study.

One-period log return of time series

It is not uncommon for time series data of a price to not be stationary, to solve this issue without the loss of interpretation of the underlying data, the return of a price is commonly used.

Definition:

Given a time series $\{Y_t\}$ then the simple return, $1 + R_t$, for one time period is given by

$$Y_t \;=\; Y_{t-1}(1 + R_t) \;\Leftrightarrow\; 1 + R_t \;=\; \frac{Y_t}{Y_{t-1}} \tag{1}$$

⧫

It is now easy to define the log return of a series by applying the natural logarithm to the expression on the right side of the equivalence arrow.

Definition:

$$r_t \;=\; \ln(1 + R_t) \;=\; \ln\left(\frac{Y_t}{Y_{t-1}}\right) \tag{2}$$

⧫

Maximum likelihood estimates

A common way to estimate parameters in time series models is the maximum likelihood method. In the general case the method is given by the following definition. [5]

Definition *([5], p.14-15)*:

The likelihood function, $L(\boldsymbol{\theta})$, for a parameter given an observed dataset $\{x_i\}$ is defined as

$$L(\boldsymbol{\theta}) \;=\; f(x_1, ..., x_n | \, \boldsymbol{\theta}) \tag{3}$$

Given the prior definition of the likelihood function, the log likelihood function is defined as the following

$$\ell(\boldsymbol{\theta}) \;=\; \ln(L(\boldsymbol{\theta})) \tag{4}$$

The maximum likelihood estimate vector, $\hat{\boldsymbol{\theta}}_{\boldsymbol{MLE}}$, of a parameter vector $\boldsymbol{\theta}$, is hence given by

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{MLE}} \;=\; \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) \tag{5}$$

A consequence of the fact that the log likelihood function is a strictly growing function, is that maximizing the likelihood function is equivalent to maximizing the log likelihood function. In other words the following holds

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{MLE}} = \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}) \tag{6}$$

♦

Akaike's information criterion (AIC)

The AIC is a widely used model selection criterion given a family of similar models that are of interest for the modelling situation. The reason to why the AIC is a good tool for model selection is due to the fact that it uses the maximum likelihood estimates for the model, which measures the goodness of fit for the model. Furthermore, it also implements a parameter which simultaneously punishes overfitting. The overfitting is an effect of the log likelihood since a more complex model yields a higher value of the log likelihood. Hence, the more complex models need to be penalized for overfitting. The definition of the AIC is as follows. [5]

Definition ***([5], p.224)*****:**

$$\text{AIC} = -2\ell\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{MLE}}\right) + 2p \tag{7}$$

Where $\ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{MLE}})$ is the log likelihood function evaluated at the maximum likelihood estimate vector and p the number of parameters that are estimated. A smaller value of the AIC indicates a better model when compared to one with a higher AIC. This is true since a better fitted model returns a higher log likelihood while also not having too many parameters.

♦

Stationarity of time series

An important aspect of modelling time series is that the process often operates under the assumption of time stationarity. The assumption of time stationarity gives the models desirable traits, with some given directly from the definition of stationarity. One of the most interesting traits among stationary time series is that they tend to fluctuate around a constant with the same variance. This means that they tend to possess the same statistical properties over time which renders the series predictable. Stationarity of time series can be divided into two types, strict and weak, where the latter is a more relaxed definition. To understand the two types of stationarity it is important to introduce the two following definitions

Definition:

A time series $\{Y_t\}$ is said to be strictly stationary if the joint distribution of $(Y_1, ..., Y_T)$ is the same as the distribution for $(Y_{1+k}, ..., Y_{T+k})$ for all k.

Definition:

Given a time series $\{Y_t\}$ then the series is said to be weakly stationary if the first moment and the covariance between all lags of the series are time independent. This is equivalent to the following mathematical properties.

$$\begin{aligned} \mathbb{E}[Y_t] &= \mu, \quad \text{where } \mu \text{ is a constant} \\ \text{Cov}(Y_t, Y_{t-l}) &= \gamma_l, \quad \text{where the value only depends on the lag, } l \end{aligned} \tag{8}$$

♦

Autocorrelation function (ACF)

In this study the term correlation will be the same as the Pearson correlation if nothing else is stated. The correlation between two stochastic variables X and Z is defined as

$$\rho_{xz} \;=\; \frac{\mathrm{Cov}(X, Z)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Z)}} \tag{9}$$

and the sample correlation between the two sample realizations $\{x_i\}$ and $\{z_i\}$ as

$$r_{xz} \;=\; \frac{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(z_i - \bar{z})^2}} \tag{10}$$

These two definitions are very general and would be more useful if modified to fit time series data and the notations that come with it. To transfer these definitions to time series data consider the series $\{Y_t\}$ and $\{Y_{t+l}\}$ for which the following holds $\{Y_{t+l}\}\mathbf{B}^l \;=\; \{Y_t\}$. To clarify the meaning of this it is easier to observe a set of time series data $\{Y_{1+l}, ..., Y_{T+l}\}$ and then apply the operator $\mathbf{B}^l$ to the set which yields the new set $\{Y_t, ..., Y_T\}$. The operator $\mathbf{B}^l$ is called the l-lag backshift operator which shifts a time series back in time. The correlation between these two series, also referred to as the autocorrelation or the autocorrelation function, is then given by the following definition.

Definition:

Given a weakly stationary time series $\{Y_t\}_{t=1}^{T}$, then the correlation, or the lag-l autocorrelation, between Y_t and Y_{t-l} is

$$\rho_l \;=\; \frac{\mathrm{Cov}(Y_t, Y_{t-l})}{\sqrt{\mathrm{Var}(Y_t)\,\mathrm{Var}(Y_{t-l})}} \;=\; \frac{\gamma_l}{\gamma_0} \tag{11}$$

It then follows that the sample lag-l autocorrelation is

$$r_l \;=\; \frac{\sum_{t=1}^{T}(y_t - \bar{y}_t)(y_{t-l} - \bar{y}_{t-l})}{\sqrt{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2 \sum_{t=1}^{T}(y_{t-l} - \bar{y}_{t-l})^2}} \tag{12}$$

Partial autocorrelation function (PACF)

The autocorrelation function is only one out of many methods to measure correlation for time series, but it has the downside of not adjusting for effects from lags other than the ones of interest. To deal with the effects from other lags, the partial autocorrelation function is a good tool. The idea behind the partial autocorrelation function is to first clear the lag series from the effect of the other lags and then perform the calculation of the ordinary autocorrelation function. In this study the clearing method will be a multiple linear regression of the lag of interest on the other lags.

Definition *([6], p.66-69)*:

Given a weakly stationary time series $\{Y_t\}_{t=1}^{T}$, then the partial auto correlation function between Y_t and Y_{t-l} can be calculated in the following manner.

Let $\hat{Y}_t$ be the multiple linear regression of Y_t on the previous lags up until the lag before Y_{t-l} and $\hat{Y}_{t-l}$ the regression of Y_{t-l} on the very same regressors. The partial autocorrelation function can now be defined as

$$\phi_{ll} \;=\; \frac{\mathrm{Cov}(Y_t - \hat{Y}_t, Y_{t-l} - \hat{Y}_{t-l})}{\mathrm{Var}\left(Y_{t-l} - \hat{Y}_{t-l}\right)} \tag{13}$$

♦

Three time series models

Definitions for the three most common time series models will be given in this section. In all three of them it is assumed that Y_t, for any t, is an observation of a time series $\{Y_t\}$ and that the series is operating under weak stationarity.

Autoregressive model (AR model)

Definition:

A time series is said to follow an autoregressive process of order p if the following holds.

$$Y_t \;=\; \phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t \tag{14}$$

Where $\{\varepsilon_t\}$ is assumed to be a white noise series, that is, a sequence of i.i.d random variables from a distribution with expected value zero and constant variance.

♦

By utilising basic properties of the expected value in combination with the definition of weak stationarity it can be shown that the following unconditional property of the AR model holds.

$$\mathbb{E}\left[Y_t\right] \;=\; \frac{\phi_0}{1 - \sum_{i=1}^{p} \phi_i} \tag{15}$$

The expression for the variance for an AR(p) process can not be derived explicitly and is hence omitted in this section. It can be shown that both the unconditional variance and expected value for the process are invariant under time, that is the processes is weakly stationary. The conditions for weak stationarity for an AR(p) processes are equivalent to the solutions to $1 - x\phi_1 - x^2\phi_2 - ... - x^p\phi_p = 0$ being outside of the unit circle. In other words, when all the solutions to the equation are positioned outside the circle with radius one and centre in origo the process is weakly stationary.

Moving average model (MA model)

Definition:

A time series is said to follow a moving average model of order q if the following holds.

$$Y_t \;=\; c_0 + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{16}$$

Where $\{\varepsilon_t\}$ is assumed to follow a white noise process with some variance and c_0 is a constant.

♦

It can without too much effort be shown that the expected value and variance for the MA(q) model are

$$
\begin{aligned}
\mathbb{E}\left[Y_t\right] &= c_0 \\
\operatorname{Var}\left(Y_t\right) &= \left(1+\sum_{i=1}^{q}\theta_i^2\right)\sigma_\varepsilon^2
\end{aligned} \tag{17}
$$

A special property of the moving average models is that they are always weakly stationary, no matter the order of it, since they are finite linear combinations of a white noise sequence. This holds true since the first and second moment always are time independent, which is seen in (17).

Autoregressive moving average model (ARMA model)

Definition:

A time series is said to follow an autoregressive moving average model of order p and q if the following holds.

$$
Y_t = \phi_0 + \sum_{i=1}^{p}\phi_i Y_{t-i} + \varepsilon_t - \sum_{i=1}^{q}\theta_i\varepsilon_{t-i} \tag{18}
$$

♦

Deriving an explicit expression for the variance is also impossible in this model. The expression for the expected value is on the other hand calculable and is in fact the same as the expected value for and AR(p) model, that is

$$
\mathbb{E}\left[Y_t\right] = \frac{\phi_0}{1-\sum_{i=1}^{p}\phi_i} \tag{19}
$$

For the ARMA models to be stationary it suffices if the solutions to $1 - x\phi_1 - x^2\phi_2 - ... - x^p\phi_p = 0$ are outside the unit circle. This is due to the fact that the ARMA(p,q) process can be viewed as an AR(p) process with error terms following an MA(q) process simultaneously. Therefore do the stationarity requirements for the AR processes hold for ARMA processes [6].

Volatility modelling in time series

In returns of financial assets, of which gold price is one, measures of dispersion is of great concern due to it being a convenient way to determine risk [2]. The standard way of measuring dispersion is by calculating the standard deviation. Large standard deviations of course imply large risks since the value of the price will vary a lot. In financial time series volatility and standard deviation are often interchangeable words describing the uncertainty in the series and if the series describes the return of an asset, uncertainty equals risk. To get a full understanding of volatility and to be able to model it, it is important to introduce the concept of conditional variance.

Definition:

Consider a time series $\{Y_t\}$ and let $\mu_t = \mathbb{E}\left[Y_t|\mathcal{F}_{t-1}\right]$ denote the conditional expected value of the series and $\sigma_t^2 = \operatorname{Var}\left(Y_t|\mathcal{F}_{t-1}\right)$ the conditional variance. The notation $\mathcal{F}_{t-1}$ is defined as the information available at time $t-1$. Using these notations the time series can be modelled in the following manner.

$$
Y_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{20}
$$

In equation (20) $\{z_t\}$ is assumed to follow a white noise series with unconditional variance one. Using equation (20) leads to the following property of the series.

$$\sigma_t^2 \;=\; \mathrm{Var}\left(Y_t|\mathcal{F}_{t-1}\right) \;=\; \mathrm{Var}\left(\varepsilon_t|\mathcal{F}_{t-1}\right) \tag{21}$$

Note: When considering $Y_t = \mu_t + \varepsilon_t$ without the underlying mathematical properties of the components, μ_t is often referred to as the mean model.

♦

The reason why the conditional variance is of interest is because it allows for powerful methods of dealing with and predicting time dependent conditional heteroskedasticity. Time dependent conditional heteroskedasticity in this case is the property of having different variances depending on the time given the information available at that time. This is a very common model characteristic of time series that changes the properties of the models. The methods used in this study lead to two part models, where one part is describes the data and the other the time evolution of the conditional variance.

The general conditional heteroskedasticity (GARCH model)

The first model for handling conditional heteroskedasticity, called the ARCH model, was proposed by Robert Engle in his publication in *Econometrica* 1982 [7]. The ARCH model can be seen as an AR model fitted on the squared error terms. The ARCH model is practical in some situation, but comes with the weakness of being restrictive in its parameter estimations. This issue becomes even more problematic when the number of parameters is large, which often is the case. Due to this the general autoregressive conditional heteroskedasticity model, also called the GARCH model, is preferred since it evades these issues. The GARCH model was first presented by Tim Bollerslev in his paper *Generalized autoregressive conditional heteroskedasticity* 1986 [8]. To continue the theory surrounding the GARCH models, a formal definition of the GARCH(p,q) model needs to be specified.

Definition:

Given the model $Y_t = \mu_t + \varepsilon_t$, then it is said that ε_t follows a GARCH(p,q) process if the following holds.

$$\begin{aligned} \varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{aligned} \tag{22}$$

where $\{z_t\}$ is assumed to be an identically distributed independent series with expected value 0 and variance 1.

♦

For the definition to hold and be in line with statistical theory $\alpha_0 > 0$ and $\alpha_i, \beta_j \geq 0$ must be true as well as $\Sigma_{s=1}^{\max(p,q)} \alpha_s + \beta_s < 1$. The constraint on the sum of α_s and β_s is imposed in order to guarantee that $\mathrm{Var}\left(\varepsilon_t\right) < \infty$. To better understand this property one may rewrite the model as an ARMA process, but since this is not the focus of the study it is left out. It is common to assume that z_t follows either a standard normal distribution or a t distribution. One important aspect of the GARCH model to be aware of is when $\Sigma_{i=1}^{p}\alpha_i + \Sigma_{i=1}^{q}\beta_i \leq 1$ is true since it ensures strict stationarity [9]. However, $\Sigma_{i=1}^{p}\alpha_i + \Sigma_{i=1}^{q}\beta_i \leq 1$ can only be used as a requirement for strict stationarity for GARCH models with standard normal white noise.

A hardship that comes with the GARCH model, and most other models of similar type, is to determine the order of the model. Due to this issue most applications only consider lower order GARCH(1,1) which has proven to be sufficient.

As mentioned before, the idea behind the GARCH models is to model the conditional heteroskedasticity in a time series model. For the GARCH model to work as intended it is important that the sum of the mean model of the series and the $\{\varepsilon_t\}$ series are stationary. When this condition is met, the underlying time series can be viewed as stationary. The importance of stationarity in the GARCH type models is due to it ensuring that the model for the conditional heteroskedasticity is well behaved. Well behaved in this case means that it has the same properties over time, that is, it behaves somewhat predictable. This aspect of the modelling process is also valid for the EGARCH model, which is covered in the next subsection. [9]

The exponential GARCH model (EGARCH model)

The second model that will be considered in this study is the exponential GARCH model. Despite the GARCH model's ability to adequately describe the conditional heteroskedasticity of some time series data, it has some weaknesses. The weakness that is of greatest concern is the fact that is treats the effects of negative volatility and positive volatility the same. More precisely do negative changes in asset prices tend to lead to larger changes in volatility than the positive changes. There are several ways to define this model, but in this study the following is used.

Definition:

Given the model $Y_t = \mu_t + \varepsilon_t$ and where the following relation is assumed $g(z_{t-i}) = \theta_i z_{t-i} + \gamma_i \left[|z_{t-i}| - \mathbb{E}\left[z_{t-i}\right]\right]$, then it is said that ε_t follows an EGARCH(p,q) if the following holds.

$$\begin{aligned}
\ln \sigma_t^2 &= c + \tfrac{1+\delta_1 \mathbf{B}+\ldots+\delta_{p-1}\mathbf{B}^{p-1}}{1+\xi_1\mathbf{B}+\ldots+\xi_q\mathbf{B}^q} g(z_{t-1}) \Rightarrow \\
\Rightarrow \ln \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i g(z_{t-i}) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 \Rightarrow \\
\Rightarrow \ln \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \left(\theta_i z_{t-i} + \gamma_i \left[|z_{t-i}| - \mathbb{E}\left[z_{t-i}\right]\right]\right) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 \Leftrightarrow \\
\Leftrightarrow \ln \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \omega_i z_{t-i} + \xi_i \left[|z_{t-i}| - \mathbb{E}\left[z_{t-i}\right]\right] + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 \Rightarrow \\
\Rightarrow \sigma_t^2 &= \exp\{\alpha_0 + \sum_{i=1}^{p} \omega_i z_{t-i} + \xi_i \left[|z_{t-i}| - \mathbb{E}\left[z_{t-i}\right]\right] + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2\}
\end{aligned} \tag{23}$$

♦

*Note: In (23) **B** is the backshift operator as defined in the section where the ACF is introduced*

To see why the model is useful to describe asymmetric effects of negative and positive changes, studying the properties of $g(z_{t-i})$ is helpful. By assuming that $\omega_i = 0$ and $\xi_i > 0$ it can be shown that the error term in the EGARCH model has a positive effect on $\ln \sigma_t^2$ if $|z_{t-i}| - \mathbb{E}\left[z_{t-i}\right] > 0$. The effect is of course negative if the opposite holds; $|z_{t-i}| - \mathbb{E}\left[z_{t-i}\right] < 0$. If by instead assuming that $\xi_i = 0$ and $\omega_i < 0$, the positive effect of a negative z_{t-i} on $\ln \sigma_t^2$ can be observed. If z_{t-i} is positive the effect on $\ln \sigma_t^2$ is negative. The first property, where $\omega_i = 0$ and $\xi_i > 0$ is assumed to hold true, accounts for the effect that larger movements in asset prices tend to lead to larger volatility. The second property reflects the idea that negative movements in asset prices lead to greater volatility than positive movements. [10]

The stationarity conditions for the EGARCH(p,q) are similar to those of the ARMA(p,q) model since it can be interpreted as an ARMA process of the $\ln \sigma_t^2$. The conditions is that the solutions for $1 - x\beta_1 - x^2\beta_2 - \ldots - x^p\beta_q = 0$ lie outside the unit circle, this can be compared with the ARMA(p,q) process in (18). [11]

The ARMA-GARCH model family

Before proceeding with methods for parameter estimation and one-step-ahead forecasting, it is essential to note that the GARCH and EGARCH models also may include a mean model. Recall the following general model $Y_t = \mu_t + \varepsilon_t$ where ε_t follows some GARCH or EGARCH model. By allowing the mean to follow an ARMA model, a new type of model family is created, namely the ARMA-GARCH type model family. This family of models includes the pure EGARCH and GARCH models, that is when the mean model is set to zero. For example, by letting μ_t follow an AR(1) process and ε_t a GARCH(1,1) process the AR(1)-GARCH(1,1) model is obtained.

Parameter estimation for GARCH, EGARCH and ARMA-GARCH type models

There are several possibilities when it comes to estimating parameters in models from the GARCH families. One of the estimation methods commonly used is the maximum likelihood estimation, but this proves to be harder than anticipated. The hard part arises from finding the form of $f(x_1, ..., x_m|\boldsymbol{\theta})$ in the following expression where $\mathcal{F}_i$ states that all variables with index less than or equal to i are known.

$$L(\boldsymbol{\theta}) \;=\; f(x_1, ..., x_n|\boldsymbol{\theta}) \;=\; f(x_n|\mathcal{F}_{n-1}) \cdots f(x_{m+1}|\mathcal{F}_m)\, f(x_1, ..., x_m|\boldsymbol{\theta}) \tag{24}$$

To arrive at an expression for the conditional likelihood simply remove the $f(x_1, ..., x_m|\boldsymbol{\theta})$ factor from the likelihood function where the index m is the amount of parameters in the model to estimate. To exemplify, for a GARCH(p,q) model m is equal to $p+q+1$, where the plus one arises from the constant term in the model. This will be the method of estimation implemented for the models in the study.

To translate this to the models, consider the general model $Y_t = \varepsilon_t = \sigma_t z_t$ with the observed sample $\{y_t\}$ and standard normal distributed z_t. The conditional likelihood is then given by the following where $\boldsymbol{\theta}$ is the parameter vector to estimate with dimension $q+p$.

$$f(\varepsilon_{p+q+1}, ..., \varepsilon_n|\boldsymbol{\theta}, \varepsilon_1, ..., \varepsilon_{p+q}) = \prod_{t=p+q+1}^{n} \tfrac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(\tfrac{-\varepsilon_t^2}{2\sigma_t^2}\right) \tag{25}$$

As mentioned in the previous section about maximum likelihood, it is more often than not easier to implement the log-likelihood. With the same conditions stated above the log likelihood can be expressed as

$$\ell(\boldsymbol{\theta}) \;=\; -\frac{1}{2}\sum_{t=p+q+1}^{n} \ln(\sigma_t^2) + \tfrac{\varepsilon_t^2}{\sigma_t^2} \tag{26}$$

If instead it is assumed that z_t follows a standardized t-distribution with an undetermined ν degrees of freedom, the conditional log likelihood is given by

$$\ell(\boldsymbol{\theta}, \nu) \;=\; (N-p-q)\left\{\ln\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \frac{1}{2}\ln[\nu\pi - 2\pi]\right\} + \ell(\boldsymbol{\theta}) \tag{27}$$

Where the last component of the sum on the right side can be specified by the following expression given the estimated ν

$$\ell(\boldsymbol{\theta}) \;=\; -\frac{1}{2}\sum_{t=p+q+1}^{n} (\nu+1)\ln\left(1 + \frac{\varepsilon_t^2}{(\nu-2)\sigma_t^2}\right) + \ln\sigma_t^2 \tag{28}$$

In the equations above the estimation can be specified for the GARCH and EGARCH models by allowing σ_t^2 to depend on a set of parameters, $\boldsymbol{\theta}$, given from either the GARCH or EGARCH model. An aspect to be aware of is that some of these likelihoods are very complicated and are in certain situations in need of further parameter specifications. For example, the GARCH(p,q) requires q pre set σ_t^2 and the rest of the σ_t^2 are recursively computed [3]. Since this can be very difficult it is left out of this paper.

It is important to note that these estimations only are valid for pure processes from the GARCH family, in other words they do not consider a mean model. If a mean model was to be introduced there are two approaches available, two-step estimation or one-step. The two-step method estimates the mean model first and then the GARCH type model, while the one-step estimates both models simultaneously. In both cases one can use conditional maximum likelihood estimations by making an assumption about the distribution for ε_t. In the one-step estimation the conditional likelihood looks like this

$$\ell(\boldsymbol{\theta}) \;=\; -\frac{1}{2}\sum_{t=s}^{n} \ln\left(\frac{(Y_t - \mu_t)^2}{\sigma_t^2}\right) + \ln(\sigma_t^2) \tag{29}$$

Note: The lower bound s in (29) is dependent on the dimension of the parameter vector, $\boldsymbol{\theta}$, containing the parameters that are estimated. The expression in (29) above is from [3]

The log likelihood in (29) is specified in such a way that $\ln \sigma_t^2$ follows a GARCH type process. A similar approach is used when the error term is assumed to follow a standardized t-distribution.

Methods of forecasting conditional variance for EGARCH(1,1) and GARCH(1,1)

To conclude the theory specific to the EGARCH and GARCH models the methods for the one-step-ahead forecast of the conditional variance need to be specified. One way of doing this is to choose a forecast estimation that minimizes the expected difference between itself and the true value given the available information. This approach is equivalent to minimizing the mean squared error of the one-step-ahead value. The estimate that minimizes it is given by the mean of the true one-step-ahead value conditioned on the information available at the forecast origin. In this study the forecast origin is the time before the step that takes the model into the future.

For the general ARMA-GARCH(1,1) model, the one-step-ahead forecast of the conditional variance with forecast origin t is

$$\hat{\sigma}_t^2(1) \;=\; \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta \sigma_t^2 \tag{30}$$

Note: $\hat{\sigma}_t^2(1)$ denotes the one-step-ahead forecast of σ_{t+1}^2 using all information up to and including time t

The same type of forecast for the ARMA-EGARCH(1,1) is given by the following expression

$$\hat{\sigma}_t^2(1) \;=\; \sigma_t^{2\beta_1} \exp\left\{\omega_1 z_t + \xi_1 \left[|z_t| - \mathbb{E}\left[z_t\right]\right]\right\} \tag{31}$$

Since the true values of the parameters in both models are unknown they must be replaced with the estimated parameters. This will of course introduce some uncertainty into the forecast due to the variance of the estimates. It is possible to adjust for this uncertainty with more complex models, but this is left for future more extensive studies.

Method of forecasting the conditional mean

In order to evaluate the quality of the conditional variance forecast it is in some cases necessary to simultaneously forecast the one-step-ahead conditional mean. The general procedure is the same as for the GARCH family models, that is minimizing the expected difference. The expression that fulfils this criterion for an ARMA(p,q) is as follows

$$\hat{Y}_t(1) \;=\; \phi_1 + \sum_{i=1}^{p} \phi_i Y_{t+1-i} - \sum_{j=1}^{q} \theta_i \varepsilon_{t+1-j} \tag{32}$$

This forecast will only be relevant when the mean model is not a constant, since this will affect the backtesting method in the next section.

Forecast evaluation: backtesting of predictive distributions

Note: This entire subsection is inspired by [3]

A very important aspect of forecasting is to formally determine whether it is good or not. As with most aspects of statistical modelling there are several possible methods of dealing with the task at hand. In this study the method is to implement backtesting of the predictive distribution. This type of backtesting requires the following theorem in order to be properly introduced.

Theorem 1 ***([3], p.222)*****:**

If the random variable X has the continuous cumulative distribution function F. Then $F(X) \sim U(0,1)$

Note: For proof of this theorem see appendix A.1

The advantage of backtesting of the predictive distribution is because it evaluates the overall prediction performance. To describe the method in detail, consider the time series $\{Y_t\}$ with conditional heteroskedasticity which for each time t can be expressed as $Y_t = \mu_t + \sigma_t z_t$. The goal is to estimate the cumulative distribution function of $Y_t|\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is the information available at time $t-1$. Let $U_t = F_{Y_t|\mathcal{F}_{t-1}}(Y_t)$, by this transformation $\{Y_t\}$ generates the new process $\{U_t\}$. As in previous sections, z_t is assumed to follow some distribution with zero variance and mean one. These assumptions imply the following property of the cumulative distribution function of $Y_t|\mathcal{F}_{t-1}$

$$F_{Y_t|\mathcal{F}_{t-1}}(y_t) \;=\; \mathbb{P}\left(\mu_t + \sigma_t z_t \leq y_t|\mathcal{F}_{t-1}\right) \;=\; F_z\left(\frac{y_t - \mu_t}{\sigma_t}\right) \tag{33}$$

The result in (33) combined with the fact that $Y_t|\mathcal{F}_{t-1}$ is continuous implies that U_t is a function dependent on ε_t. By Theorem 1 it now follows that the distribution that generates $\{U_t\}$ is the standard uniform distribution.

Applying this in reality is a matter of estimating values for $F_{Y_t|\mathcal{F}_{t-1}}$ and then forming $\hat{U}_t = F_{Y_t|\mathcal{F}_{t-1}}(Y_t)$ for all the forecasts by replacing μ_t and σ_t with their forecasted values. If the number of forecasts is n then the estimations create a sample of $\hat{U}_1, ..., \hat{U}_n$, which should be close to a sample from the standard uniform distribution if the forecasts are good. This leads to the following property of the sample: The higher the quality of the forecast the more the sample will behave like a sample from the standard uniform distribution. One way to assess whether the method is good for forecasting is to look at the histogram of the sample. The more formal approach is to apply the Kolmogorov-Smirnov test with the null hypothesis that the sample, $\{\hat{U}_i\}$, is from the standard uniform distribution. A detailed description of this test can be found in the subsection concerning all formal tests in this study.

Note: This test is of course affected by whether or not a forecast of the mean model is required. To get an idea of its effect it can be helpful to compare the same type of backtesting for the GARCH family forecast with a mean model and one without a mean model

Statistical tests

Kolmogorov-Smirnov test

Note: This entire subsection is based on [12]

Evaluating the backtesting of the predictive distribution involves determining whether or not a sample can be considered standard uniform distributed. The informal way of doing this, which necessarily is not wrong, is to look at plots such as histograms. A more formal approach is the Kolmogorov-Smirnov test, a test that measures the discrepancy between two distributions. The test concerns the two hypotheses: the null which states that the two distributions are the same and the alternative that they are not the same. To test the null hypothesis the following test statistic is used where $F(x)$ is the known cumulative distribution to test for.

$$\begin{aligned} D_n &= \sup_{x\in\mathbb{R}}\{\left|\hat{F}_n(x) - F(x)\right|\} \\ \hat{F}_n(x) &= \sum_{i=1}^{n} \frac{\mathbb{I}\{x_i : x_i \leq x\}}{n} \end{aligned} \tag{34}$$

The distribution of the test statistic (when the random variable to test for is continuous) can be proved to be independent of $F(x)$ implying that the result is only affected by $F_n(x)$. For proof of this see appendix A.3. The null hypothesis will be rejected with significance level α if $D_n > D_\alpha(n)$ where $D_\alpha(n)$ is the α-critical value of the test statistic.

Ljung-Box test for autocorrelation

In time series, autocorrelation is an important aspect to study in order to understand the effects of models and whether they are sufficient or not. To test for autocorrelation in a formal manner, the Ljung-Box test for autocorrelation is often implemented. The test allows for hypothesis testing of the kind where the null hypothesis might be $H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$ and the alternative hypothesis $H_1 : \rho_i \neq 0$ for at least some i. The ρ_i is the auto-correlation of lag i. The test statistic can be expressed in like this

$$Q(m) \;=\; T(T+2)\sum_{l=1}^{m} \tfrac{r_l^2}{T-l} \overset{a}{\sim} \chi^2\left(m\right) \tag{35}$$

The notation $\overset{a}{\sim}$ states that the test statistic is asymptotically chi-squared distributed with m degrees of freedom, where m can be modified in different ways. In this study m is modified by subtracting the amount of parameters in the model. An asymptotic result means that it holds by letting the samples size approach infinity, that is it converges to the result as the sample size approaches infinity, for further reading see [5]. One thing to be aware of when implementing this test is that, depending on the choice of degrees of freedom, the test might not work for the lower lags. For example, the degrees of freedom with $m = 2$ and adjusting for a model of three parameters will be -1 which is not sensible for a chi-squared distribution.

McLeod-Li test

When using volatility models it is interesting to test for autocorrelation between the squared residuals of the mean model. To do this the McLeod-Li test can be implemented since it is the same as the Ljung-Box test with the only difference being that it is applied to squared data. The null hypothesis for this test is $H_0 : \rho_1 = \rho_2 = ... = \rho_m = 0$ and the alternative hypothesis $H_1 : \rho_i \neq 0$ for at least some i. Due to the tests being very closely related to the Ljung-Box test, it will not be discussed in more detail.

ACF and PACF test

When calculating the sample autocorrelation, r_l, of a time series data it is often of interest to test whether it is significantly separated from zero. In order to test this, use the property that r_l is considered to be asymptotically normal distributed with mean zero and variance $\frac{1}{T}$ according to the central limit theorem [6]. Hence, testing the null hypothesis that r_l is zero is done by rejecting it when r_l is outside the interval $\pm\dfrac{1.96}{\sqrt{T}}$ where ± 1.96 are known as the 2.5% and 97.5% quantiles. The use of these quantiles implies that the significance level is at five percent. A similar result can be derived for the PACF under certain assumptions. If it is assumed that the underlying series of the data that the PACF is applied to follows an AR(p) process, then the PACF of lags larger than $1+p$ are normal distributed with mean zero and variance $\frac{1}{T}$ [6]. Due to the demanding requirements for the distribution of the PACF to hold it is important to be wary of this and be careful not to rely on this and rather use it as a supplementary test.

Exploratory data analysis

Before any modelling is done on the dataset it is important to study the characteristics of it. The data in this study is the London Bullion Market Association's (LBMA) daily gold price set at half past ten between the dates 2000-01-04 to 2020-01-04. All the data was retrieved via the Quandl package in R studio. Furthermore, when implementing hypothesis testing in this thesis the significance level is set to be five percent. In Figure 1 the non-transformed price data is plotted. The plot shows that the price does not vary around a constant, a common characteristic of non-stationary time series, but instead steadily increases for most of the time periods. In Figure 2 the difference in behaviour between non stationary and weakly stationary models can be better observed. The process that oscillates around the start-value zero is a stationary AR(1) model while the two others are random-walks without drift, that is AR(1) models with no constant and the single coefficient equal to one. The latter two do not at all behave like the stationary model, but instead tend to drift away from the start value. Since it is likely that the original gold price is not weakly stationary, which violates the model requirements in this study, the data needs to be transformed into something that looks more promising. A common way of transforming assets price data, to acquire stationarity, is by calculating the log return of it, a value that often varies around zero [4].

Figure 1: Daily data of the LBMA gold price

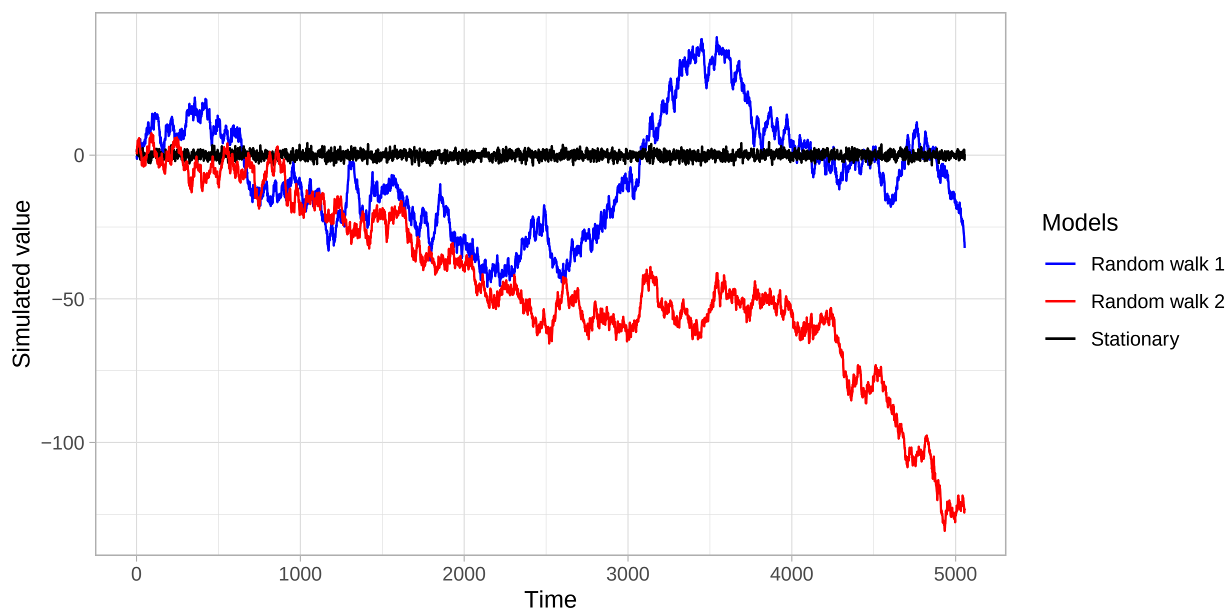

Figure 2: Three simulated time series models of 5055 observations with start point zero

Figure 3 shows the plot of the daily log return of the gold price, in the figure it is observable that the data seems to vary around zero. The property of varying around a constant is a common trait of stationary time series. Therefore, is it likely that the transformed price is at least weakly stationary. The summarising measurements in Table 1 further confirm this property since the maximum and minimum is approximately the same small distance from zero; the mean close to zero and the standard deviation low compared to the mean.

Table 1: Summarising measurements of the log-return of the LBMA gold price

Maximum	Minimum	Sample standard deviation	Mean value
0.0964163	-0.0891278	0.0109744	0.0003404

Since this study's main focus is conditional heteroskedasticity it is interesting to examine the volatility of the log-returns closer. By looking closer at the plot in Figure 3 it is evident that points of high volatility are followed by other points of high volatility, in other words does the data exhibit volatility clustering. Volatility clustering is the trait of having a large fluctuation followed by other large fluctuations in the data. This property is reflected by the GARCH model presented in the previous section [4]. Moreover, does it seem like the data also exhibits the leverage effect (with reservation for the most extreme spikes) reflected in the EGARCH model [4]. The leverage effect is the same as the data exhibiting larger fluctuations after a large negative change compared to the fluctuations caused by a positive change. In conclusion does the data indicate existence of conditional heteroskedasticity and leverage effects, which entails that the GARCH type models in the previous section may be suitable.

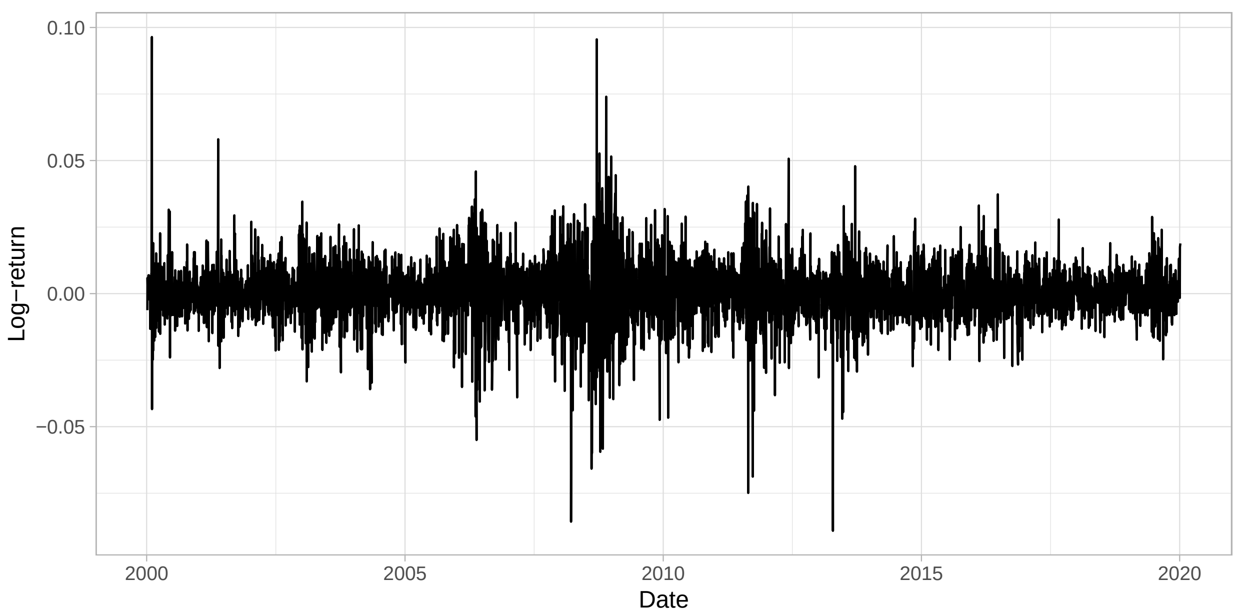

Figure 3: Daily data of the log return of the LBMA gold price

The compatibility of the arma-garch type models with the LBMA gold price

Since the log-return of the LBMA gold price shows proof of existence of conditional heteroskedasticity, leverage effects in volatility and stationarity, the next step is to apply the models. Recall that this study will only focus on joint models of ARMA and EGARCH(1,1) or GARCH(1,1). The justification for the inclusion of this section is to get an idea of what models might be suitable for forecasting instead of guessing wildly which models to try.

In this section the modelling approach is of the two-step type since an appropriate mean model had to be selected before combining it with a preselected GARCH type model, for further reading see [4]. Selecting the mean model independently of the GARCH model is more appropriate since the GARCH models are already preselected and would change the selection of the mean model if selected simultaneously. By selecting the mean model independently, the model is chosen in a way that best captures the first order auto-correlation between the error terms as adequately as possible without considering the second order auto-correlation. The second order auto-correlation is accounted for by the preselected GARCH model if it is appropriate for the data. One downside of this two-step approach is that the parameters are not estimated at the same time. In other words, the ARMA model is selected and estimated independently and the estimation of the GARCH type model is then based on the result from the ARMA model. This might affect the reliability of the model estimation, since considering both models at the same time in the maximum likelihood estimation would probably yield a more correct fit since, in reality, it is one model not two separate. In conclusion is this approach a trade-off between being able to control and understand the model selection and getting a theoretically superior model (one-step) that is hard to control and understand. Despite the inferiority of the two-step model is often not too different from the one-step approach [4].

To select the best mean-model amongst all the ARMA models the AIC model selection approach is used, that is selecting the mean model with the lowest AIC. Out of all the ARMA models the MA(1) model with a zero separated mean has the lowest AIC and is thus selected. The parameter estimations and AIC value can be found in Table 5 in appendix B.1. If the mean model appropriately captures the trends in the data, the residuals should be uncorrelated [4]. The residuals in this case are the residuals from the mean model. By calculating the p-value of the Ljung-Box tests for lags 1-30 based on the residuals one can see that the residuals are not uncorrelated for all lags, see Figure 4. That is, with significance level of five percent, the null hypotheses for the tests are rejected at a majority of the lags. Hence, does the mean model not capture the trends in the data well enough since there is first order correlation between the residuals which should not present in well fitted models. Hopefully this can be fixed by combining the mean model with the GARCH type model. Based on this the modelling continues while keeping in mind that the mean model is not perfect.

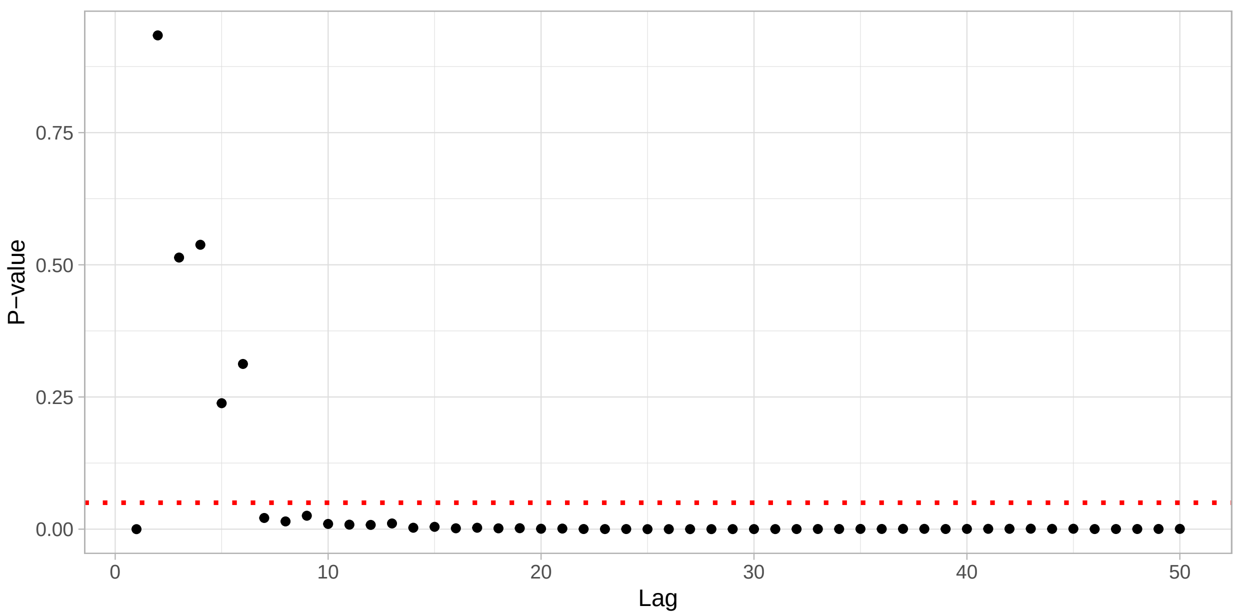

Figure 4: P-values for the Ljung-Box tests on the residuals from the MA(1) model from the AIC selection. The red dotted line represents the 5% level

To see if the data exhibits any conditional heteroskedasticity after the mean model is selected, the autocorrelations of the squared residuals are tested using the McLeod-Li test [4]. In Figure 5 the p-values for the tests are shown and they are lower than five percent, meaning that the null hypothesis is rejected for all lags at the five percent level. This leads to the conclusion that there is auto-correlation between the squared residuals. It is important to remember that the null hypothesis for lag i is that all autocorrelations between all lags up to lag i are zero. This is a good way to check for conditional heteroskedasticity since the squared residuals are estimates of the squared error terms. Due to this result it is suitable to combine the mean model with an EGARCH(1,1) and GARCH(1,1).

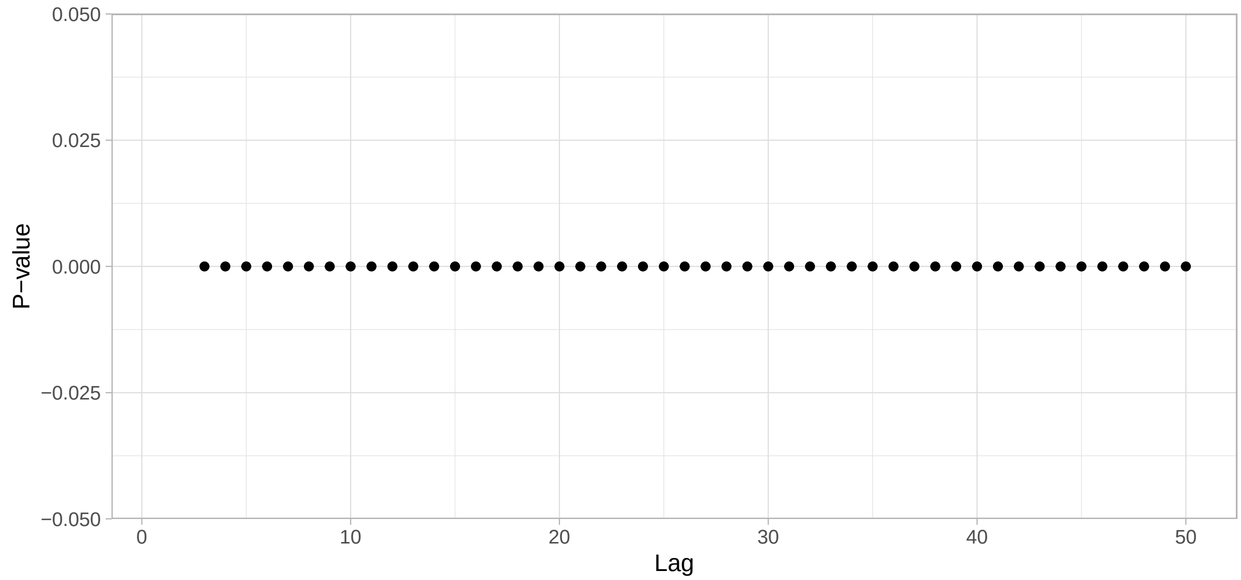

Figure 5: P-values for the McLeod-Li tests on the residuals from the MA(1) model from the AIC selection.

Joining the mean model with the model for the conditional heteroskedasticity yields the parameter estimations in Tables 6-7 *(see appendix B.2-B.3)*. It is easy to see that these models fulfil their respective stationarity conditions since the sum of α_1 and β_1 in the MA(1)-GARCH(1,1) model is less than one and the single solution to $1 - 0.9890895x = 0$ in the MA(1)-EGARCH(1,1) model is positioned outside the unit circle. The mean model which is a MA(1) process is always weakly stationary as stated in earlier sections. The last thing to check in order to confirm the validity of these models is the correlation amongst the squared and ordinary standardised residuals. Standardising in this case is done by dividing the residuals with their respective estimated variances. If the standardised residuals are uncorrelated the mean model is satisfying and if the squared standardised residuals are uncorrelated the conditional heteroskedasticity model is satisfying [4]. To examine this, the Ljung-Box test and McLeod-Li test are applied to the ordinary standardised residuals. By looking at the results from the tests *(see Figure 6-7)*, the conclusion is that both of the models, MA(1)-EGARCH(1,1) and MA(1)-EGARCH(1,1), explain the conditional heteroskedasticity really well since the p-values are high. High p-values mean that the null hypothesis fails to be rejected in each case at the five percent level. Despite the good results in the squared standardised residuals, do neither of the models explain the linear relationship in the data. The reason is because the p-values from the Ljung-Box test for the ordinary standardised residuals are below five percent for the lower lags. Another observation that can be made is that the results from the Ljung-Box test on the ordinary standardised residuals is contradicted by the ACF and PACF plots in Figure 14 *(see appendix B.4)*. The ACF and the PACF plots do show signs of correlation at lag one but insignificant signs of correlation for the rest of the lags since all values are positioned within the dashed lines. The somewhat contradicting results make them hard to interpret, since they indicate good and bad fitting of the models at the same time. However, this contradiction is not present in the ACF and PACF plots in Figure 15 for the squared standardised residuals since all bars are inside the confidence interval. This means that the null hypothesis of them being zero is not rejected *(see appendix B.5)*. This entails that there is no doubt that the conditional heteroskedasticity is well explained by the models. Due to the ambiguous result in the ordinary standardised residuals a model attempt without any mean model will also be done in order to check if the GARCH type models on their own are good fits.

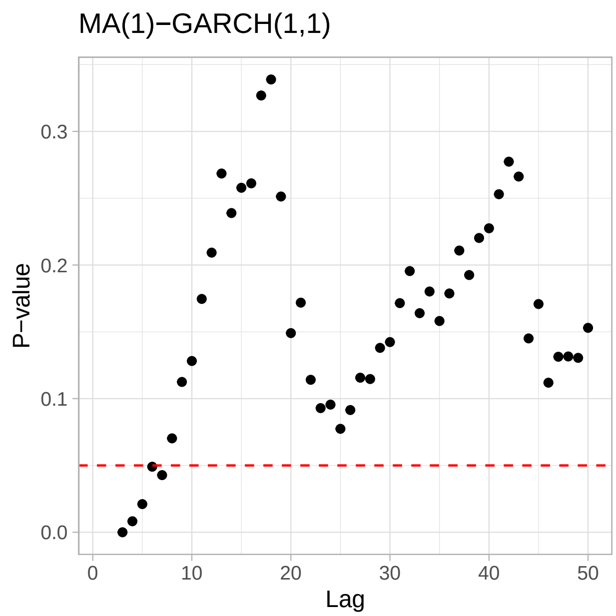

Figure 6: P-values from Ljung-Box test applied to ordinary standardised residuals for both MA(1)-GARCH(1,1) and MA(1)-EGARCH(1,1). The white noise series is set to follow a standard normal distribution. The red dotted line represents the five percent level.

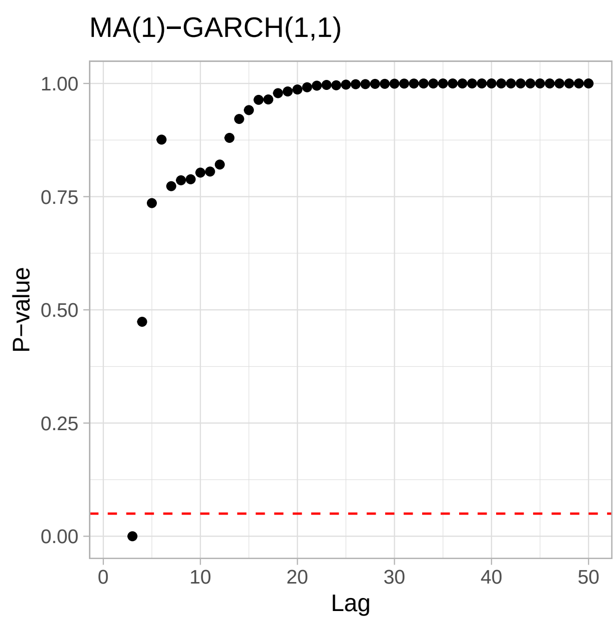

Figure 7: P-values for McLeod-Li test applied to standardised residuals for both MA(1)-GARCH(1,1) and MA(1)-EGARCH(1,1). The white noise series is set to follow a standard normal distribution.

Having no mean model is of course the same as setting it to zero which yields the pure conditional heteroskedasticity model $Y_t = \varepsilon_t = \sigma_t z_t$ where ε_t follows the GARCH(1,1) model in one of the models and the EGARCH(1,1) in the other. An implication of having no mean is that the log return of the gold price is viewed as the mean model residuals, which are used to model the conditional heteroskedasticity part of the model. To study these simpler models, the same method is implemented as for the combined ARMA-GARCH type models. Without any mean-model the Ljung-Box tests for the non-standardised residuals from the pure GARCH(1,1) and EGARCH(1,1) show strong indications of auto-correlations. This is due to the very low p-values which means that the null hypothesis is rejected at all lags with a significance level of five percent, see Figure 8. The results from the McLeod-Li tests for the residuals show that there is conditional heteroskedasticity present, since the null hypothesis can be rejected for all lags with significance level at five percent *(see Figure 9)*. This indicates that the pure GARCH(1,1) and EGARCH(1,1) models might be suitable. Despite the really poor result in auto-correlation among the residuals in both models, the standardised residuals will be studied closer to make sure nothing is left unchecked.

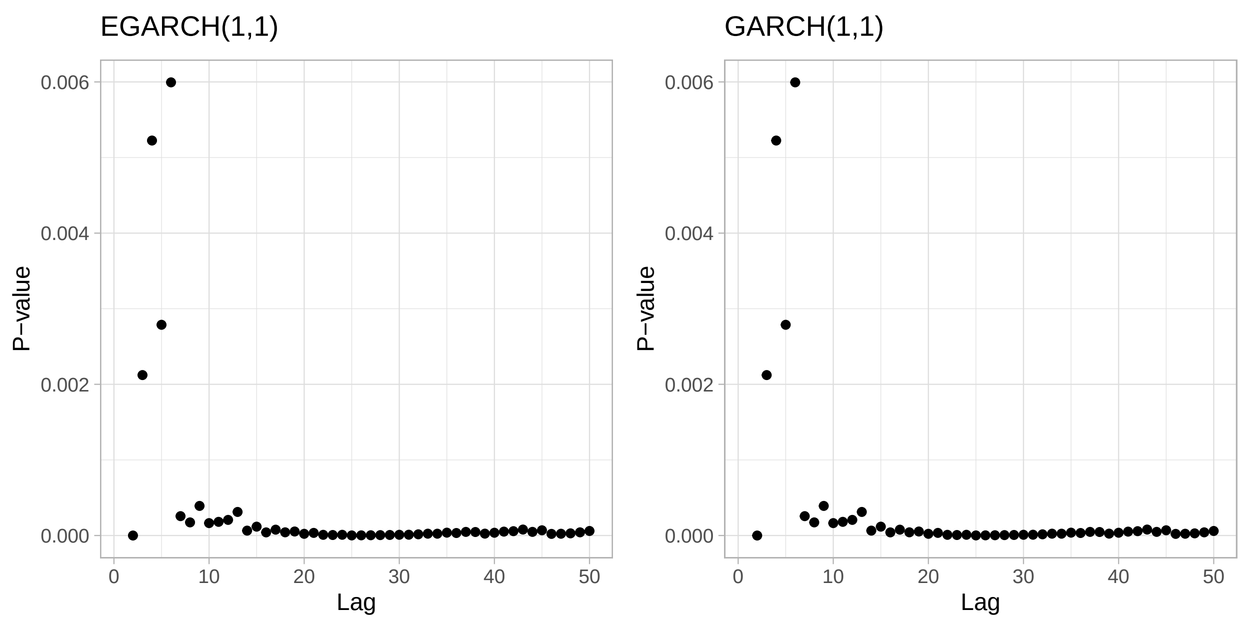

Figure 8: P-values for the Ljung-Box test on the residuals from the pure GARCH(1,1) and EGARCH(1,1). The white noise series is set to follow a standard normal distribution.

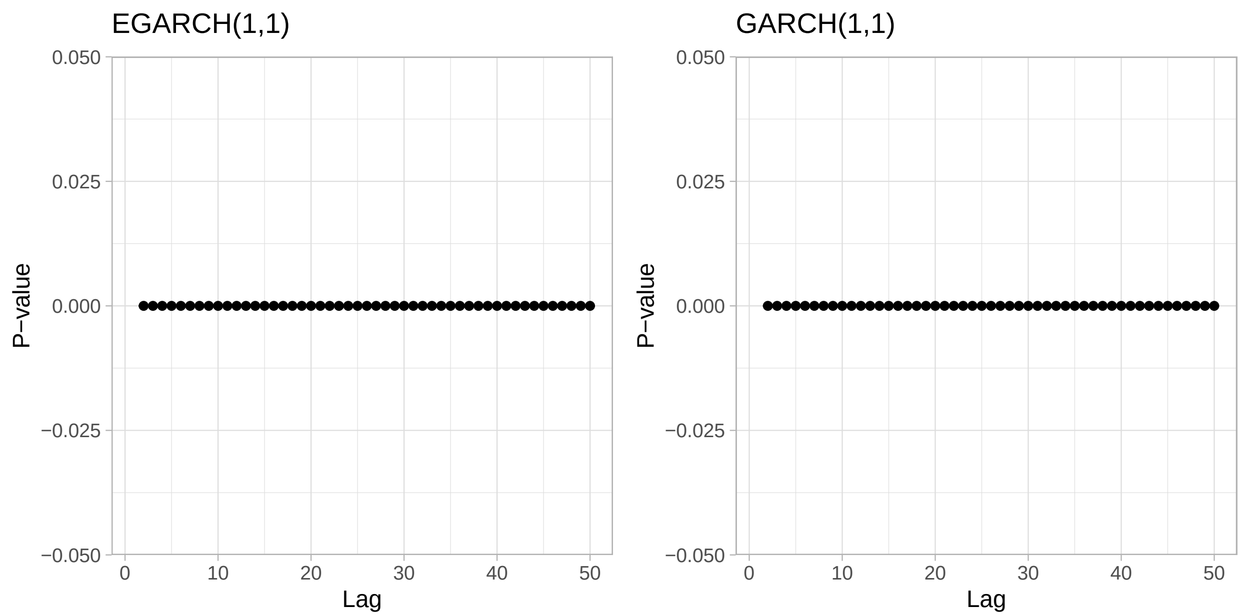

Figure 9: P-values for the McLeod-Li test on the residuals from the pure GARCH(1,1) and EGARCH(1,1). The white noise series is set to follow a standard normal distribution.

As stated before it is more important to study the standardised residuals because they can show whether the models are adequate after the whole model fitting is done. In other words, the standardised residuals hold information of what was left unexplained by the model. When doing the same test on the standardised residuals the results from the Ljung-Box tests show that there are no signs of auto-correlation in neither model due to the very high p-values on all lags *(see Figure 10)*. That is, the null hypothesis can not be rejected for all lags at the five percent level. This clearly shows that no mean model is better than having a MA(1) mean model, which might not be too surprising since the log return of the LBMA gold price seems to vary around zero with time varying conditional variance. The same result is attained when looking at the PACF and ACF plots for the standardised residuals of these models since all bars are inside the dotted lines *(see Figure 16 in B.6)*. Checking whether the ordinary standardised residuals are correlated or not is not enough, the squared standardised residuals also have to be deemed uncorrelated for the model to be a good fit. The p-values for the McLeod-Li tests in Figure 11 are very high for all lags which indicates that the conditional heteroskedasticity in the data is captured by both the pure GARCH(1,1) and EGARCH(1,1) models. Similar to the ordinary standardised residuals is this result supported by the PACF and ACF plots for the squared standardised residuals due to all bars being positioned inside the dotted lines *(see Figure 17 in B.7)*.

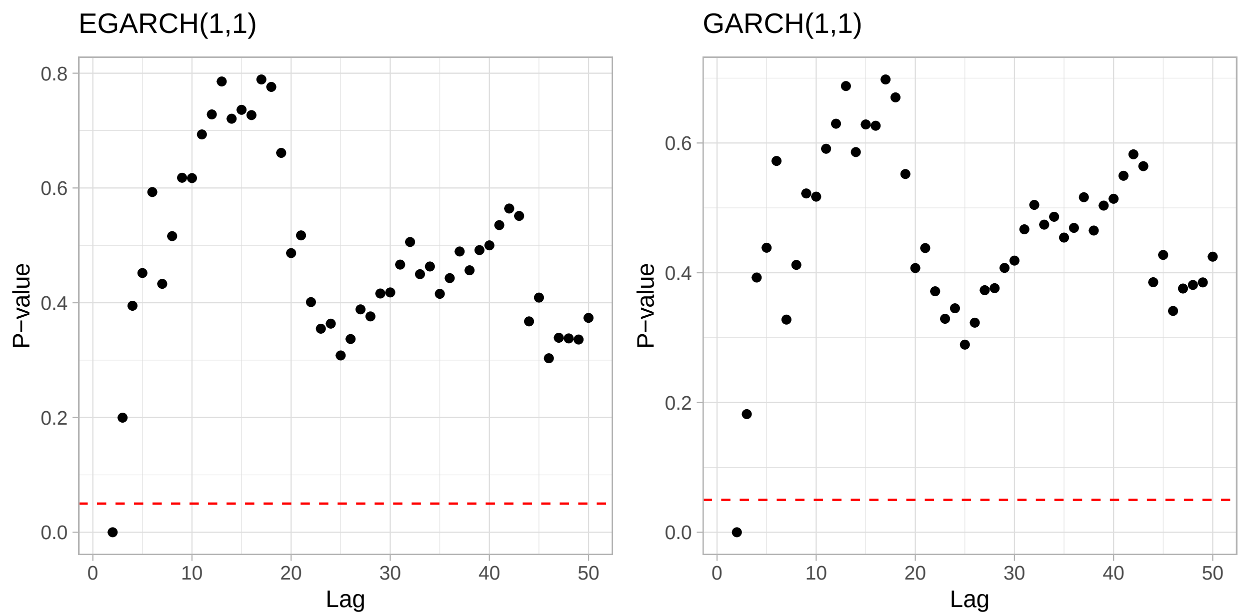

Figure 10: P-values for the Ljung-Box test on the standardised residuals from the pure GARCH(1,1) and EGARCH(1,1). The white noise series is set to follow a standard normal distribution. The red dotted line represents the five percent level

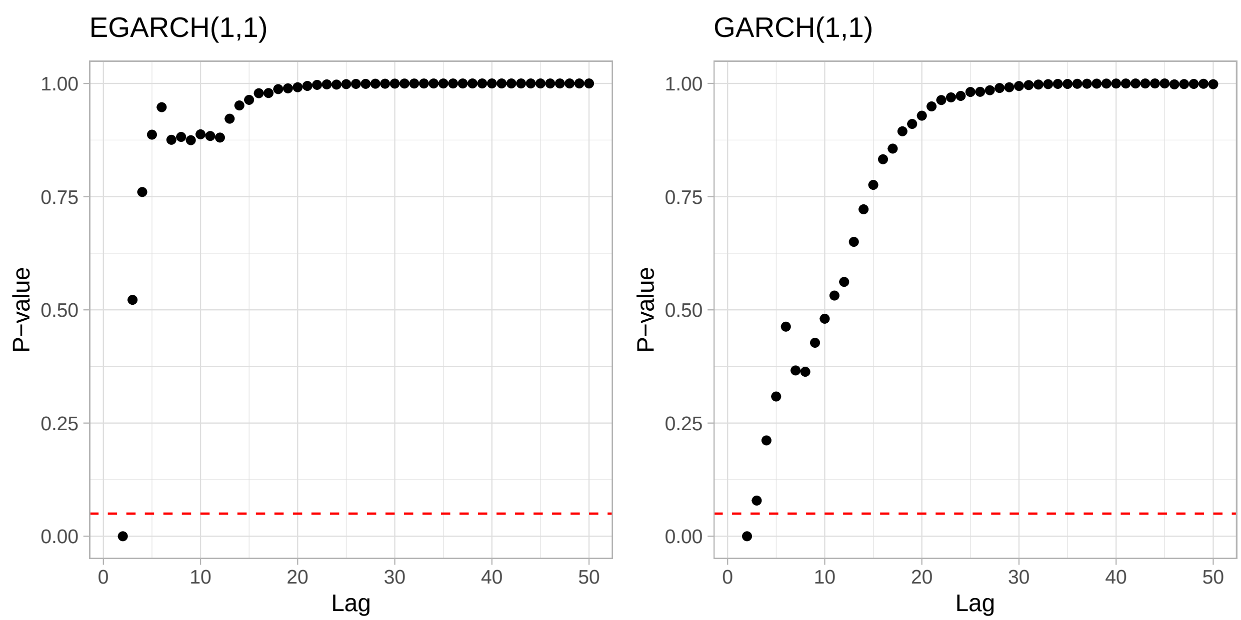

Figure 11: P-values for the McLeod-Li test on the standardised residuals from the pure GARCH(1,1) and EGARCH(1,1). The white noise series is set to follow a standard normal distribution.

To conclude this section, the best fittings for the data as a whole is the pure GARCH(1,1) and EGARCH(1,1) models since they show no signs of auto-correlations in the squared and ordinary standardised residuals. However, all the models in this section will be used in the one-day-ahead forecast in the next section. The reasoning behind this is the following: Since the model fitting in this section only is valid for the whole dataset it might be the case that some of the models work better on smaller partitions of the dataset. This implies that the set of models that are good fits for the whole dataset might be different from the set of models with an adequate forecasting ability. This is because the parameter estimations in the forecasting procedure are done on smaller partitions of the dataset. It is important to note and remember that all of the models in this section use the standard normal distribution for the white noise series $\{z_t\}$. This will be changed slightly in the next section where the t distribution also will be tested to see if it performs better or worse.

Rolling window forecast and forecast evaluation

Before performing the forecast part of this study, it is important to note that the one-step-ahead forecast method used is the rolling window method. The rolling window method is done by selecting a window size (in-sample size) for which the model estimations will be based on. This model is then used to forecast the next day's conditional variance as described in the theory section. When a forecast point has been calculated the windows rolls one step ahead, leaving one observation out whilst including a new. This procedure is repeated until the end of the time series data has been reached [13]. When testing the forecasting ability of the ARMA-GARCH type models, both the mean and the conditional variance will have to be estimated in order for the backtesting to be applicable. This is because the sample $\{F_z\left(\frac{Y_t-\mu_t}{\sigma_t}\right)\}$ is what the backtesting is based on and in order to create that sample, both the forecasted conditional mean and variance are needed. Since there is no praxis when it comes to setting window sizes, the following window sizes of 365, 912 and 1825 days will be used [14]. These sizes are not entirely selected in an arbitrary way since it represents a one year, two and a half year and a five-year period, where the last one is large enough to capture periods of uncertainty and the calmer periods before and after. There are in total eight models whose forecast proficiency are to be tested. The models are: MA(1)-GARCH(1,1), MA(1)-EGARCH(1,1), GARCH(1,1) and EGARCH(1,1) where the error terms are either assumed to be standard normal distributed or t distributed where the degrees of freedom are re-estimated for each window.

As stated in the theory section of the study, the forecasting ability will be evaluated by forming the forecasting sample $\{F_z\left(\frac{Y_t-\mu_t}{\sigma_t}\right)\}$ for each forecast. It is important to recall that in this study z_t, the white noise, can be either standard normal distributed or t distributed with estimated degrees of freedom. When forming this sample when z_t is t distributed, it has to be multiplied with $\sqrt{\frac{\nu}{\nu-2}}$. The argument for this is provided in the discussion section of this study and is a matter of programming. If the forecast is good, the sample should be the same as a sample from the standard uniform distribution, due to Theorem 1 as mentioned in the theory section. In order to to assess whether the formed samples are similar to a standard uniform sample, the Kolmogorov-Smirnov test is implemented alongside with histograms of the samples. If the histogram shows bars that have close to the same height for all x-axis values the result indicates that the sample might be standard uniform. In other words does a sample of standard uniform random variables form a jagged lined between the tops of the histogram bars. This procedure is very informal and is therefore supplemented with the Kolmogorov-Smirnov test using the null hypothesis that the sample is from a standard uniform distribution.

First, the largest window size of 1825 days is implemented and the results interpreted. The histogram plots in Figure 12 show that the forecast sample from the MA(1)-EGARCH(1,1) model with standard normal white noise has too many observations around the centre, which is not a trait of the standard uniform distribution as mentioned above. However, this is no true for the same model but with white noise that is t distributed. The histogram for that model does not have any extreme amounts of observations in any interval, they are instead, to some degree, evenly spread out which is a property of the standard uniform distribution. The pattern in the histogram with standard normal white noise is also present for the histograms of the rest of the models with the same distribution *(see appendix B.8)*. The rest of the models that have white noise that follows a t distribution have histograms that exhibit the same pattern as the model with t distribution in Figure 12 *(see appendix B.8)*. This indicates that forecasts using a window size of 1825 are adequate when the distribution of the white noise terms are assumed to follow a t distribution. This is not the case when the distribution is instead assumed to be the standard normal. For a more formal investigation the Kolmogorov-Smirnov test is applied to the forecasting sample for each model. The results of the tests shown in Table 2 support the results from the histograms. This is due to the rejection of the null hypothesis on a level of five percent in the case of the white noise following a standard normal distribution. The result is the opposite when the distribution instead is the t distribution, in this case the null hypothesis is not rejected on a 5% level. From both of these results the conclusion can be drawn that the models with t distributed white noise using a window size of 1825 days are satisfying in terms of forecasting ability. The models with standard normal distributed white noise are not adequate, as the histograms did not exhibit that jagged line which is a characteristic of standard uniform samples. Furthermore, did the Kolmogorov-Smirnov test reject the null hypothesis on a five percent level.

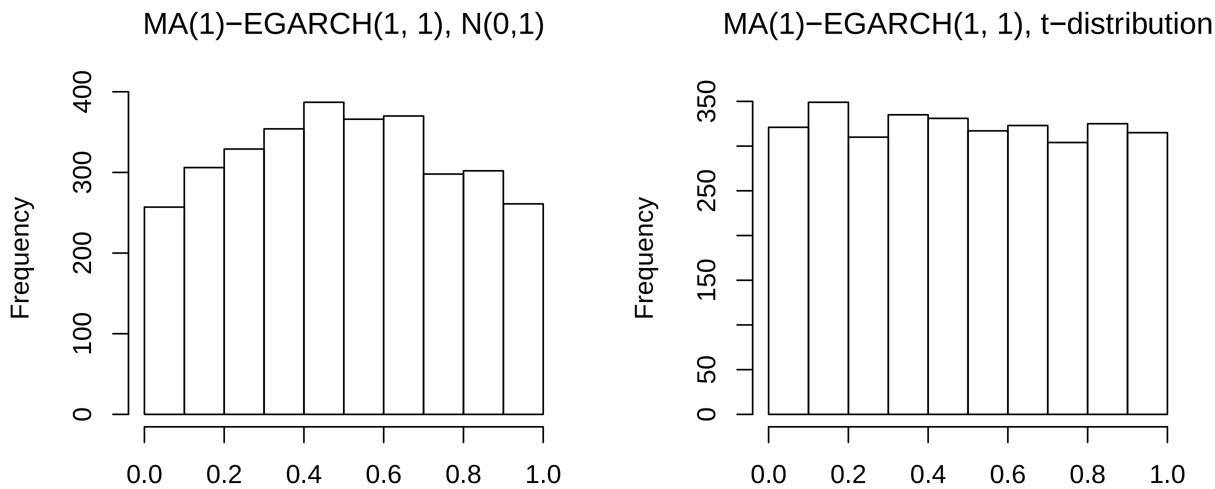

Figure 12: The plot shows the histograms for the sample formed from the forecast using the MA(1)-EGARCH(1,1) models with both standard normal and t distributed white noise terms. Forecast window is of size 1825 days (five years).

Table 2: P-values from the Kolmogorov-Smirnov tests on the forecast samples with window size 1825 days

Model	P-value
MA(1)-GARCH(1,1),N(0,1)	0.000267764381400193
MA(1)-EGARCH(1,1),N(0,1)	0.000116871503193838
GARCH(1,1),N(0,1)	7.593685104057e-07
EGARCH(1,1),N(0,1)	2.17211250441718e-06
MA(1)-GARCH(1,1),t-distr.	0.704405564247135
MA(1)-EGARCH(1,1),t-distr.	0.386272721795066
GARCH(1,1),t-distr.	0.102674494444867
EGARCH(1,1),t-distr.	0.101228860616781

Since the models are fixed changing the window size might lead to different and hopefully better results. When changing the window size from 1825 days to the smaller size of 365 days, there are two models that are not fit for forecast at all since the estimations fail in some windows. The two models that fail are the MA(1)-GARCH(1,1) and GARCH(1,1) model with white noise following a t distribution. This means that even before analysing the quality of the forecasts two models are already disregarded. For the rest of the forecasts the same procedure for studying their adequacy as in the forecast with window size 1825 days is applied. In Figure 13 the histograms of the forecast samples for the MA(1)-EGARCH(1,1) models are shown. The histograms indicate that the sample with t distributed white noise behaves like a random sample from the standard uniform distribution, while the other sample with standard normal white noise does not. This is due to the histogram for the model with standard normal white noise having too many observation around the centre of the interval from zero to one. The histogram for the sample with the t distribution is very evenly spread out across the interval, which is a signum of the standard uniform sample. The histogram for the MA(1)-EGARCH(1,1) with standard normal distributed white noise is representative for all forecast models with the same white noise distribution and the window size of 365 days *(see appendix B.9)*. The pattern exhibited in the histogram with the t distribution is also present in all the other models with the t distribution. For a more formal verification of the results in the histograms, the Kolmogorov-Smirnov test is implemented. The p-values from the tests can be seen in Table 3. The table shows that the p-values are all lower than five percent except for the MA(1)-EGARCH(1,1) with t distributed white noise, which has an extremely high p-value. For the models with the low p-values, the null hypothesis of the samples being from a standard uniform distribution is rejected at a level of five percent. However, for the MA(1)-EGARCH(1,1) with t distributed white noise the null hypothesis can not be rejected at the five percent level. Thus far do the results change a bit when adjusting the window size to 365 from 1825 days. Two models with the t distribution and window size 365 are non-estimable and the EGARCH(1,1) model with t distribution rejects the null hypothesis in the Kolmogorov-Smirnov test when the window is narrowed down. Non-estimable in this case means that the algorithm used for estimating and forecasting with these two models did not work as it failed in its computations somewhere along the way.

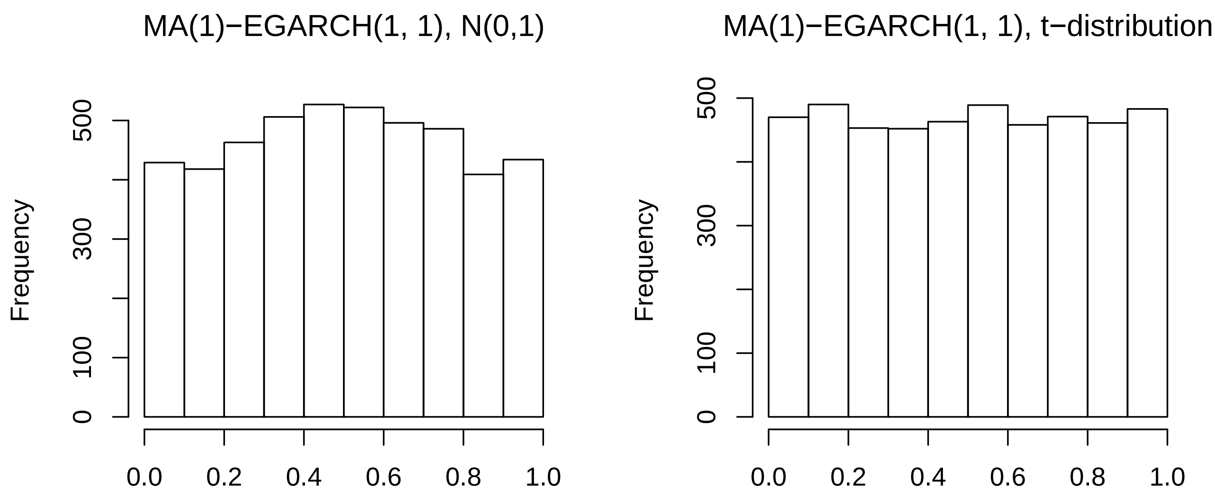

Figure 13: The plot shows the histograms for the sample formed from the forecast using the MA(1)-EGARCH(1,1) models with both standard normal and t distributed white noise terms. Forecast window is of size 365 days (one year).

Table 3: P-values from the Kolmogorov-Smirnov tests on the forecast samples with window size 365 days

Model	P-value
MA(1)-GARCH(1,1),N(0,1)	0.000641132404412437
MA(1)-EGARCH(1,1),N(0,1)	0.00245002587798637
GARCH(1,1),N(0,1)	2.43433709079532e-07
EGARCH(1,1),N(0,1)	9.65124946628038e-09
MA(1)-EGARCH(1,1),t-distr.	0.976361599938559
EGARCH(1,1),t-distr.	0.00290043943771423

Analysing the forecasting ability with a window size that is of size 912 days is analogous with the method for the other window sizes. The histograms for the models with this window size follow the same pattern as the histograms from the models with a window size of 1825 days. The samples from the models with t distributed white noise are evenly spread out over the interval with their tops forming a jagged line. This is, with the risk of over repetition, again a property of the standard uniform distribution. As for the models with standard normal distributed white noise, they are too heavy around the centre of the interval to be considered standard uniform samples. Despite the similarities in histograms do the Kolmogorov-Smirnov tests have different results. The p-values for the tests are shown in Table 4. Based on the p-values, the null hypothesis is rejected at the five percent level, except for the MA(1)-GARCH(1,1) and MA(1)-EGARCH(1,1) models with t distributed white noise. For these two models the null hypothesis can not be rejected at the five percent level.

Table 4: P-values from the Kolmogorov-Smirnov tests on the forecast samples with window size 912 days

Model	P-value
MA(1)-GARCH(1,1),N(0,1)	0.000129307800399991
MA(1)-EGARCH(1,1),N(0,1)	0.000504312907966042
GARCH(1,1),N(0,1)	7.15838477383102e-09
EGARCH(1,1),N(0,1)	6.44961983820025e-09
MA(1)-GARCH(1,1),t-distr.	0.90003591817741
MA(1)-EGARCH(1,1),t-distr.	0.989269168638489
GARCH(1,1),t-distr.	0.00440907277702962
EGARCH(1,1),t-distr.	0.00405124980739546

To summarise, the forecasting ability of the eight different models does vary when changing the size of the window. The window size of 1825 days can be considered to be the best since it has the least amount of rejected null hypotheses in the Kolmogorov-Smirnov tests that also agree with their respective histograms. The model that is optimal for the data, out of the possible ones, is the MA(1)-EGARCH(1,1) model with t distributed white noise, since it consistently did not reject the null hypothesis in the Kolmogorov-Smirnov test and exhibited the desired traits in the histograms. Why the result is the way it is will be discussed in the discussion section.

Results

When fitting an ARMA-GARCH type model to the data two mean models were used, the MA(1) and no mean model. When combining these mean models with the GARCH(1,1) and the EGARCH(1,1) models it was found that the pure GARCH(1,1) and EGARCH(1,1) were suitable for the data as a whole. Despite the presence of auto-correlation in the models with the MA(1) mean model, they were included when studying whether the conditional variance was forecastable. When evaluating the forecasting ability of the 8 models three sizes of the rolling window were chosen: 365, 912 and 1825 days. To test whether the models were capable of forecasting the conditional variance, the histograms of the $\{F_z\left(\frac{Y_t-\mu_t}{\sigma_t}\right)\}$ samples were studied. For a model to be good the histogram had to display an evenly distributed sample. Furthermore, the Kolmogorov-Smirnov test was implemented using the null hypothesis that the sample is from a standard uniform distribution. The histograms for the models with standard normal distributed white noise were consistently, through all window sizes, found to be too heavy around the centre in the interval from zero to one. The p-values for the Kolmogorov-Smirnov tests for these models were always less than five percent which lead to rejection of the null hypothesis at the five percent level. Putting these two results together lead to the conclusion that the models with standard normal distributed white noise were not suitable for forecasting. When evaluating the forecasting ability of the models with t-distributed white noise, the top of the bars in the histograms formed jagged lines. This is a trait which is shared with the realisation of a random sample from the standard uniform distribution. However, the good results in the histograms were not always shared by the Kolmogorov-Smirnov tests: In the window size of 1825 days no null hypothesis was rejected; in the window size of 365 days the null hypothesis was rejected for the EGARCH(1,1) model; in the 912 days window the null hypothesis was rejected for the GARCH(1,1) and EGARCH(1,1) models. The model that was consistently found to have an adequate forecasting ability was the MA(1)-EGARCH(1,1) model with t distributed white noise. Furthermore, the standard normal distribution was found to be a poor choice of white noise distribution since these models often had poor looking histograms and rejected null hypotheses.

Discussion

The main result of this study is that the conditional variance of the log return of the LBMA gold price is indeed forecastable for some of the presented models. However, this only holds true when the the eight models in the study have t distributed white noise. One interpretation of this is that the standard normal distribution, as opposed to the t distribution, does not have tails heavy enough to appropriately model the white noise. This can be seen in the histograms of the forecasting samples from the models with the standard normal distribution. In these histograms the mass is too large at the centre of the interval one to zero which indicates that the sample is lacking in observations farther out in the tails. A point worth making is that throughout the rolling window forecast the degrees of freedom of the t distribution is re-estimated for each roll. This will probably make the distribution fit the data better, than if the degrees of freedom would have been selected pre-forecasting. That is, before doing the actual forecasting select one degrees of freedom which is set to be the same for every forecast point. This means that the forecast of the models might not have performed as well if the degrees of freedom were pre-selected.

Another aspect concerning the t distributed white noise is the forming of the samples $\{F_z\left(\frac{Y_t-\mu_t}{\sigma_t}\right)\}$. Since these are created using the *pt*() function in R each $\frac{Y_t-\mu_t}{\sigma_t}$ has to be multiplied with $\sqrt{\frac{\nu_t}{\nu_t-2}}$ where ν_t is the estimated degrees of freedom. The reason for doing this is because the white noise is assumed to follow a t distribution with variance one. This becomes problematic when evaluating the cumulative distribution function using ν_t degrees of freedom and the *pt*() function, which assumes that the sample has the variance $\frac{\nu_t}{\nu_t-2}$. This can obviously not hold since the model assumes that the variance is one. To solve the problem the variance is corrected by the multiplication, since it alters the variance from one to $\frac{\nu_t}{\nu_t-2}$. When first checking the forecasting ability of the models this approach was not implemented and wrongly so. Ignoring this variance correction gives the result that no model is capable of forecasting the conditional variance.

One thing worth noting is that when using this type of backtesting and the model of interest has a zero separated mean model, it is very hard to determine how much the forecasting of the mean model affects the result. Recall that the forecast samples are created by evaluating the cumulative distribution function of the white noise variable at $\dfrac{y_t-\mu_t}{\sigma_t}$. Since the μ_t and σ_t are replaced with their forecasted values, the sample values will be influenced by the accuracy of the forecast for both the conditional mean and variance. This might be an explanation to why there are some discrepancies in the forecasting abilities in the models that differ only on having a mean model or not. This is especially relevant when the backtesting is applied to the models with window size of 912 days. In this window only the MA(1)-GARCH(1,1) and MA(1)-EGARCH(1,1) with t distribution are able to forecast the conditional variance. Since neither the MA(1)-GARCH(1,1) or MA(1)-EGARCH(1,1) with standard normal distribution or the pure GARCH and EGARCH models (regardless of distributional assumption) were able to forecast the conditional variance, it appears like the MA(1) mean model has influenced the results. It could be the case that the models with standard normal distributed white noise are more affected by the distribution of white noise rather than the mean model. However, on a cautionary note it might not always be the case that the effect of a mean model is good.

Something that the result does not imply is which of the two volatility models are best. Whenever the histogram of the forecasting samples is good for the model using the GARCH(1,1) the corresponding model with the EGARCH(1,1) is also good. The same tendency can be observed in the case when the histogram does not display the desired traits of a standard uniform sample. This pattern is also exhibited by the Kolmogorov-Smirnov test in terms of rejection or failure to reject the null hypothesis at the five percent level. To be clear, it is not possible with the results at hand to determine which of the two models that generally performed the best. In order to determine this, further studies have to be conducted on this subject involving other methods than the ones used in this.

One weakness in the modelling approach used in this study is the selection of the volatility models. It is in the beginning stated that the models that this study will implement are the GARCH(1,1) and EGARCH(1,1). These are chosen based on the fact that they are commonly used in similar studies and have in many occasions been found to perform well. For a more general approach and interesting result, more and very different models could have been used. The easiest way to achieve some sort of variability is to use some model selection criterion to determine the most appropriate choice of order in the GARCH and EGARCH models.

The results of this study might be used when speculating in the gold market, but it is important to be aware of the fact that the data used in this study is only for one particular price. Applying these models to similar prices does not automatically mean that they will perform well. This is due to the possibility of similar prices having vastly different data. To get a more accurate result for the gold market in general it might be better to use models based on an index for the gold market as a whole.

Further research

There are several areas of improvement for this study one being the backtesting. The backtesting used is not wrong in any way, but it would be interesting to evaluate whether different backtesting methods yield different results. One could extend the testing even further to see which backtesting method is the most reliable. For other approaches to backtesting see [3]. Another aspect left out in in this study is the time evolution of stationarity during the rolling window forecast. In other words, it would be interesting to examine if models tend to stay stationary or if they become non-stationary after some point in time. A topic of further investigations concerning the data, would of course be to study how the inclusion of the financial crisis has impacted the result of this study. A simple way of doing this might be to divide the data into four sections and study the differences: One with the crisis at the end; one with it at the beginning; one after while excluding it and lastly one before while excluding it. Lastly it would be interesting to see how the models perform when including the data from the whole of spring 2020. This period should be pretty volatile due to the spread of the pandemic disease COVID-19 which has caused great shocks on the stock market.

References

[1] D.G. Baur, B.M. Lucey, Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold, Financial Review. 45 (2010) 217–229.

[2] M. Capinski, T. Zastawniak, Mathematics for finance : An introduction to financial engineering, 2nd ed., Springer, New York, 2011.

[3] A.J. McNeil, R. Frey, P. Embrechts, Quantitative risk management: Concepts, techniques and tools - revised edition, Princeton University Press, 2015.

[4] R.S. Tsay, Analysis of financial time series, 3rd ed., Wiley, 2010.

[5] L. Held, B.D. Sabanés, Applied statistical inference likelihood and bayes, Springer Berlin Heidelberg, 2014.

[6] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time series analysis: Forecasting and control, John Wiley & Sons, 2015.

[7] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica: Journal of the Econometric Society. (1982) 987–1007.

[8] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics. 31 (1986) 307–327.

[9] S.T. Rachev, T. Jasic, S. Mittnik, S.T. Rachev, Financial econometrics : From basics to advanced modeling techniques, John Wiley & Sons, Inc., Hoboken, 2007.

[10] D.B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica: Journal of the Econometric Society. (1991) 347–370.

[11] T.L. Lai, H. Xing, Statistical models and methods for financial markets, Springer, New York, 2008.

[12] V. Bagdonavičius, K. Julius, M.S. Nikulin, Non-parametric tests for complete data, ISTE/Wiley, London, 2011.

[13] E. Zivot, J. Wang, Modeling financial time series with s-plus® [electronic resource], Second Edition., Springer Science+Business Media, Inc., New York, NY, 2006.

[14] T. Angelidis, A. Benos, S. Degiannakis, The use of garch models in var estimation, Statistical Methodology. 1 (2004) 105–128.

Appendix

Appendix A: Theoretical results

A.1: Distribution of the cumulative function

Theorem 1 ***([3], p.222)*:**

If the random variable X has the continuous cumulative distribution function F. Then $F(X) \sim U(0,1)$

Proof ***([3], p.222)*:**

Let $F^{\leftarrow}(u) = \inf\{x: \ F(x) \geq u\}$ denote the generalized inverse of F.

Let $u \in [0,1]$.

$$\mathbb{P}\left(F(X) \leq u\right) \ = \ \mathbb{P}\left(X \leq F^{\leftarrow}(u)\right) \ = \ F(F^{\leftarrow}(u)) \ = \ u \tag{36}$$

Since the cumulative distribution function is the same as the one for a standard normal random variable $F(X) \sim U(0,1)$ holds. The second equality holds since it can be shown that all cumulative distribution functions are right continuous and then, by definition, it has to hold that $F(x) \geq y \ \Leftrightarrow \ x \geq F^{\leftarrow}(y)$.

■

A.2: Independence of the Kolmogorov-Smirnov test statistic

Theorem 2 ***([12], p.78-79)*:**

Suppose that the $X_1, ..., X_n$ is a sample of a continuous random variable with the continuous distribution function F. The test statistic for the Kolmogorov-Smirnov test is then independent of the cumulative distribution function F.

Proof ***([12], p.78-79)*:**

By Theorem 1 $Y_i = F(X_i) \sim U(0,1)$. Thus, the distribution of the empirical distribution function

$$\hat{G}_n(y) = \sum_{i=1}^{n} \frac{\mathbb{I}\{Y_i : Y_i \leq y\}}{n}$$

is independent of F. Set $u(x) = \sup\{v : F(v) = F(x)\}$ and since F is increasing on $u(x)$ the following holds.

$$\mathbb{P}\left(X_i \leq x\right) \ = \ \mathbb{P}\left(X_i \leq u(x)\right) \ = \ \mathbb{P}\left(F(X_i) \leq F(x)\right) \ = \ \mathbb{P}\left(Y_i \leq F(x)\right) \tag{37}$$

This leads to the following equality $\hat{F}_n(x) \stackrel{d}{=} \hat{G}_n(F(x))$. If the support for X is $(-\infty, \infty)$ then the support for Y must be $[0,1]$. Due to this the next chain of equalities hold true.

$$D_n = \sup_{x\in\mathbb{R}}\{\left|\hat{F}_n(x) - F(x)\right|\} \stackrel{d}{=} \sup_{x\in\mathbb{R}}\{\left|\hat{G}_n(F(x)) - F(x)\right|\} = \sup_{y\in[0,1]}\{\left|\hat{G}_n(y) - y\right|\} \tag{38}$$

The last equality in (38) shows that the test statistic is indeed independent of F

■

Appendix B: Graphs and tables

B.1: AIC-selected mean model

Table 5: Selected mean model based on AIC value in the two-step method

$\hat{\theta}_1$	$\hat{c}_0$	AIC
-0.0399357	0.0003404	-31275.8

B.2: MA(1)-GARCH(1,1)

Table 6: Parameter estimations for the MA(1)-GARCH(1,1) model

$\hat{\mu}$	$\hat{\theta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
0.0003404	-0.0399357	1.2e-06	0.0491501	0.9407756

B.3: MA(1)-EGARCH(1,1)

Table 7: Parameter estimations for the MA(1)-EGARCH(1,1) model

$\hat{\mu}$	$\hat{\theta}_1$	$\hat{\alpha}_0$	$\hat{\omega}_1$	$\hat{\beta}_1$	$\hat{\xi}_1$
0.0003404	-0.0399357	-0.0950102	0.0250871	0.9890895	0.1150995

B.4: PACF and ACF for standardised residuals with mean model

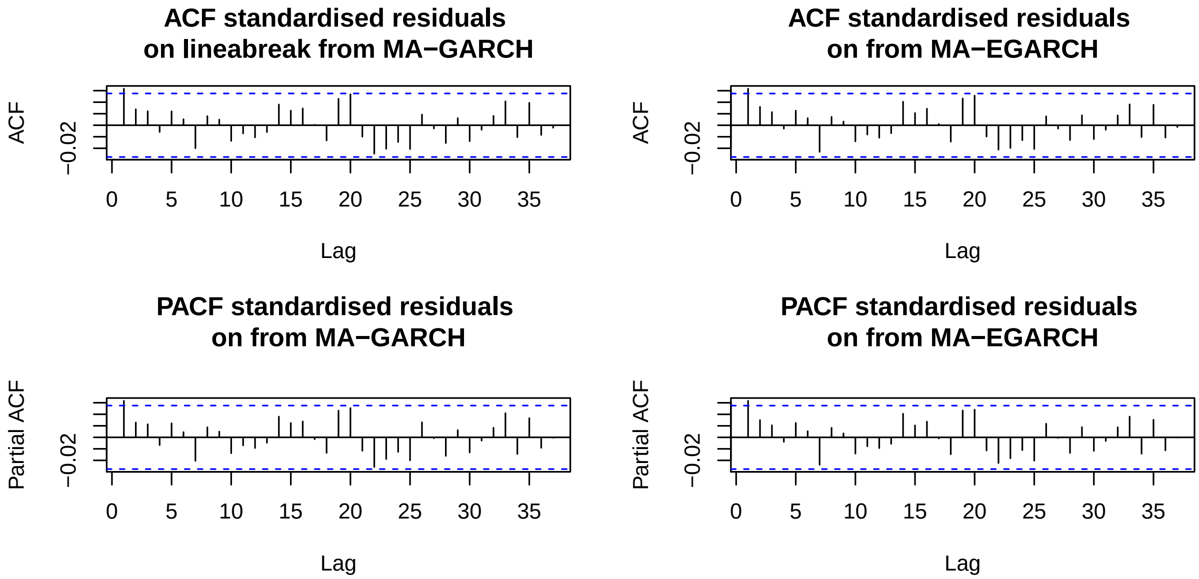

Figure 14: ACF and PACF plots for the standardised residuals from the fitted MA(1)-GARCH(1,1) and MA(1)-EGARCH(1,1)

B.5: PACF and ACF for standardised squared residuals with mean model

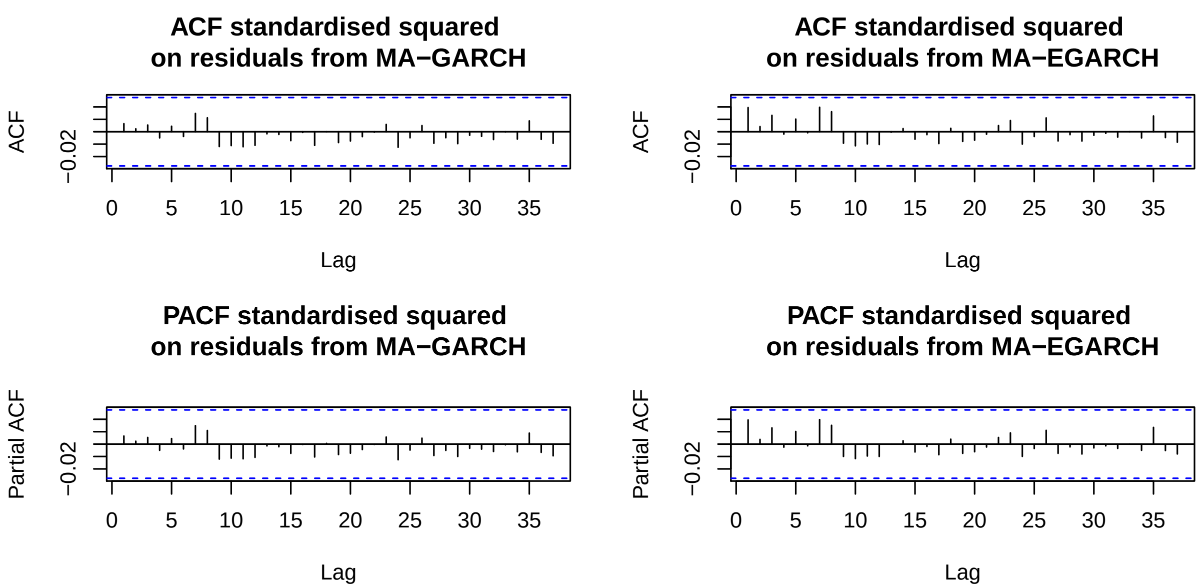

Figure 15: ACF and PACF plots for the standardised squared residuals from the fitted MA(1)-GARCH(1,1) and MA(1)-EGARCH(1,1)

B.6: PACF and ACF for standardised residuals without mean model

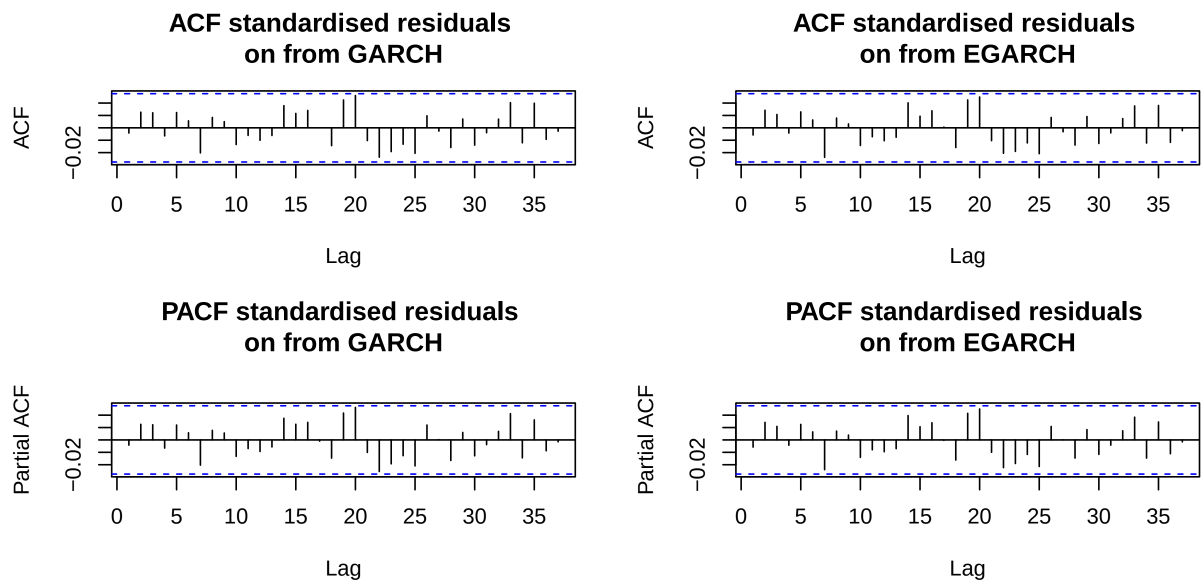

Figure 16: ACF and PACF plots for the standardised residuals from the fitted GARCH(1,1) and EGARCH(1,1)

B.7: PACF and ACF for standardised squared residuals without mean model

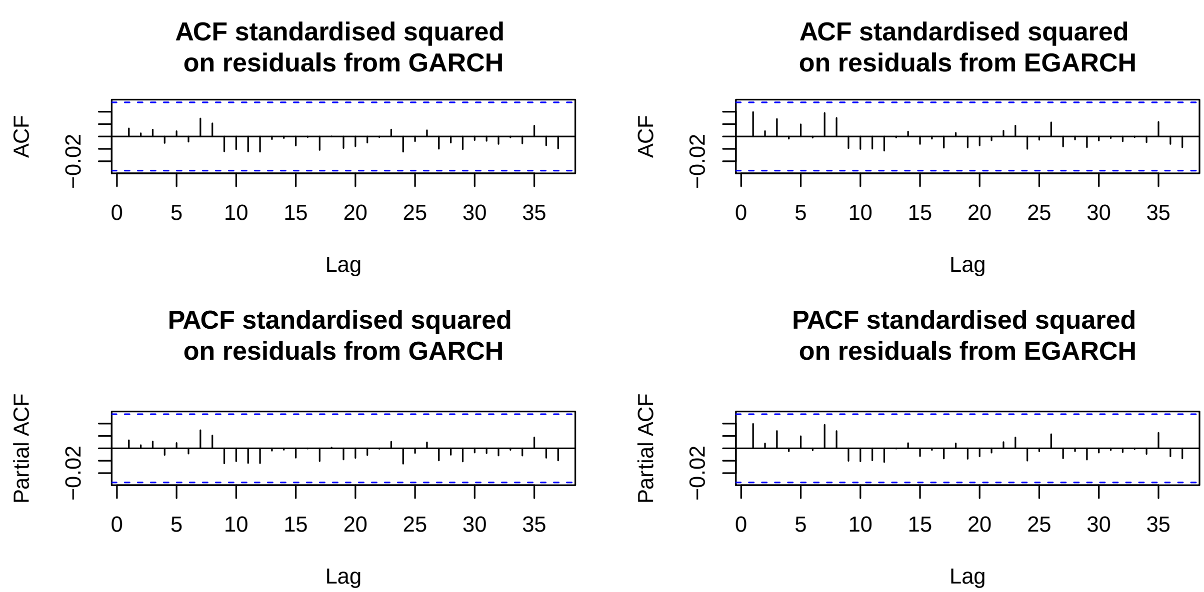

Figure 17: ACF and PACF plots for the standardised squared residuals from the fitted GARCH(1,1) and EGARCH(1,1)

B.8: Histograms for forecasting samples with window size of 1825 days

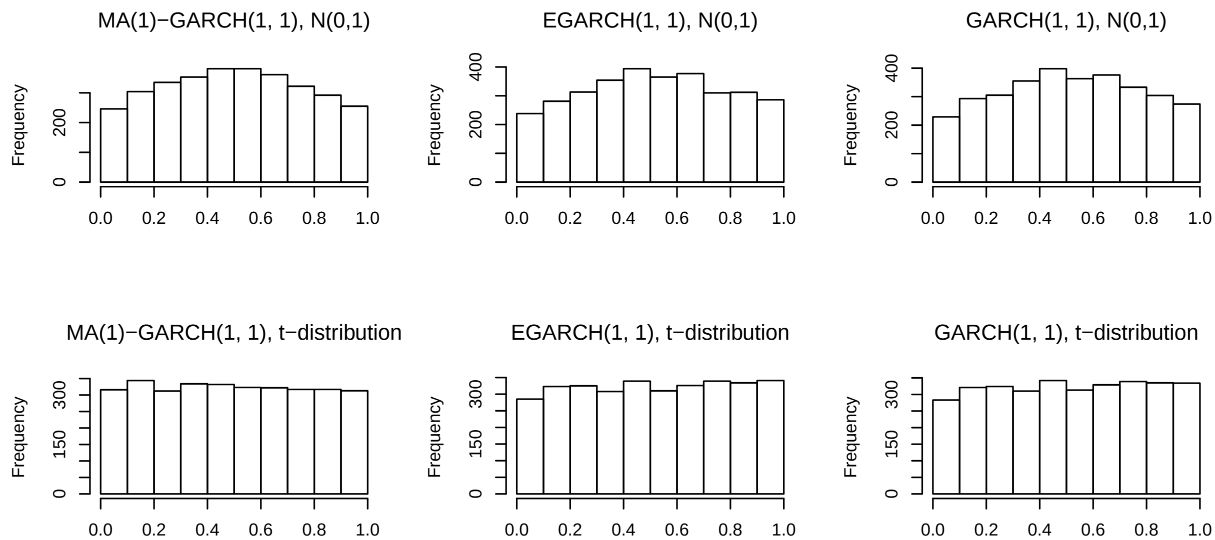

Figure 18: Histograms for the forecast samples with window size 1825 days. The white noise is either standard normal or t distributed.

B.9: Histograms for forecasting samples with window size of 365 days

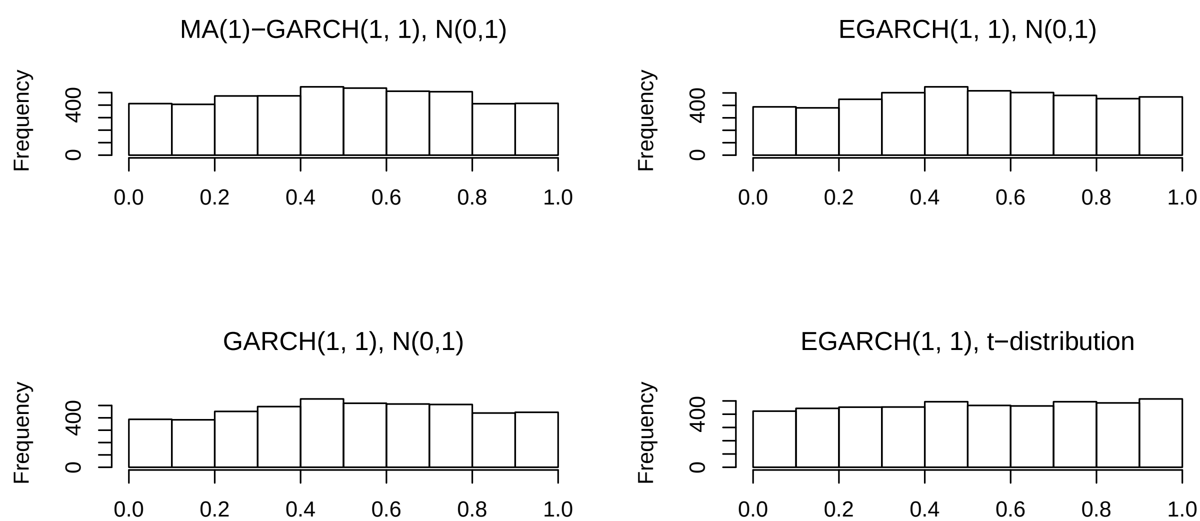

Figure 19: Histograms for the forecast samples with window size 365 days. The white noise is either standard normal or t distributed.

B.10: Histograms for forecasting samples with window size of 912 days

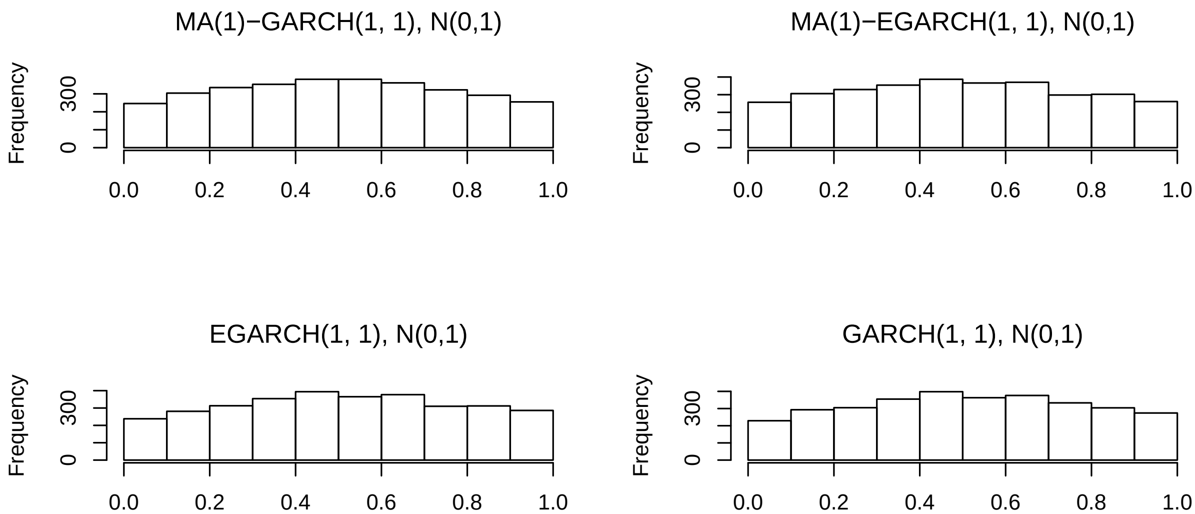

Figure 20: Histograms for the forecast samples with window size 912 days. The white noise is either standard normal or t distributed.

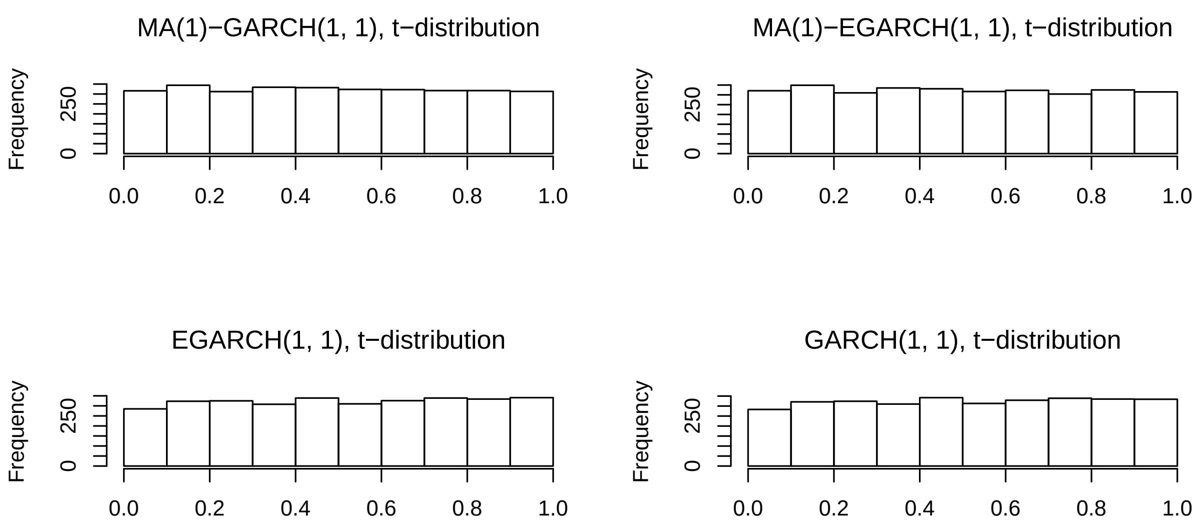

Figure 21: Histograms for the forecast samples with window size 912 days. The white noise is either standard normal or t distributed. Continuation of Figure 20