MODELING FUEL COMMODITIES FORWARD CURVE PRICES

July 7, 2019

Jurn Ullrich (2599333)
Vrije Universiteit Amsterdam
Faculty of Science
Business Analytics
De Boelelaan 1081a
1081 HV Amsterdam

Preface

This thesis is my graduation assignment and fits within the topic of Financial risk management for the master Business Analytics. The problem discussed in the thesis is as follows: Is modelling the margin as a univariate time series better than modelling the underlying products and their interaction?

The assignment is done for Varo energy which I want to thank for the opportunity to do this interesting assignment. I also want to thank my company supervisor M. Arias for helping me with the introduction to a lot of new topics learned while writing my thesis. I also want to thank my university supervisor E. Belitser for the discussion sessions and comments which helped me develop and improve my thesis and lastly I want to thank my second reader A. Ran for taking the time to read and grade my thesis.

Abstract

Varo Energy is a company which specializes in international trading, refining, sales, storage and distribution of (bio) fuels. For the two refinery's Varo requires crude oil for production. The exposure on the margin of the refinery's is a process which fluctuates through time because of the moving crude and product prices. Varo handles this exposure by hedging the margin of the refineries. Varo wants to understand how this margin behaves. There are two ways to model the margin, the first is modelling the margin as a univariate series and the second is modelling the margin as a multivariate portfolio consisting of all the individual products. This thesis discusses two types of time series models to describe the univariate and multivariate model, the first one are the GARCH models and the second the continues models. For the multivariate model, the interaction between the products is modelled via a vine copula. Both the GARCH and continuous models are used to forecast one month ahead. The accuracy of the multivariate model is better but it does not provide a consistent model. The univariate model is a consistent model but lacks in accuracy. The univariate model is preferred over the multivariate model since the slight increase in accuracy is less important than the consistency of the model. This is because the multivariate model underestimates the volatility of the margin which results in a tight interval which does not capture the full extend of the margin volatility.

Contents

1   Introduction

Varo Energy is a company which specializes in international trading, refining, sales, storage and distribution of (bio) fuels. With two refineries (Bayernoil and Cressier) and 47 inland terminals, it is one of the major suppliers of (bio) fuels in north-west Europe. The refineries combined produce about 165.000 barrels per day.

From a refinery perspective Varo has an exposure which is the difference between the value of the products and the crude from which the products are produced. At the moment Varo handles this exposure by hedging the margin of the refinery (The margin is the difference between the price of one barrel of crude and a combination of the price of the produced products from one barrel). Varo has constructed margin baskets to hedge its margins. The tool Varo uses to hedge are forward contracts. Forward contracts are contracts where one party buys a product in the future for a fixed price. The baskets are a weighted average of product forward prices minus the cost of forward crude price (for example, a margin basket could be 2 barrels of gasoline, 1 barrel of diesel, 3 barrels of crude; in this case a margin basket sale would be selling 2 barrels of gasoline, sell 1 barrel of diesel, buy 3 barrels of crude). With these baskets, Varo makes sure that they receive a steady cash flow and fix their exposure to the market. The ratios in the basket are now pre-determined by historical refinery yields (yields are product ratios in terms of output from one barrel of crude). These ratios might not be optimal. To determine what the optimal hedge ratios are, it is possible to minimize certain risk measures. These risk measures depend on the distribution of the forward prices and their interaction, or the univariate refinery margin distribution.

Ultimately Varo is interested in the distribution and behaviour of the refinery margin. The refinery margin can be modelled in two ways:

1. Model the forward prices per product and the interaction between the products. From these models construct the margin as a multivariate model.

2. Model the margin as a univariate time series.

This thesis discusses both methods and compares the results to see which approach achieves the best results. The focus of this paper lies on the following products: Mogas (Benzine), Jet fuel, High sulfur fuel oil (HSFO), Low sulfur gasoil (LSGO also known as Diesel), Naphtha, Brent (Crude oil) and the Bayernoil refinery. Another important point is that there is only one type of forward contract analyzed in the thesis: the one month ahead contract.

Other applications of these models consist of:

- Calculating option prices to compare to the market.

- Calculate forward curves.

- Optimizing the amount of barrels hedged.

- Maximizing Profit.

1.1   Problem statement

Is modelling the margin as a univariate time series better than modelling the underlying products and their interaction?

1.2   Thesis organization

The thesis contains the following sections. Section two contains a list of symbols. Section three contains the literature review here the related literature and models are discussed. Section four contains the data analysis, the analysis consist of general statistics, checking data requirements and preparing data for the models discussed in section three. Section five discusses the estimation methods used and some

proofs are given to see if the parameters converge to the true parameters. Section six is about fitting and evaluating the models to the data and interpreting the output. Section seven contains some of the packages and functions used to get the results. Section eight contains the conclusions and the discussion about the results. Section nine contains possible extensions to the approach discussed in this thesis.

2   List of symbols

T	Length of time series
$\hat{\rho}_j$	j-lag sample autocorrelation
r_t	tth return of a time series defined as $ln(P_t) - ln(P_{t-1})$
r_t^2	tth return volatility
$\hat{\epsilon}_t^2$	tth sqaured residual
$\hat{L}$	Likelihood function
k	Number of estimated parameters
θ	True parameter vector
$\hat{\theta}$	Estimated parameter vector
d	Number of products (variables)
c_e	Copula density of edge e
B_e	Type of Copula of edge e
$j(e), k(e)$	Conditional nodes of edge e
$D(e)$	Conditional set of edge e
τ	Kendall's tau coefficient
F	Marginal distribution
f	probability distribution function
χ_l^2	Chi-squared distribution with l degrees of freedom
m	Number of AR terms in the ARMA model
n	Number of MA terms in the ARMA model
p	Number of past volatility's in the GARCH model
q	Number of error terms in the GARCH model
X, Y	Random variables
$\bar{x}, \bar{y}$	Mean of variables X, Y
$r_{X,Y}$	Pearson correlation between variables X, Y
$E(X), V(X), S(X), K(X)$	Central moments of random variable X (Mean, Variance, Skewness, Kurtosis)

3 Literature overview and methodology

The topic of financial time series modelling contains a lot of information and methods which can be used to model prices of stocks and commodities. According to [4] this topic of research is rather new has been developing rather fast since data is widely available nowadays. In this thesis, two types of models are discussed. First, the generalized autoregressive conditional heteroskedasticity (GARCH) type of models and second the continuous time processes are discussed. An important note is that these two models model the log-return of the financial time series. This is a widely used technique due to making the financial time series a white noise series which makes the financial time series a stationary process. A stationary process is easier to model; c.f. [4].

The goal of this thesis is not only to find a univariate model for the margin but also a combined portfolio containing all the prices. To model the portfolio, it is necessary to model the dependency between the prices.

3.1 GARCH Models

There is a lot of literature available on GARCH models, such as [11] which introduces the GARCH model for the first time and [4] which gives a good introduction into all type of GARCH models and what they try to model. In this paragraph a small introduction is made into GARCH and some properties are discussed. GARCH models are discrete-time models which have been developed to reflect the so called stylized features of financial time series. Stylized features are features such as tail heaviness, volatility clustering and asymmetric leverage effects; c.f. [8]. GARCH models are a type of models which are derived from the ARCH models. The ARCH model is developed by Engle in 1982. Let ϵ_t be a random variable conditionally on the information set $\mathcal{F}_{t-1}$ (the σ-field created by by $\epsilon_{t-j}, j \geq 1$). Engle assumed that $\epsilon_t = z_t \sqrt{h_t}$ where z_t is a sequence of independent, identically distributed (iid) random variables with zero mean and unit variance. This implies that $\epsilon_t | \mathcal{F}_{t-1} \sim D(0, h_t)$ where D stands for the distribution. Distributions as normal, student t or leptokurtic ones are generally assumed. An ARCH model of the order q looks as follows:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2, \tag{1}$$

where $\alpha_0 > 0, \alpha_j \geq 0, j = 1, ..., q - 1$ and $\alpha_q > 0$. The process $(\epsilon_t, t \in \mathbb{Z})$ is assumed to be strictly stationary, ergodic and nonanticipative solution of (1) and (2). An extension to (1) was introduced by Bollerslev (1986) and is as follows:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \tag{2}$$

where in addition $\beta_i > 0$. in comparison to (1) Bollerslev model can have a slow decay of ϵ_t^2 which is often observed in financial time series; c.f. [27]. The GARCH model is one of the popular ARCH models there are other extensions such as exponential GARCH (EGARCH), nonlinear GARCH and time varying GARCH model; c.f. [27]. The choice of GARCH model is dependent on the data set and some of the characteristics of the data set. For example, there are test/diagnostics to find out if there is are asymmetric leverage effects present in the time series, asymmetric leverage effects are when time series react more to negative shocks then to positive shocks (shocks can be news are the market suddenly collapsing). When there are asymmetric leverage effects present one would prefer to use a model which can handle these types of effects such as EGARCH or threshold GARCH (TGARCH). The form of the TGARCH model is as follows:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{q} \gamma_j S_{t-j} \epsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \tag{3}$$

where

$$S_{t-j} = \begin{cases} 1 \text{ if } \epsilon_{t-j} < 0 \\ 0 \text{ if } \epsilon_{t-j} \geq 0. \end{cases}$$

So, when ϵ_{t-j} is less than zero the effect is larger than when it is greater than zero; c.f. [1].

Further, we discuss test/diagnostics in section 4. The results of these test/diagnostics determine the most probable model.

The GARCH models discussed do not have any explanatory variables but it is possible to extend the models such that they can incorporate explanatory variables. This is accomplished by adding dummy variables. Which yields the following expression:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{k=1}^{K} \delta_k e_{t-k}, \tag{4}$$

where e_t is the vector with the explanatory variables. Variables which might improve the fit of the models are trading volume, news announcements or overnight return; c.f. [1].

3.1.1 ARMA-GARCH Models

The GARCH models discussed in paragraph 3.1 model the volatility of the log return series. GARCH models are called conditional variance models so they model variance or volatility of a time series. An extra model is needed to predict the mean process of the log returns. The log returns can be viewed in the following way:

$$r_t = y_t + \varepsilon_t,$$
$$\varepsilon_t = z_t \sqrt{h_t},$$

Where h_t and z_t are as defined in paragraph 3.1. The mean process y_t can be described by an Auto Regressive Moving Average (ARMA) model. An ARMA model of the order (m, n) is defined as follows:

$$y_t = \mu + \sum_{i=1}^{m} a_i r_{t-i} + \sum_{j=1}^{n} b_j \epsilon_{t-j}.$$

The mean process consists of a combination of past log returns and errors.

3.2 Continuous time models

Continuous models try to describe an infinitely small step in X in time interval dt. The models are described in the form of diffusion equations. The general model consists of a drift term and a volatility term see for example (5). In (5) the first term defines the drift term and the second term the volatility term. Equation (5) is called the arithmetic Brownian motion.

$$dX_t = \mu dt + \sigma dW_t. \tag{5}$$

Another continuous model which is considered is a model with mean reversion. This model is called the Ornstein-Uhlenbeck process and is defined as follows:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t,$$

where κ is the rate of convergence to α and α is the mean of the time series; c.f. [15].

These continues models assume that the volatility is constant through time. In paragraph 4.3.1 we demonstrate that this is not the case for the data used in this research. The Heston model introduced by [16] is a model which can model time-varying volatility. Such models have the following form:

$$dX_t = \mu X_t dt + \sqrt{\sigma_t} X_t dW_{1_t},$$
$$d\sigma_t = \kappa(\alpha - \sigma_t)dt + v\sqrt{\sigma_t} dW_{2_t}, \tag{6}$$

where dW_{1_t} and dW_{2_t} are Brownian motions. dW_{1_t} and dW_{2_t} may be correlated via the following equation:

$$dW_{2_t} = \rho dW_{1_t} + \sqrt{1 - \rho^2} dW_{3_t},$$

where dW_{3_t} is a independent of dW_{1_t}

3.3 Copulas and dependency

A portfolio of more than one model consists of two parts. The first part being the models for the prices and the second part being the dependency between the prices. Dependency between prices is often expressed in correlations like the Spearman rank or the Pearson correlation. These types of dependencies are on a linear scale and are based on the assumption that the prices are jointly elliptically distributed; c.f. [10]. Since for many prices the dependency structure is non normal another method is needed to model the dependency; c.f. [24].

Copula's are functions that link univariate distributions functions to form a multivariate distribution function. Consider a vector with random variables $X = [X_1, X_2, ..., X_n]$ with joint distribution F and marginal distributions $F_1, F_2, ..., F_n$. Sklar (1959) introduces Sklar's theorem, which provides the mapping from the individual marginal distribution functions to the joint function which is as follows:

$$F(x) = C(F_1(x_1), ..., F_d(x_d)), \forall x \in \mathbb{R},$$

where C is the chosen copula. If the joint cumulative distribution function is absolutely continuous and marginal cumulative distribution functions are strictly increasing and continuous, then the copula density can be expressed as; c.f. [23]:

$$f(x_1, ..., x_d) = \prod_{i=1}^{n} f_i(x_i) c(F_1(x_1), ..., F_d(x_d)). \tag{7}$$

An important feature of this approach is that the marginal distributions do not need to be similar nor does the choice of the copula depend on the marginal distributions; c.f. [24]. Another important feature of copulas is that they are very flexible and there exist many types of copulas. Examples of copula are shown in figure 1. Copulas cannot model trends, seasonality or volatility clustering which might all be present in the data; c.f. [23]. The copulas are used to model the remaining errors which cannot be explained by the models described in paragraphs 3.1 and 3.2.

3.3.1 Vine Copulas

It is highly probable that the dependency structure between the prices is not identical this means that a single type of copula cannot accurately model the dependency between all the prices. To account for this problem we introduce the notation of Vine copulas. Vine Copulas make it possible to have different dependency structures between prices. This means that the joint distribution in equation (7) needs to be expressed using only the marginal distributions and single copulas; c.f. [2]. For example the density of a three dimensional vine looks as follows:

$$\begin{aligned} f(x_1, x_2, x_3) =& f_1(x_1) f_2(x_2) f_3(x_3) \\ & \cdot c_{12}[F_1(x_1), F_2(x_2)] c_{23}[F_2(x_2), F_3(x_3)] \\ & \cdot c_{13|2}[F(x_1|x_2), F(x_3|x_2)], \end{aligned} \tag{8}$$

where c can be different copulas. The amount of different combinations of the pair copulas increases drastically when having more dimensions. With a problem of six different prices the possible combinations reaches 23,040; c.f. [23]. To handle this appropriately [5] introduces a graphical model denoted as the regular vine model. A d-dimensional graphical model can be represented as $d - 1$ linked trees with nodes N and edges E. The definition is as follows:
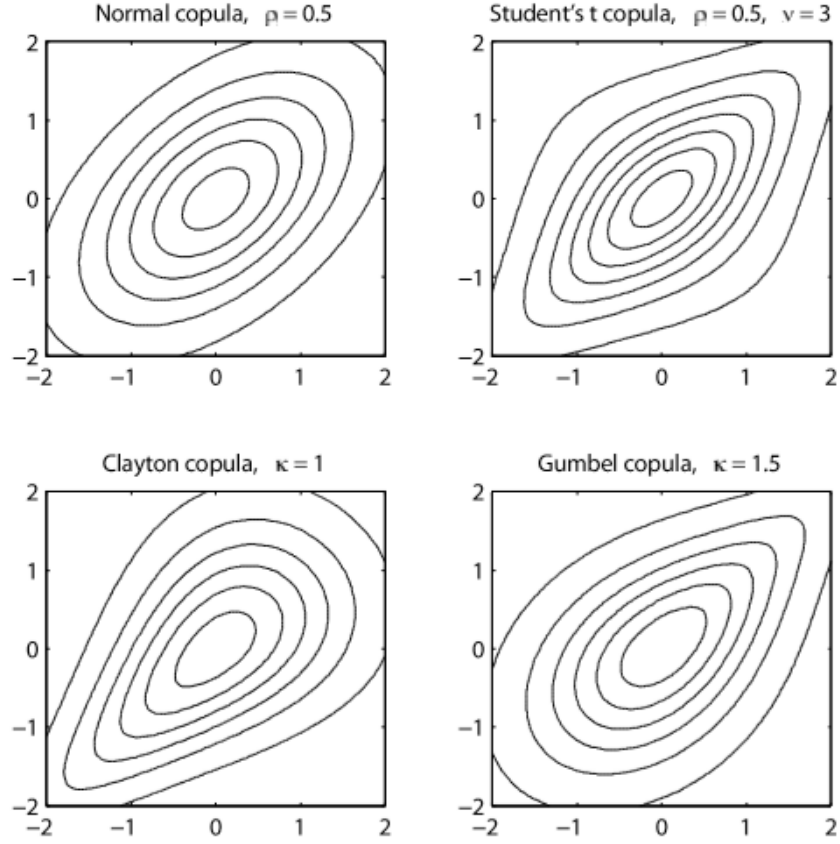
Figure 1: Different types of copulas

1. Let $V = \{Tr_1, Tr_2, ..., Tr_{d-1}\}$ be a set of $d-1$ trees,

2. Tr_1 has nodes $N_1 = \{1, 2, ..., d\}$ and edges E_1,

3. For $i \leq Tr_i$ has nodes $N_i = E_{i-1}$,

4. for $i = 2, 3, ..., d-1$ and $\{a, b\} \in E_i$ it must hold that $|a \cap b| = 1$.

$|a \cap b| = 1$ Is the proximity condition which states that an edge in Tr_i (where $i \geq 2$) connecting two nodes then those two nodes are edges in Tr_{i-1} and they share a common node. This tree structure can be translated into copulas and marginal distributions as follows:

1. (F, V, B) is a regular vine copula,

2. $F = (F_1, F_2, ..., F_d)$ (the marginal distributions),

3. V is a set of $d-1$ trees defined as above,

4. $B = \{B_e | i = 1, 2, ..., n-1; e \in E_i\}$ where B_e is a copula and E_i is the edge set of Tr_i of V.

With these definitions it is possible to create an regular vine copula; c.f. [14]. The vine structure of example (8) is shown in figure 2. This is one of the ways to order the x's there are six different ways

of structuring the density but only three are unique. The uniquely determined density of the regular vine is as follows:

$$c\big(F_1(x_1), ..., F_d(x_d)\big) = \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)|D(e)}\Big(F\big(x_{j(e)}|x_{D(e)}\big), F\big(x_{k(e)}|x_{D(e)}\big)\Big),$$

where $e = j(e), k(e)|D(e)$ are edges in E_i, $j(e), k(e)$ are conditional nodes, $D(e)$ is the conditioning set and $c_{j(e),k(e)|D(e)}$ is the copula density; c.f. [14]. The density of example (8) is as follows:

$$c(F_1(x_1), F_2(x_2), F_3(x_3)) = c_{1,2}(F_1(x_1), F_2(x_2))$$
$$\cdot c_{2,3}(F_2(x_2), F_3(x_3))$$
$$\cdot c_{1,3|2}\big(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\big).$$
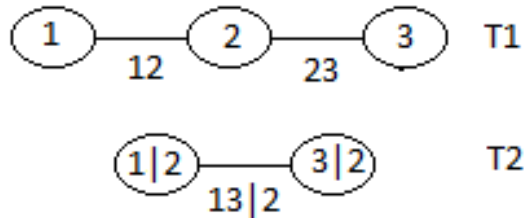
Figure 2: Example vine structure

4 Data analysis

The data used for the analysis is from one of the systems within Varo. This system stores the prices of the products and the forward curves. The data range used in this analysis is from 1-1-2013 until 31-12-2018. About 252 values per year which is nearly equal to the pricing days per year (No weekends and holidays). The daily prices for the products are in dollars per metric ton except for the dated Brent which is in dollars per barrel.

Forward contracts have a maturity date, this is when the contract ends and a new contract begins. Since this step from one contract to the other is not a continuous process the data needs to be transformed such that this step can be assumed to be continuous. By transforming the data it becomes possible to see all the contracts as one time series. The method used is the perpetual series method, this method uses a rollover period where each day has a different weight for the old and new contract. For example, a rollover period will look like this: For day one take 80% of the old contract and 20% of the new contract, next day take 60% of the old contract and 40% of the new contract etc. The advantage of this method over other methods is that it is better for statistical purposes; c.f. [22]. The rollover period used in the thesis is four days. As mentioned in section 3 the prices are transformed in to log returns. The margin is defined as follows:

$$Margin = Naphtha/8.9 * 0.13 + Mogas/8.33 * 0.21 + Jet/7.88 * 0.09 + LSGO/7.45 * 0.495$$
$$+ HSFO/6.35 * 0.075 - DatedBrent.$$

The product prices are first divided by their density such that the price is per barrel and then multiplied by the hedge ratio which is an average representation of the yield of a barrel of crude oil.

There exists some extreme behaviour in the time series of all the prices between 2016 and 2017. This extreme behaviour was caused by a hurricane in the USA which flooded some of the refinery's causing a shortage.

Table 1 show some general statistics of the log returns. One thing to notice is that the mean and the standard deviation of the log returns are all within the same range but when looking at the skewness and kurtosis the log returns differ a lot. A rather simple test to see whether the log returns follow the same skewness and kurtosis as the normal distribution is the Jarque-Bera test. Under the null hypothesis that the data is normally distributed the Jarque-Bera test follows χ^2; c.f. [1]. The Jarque-Bera test is calculated as follows:

$$JB = \frac{T}{6}\left(\widehat{Skew}^2 + \frac{(\widehat{Kurt} - 3)^2}{4}\right).$$

The test is significant for all the products and the margin indicating that r_t does not have normal skewness and kurtosis.

Product	Mean	Stdv	Min	Max	Skew	Kurt	JB	JB (p-value)
Mogas	-0.0005	0.0204	-0.1481	0.1406	-0.3456	6.9805	3138	($p < 0.01$)
HSFO	-0.0004	0.0243	-0.1190	0.1497	0.2671	4.9435	1578	($p < 0.01$)
Jet	-0.0004	0.0165	-0.0678	0.0956	0.2207	2.7950	512	($p < 0.01$)
Naphtha	-0.0005	0.0209	-0.1140	0.1245	0.0133	3.8638	953	($p < 0.01$)
LSGO	-0.0004	0.0173	-0.0687	0.1016	0.1901	2.7868	505	($p < 0.01$)
Dated_Brent	-0.0005	0.0203	-0.0827	0.1042	0.1734	3.0244	592	($p < 0.01$)
Margin	-0.0001	0.0354	-0.1960	0.2782	0.0803	4.7027	1413	($p < 0.01$)

Table 1: General statistics of daily log returns

The sections 4 and 5 show only the plots for the product Mogas the other product plots are in the according appendix.
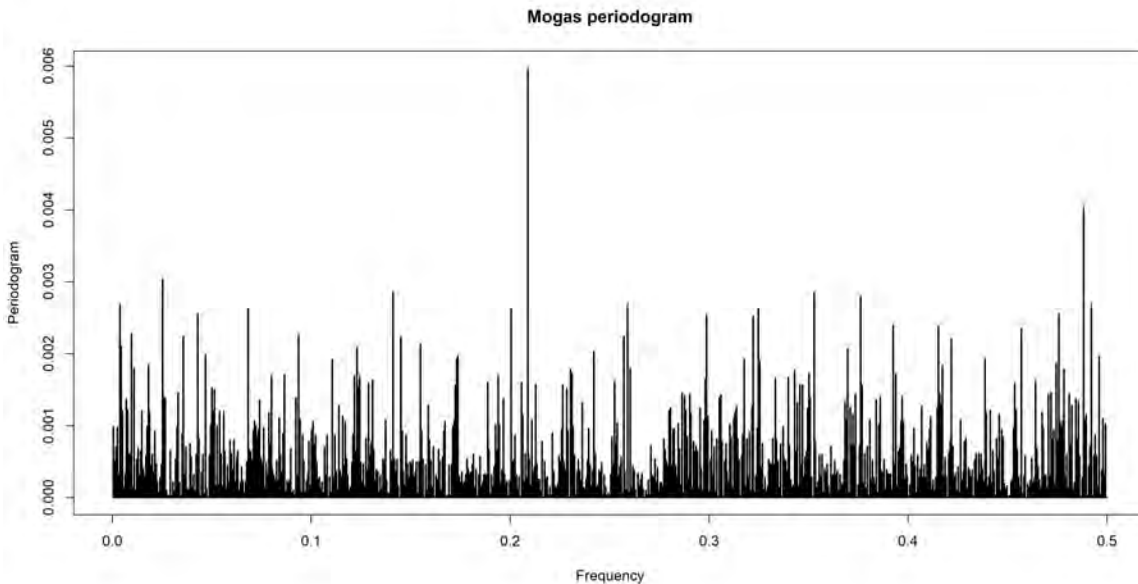
Figure 3: The Mogas periodogram

4.1 Seasonality

Seasonality in time series is a concept of patterns which repeat themselves after a certain period. For example each year there are lower prices in the summer than there are in the winter. Seasonality is not part of the models described in section 3 this needs to be modelled separately. So the seasonality has to be extracted from the time series. It is possible to check seasonality with a Fourier analysis. A Fourier analysis tries to decompose the time series in a sum of sines and cosines; c.f. [15]. A large spike in any of the periodogams in figure 3 would indicate a seasonal pattern with that specific frequency. There are no major spikes in the periodograms this indicates that there are most likely no seasonal patterns in the log returns. The other product and margin periodograms are in Appendix A.

4.2 Pearson correlation

An important feature for the determination of the coherence of the log returns is the correlation between the log returns. Table 2 shows the Pearson correlations between the log returns and figure 4 shows the log returns plotted against each other. The Pearson correlation is calculated as follows:

$$r_{X,Y} = \frac{\sum_{i=1}^{T}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{T}(x_i - \bar{x})}\sqrt{\sum_{i=1}^{T}(y_i - \bar{y})}}.$$

Table 2 shows high correlations between the log returns. Figure 4 shows that some log returns have a non linear correlation such as the Mogas and Naphtha. As discussed in paragraph 3.3 copulas are used to model these interactions such that they can be modelled properly.

4.3 Stylized features

Features such as volatility clustering, fat tails in the distribution and long memory are called stylized features of a time series; c.f. [27]. These features have an impact on the GARCH model choice. In the following paragraphs each feature is discussed. These analyses need to be done on a "clean" time series
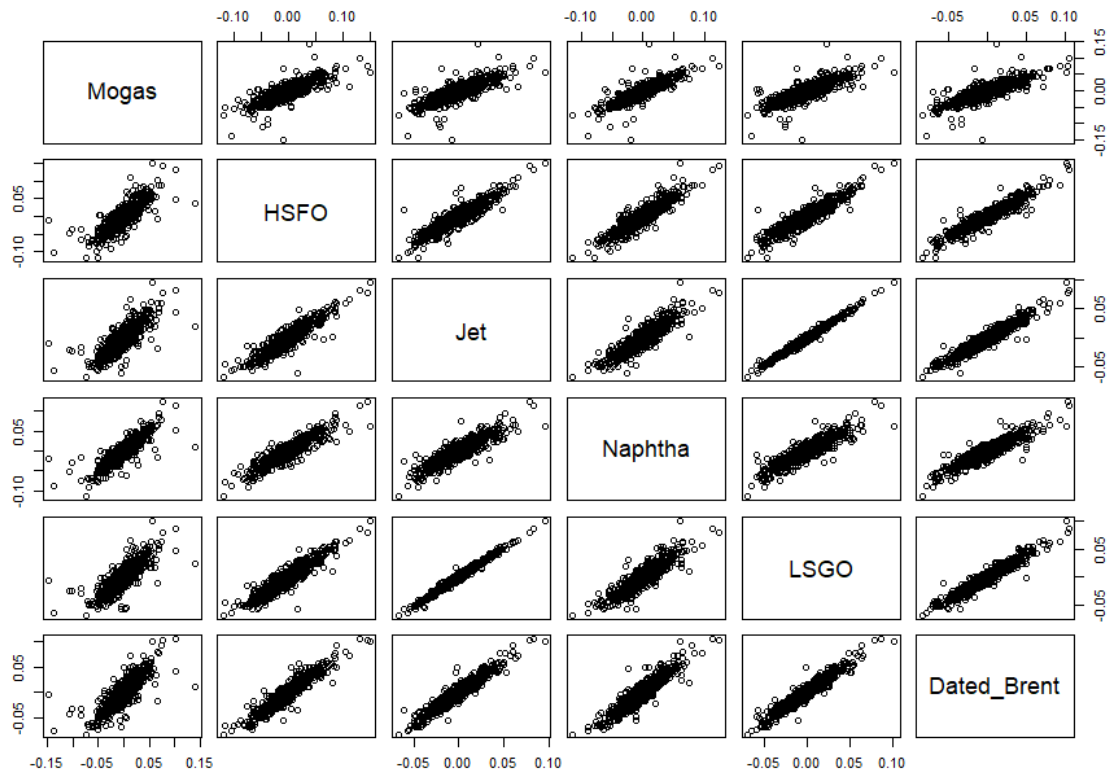
Figure 4: Log returns plotted against each other

	Mogas	HSFO	Jet	Naphtha	LSGO	Dated_Brent
Mogas	1.0000	0.8107	0.8053	0.8561	0.8142	0.8316
HSFO	0.8107	1.0000	0.9042	0.8995	0.9140	0.9369
Jet	0.8053	0.9042	1.0000	0.8911	0.9921	0.9517
Naphtha	0.8561	0.8995	0.8911	1.0000	0.8976	0.9218
LSGO	0.8142	0.9140	0.9921	0.8976	1.0000	0.9579
Dated_Brent	0.8316	0.9369	0.9517	0.9218	0.9579	1.0000

Table 2: Correlations between the log returns
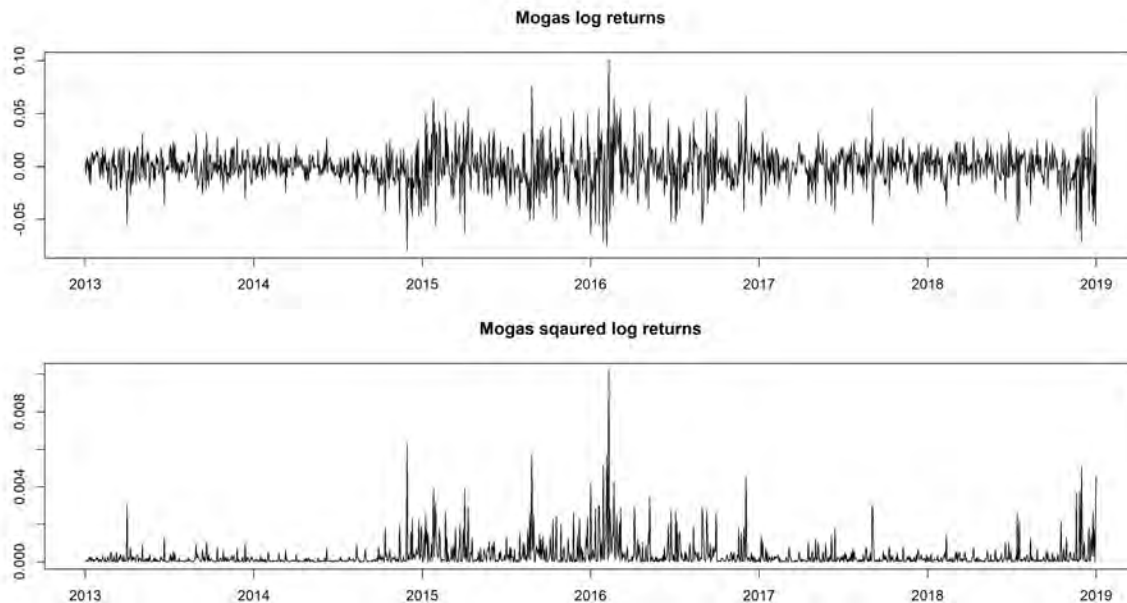
Figure 5: r_t and r_t^2 for Mogas

which means that there should be no trend or seasonal component present. Paragraph 4.1 indicates that there are no seasonal patterns in the product time series this means that the analysis of the stylized features can be directly applied on the log returns.

4.3.1 Volatility modeling

Before starting the modelling a few characteristics of the data need to be checked. The first two characteristics are the autocorrelation of r_t and r_t^2. Autocorrelation indicates the correlation between different lags, this means that if there is a correlation between lags that measures in time depend on each other. Figure 5 shows the plots for the log returns for the Mogas (see appendix B for the other log return series). The r_t^2 plot shows that periods of low volatility are followed by low volatility and high volatility is followed by high volatility. This indicates that there is a serial correlation between the r_t^2 lags. This is confirmed by figure 6 where the autocorrelation function (ACF) plot for Mogas is shown for the r_t and r_t^2 (See appendix C for the other log return series). The plots indicate a slow decay of the autocorrelations of r_t^2 which might also indicate long memory this characteristic is discussed in paragraph 4.3.2. The Margin ACF plots show that there might be an autoregressive term in the mean process. We will come back to this in section 5 where different lags are tested in ARIMA-GARCH models.

Table 3 shows the Ljung-Box statistics which is calculated according to (9). Where $MQ(p)$ follows
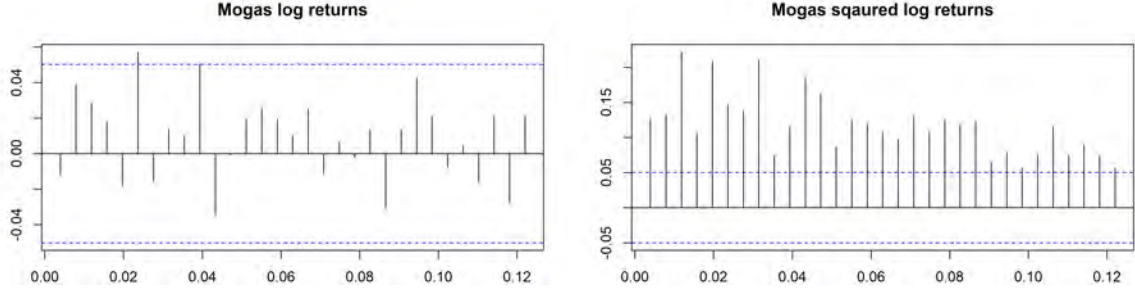
Figure 6: r_t and r_t^2 ACF plot for Mogas

a χ^2 if the r_t^2 are white noise (null hypothesis). A significant p-value indicates that there is time varying conditional volatility; c.f. [20]. The first two lags statistics are shown in table 3 these p-values indicate that there is time varying conditional volatility in the product time series. Further lags are also tested and showed the same results.

$$MQ(p) = T(T+2) \sum_{j=1}^{p} \frac{\hat{\rho}_j}{T-j}. \tag{9}$$

The log returns for all the products and the margin experience volatility clustering and have time-varying conditional volatility.

Product	Lag 1		Lag 2	
Mogas	24.9592	$(p < 0.01)$	51.6614	$(p < 0.01)$
HSFO	123.1432	$(p < 0.01)$	172.0811	$(p < 0.01)$
Jet	36.2859	$(p < 0.01)$	90.8007	$(p < 0.01)$
Naphtha	74.2597	$(p < 0.01)$	119.6392	$(p < 0.01)$
LSGO	41.1645	$(p < 0.01)$	90.6539	$(p < 0.01)$
Dated_Brent	88.9963	$(p < 0.01)$	141.6822	$(p < 0.01)$
Margin	59.1497	$(p < 0.01)$	97.4017	$(p < 0.01)$

Table 3: Ljung-Box results per product and lag 1, 2 (r_t^2)

4.3.2 Testing long memory

As mentioned in paragraph 4.3.1 the ACF plots in figure 6 indicate that the autocorrelations do not decay exponentially which should be the case in the standard GARCH model.

Testing for long memory can be done with the modified rescaled range (R/S) statistic. The R/S statistic is defined as follows:

$$Q_t = \frac{1}{s_T(q)} \Big(\max_{1 \le k \le T} \sum_{j=1}^{k} (x_j - \bar{x}) - \min_{1 \le k \le T} \sum_{j}^{k} (x_j - \bar{x}) \Big),$$

$$s_T(q) = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (x_t - \bar{x})^2 + \sum_{j=1}^{q} w_j(q) \hat{\rho}_j},$$

where $w_j = 1 - \frac{j}{q+1}$. If r_t is iid with finite variance then $\frac{1}{\sqrt{T}} Q_T$ weakly converges to V. V is the range of a Brownian bridge on the unit interval; c.f. [21]. The critical values for the Brownian brigde can be

found in table 2.a in [21]. Table 4 shows the results for the modified R/S statistic per product and the margin. The significance level for 1% is 2.098. All the products are significant which indicates that the products show long memory properties. The margin is not significant when looking at the r^2 this indicates that the volatility of the margin does not have long memory.

Product	r_t^2	$abs(r_t)$
Mogas	5.0392	5.8821
HSFO	5.9114	8.1702
Jet	5.4872	6.6109
Naphtha	5.1475	6.6366
LSGO	5.5593	6.6918
Dated_Brent	5.7470	7.0903
Margin	1.8322	2.4992

Table 4: Modified R/S statistic per product

4.3.3 Testing asymmetry

Asymmetric effects or leverage effects are effects in time series where the volatility has a bigger increase or decrease with a certain event than with the opposite of that event. For example, bad news tends to have a larger impact on volatility than good news; c.f. [1].

An indication of asymmetric effects can be found by a negative correlation between r_t^2 and r_{t-1}. Which would indicate that when the return of the last day is up then the volatility of the next day is down. This is exactly what the example of the news indicates. If there is an indication of asymmetric effects one can test this by testing if β_1 is significant in the following regression; c.f. [11].

$$\hat{\epsilon}_t^2 = \beta_0 + \beta_1 \hat{w}_{t-1} + \xi_t,$$

Where ϵ_t is the estimated residual from the conditional mean and $\hat{w}_t$ is a variable constructed from ϵ_{t-1}. By setting different values for ϵ_{t-1} the following test can be executed:

- Sign Bias test: $\epsilon_{t-1} = S_{t-1}^-$,

- Negative Sign Bias test: $\epsilon_{t-1} = S_{t-1}^- \hat{\epsilon}_{t-1}$,

- Positive Sign Bias test: $\epsilon_{t-1} = S_{t-1}^+ \hat{\epsilon}_{t-1}$,

where S_{t-1}^- is one if $\epsilon_{t-1} < 0$ and S_{t-1}^+ is one if $\epsilon_{t-1} > 0$. The test results are shown in table 5. The negative correlations are rather small which is also reflected in the Sign tests. Only the LSGO has a significant p-value so models with asymmetric effects would not be a good choice for these log returns.

Product	Corr(r_t^2,r_t)	Sign bias	Negative Sign Bias	Positive Sign Bias
Mogas	-0.0350	0.6364	0.4013	0.9229
HSFO	0.0044	0.4665	0.3054	0.5772
Jet	0.0198	0.1528	0.6325	0.5165
Naphtha	-0.0547	0.1490	0.4314	0.8921
LSGO	0.0256	0.0311*	0.4752	0.2421
Dated_Brent	0.0138	0.0521	0.6065	0.3928
Margin	-0.1098	0.9870	0.9747	0.3709

Table 5: Correlation and p-values for the Sign tests

4.3.4 Error distribution

The standard GARCH model depends on the assumption that the standardized residuals are normally distributed. The normal distribution might not be the best fit for these log returns. There is no direct test available to test for the appropriate error distribution, it is possible to fit the models with different distributions and select the best fit via a model selection criteria such as AIC or BIC. Another verification is done by making a quantile-quantile plot (QQ-plot) of the standardized residuals and the error distribution. If the data points in the QQ-plot follow the linear line the error distribution is a good fit.

The distributions fitted are distributions which have nice property's (in convergence) when estimating the parameters which are further discussed in paragraph 5.1. The distributions are as follows:

- Normal (norm),

- Generalized Error (ged),

- Student (std),

- Skew Student ,

- Normal inverse Gaussian distribution (nig).

AIC is a selection criterion for general parametric models. The model with the lowest value AIC value is selected to be the best fit for the time series; c.f. [19]. AIC makes a trade-off between goodness of fit and number of parameters of the model.

$$AIC = 2k - 2\log(\hat{L}(\hat{\theta})).$$

The residuals of the ARMA-GARCH models depend on the lags q, p, m, n. The lags need to be determined before the error distribution is fitted. Determining the lags is done in a similar fashion as determining the error distribution the lags $q, p, m, n \leq 2$ are all fitted and the model with the lowest AIC value is used. Based on the test done in the previous paragraphs the GARCH models are selected and shown in table 6. Table 7 shows the best lags based on the lowest AIC value per GARCH model.

Product	Volatility clustering	Long memory	Asymmetry	GARCH model
Mogas	Yes	Yes	No	S, IGARCH
HSFO	Yes	Yes	No	S, IGARCH
Jet	Yes	Yes	No	S, IGARCH
Naphtha	Yes	Yes	No	S, IGARCH
LSGO	Yes	Yes	Yes	S, I, EGARCH
Dated_Brent	Yes	Yes	Yes	S, I, EGARCH
Margin	Yes	No	No	SGARCH

Table 6: Summary of the stylized features per product and the proposed GARCH models

Based on the lags from table 7 the most appropriate error distribution is chosen. The AIC results are in table 8. In bold and underlined are the minimal AIC values per product which are the error distribution used in the models. The QQ-plots for the product Mogas and the error distributions are shown in figure 7. For all the products the best error distribution is the Normal Inverse Gaussian distribution. When looking at the QQ-plots of Mogas it is clear that the other distributions do not capture the tails of the distribution as good as the nig distribution. The other product QQ-plots are in appendix D.

Product	GARCH model	lags(q, p, m, n)
Mogas	SGARCH/IGARCH	(1, 1, 2, 2)/(1, 1, 0, 0)
HSFO	SGARCH/IGARCH	(1, 2, 2, 2)/(1, 1, 2, 1)
Jet	SGARCH/IGARCH	(2, 2, 2, 2)/(2, 1, 2, 2)
Naphtha	SGARCH/IGARCH	(2, 2, 2, 2)/(1, 1, 2, 2)
LSGO	SGARCH/IGARCH/EGARCH	(1, 2, 2, 2)/(1, 1, 1, 2)/(2, 1, 2, 2)
Dated_Brent	SGARCH/IGARCH/EGARCH	(1, 2, 2, 1)/(1, 1, 1, 0)/(2, 1, 1, 0)
Margin	SGARCH	(2, 1, 0, 2)

Table 7: GARCH models and the best lags per product

Product	norm	ged	std	sstd	nig
Mogas	-5.3985	-5.4325	-5.4376	-5.4404	**-5.4414**
HSFO	-5.2446	-5.2920	-5.2941	-5.2948	**-5.3044**
Jet	-5.7180	-5.7534	-5.7564	-5.7574	**-5.7618**
Naphtha	-5.3572	-5.3862	-5.3890	-5.3895	**-5.3912**
LSGO	-5.6397	-5.6753	-5.6805	-5.6807	**-5.6812**
Dated_Brent	-5.2519	-5.3042	-5.3068	-5.3095	**-5.3104**
Margin	-3.6486	-3.7308	**-3.7433**	-3.7433	-3.7399

Table 8: AIC values for the products and the margin per error distributions
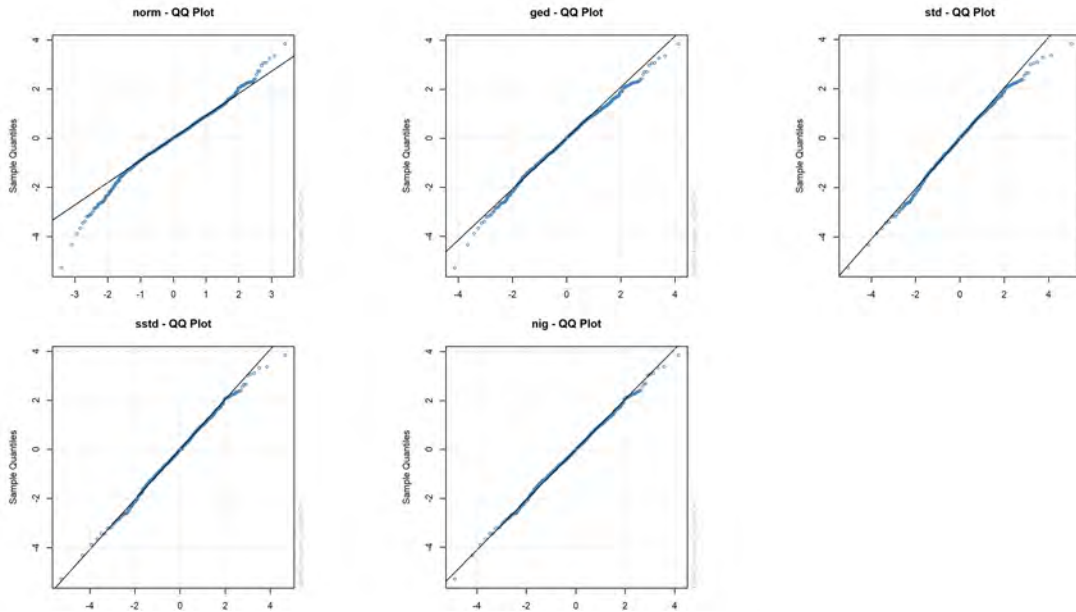
Figure 7: QQ-plot per error distribution for the Mogas GARCH(1,1) model

5 Estimation of the model parameters

This section discusses the estimation methods for the GARCH and continues models.

5.1 Parameter estimation GARCH models

Here we follow [12] to estimate the parameters. An advantage of the GARCH models is that estimating the parameters is rather simple. Estimation of the parameters can be done via least squares (LS) or maximum likelihood methods, where the maximum likelihood (ML) methods have a preference over the least squares since they estimate the parameters more accurately; c.f. [12]. ML estimation is the method of maximizing the likelihood function which is as follows:

$$\hat{L}(\theta) = \prod_{t=1}^{T} \frac{1}{\sigma_t} f\left(\frac{\epsilon_t}{\sigma_t}\right),$$

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \hat{L}(\theta), \tag{10}$$

where f is the conditional probability distribution function of ϵ_t on past ϵ's and σ's. Under certain conditions, the MLE is strongly consistent meaning:

$$\hat{\theta} \to \theta, \qquad \text{a.s. as } T \to \infty.$$

The estimators are consistent, asymptotically normally distributed and asymptotically efficient. Distributions which satisfy these conditions are standard Gaussian distributions and Student distributions; c.f. [12]. The next paragraph shows and eloborates on the conditions for the nig distribution. The nig density function is as follows:

$$f_{nig}(x, \upsilon, \xi, \varrho, \eta) = \upsilon \eta e^{\eta\gamma + \xi(x-\varrho)} \frac{K_1(\upsilon\sqrt{\eta^2 + (x-\varrho)^2})}{\pi\sqrt{\eta^2 + (x-\varrho)^2}},$$

where

$$K_1(w) = \frac{1}{2} \int_0^{\infty} e^{-\frac{1}{2}w(x+\frac{1}{x})} dx \text{ is the modified Bessel function of the third kind,}$$

$$\gamma = \sqrt{\upsilon^2 - \xi^2}.$$

The parameters need to satisfy $0 \le |\xi| < \upsilon$ and $\eta > 0$; c.f. [17]. For the products the following is assumed to be true $\epsilon \sim nig(\upsilon, \xi, \varrho, \eta)$

5.1.1 Parameter convergence conditions

Here we follow [6] which states that the following conditions need to be met to ensure that the parameters of the GARCH model converge.

C1 $\theta \in \Theta$ and Θ is compact.

C2 $\gamma(A_0) < 0$ and $\forall \theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$.

C3 The polynomials $\alpha_1 x + \alpha_2 x^2 + ... + \alpha_p x^p$ and. $1 - \beta_1 x - \beta_2 x^2 - ... - \beta_q x^q$ are coprimes in the set of polynomials with real coefficients.

C4 ϵ^2 is a nondegenerate random variable.

C5 $\lim_{t\to 0} t^{-\mu} P(\epsilon^2 \le t) = 0$ with some $\mu > 0$.

C6 $E|\epsilon^2|^\kappa < \infty$ with some $\kappa > 0$.

C7 $E|\log[h(\epsilon t)]| \le C_0(t^{\nu_0} + 1)$ for all $t > 0$, with some $0 \le \nu_0 \le 2\kappa$,

where $0 < C_0 < \infty$.

$$g(y,t) = \log\big(th(yt)\big), \qquad -\infty < y < \infty, t > 0,$$
$$g_1(y,t) = \frac{\partial}{\partial t}g(y,t), \qquad -\infty < y < \infty, t > 0.$$

The following two conditions assume that there is a function $C_1(y)$ such that:

C8 $|g_1(y,t)| \le C_1(y)(t^{\nu_1} + 1)/t$, for all $0 < t < \infty$ and $y \in \mathbb{R}$, with some $0 \le \nu_1 \le 2\kappa$.

C9 $E(C_1(\epsilon)) < \infty$.

C10 $E\big(g(\epsilon,t)\big) < E\big(g(\epsilon,1)\big)$ for all $0 < t < \infty$, $t \ne 1$.

5.1.2 Parameter convergence for the Normal Inverse Gaussian distribution

In this paragraph, we analyze the conditions for the Normal Inverse Gaussian distribution from paragraph 5.1.1.

C1 is a condition which has to be assumed since the true parameters are unknown and so is the parameter space Θ. In C2 $\gamma(A_0)$ is the Lyapunov exponent of the series A_0. Where A_0 is as follows:

$$A_{0t} = \begin{pmatrix} \tau_t^{'} & \beta_{0p} & \alpha_{02:q-1}^{'} & \alpha_{0q} \\ \mathbb{I}_{p-1} & 0 & 0 & 0 \\ \xi_t^{'} & 0 & 0 & 0 \\ 0 & 0 & \mathbb{I}_{q-2} & 0 \end{pmatrix},$$

with

$$\tau_t = (\beta_{01} + \alpha_{01}\epsilon_t^2, \beta_{02}, ..., \beta_{0p-1})^{'} \in \mathbb{R}^{p-1},$$
$$\xi_t = (\epsilon_t^2, 0, ..., 0)^{'} \in \mathbb{R}^{p-1},$$
$$\alpha_{02:q-1} = (\alpha_{02}, ..., \alpha_{0q-1})^{'} \in \mathbb{R}^{q-2},$$

and

$$\gamma(A_0) = \lim_{t \to \infty} \frac{1}{t} \log ||A_{0t}A_{0t-1}...A_{01}||.$$

In [7] it is established that the following statements are equivalent in case of a GARCH model:

1. $\gamma(A_0) < 0$.

2. The model used is irreducible.

3. There is an unique stationary solution.

In paragraph 3.1 it is assumed that the process $(\epsilon_t, t \in \mathbb{Z})$ is strictly stationary, ergodic and nonanticipative solution of the SGARCH model. Just like in paragraph 4.3.1 it is possible to test if ϵ_t is strictly stationary, this is done in paragraph 5.1.3.

Condition C3 can be checked if the parameters α and β are estimated, estimating these parameters and showing if this condition is satisfied is shown in paragraph 5.1.3.

Condition C4 means that the probability distribution of ϵ^2 is not concentrated at one point; c.f. [3]. This is proved by checking if $Var(\epsilon^2) = 0$ if this is false condition C4 is satisfied. The central moments of the nig distribution are given by [17]:

18

$$E(\epsilon) = \varrho + \eta\frac{\xi}{\gamma},$$

$$V(\epsilon) = \eta\frac{\upsilon^2}{\gamma^3},$$

$$S(\epsilon) = 3\frac{\xi}{\upsilon\sqrt{\gamma\eta}},$$

$$K(\epsilon) = 3 + 3\big(1 + 4(\frac{\xi}{\upsilon})^2\big)\frac{1}{\gamma\eta}.$$

From the online appendix M page 5 from [18] expressions for $E(x^4)$ and $E(x^3)$ are given.

$$E(\epsilon^3) = S(\epsilon) + 3V(\epsilon)E(\epsilon) + E(\epsilon)^3,$$
$$E(\epsilon^4) = K(\epsilon) + 4S(\epsilon)E(\epsilon) + 6V(\epsilon)E(\epsilon)^2 + E(\epsilon)^4.$$

$$
\begin{aligned}
V(\epsilon) &= E(\epsilon^2) - E(\epsilon)^2, \\
E(\epsilon^2) &= Var(\epsilon) - E(\epsilon)^2, \\
&= \eta\frac{\upsilon^2}{\gamma^3} - (\varrho + \eta\frac{\xi}{\gamma})^2, \\
&= \eta\frac{\upsilon^2}{\gamma^3} - \eta^2\frac{\xi^2}{\gamma^2} - 2\varrho\eta\frac{\xi}{\gamma} - \varrho^2.
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
V(\epsilon^2) &= E(\epsilon^4) - E\big((\epsilon^2)\big)^2, \\
&= K(\epsilon) + 4S(\epsilon)E(\epsilon) + 6V(\epsilon)E(\epsilon)^2 + E(\epsilon)^4 - \big(E(\epsilon^2)\big)^2, \\
&= 3 + 3\big(1 + 4(\frac{\xi}{\upsilon})^2\big)\frac{1}{\gamma\eta} + 12\frac{\xi}{\upsilon\sqrt{\gamma\eta}} + 6V(\epsilon)E(\epsilon)^2 + E(\epsilon)^4 - \big(E(\epsilon^2)\big)^2.
\end{aligned}
\tag{12}
$$

Equation (12) is a complicated equation so it is hard to prove analytically that $V(\epsilon^2) \neq 0$. To bypass this problem condition C5 is checked with estimated parameters, this is done in paragraph 5.1.3.

Based on the first two moments of the nig distributions $E|\epsilon^2|$ is worked out. Condition C6 is satisfied if γ in (11) does not go to 0. This is satisfied since $0 \leq |\xi| < \upsilon$ is one of the conditions of the nig distribution. This requirement makes sure that $\gamma = \sqrt{\upsilon^2 - \xi^2} > 0$. To be sure this condition is checked with the estimated parameters.

For conditions C7, C8, C9 and C10 the function $h(x) = f_{nig}(x)$.

$$
\begin{aligned}
\log(h(x)) &= \log(\upsilon\eta) + \eta\gamma + \xi(x - \varrho) + \log(K_1(\upsilon\sqrt{\eta^2 + (x - \varrho)^2})) - \log(\pi\sqrt{\eta^2 + (x - \varrho)^2}), \\
g(y, t) &= \log(t\upsilon\eta) + \eta\gamma + \xi(yt - \varrho) + \log(K_1(\upsilon\sqrt{\eta^2 + (yt - \varrho)^2})) - \log(\pi\sqrt{\eta^2 + (yt - \varrho)^2}), \\
g_1(y, t) &= \frac{1}{t} + \xi y + K_1(\upsilon\sqrt{\eta^2 + (yt - \varrho)^2})\frac{2\upsilon y(yt - \varrho)}{\sqrt{\eta^2 + (yt - \varrho)^2}} - \frac{y(yt - \varrho)}{\eta^2 + (yt - \varrho)^2}.
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
g(\epsilon, t) &= \log(t\upsilon\eta) + \eta\gamma + \xi(\epsilon t - \varrho) + \log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon t - \varrho)^2})) - \log(\pi\sqrt{\eta^2 + (\epsilon t - \varrho)^2}), \\
g(\epsilon, 1) &= \log(\upsilon\eta) + \eta\gamma + \xi(\epsilon - \varrho) + \log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon - \varrho)^2})) - \log(\pi\sqrt{\eta^2 + (\epsilon - \varrho)^2}).
\end{aligned}
$$

Conditions C7 until C10 are to complicated to prove analytically with the parameters of the nig distribution unknown. So conditions C7 until C10 are checked by estimating the parameters and checking whether the conditions hold for the estimated model.

5.1.3 Parameter convergence conditions with estimated parameters

We were unable to prove conditions C2 until C10 analytically due to the complicated distribution. Instead, we analyze the conditions with estimated parameters to show that the conditions are not unreasonable for the values of the parameters. In this paragraph, one example is discussed (Mogas IGARCH model with lags $q = 1, p = 1$) the other models can be done in a similar fashion. The estimated parameters are in table 9.

	υ	ξ	ϱ	η
Mogas	49.4065	-2.4304	0.0007	0.01614

Table 9: Nig distribution parameters for the Mogas IGARCH model with $p, q = (1, 1)$

Now filling in these values for the conditions we get the following results:

C2 The p-values for the Ljung-box test at lag 1 and 2 are 0.6406 and 0.2738 further lags were also tested and showed have a $p > 0.05$. So the null-hypothesis can not be rejected (null-hypothesis is that the data behaves like white noise). Which means that C2 is satisfied.

C3 Two functions are coprime if there exist $u(x)d(x) + v(x)e(x) = 1$, see [29], with $d(x) = \alpha_1 x + \alpha_2 x^2 + ... + \alpha_p x^p$ and $e(x) = 1 - \beta_1 x - \beta_2 x^2 - ... - \beta_q x^q$. The model parameters are: $\alpha = 0.05$ $\beta_1 = 0.95$. So set functions $u(x) = 19$ and $v(x) = 1$ and the desired results is achieved. So the functions are coprime and C3 is satisfied.

C4 $Var(\epsilon^2) = 6.8025$ which means condition C4 is satisfied.

C5 After $t = 0.9000 * 10^{-11}$ the term $P(\epsilon^2 \leq t)$ becomes zero which results in the limit being zero so the condition is satisfied.

C6 $E|\epsilon^2|^\kappa = 0.0003^\kappa$ so condition C6 is satisfied.

C7: The expectation of constants are constant so these are all added in to c_1 and c_2 and what remains is the following:

$$E|\log[h(\epsilon t)| = -c_1(c_2 + t) + E\big(\log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon - \varrho)^2}))\big) - E\big(\log(\pi\sqrt{\eta^2 + (\epsilon - \varrho)^2})\big).$$

Jensen's inequality states that if f is a convex function and X a random variable then the following holds:

$$E\big(f(X)\big) \leq f\big(E(X)\big).$$

This can also be applied to C7. The last term contains a log which is a convex function so Jensen's inequality can be applied which then also gives the required inequality needed for the condition.

$$E|\log[h(\epsilon t)| = -c_1(c_2 + t) + E\big(\log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon - \varrho)^2}))\big) - E\big(\log(\pi\sqrt{\eta^2 + (\epsilon - \varrho)^2})\big),$$
$$\leq -c_1(c_2 + t) + E\big(\log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon - \varrho)^2}))\big) - \log\big((\pi E\sqrt{\eta^2 + (\epsilon - \varrho)^2})\big),$$

where constants $c_2 = \frac{\log(\upsilon\eta) + \eta\gamma - \xi\varrho}{c_1}$ and $-c_1 = \xi E(\epsilon)$.

C8: If the last term is always negative in equation (13) then the following result will hold:

$$g_1(y, t) = \frac{1}{t} + \xi y + K_1(\upsilon\sqrt{\eta^2 + (yt - \varrho)^2})\frac{2\upsilon y(yt - \varrho)}{\sqrt{\eta^2 + (yt - \varrho)^2}} - \frac{y(yt - \varrho)}{\eta^2 + (yt - \varrho)^2},$$
$$\leq \frac{1}{t} + \xi y + K_1(\upsilon\sqrt{\eta^2 + (yt - \varrho)^2})\frac{2\upsilon y(yt - \varrho)}{\sqrt{\eta^2 + (yt - \varrho)^2}}.$$

(14)

The last term can, unfortunately, be positive when for example $yt - \varrho < 0$ and $y > 0$. Condition C8 is not satisfied for a certain range of y and t.

C9: In principle one needs to check if $E(C_1(\epsilon)) < \infty$. Since it is not possible to construct a function C_1 from condition C8 is not possible to verify condition C9.

C10: In principle one would check $E\big(g(\epsilon,t)\big) < E\big(g(\epsilon,1)\big)$ for all $0 < t < \infty$, $t \neq 1$. Both expectations are as follows:

$$E\big(g(\epsilon,t)\big) = \log(t) - c_1(c_2 + t) + \log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon t - \varrho)^2})) - \log(\pi\sqrt{\eta^2 + (\epsilon t - \varrho)^2}),$$
$$E\big(g(\epsilon,1)\big) = -c_1(c_2 + 1) + \log(K_1(\upsilon\sqrt{\eta^2 + (\epsilon - \varrho)^2})) - \log(\pi\sqrt{\eta^2 + (\epsilon - \varrho)^2}).$$

This condition is still too complicated to verify even for the estimated parameters.

Conditions C1 until C7 are satisfied for the Mogas model but the conditions C8 until C10 can either not be proved or are not satisfied for a certain range. In the following sections conditions C8, C9 and C10 are assumed to be satisfied such that the properties of convergences apply to the models used. The other products and the margin can be proved in a similar fashion but this is not done since this is not the goal of the thesis. It is assumed that the same conclusions for Mogas also hold for the other models which means that we assume that the parameters converge to their true values.

5.2 Parameter estimation continuous models

Parameter estimation for the continuous models is done via the same way as the GARCH models, i.e. ML. The likelihood function of the continuous models is based on the transition density $p(X_i|X_{i-1},\theta)$. The likelihood function is as follows:

$$\hat{L}(\theta) = \prod_{i=1}^{T} p(X_i|X_{i-1},\theta).$$

By rewriting the Heston model from paragraph 3.2 in the form of (15) it is possible to estimate the parameter via some transition density; c.f. [28].

$$d\sigma_t = \kappa(\alpha - \sigma_t)dt + \upsilon\sqrt{\sigma_t}dW_{2_t},$$
$$dr_t = \big(\mu - \frac{1}{2}\sigma_t\big)dt + \sqrt{(1-\rho^2)\sigma_t}dW_{1_t} + \rho\sqrt{\sigma_t}dW_{2_t}. \tag{15}$$

The transition density of the volatility of the Heston model described in (15) is given by [28]:

$$\sqrt{\sigma_t}|\sqrt{\sigma_{t-1}} \sim N\Big(m_{t-1}, \frac{\upsilon^2}{4}\delta\Big),$$

where $m_{t-1} = \sqrt{\sigma_{t-1}} + \frac{1}{2\sqrt{\sigma_{t-1}}}\big(\kappa\alpha - \kappa\sigma_{t-1} - \frac{1}{4}\upsilon^2\big)\delta$. Filling the transition density into the likelihood function results in the following equation:

$$\hat{L}(\kappa,\alpha,\sigma) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi}\sqrt{\frac{\upsilon^2}{4}}} exp\Big(-\frac{(\sqrt{\sigma_t} - m_{t-1})^2}{\frac{\upsilon^2}{2}\delta}\Big).$$

The resulting MLE estimators of the parameters κ, α, σ are given and proven by [28].

$$\hat{\kappa} = \frac{2}{\delta}\left(1 + \frac{\hat{P}\delta}{2}\frac{1}{T}\sum_{k=1}^{T}\frac{1}{\sigma_{k-1}} - \frac{1}{T}\sum_{k=1}^{T}\sqrt{\frac{\sigma_k}{\sigma_{k-1}}}\right),$$

$$\hat{v} = \sqrt{\frac{4}{\delta}\frac{1}{T}\sum_{k=1}^{T}\left(\sqrt{\sigma_k} - \sqrt{\sigma_{k-1}} - \frac{\delta}{2\sqrt{\sigma_{k-1}}}(\hat{P} - \hat{\kappa}\sigma_{k-1})\right)^2},$$

$$\hat{\alpha} = \frac{\hat{P} + \frac{1}{4}\hat{v}^2}{\hat{\kappa}},$$

$$\hat{\rho} = \frac{1}{T\delta}\sum_{k=1}^{T}\Delta W_{1_t}\Delta W_{2_t},$$

$$\hat{\mu} = \frac{1}{T}\sum_{k=1}^{T}r_k,$$

where

$$\hat{P} = \frac{\frac{1}{T}\sum_{k=1}^{T}\sqrt{\sigma_{k-1}\sigma_k} - \frac{1}{T^2}\sum_{k=1}^{T}\sqrt{\frac{\sigma_k}{\sigma_{k-1}}}\sum_{k=1}^{T}\sigma_{k-1}}{\frac{\delta}{2} - \frac{\delta}{2}\frac{1}{T^2}\sum_{k=1}^{T}\frac{1}{\sigma_{k-1}}\sum_{k=1}^{T}\sigma_{k-1}},$$

$$\Delta W_{1_t} = \frac{\log X_t - \log X_{t-1} - (\mu - \frac{1}{2}\sigma_{t-1})\delta}{\sqrt{\sigma_{t-1}}},$$

$$\Delta W_{2_t} = \frac{\sigma_t - \sigma_{t-1} - \kappa(\alpha - \sigma_{t-1})\delta}{\sigma\sqrt{\sigma_{t-1}}},$$

$$\delta = dt = 1.$$

5.3 Vine copula selection

As mentioned in paragraph 3.3.1 there are 23,040 different possible regular vine combinations which can be considered. It takes to much time to estimate every regular vine and conclude which has the best fit. In [14] three different methods are reviewed, the stepwise frequentist selection, stepwise Bayesian selection and the fully Bayesian selection. The latter two have intensive computation times ranging from two to ten hours but perform better when comparing the likelihood. These computation times are too high considering that the models need to be recalculated many times. So the frequentist selection is used. The frequentist selection algorithm is described according to [14].

The first tree is determined via a maximum spanning tree algorithm (a maximum spanning tree is a tree where all nodes can reach all other nodes but there exists no cycle) on a complete graph (a complete graph is a graph where all nodes are connected to each other via an edge). The weight on an edge is a measure of dependence between the two products connected via that edge. The measure used is the Kendall's τ coefficient:

$$\tau = P\big((X - X')(Y - Y') > 0\big) - P\big((X - X')(Y - Y') < 0\big),$$

where $(X, Y) \sim F$ and $(X', Y') \sim F$ are independent random variables; c.f. [14]. When the structure of the tree is determined the tree's that follow can be determined from the first one. The next step is to fit copulas between the products connected via the edges in tree one. This is done by fitting different copulas and selecting the model with the lowest AIC value. Iterate this procedure overall tree's. When the copulas are selected the parameters of the copulas have to be estimated. This is also done with a maximum likelihood method:

$$\hat{L}(B_e, \theta_e) = \prod_{t=1}^{T} c_{B_e}\big(F(x_{j(e)}|x_{D(e)}), F(x_{k(e)}|x_{D(e)}); \theta_e\big),$$

where θ_e is the parameter vector of the according copula.

The complete procedure is as follows (for an example of trees see figure 2):

1. Initiate $G(N, E) =$ complete graph with $N =$ products and $e_i \in E =$ Kendall's τ coefficient between the two products connected via e_i with $i = 1, ..., \frac{d(d-1)}{2}$.

2. Run a maximum spanning tree algorithm on $G(N, E)$ to get $Tr_1(N, E)$ which is a spanning tree. This means that all nodes are connected and there exists no cycle.

3. Construct Tr_d for $d > 1$ from Tr_1 where $d = 1, 2, 3, 4, 5$.

4. Estimate and fit copulas to the edges in Tr_d and select the copula with the lowest AIC.

6 Model Selection and Evaluation

The model selection is based on the evaluation metric: Mean Squared Error method (MSE). The MSE is a widely used metric which looks as follows:

$$MSE = \frac{1}{T-N} \sum_{i=N+1}^{T} (h_i - \sigma_i)^2,$$

where σ_i is an proxy for the observed volatility. Two proxy's are evaluated r^2 and $|r|$. The models are evaluated using a rolling window approach. The first model is estimated on a sample of the data from $t = 1, ..., N$ then an g-ahead forecast is calculated from that point on. The model is refitted every day so the next data set is as follows $t = 2, ..., N + 1$. The fitting period is three years.

6.1 Fitting GARCH models

The results from section 4 are shown in table 10. For each product different types of models are given each of these models is appropriate for the features of the time series. The best lags based on the AIC value are shown in table 7. The GARCH models in table 10 have the following forms:

Product	Volatility clustering	Long memory	Asymmetry	Error distribution	GARCH model
Mogas	Yes	Yes	No	nig	S, IGARCH
HSFO	Yes	Yes	No	nig	S, IGARCH
Jet	Yes	Yes	No	nig	S, IGARCH
Naphtha	Yes	Yes	No	nig	S, IGARCH
LSGO	Yes	Yes	Yes	nig	S, I, EGARCH
Dated_Brent	Yes	Yes	Yes	nig	S, I, EGARCH
Margin	Yes	No	No	std	SGARCH

Table 10: Summary of the stylized features per product and the proposed GARCH models

$$SGARCH = h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i},$$

$$IGARCH = h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \text{ where } 1 = \sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i,$$

$$EGARCH = h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \frac{|\epsilon_{t-j}| + \gamma_j \epsilon_{t-j}}{\sigma_{t-j}} + \sum_{i=1}^{p} \beta_i h_{t-i},$$

$$ARMA = y_t = \mu + \sum_{i=1}^{m} a_i r_{t-i} + \sum_{j=1}^{n} b_j \epsilon_{t-j},$$

$$Return = r_t = y_t + \epsilon_t,$$

$$Error = \epsilon_t = z_t \sqrt{h_t}.$$

(16)

Forecasting h_{N+1} with a GARCH model is done with with the known ϵ_N^2 but forecasting h_{N+g} $(g > 1)$ is different since ϵ_{N+g-1} are unknown. For $g > 1$ the following is used $E(\epsilon_{N+g-1}^2) = E(z_{N+g-1}^2 h_{N+g-1}^2) = E(h_{N+g-1}^2)$.

Since volatility cannot be measured a proxy is needed which is assumed to be the volatility. Two proxies are used r^2 and $|r|$. The MSE results for the return r, volatility proxy r^2 and volatility proxy

$|r|$ are shown in table 11. The best model fit per model is determined by the volatility proxy MSE. This is because the volatility is an important result from these models. In four out of seven cases the best MSE score for the volatility goes together with the best MSE score for the mean process. In the other cases, the mean process does not perform much worse than the best MSE score for the mean process.

Figure 8 shows the forecasted returns versus the actual returns. The blue line represents the forecasted mean and the red lines are the 95% confidence interval for the forecasted returns. Figure 9 and 10 show the forecasted volatility versus the proxy volatility r^2, $|r|$ and the confidence interval of the volatility. Since the distribution of ϵ^2 is unknown the confidence interval is obtained by running $10,000$ simulations and taking the 0.05 and 0.95 quantiles of the simulation results. Appendix E contains the other products and margin plots for the models with the best fit (The best model scores are underlined and bold in table 11).
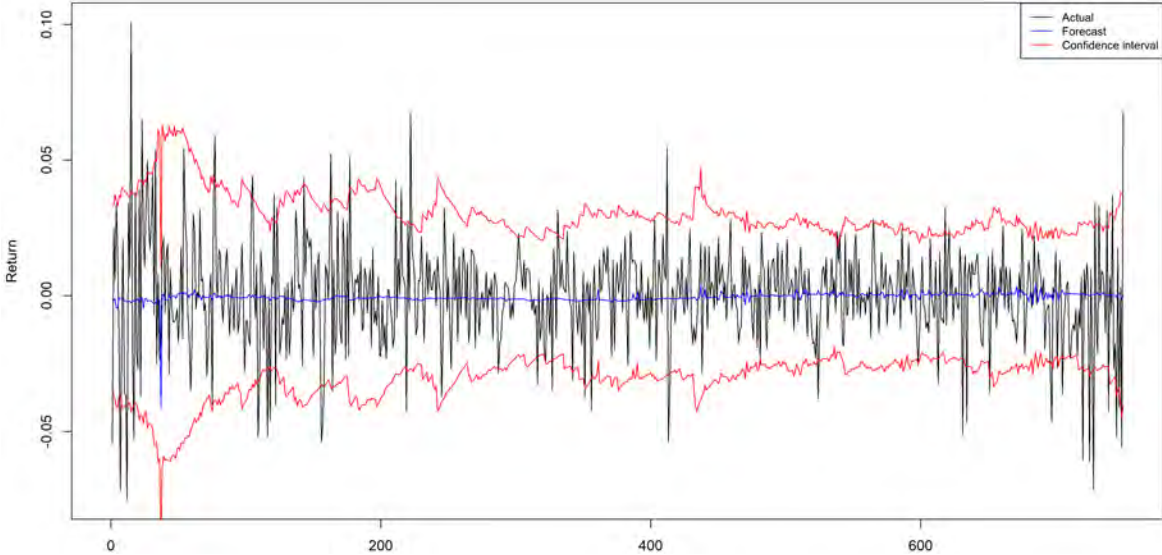
Figure 8: Return forecast (r) Mogas (1,1,2,2) SGARCH model with confidence interval
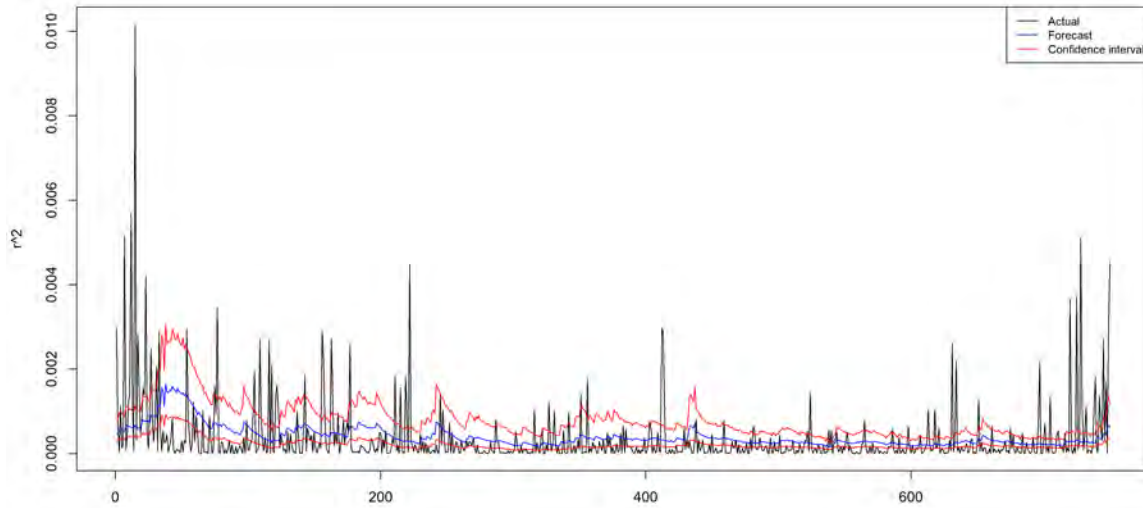
Figure 9: Return Volatility forecast (r^2) Mogas (1,1,2,2) SGARCH model with confidence interval
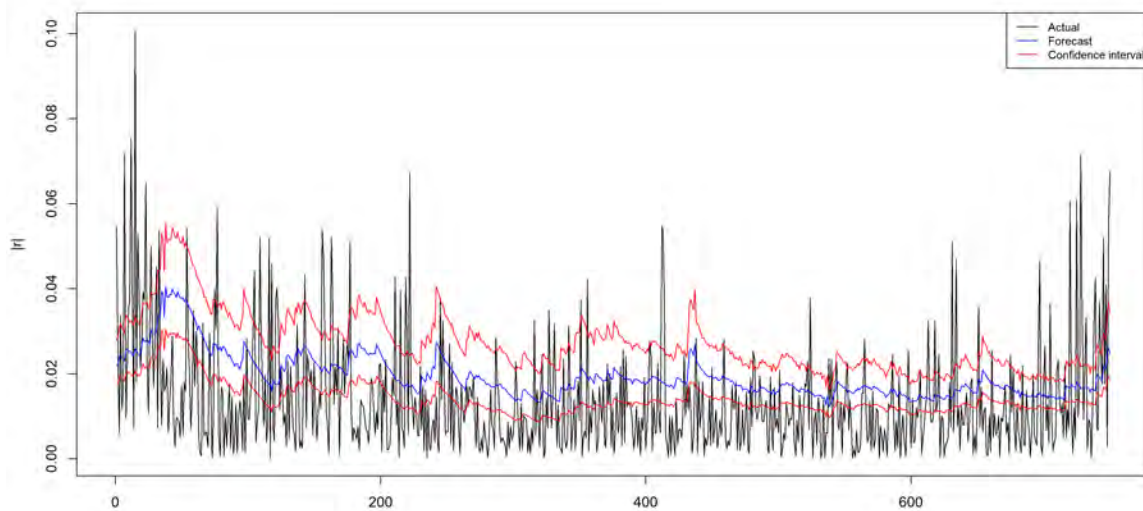
Figure 10: Return Volatility forecast ($|r|$) Mogas (1,1,2,2) SGARCH model with confidence interval

6.2 Fitting continues models

One continues model is fitted to the data. The model is the Heston model with time varying volatility as described in (6). This model uses an Orstein-Uhlenbeck process for volatility this is due to the mean reverting nature of the volatility. From figure 5 it is clear the volatility returns to a standard level even after periods of high volatility. The ML estimators for the parameters are given in 5.2.

These parameters are used to do the g-ahead forecast. The return and the volatility are forecasted via equation 17. Figure 11 shows the forecasted returns and the confidence interval. Figure 12 and 13 show the volatility proxy plots.

$$
\begin{aligned}
h_t &= h_{t-1} + \hat{\kappa}(\hat{\alpha} - h_{t-1}), \\
r_t &= r_{t-1} + \hat{\mu} - \frac{1}{2}h_t.
\end{aligned}
\tag{17}
$$

Just like with the discrete case an proxy for the volatility is needed since this can not be measured so r^2 and $|r|$ are used as an proxy. The evaluation metric results are in table 11. Appendix E contains the other products and margin plots for the models with the best fit (The best model scores are underlined and bold in table 11).
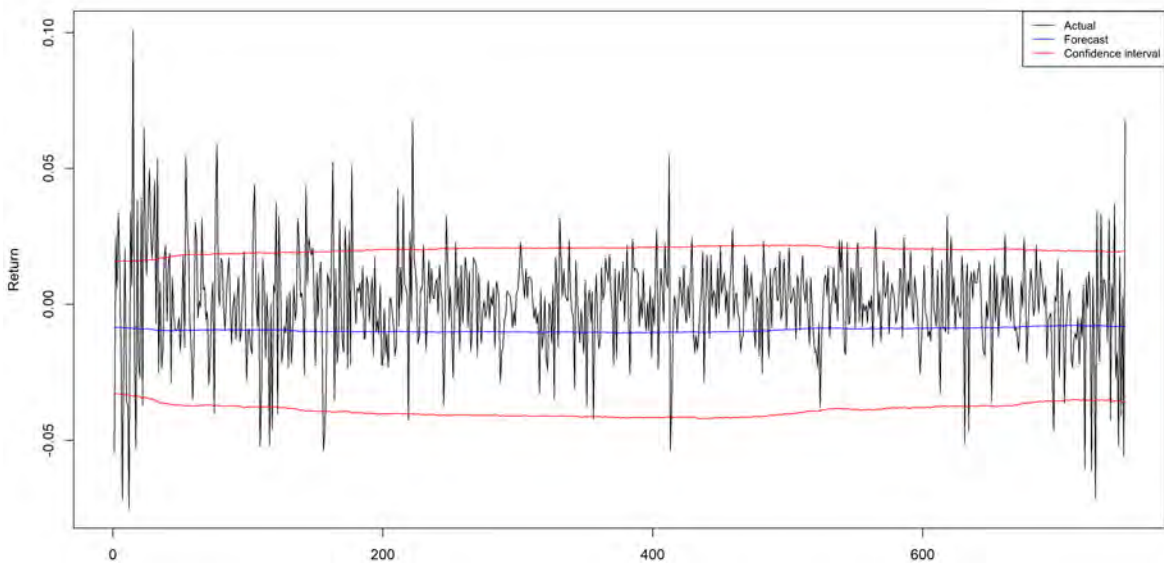
Figure 11: Return forecast (r^2) continues Mogas model with confidence interval
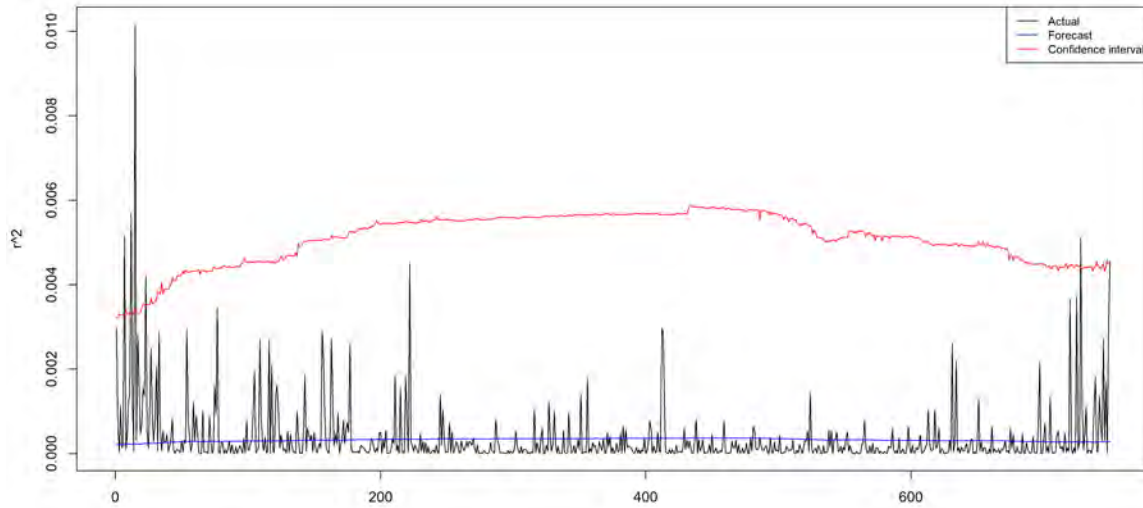
Figure 12: Return Volatility forecast continues Mogas model with confidence interval
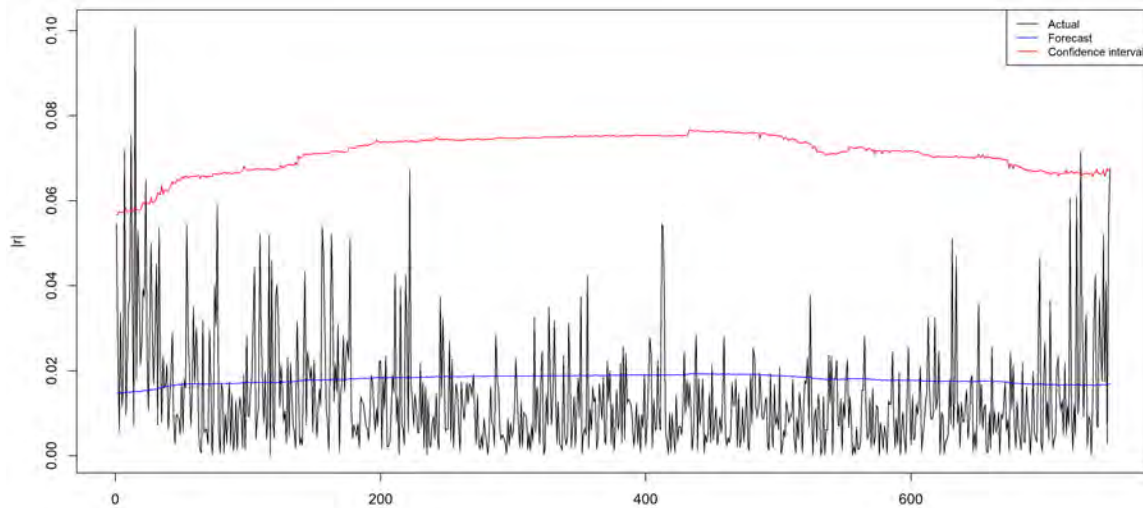
Figure 13: Return Volatility forecast ($|r|$) continues Mogas model with confidence interval

6.3 Back testing

In table 11 the best performing models are underlined and bold. To check whether these models perform well with regard to coverage and number of exceedances one can do the conditional coverage test introduced by [9]. Exceedances are returns which go outside the volatility forecast boundary and

Product		SGARCH	IGARCH	EGARCH	Continues model
Mogas	(r)	**3.3052**	3.5134	-	10.6982
	(r^2)	**0.0056**	0.0060	-	0.0063
	$(\lvert r \rvert)$	**1.9118**	2.0049	-	1.9843
HSFO		**5.5403**	5.5281	-	18.7382
		0.0170	0.0172	-	0.0181
		3.0496	3.1291	-	3.8216
Jet		4.90439	**2.7929**	-	9.6397
		0.0600	**0.0032**	-	0.0033
		4.3009	**1.5389**	-	1.5222
Naphtha		**4.2050**	4.2357	-	12.7702
		0.0108	0.0109	-	0.0109
		2.4394	2.5464	-	2.5203
LSGO		**3.0766**	3.0511	3.3616	10.4573
		0.0037	0.0038	0.0037	0.0040
		1.6671	1.6933	1.4708	1.7008
Dated_Brent		4.3078	4.2921	**4.2955**	14.2166
		.0085	0.0085	**0.0074**	0.0087
		2.4654	2.4851	**1.9207**	2.4990
Margin		**15.3489**	-	-	36.2044
		0.1731	-	-	0.1754
		10.0917	-	-	8.6651

Table 11: MSE per model (values are of the order 10^{-4})

are defined as follows:

$$
I_t = \begin{cases} 1, & \text{if } r_t \in [L_{t|t-1}, U_{t|t-1}] \\ 0, & \text{otherwise.} \end{cases} \sim Bern(p),
$$

where $L_{t|t-1}$ and $U_{t|t-1}$ are the lower and upper bound respectively.

The testing procedures are given by [9]. The conditional coverage test is a test which checks whether the exceedances going outside the volatility boundary happen independently and proportionately. The exceedances need to be independent since patterns might indicate something that the model cannot model itself such as seasonal patterns which in turn need to be modelled via some other model. The exceedances need to be proportional otherwise the model is under or over estimating the volatility.

The conditional coverage test is constructed from the unconditional coverage and independence test. Each of the test can be formulated as a likelihood ratio test. The coverage test is used to test if $E(I_t) = p$ given that I_t is independent. The likelihood under H_0 and the alternative are as follows:

$$
L(j; I_N, ..., I_T) = (1 - j)^{n_0} j^{n_1},
$$
$$
L(\hat{\pi}; I_N, ..., I_T) = (1 - \hat{\pi})^{n_0} \hat{\pi}^{n_1},
$$

where $j = 0.05$ is the test quantile, $\hat{\pi} = \frac{n_1}{n_0 + n_1}$ is the estimate of the exceedance probability, n_0 is amount of non exceedances and n_1 is the amount of exceedances. Then the coverage likelihood ratio test is as follows:

$$
LR_{coverage} = -2 \log \left(\frac{L(j; I_1, ..., I_N)}{L(\hat{\pi}; I_1, ..., I_N)} \right) \sim \chi_1^2.
$$

The coverage test does not say anything about dependence of the exceedances this is tested with the LR test of dependence. The independence is tested via a explicit first order Markov chain. Consider a Markov chain, I_t with transition matrix:

$$
\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix},
$$

where $\pi_{ij} = P(I_t = j | I_{t-1} = i)$. The approximate likelihood is as follows:

$$L(\Pi_1; I_N, ..., I_T) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}},$$

where n_{ij} is the number of observations with value i followed by j. The approximation of Π_1 is as follows:

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \frac{n_{10}}{n_{10}+n_{11}} & \frac{n_{11}}{n_{10}+n_{11}} \end{bmatrix},$$

Then Π_2 is an independent Markov chain and the likelihood under H_0 becomes as follows:

$$\Pi_2 = \begin{bmatrix} 1 - \pi_2 & \pi_2 \\ 1 - \pi_2 & \pi_2 \end{bmatrix},$$

$$L(\Pi_2; I_N, ..., I_T) = (1 - \pi_2)^{n_{00}+n_{10}} \pi_2^{n_{01}+n_{11}},$$

where $\hat{\Pi}_2 = \hat{\pi}_2 = \frac{n_{01}+n_{11}}{n_{01}+n_{01}+n_{01}+n_{11}}$. Then the likelihood ratio test is again χ_2^2 distributed.

$$LR_{independent} = -2 \log \left(\frac{L(\hat{\Pi}_2; I_1, ..., I_N)}{L(\hat{\Pi}_1; I_1, ..., I_N)} \right) \sim \chi_1^2.$$

These two test can be used separately to test their H_0 but the test can also be combined to form the joint test of coverage and independence (conditional coverage test) which has a likelihood ratio as follows:

$$LR_{conditional} = -2 \log \left(\frac{L(p; I_1, ..., I_N)}{L(\hat{\Pi}_1; I_1, ..., I_N)} \right) \sim \chi_2^2.$$

This ratio test follows a χ_2^2 distribution under the H_0.

Figure 14 shows the exceedances for the Mogas SGARCH model. The results of the coverage test and conditional coverage test are shown in table 12. The results show that all the models except the Dated_Brent have a p-value larger than 0.05 so the H_0, exceedances are independent and the number of exceedances is within the boundary's cannot be rejected. These results show that the forecast interval achieved from the models is a good approximation of the actual returns.
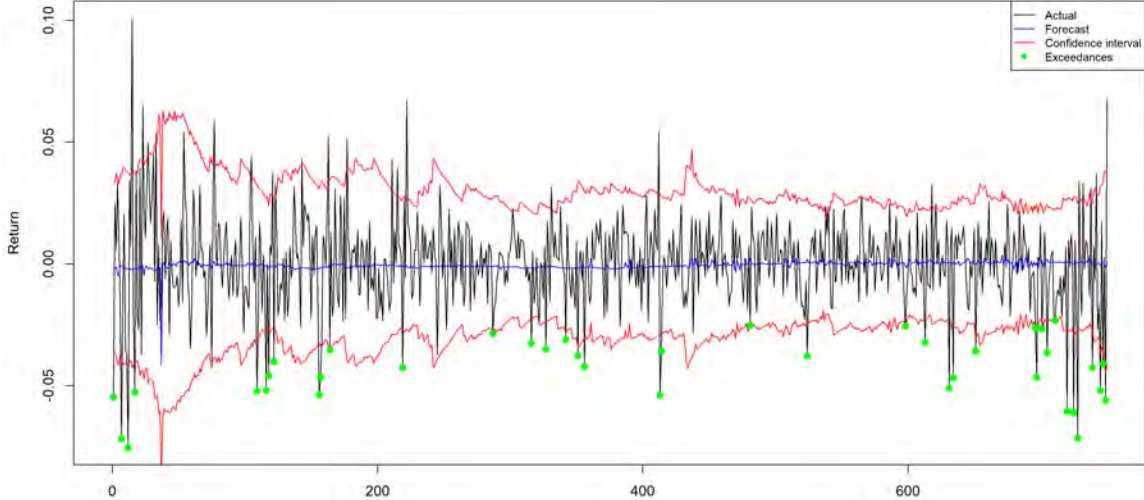
Figure 14: Exceedances for the Return forecast Mogas (1,1,2,2) SGARCH model

Product	Expected exceedances	Actual exceedances	Coverage test (p-value)	Conditional coverage test (p-value)
Mogas	37	39	0.8028	0.7497
HSFO	37	47	0.1250	0.3076
Jet	37	39	0.8028	0.7018
Naphtha	37	41	0.5632	0.7268
LSGO	37	37	0.9331	0.7993
Dated_Brent	37	54	0.0093*	0.0304*
Margin	37	30	0.1936	0.1483

Table 12: Coverage and conditional coverage test

6.4 Vine Copula fitting

Now that all the MSE scores are calculated the best models are selected to construct a margin portfolio. This portfolio of six products consists of the separate models and their interaction which is described by a vine copula. As mentioned in paragraph 3.3 the fitting of the copula is done on the errors of the best models. The fitting scheme is described in paragraph 5.3. The copula structure for the first period is shown in figure 15 and the types of copula are shown in figure 16. The number representation is in table 13.

Product	Number
Mogas	1
HSFO	2
Jet	3
Naphtha	4
LSGO	5
Dated_Brent	6

Table 13: Products and their corresponding number for the copula plots

The fitting and refitting scheme is the same as with the GARCH and continues models but the forecasting scheme changes. To get the g-ahead forecast g simulations are done and the quantiles are taken to get the desired errors. These errors can be filled in (16) and the return can be forecasted.

Table 14 shows the MSE results for the univariate and multivariate models for the margin. Just like in paragraph 6.3 the interval of the model can be evaluated with the coverage and unconditional coverage test. Table 15 shows that the vine copula model for the margin is not a good approximation for the interval of the actual margin returns. Figure 17 and 18 show the exceedance plots for the univariate and multivariate models. Figure 19 and 20 show the volatility proxy's forecast and the confidence interval is missing since this can not be computed for the portfolio.

	r	r^2	$|r|$
Margin (Univariate)	15.3489	0.1731	10.0917
Margin (Copula model)	15.3159	0.1781	8.4194

Table 14: MSE for the uni- and multivariate models for the margin (values are of the order 10^{-4})
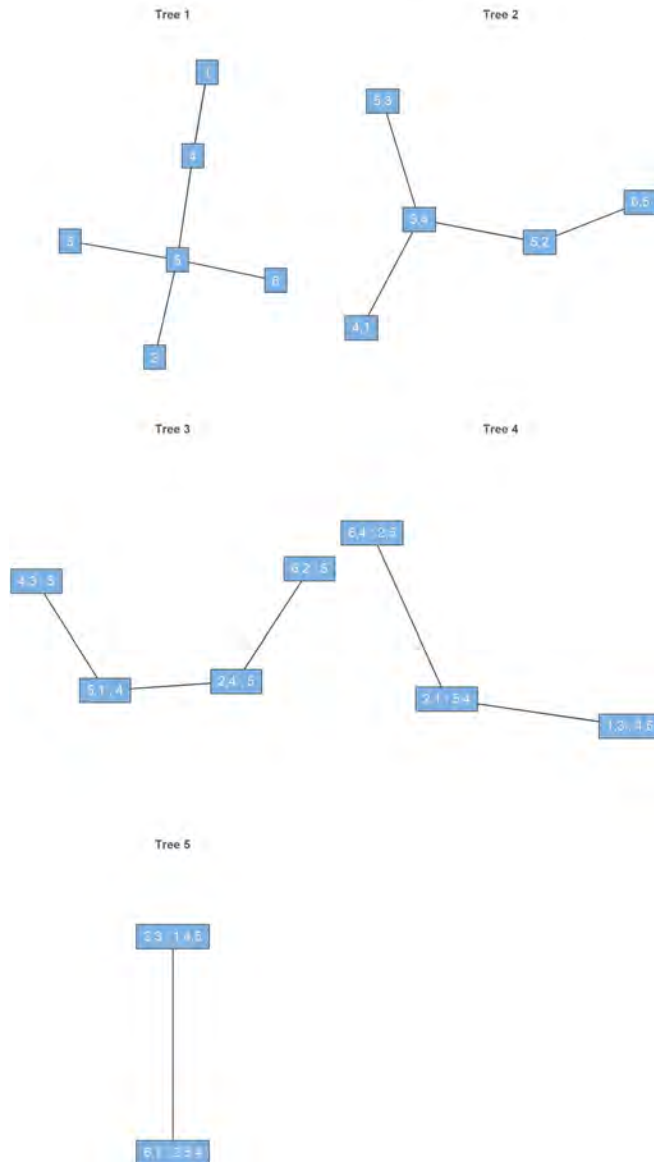
Figure 15: Copula tree structure
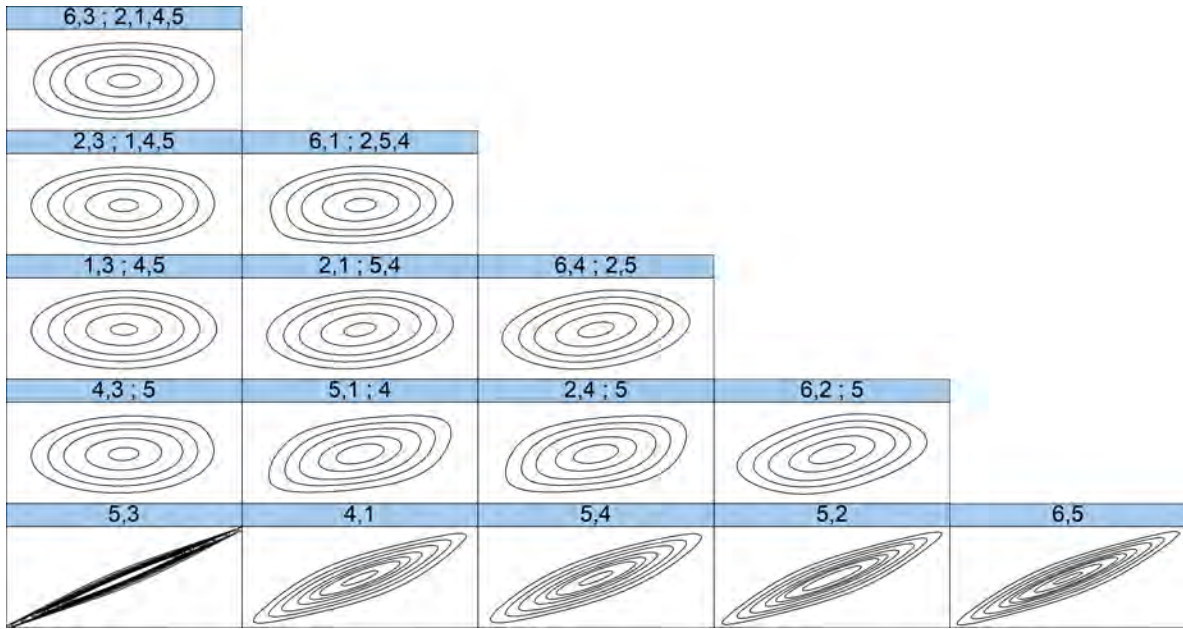
Figure 16: Copulas for each edge in the corresponding tree
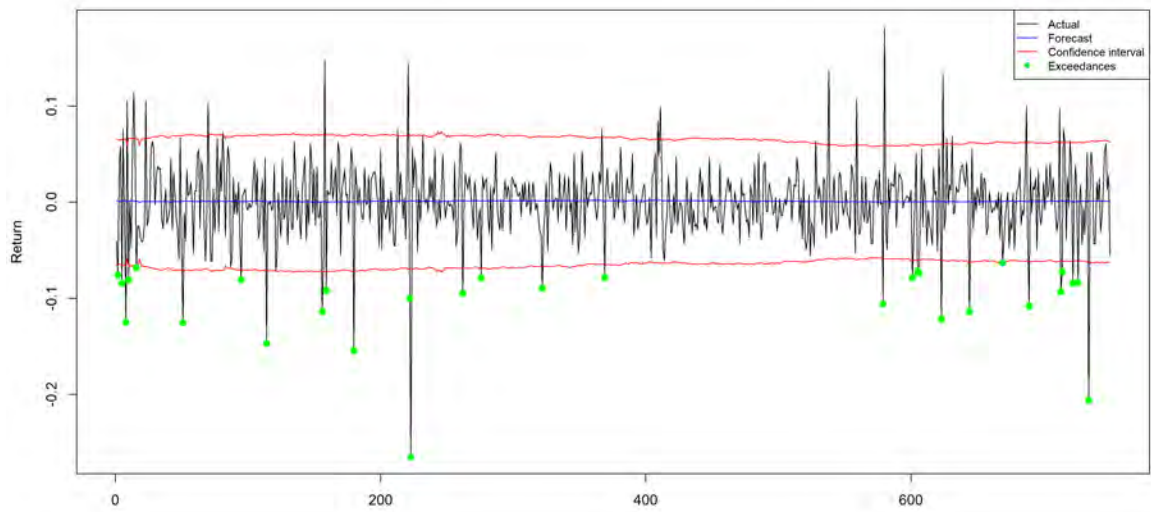
Figure 17: Exceedances for the Return forecast Margin (2,1,0,2) SGARCH model
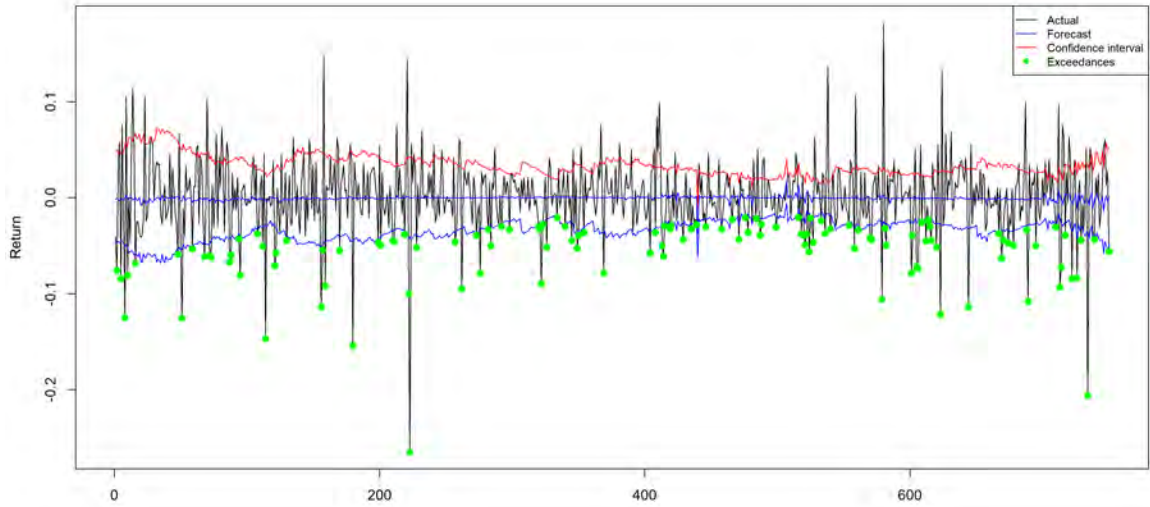
Figure 18: Exceedances for the Return forecast Margin vine copula model
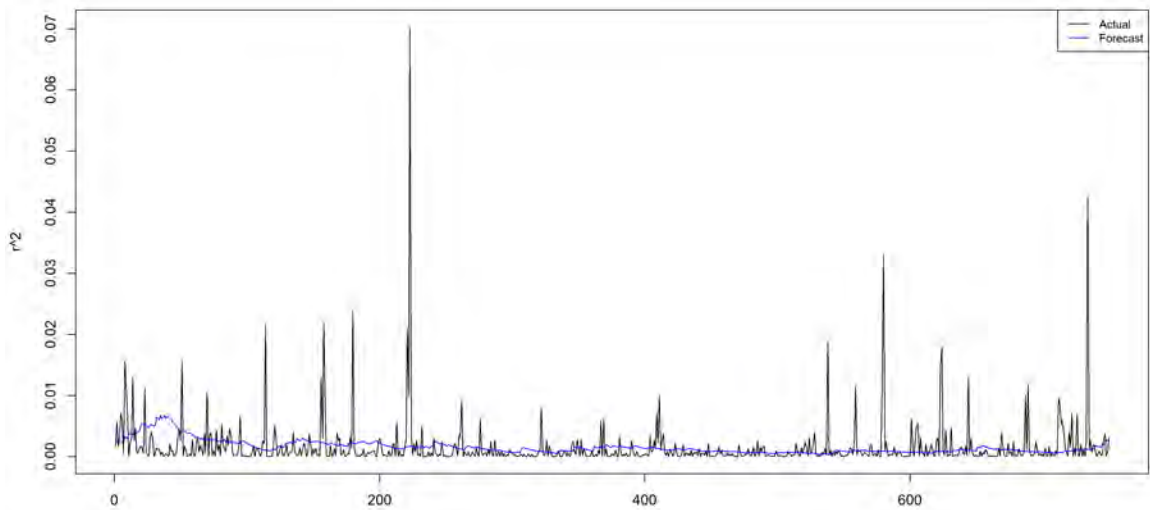
Figure 19: Return Volatility forecast (r^2) Margin multivariate model with confidence interval

Product	Expected exceedances	Actual exceedances	Coverage test (p-value)	Conditional coverage test (p-value)
Margin (Univariate)	37	30	0.1936	0.1483
Margin (Copula model)	37	123	0	0

Table 15: Coverage and conditional coverage test for the margin models
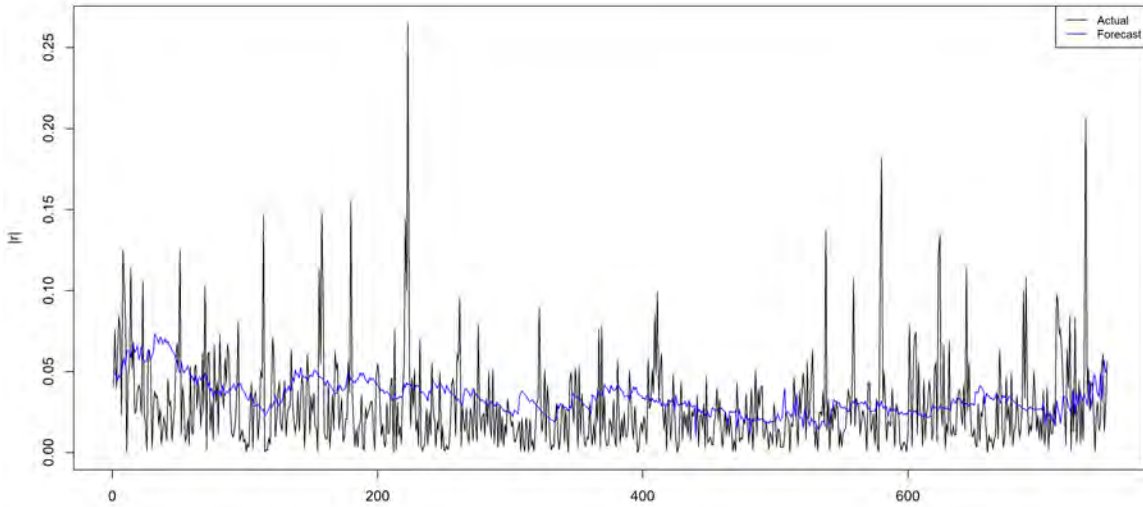
Figure 20: Return Volatility forecast ($|r|$) Margin multivariate model with confidence interval

7 R implementation

The models and test used in this thesis are implemented in RStudio. This section shows the packages and the important functions used to obtain the results.

- Rugarch package contains functions which estimate ARMA-GARCH parameters. We used the following functions:

    - ugarchspec() is used to set the model specifications such as error distribution and GARCH model type.
    - ugarchfit() is used to find the model parameters and details such as AIC values and test resutls for asymmetry.
    - fitdist() is used to find the parameters for the error distribution.
    - qdist() is used to calculate the quantiles of a certain error distribution.
    - VaRTest() is used to perform the coverage and unconditional coverage test.

    For further information about the package and the functions see [13].

- VineCopula package contains functions which estimate copula parameters and fits vine structures to the data. We used the following functions:

    - RVineStructureSelect() is used to determine the vine structure and estimate the copula parameters.

    For further information about the package and the functions see [26].

- TSA package contains all kinds of time series functions, we only use the periodogram() function to determine if there is an seasonal component present.

The following functions are normally present in R:

- acf() is used to make the acf plots.

- Box.test() is used to perform the Ljung-Box.

- cor() is used to calculate the Pearson correlations.

8 Conclusion

We have described how for each product and the margin certain types of models are preferred over the standard GARCH models. Each model has its own properties which also apply to the products and the margin. For example, the best model for the Mogas was the standard GARCH model which indicates that the returns of the Mogas have time-varying volatility, are not likely to be impacted by the news, show no periodicity and the volatility does not have a long memory. Each of these models is then fitted to their respective return series and the MSE is calculated for the mean process and the proxy's for the volatility. Table 11 shows that the discrete models perform better than the continues models. A cause is that the continues models have a rather fast convergence rate towards the mean volatility. In the Mogas case this took three days. The GARCH models can handle a less steep decay of the volatility which in this case is the better fit.

The best models per product are selected based on the MSE results. The best models are combined in a portfolio which connects the model errors via some copula. With the univariate and multivariate model it is possible to answer the problem statement of this thesis:

- Is modelling the margin as a univariate time series better than modelling the underlying products and their interaction?

The MSE results showed that the multivariate model slightly outperforms the univariate series. This indicates that when modelling the margin as a univariate series the individual information of the products is lost which results in a worse performance. The backtest on the models in the portfolio showed that the individual models for the products are good approximations of the actual return interval, except for the Dated_Brent which showed a significant result. The forecast interval of the portfolio done with the vine copula model was significant as well which indicates that the portfolio does not forecast the interval properly. So the slight increase in accuracy of the forecast MSE is not translated into better coverage which means that the multivariate model is too confident in his forecast. The univariate model would be a better choice since this model covers the forecast interval at an 0.05 level which results in a better approximation of the margin.

While the univariate model captures the interval better it does limit the number of applications which can be done with the model. Applications such as optimizing certain aspects of the hedging process such as the hedging ratios and how many barrels to hedge will not be possible because the individual models are needed for this application. In section 9 improvements to the estimating process and fitting process are suggested these improvements might help the multivariate model perform better. Other possibilities such as option simulations are possible with the univariate model. Option simulation can be done by running a certain amount of simulations and see what the option price will do versus what the banks or other institutes offer.

9 Possible extensions

The current models are based on one type of approach. The following methods could improve the models and their fit. Small changes such as the length of the fitting period could improve the fit. The current fitting period was chosen as an estimate of a certain pricing cycle. This cycle estimate might not be accurate. There are also some more complicated improvements which are mentioned further in this section.

9.1 Estimation method

The preferred method of parameter estimation is the ML method which is also used in this thesis. The preference for the ML method is due to its nice properties such as consistency and efficiency. However, it might be possible that the distribution of the ML estimators is a poor approximation of the sample ; c.f. [25]. [25] shows that with techniques such as the Jackknife estimation method one can reduce the bias in the estimation results. The Jackknife method divides the data of T observations into o subsets with an equal amount of observations and then uses the parameter estimation results of those subsets to reduce the bias via the following formula:

$$\hat{\theta}_{Jack} = \frac{o}{o-1}\hat{\theta}_T - \frac{\sum_{i=1}^{o}\hat{\theta}_i}{o^2 - o}.$$

Estimators $\hat{\theta}_T$ and $\hat{\theta}_i$ are estimates of θ obtained by via ML. This method might improve the parameter estimation of the GARCH and continues models which then might improve the fit of the copula model.

9.2 Extra variables

The current models only use the returns as input for their parameter estimation. The data can be expended using other variables such as News sentiment, number of trades, the volume traded. These types of variables can improve the fit of the model for both processes the mean and volatility. Equation (4) shows how extra variables can be added to the volatility process in a similar fashion this can be applied to the mean process.

References

[1] (2009). *Handbook of Financial Time Series*, chapter Practical Issues in the Analysis of Univariate GARCH Models, pages 113–151. Springer.

[2] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198.

[3] Allen, A. O. (1990). *Probability, Statistics, and Queueing Theory with Computer Science Applications*. Academic Press Professional, Inc., San Diego, CA, USA.

[4] Andersen, T. G., Davis, R. A., P., K. J., and T., M. (2009). *Handbook of Financial Time Series*, chapter Introduction, pages 1–13. Springer.

[5] Bedford, T. and Cooke, R. M. (2002). Vines-a new graphical model for dependent random variables. *Annals of Statistics*, 30(1):1031–1068.

[6] Berkes, I. and Horvath, L. (2004). The efficiency of the estimators of the parameters in garch processes. *The Annals of Statistics*, 32(2):633–655.

[7] Bougerol, P. and Picard, N. (1992). Strict stationarity of generalized autoregressive processes. *Annals of Probability*, 20(4):1714–1730.

[8] Brockwell, P. J. (2009). *Handbook of Financial Time Series*, chapter An Overview of Asset-Price Models, pages 403–416. Springer.

[9] Christoffersen, P. F. (1995). Evaluating interval forecast. *International Economic Review*, 39(4):841–862.

[10] Embrechts, P., Lindskog, F., and A., M. (2001). Modelling dependence with copulas and applications to risk management. Master's thesis, Federal Institute of Technology (ETH) Zurich, Zurich.

[11] Engle, R. F. and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778.

[12] Francq, C. and Zakoïan, J.-M. (2009). *Handbook of Financial Time Series*, chapter A Tour in the Asymptotic Theory of GARCH Estimation, pages 85–111. Springer.

[13] Ghalanos, A. (2017). Introduction to the rugarch package.

[14] Gruber, L. F. (2017). Review of dependence modeling with regular vine copulas and current methods for inference and model selection. Working paper, Research group Mathematical Statisticss.

[15] Havran, J. (2013). Forward curve modeling for natural gas. Master's thesis, Swiss Federal Institute of Technology (ETH) Zurich, Zurich.

[16] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of Financial Studies*, 6(2):327–343.

[17] Kalemanova, A., Schmid, B., and Werner, R. (2007). The normal inverse gaussian distribution for synthetic cdo pricing. *Journal of Derivatives*, 14(3):80–94.

[18] Klemens, B. (2009). *Modeling with Data: Tools and Techniques for Scientific Computing*. Princeton University Press.

[19] Leeb, H. and Pötscher, B. M. (2009). *Handbook of Financial Time Series*, chapter Model selection, pages 889–917. Springer.

[20] Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

[21] Lo, A. W. (1989). Long-term memory in stock market prices. Working Paper 2984, National Bureau of Economic Research.

[22] Masteika, S., Rutkauskas, A. V., and Alexander, J. A. (2012). Continuous futures data series for back testing and technical analysis. Working paper, International Conference on Economics, Business and Marketing Management.

[23] Monstvilaite, M. (2016). Portfolio value-at-risk using regular vine copulas. Master's thesis, Stockholms Universitet, 106 91 Stockholm.

[24] Patton, A. J. (2009). *Handbook of Financial Time Series*, chapter Copula-Based Models for Financial Time Series, pages 767–781. Springer.

[25] Phillips, P. C. B. and Yu, J. (2009). *Handbook of Financial Time Series*, chapter Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance, pages 497–528. Springer.

[26] Schepsmeier, U. (2015). Package vinecopula.

[27] Teräsvirta, T. (2009). *Handbook of Financial Time Series*, chapter An Introduction to Univariate GARCH models, pages 17–39. Springer.

[28] Wang, X., He, X., Y., B., and Yanlong, Z. (2018). Parameter estimates of heston stochastic volatility model with mle and consistent ekf algorithm. *SCIENCE CHINA Information Sciences*, 61(1):1–17.

[29] Yan, M. (2014). Appendix: Polynomial.

Appendices

A Appendix 1: Periodograms

HSFO periodogram
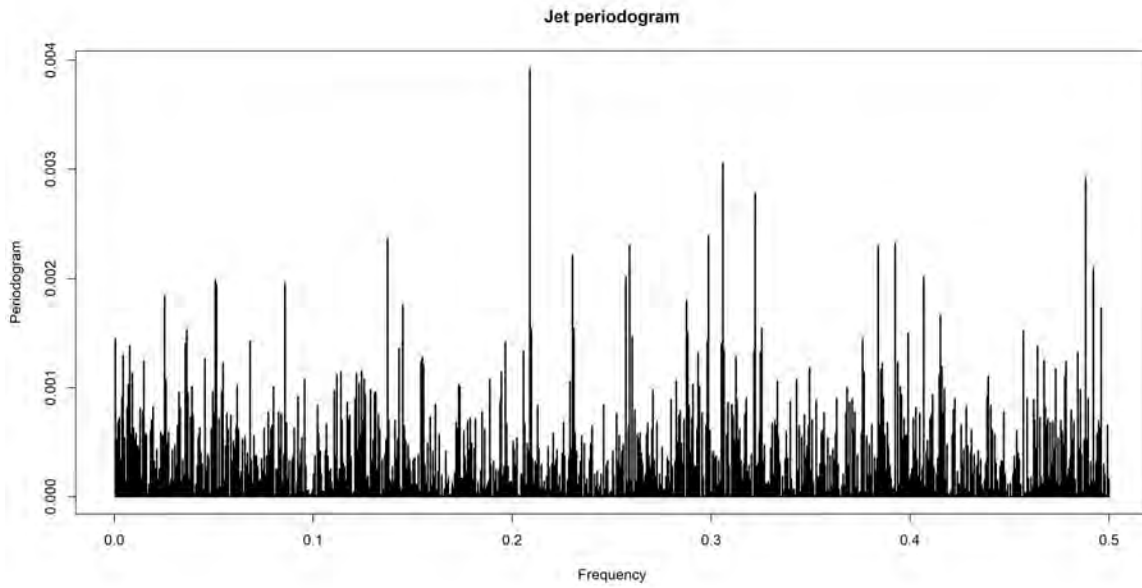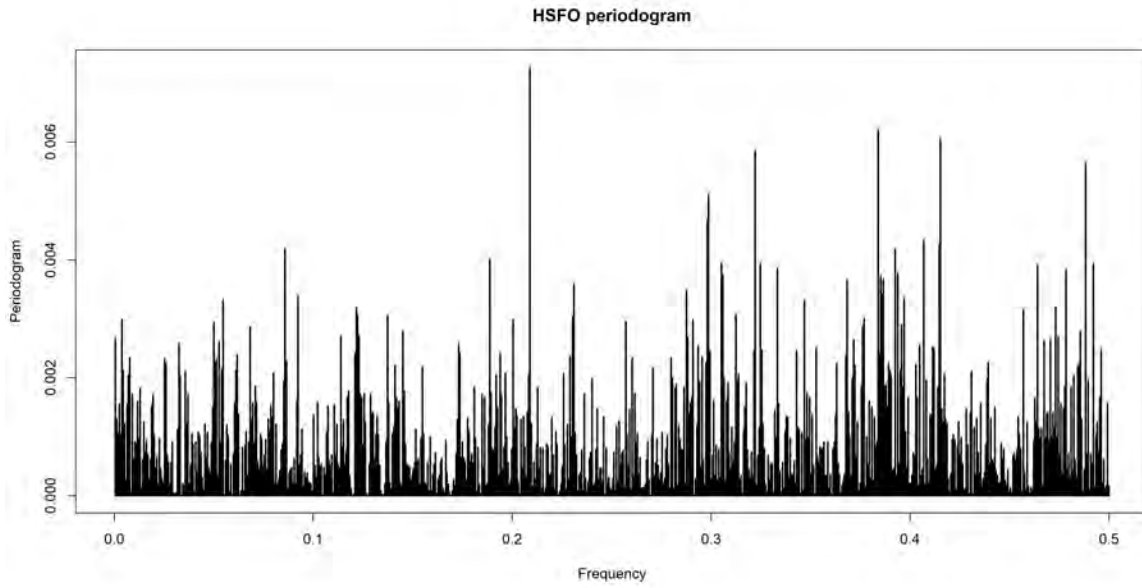
Jet periodogram

Naphtha periodogram
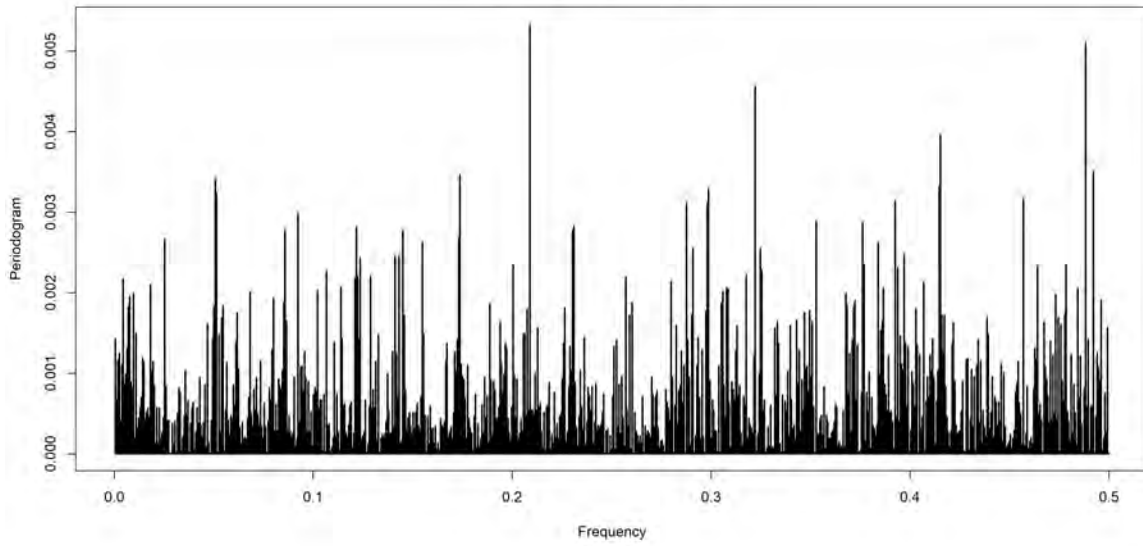
LSGO periodogram
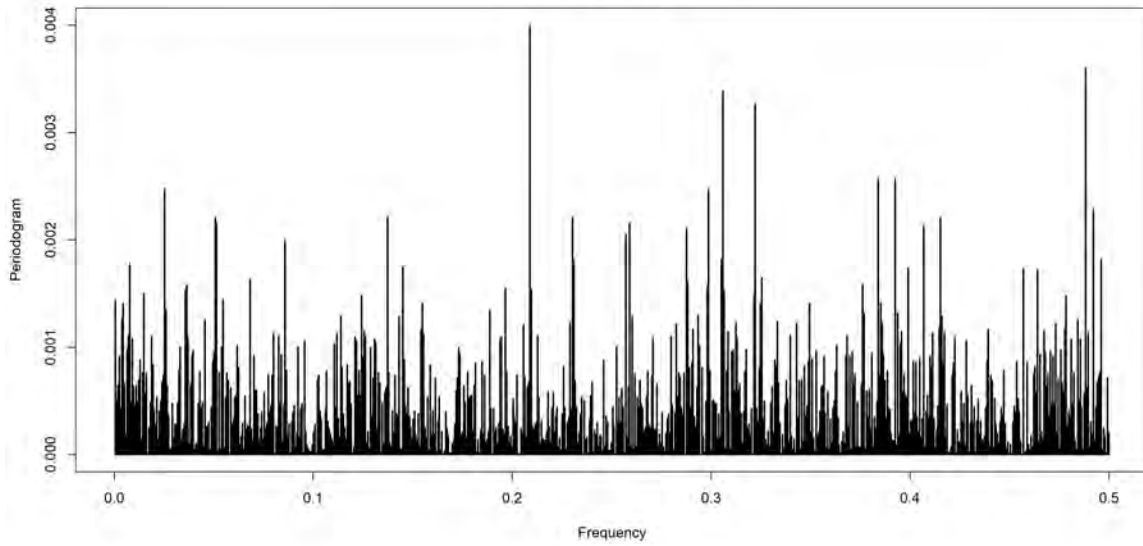
Dated_Brent periodogram
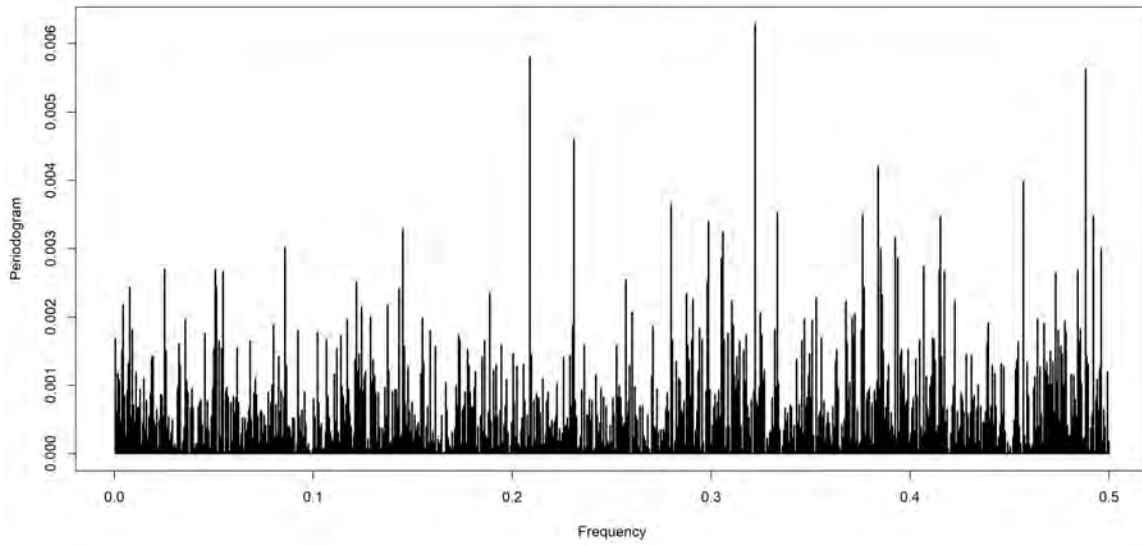
Margin periodogram
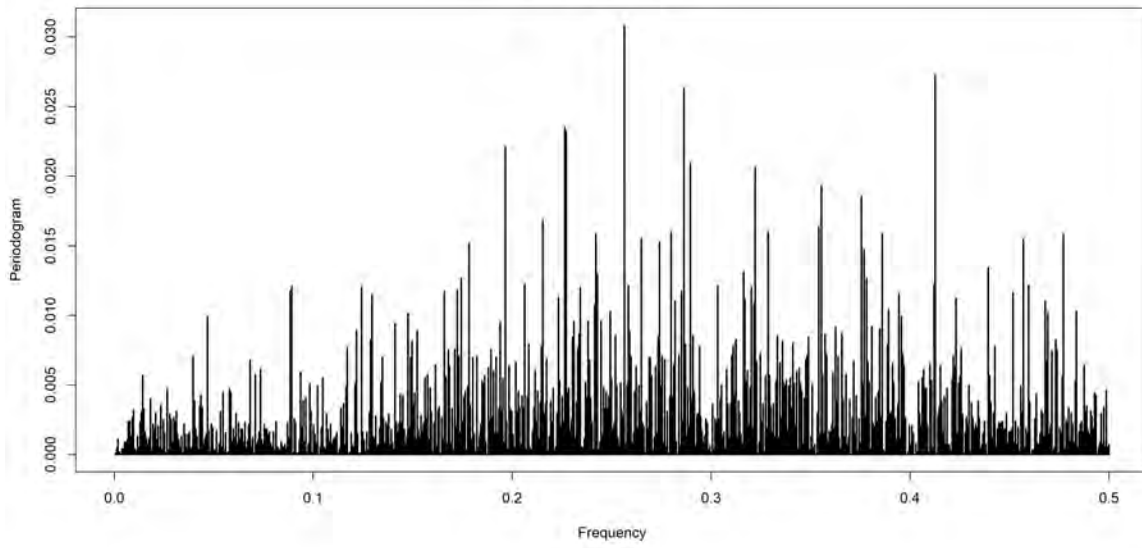
B   Appendix 2: r_t and r_t^2 plots

HSFO log returns
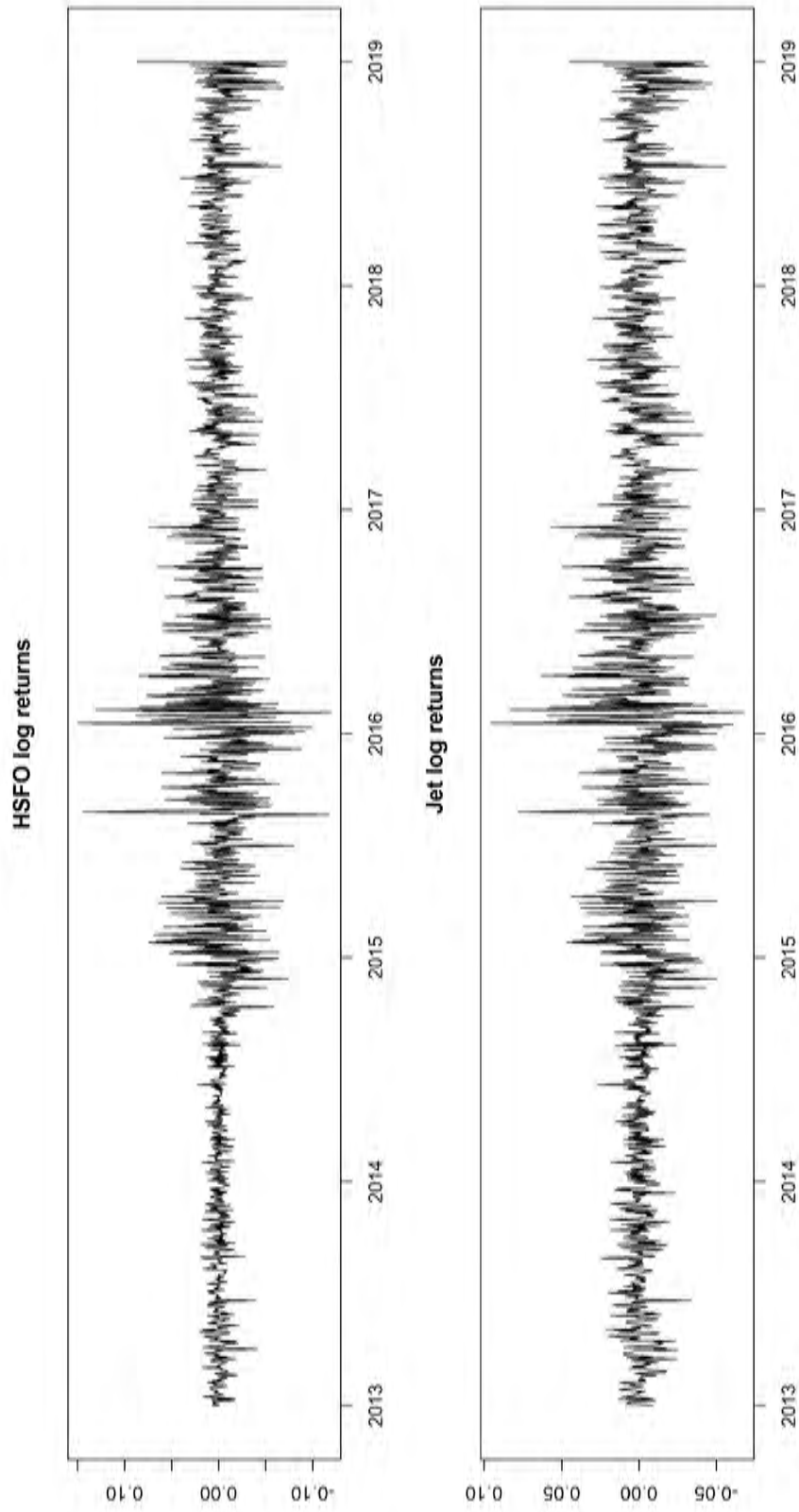
Jet log returns

Naphtha log returns
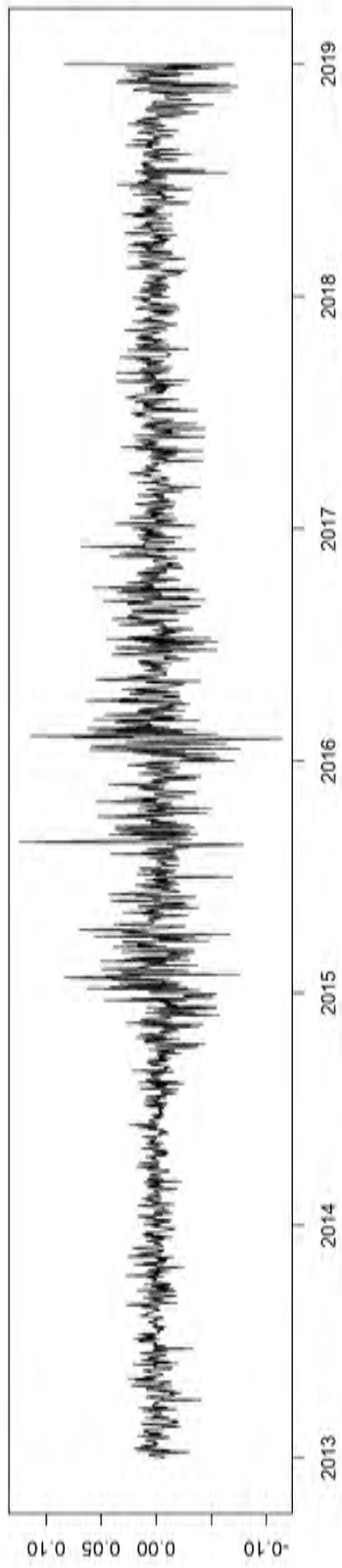
LSGO log returns
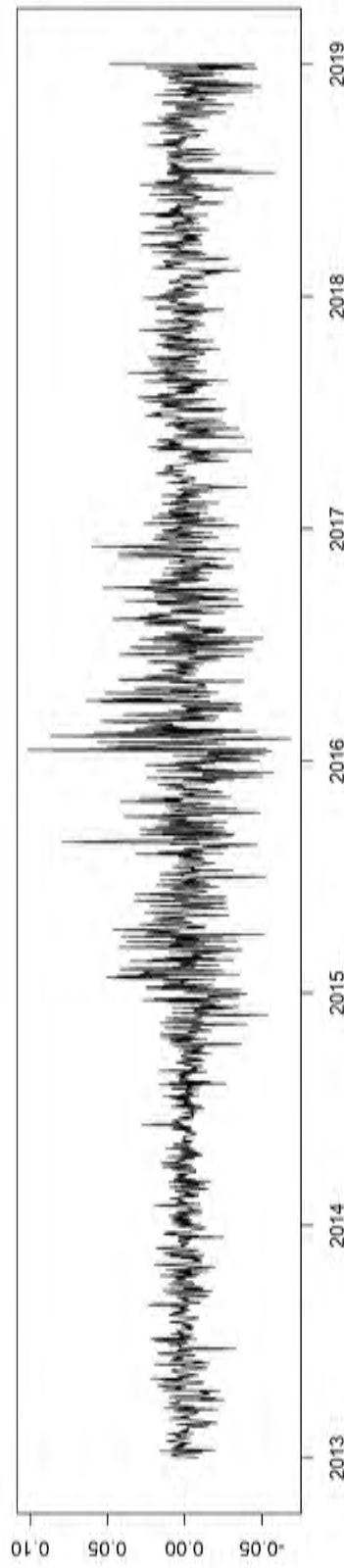
Dated_Brent log returns
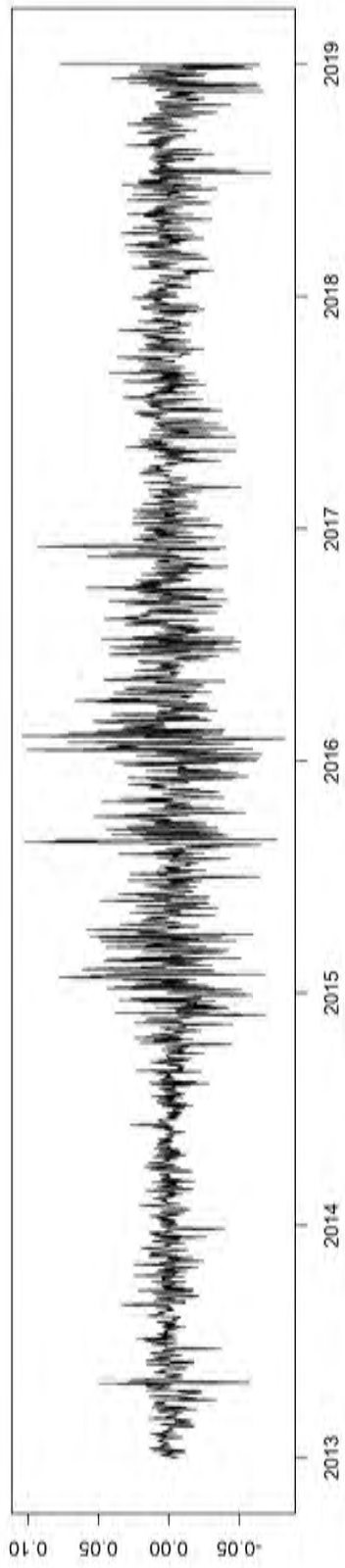
Margin log returns
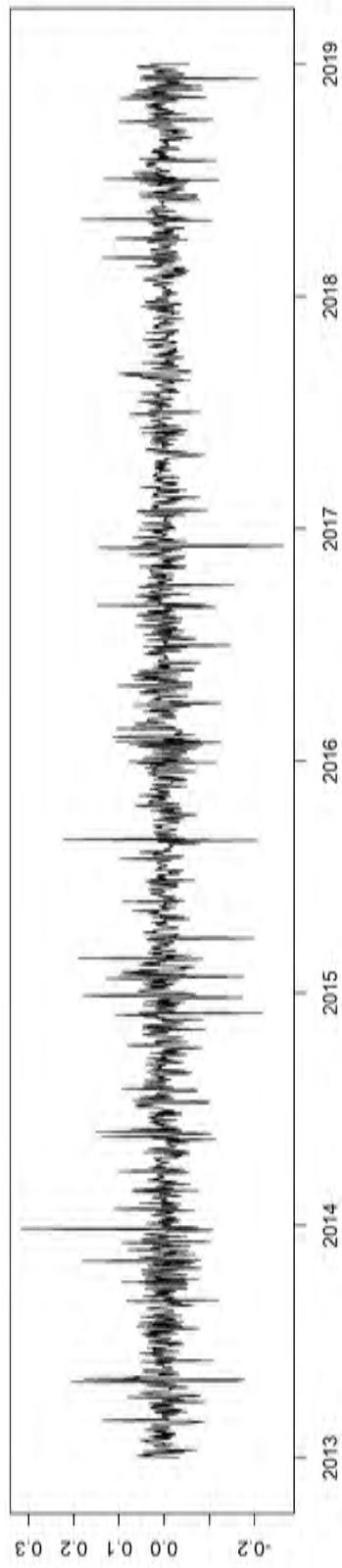
HSFO sqaured log returns
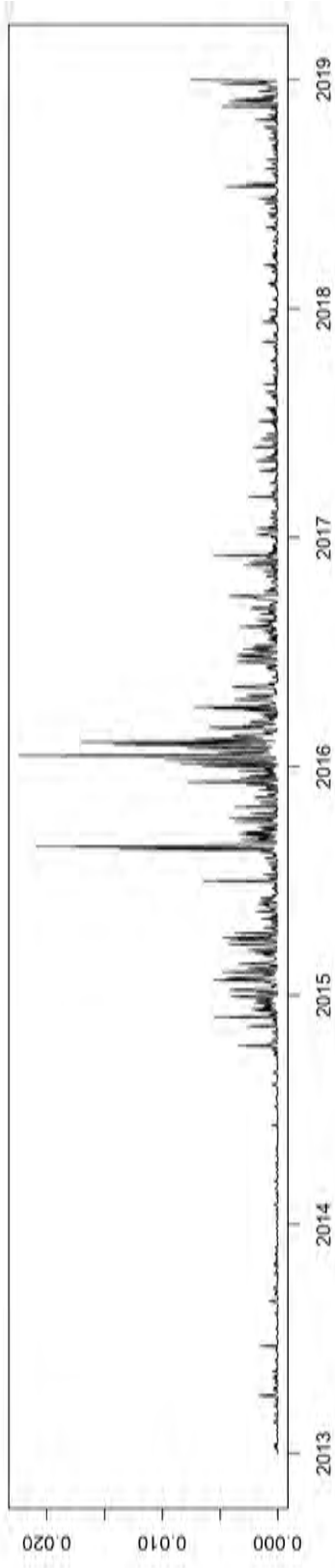
Jet sqaured log returns
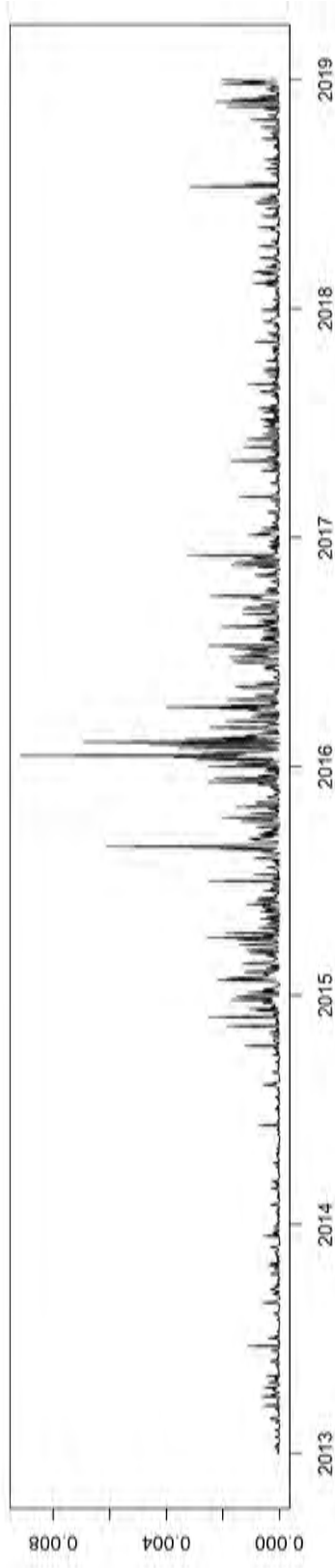
Naphtha sqaured log returns
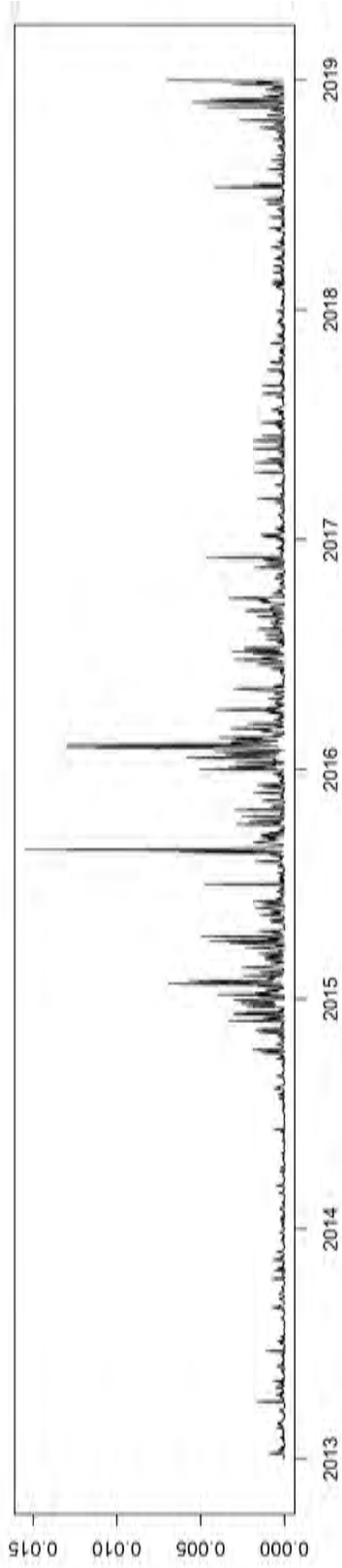
LSGO sqaured log returns
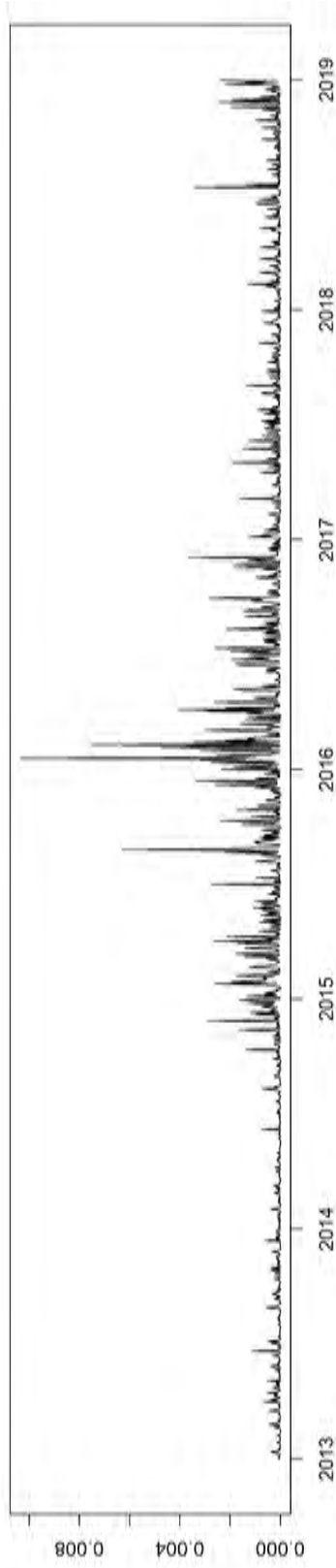
Dated_Brent sqaured log returns
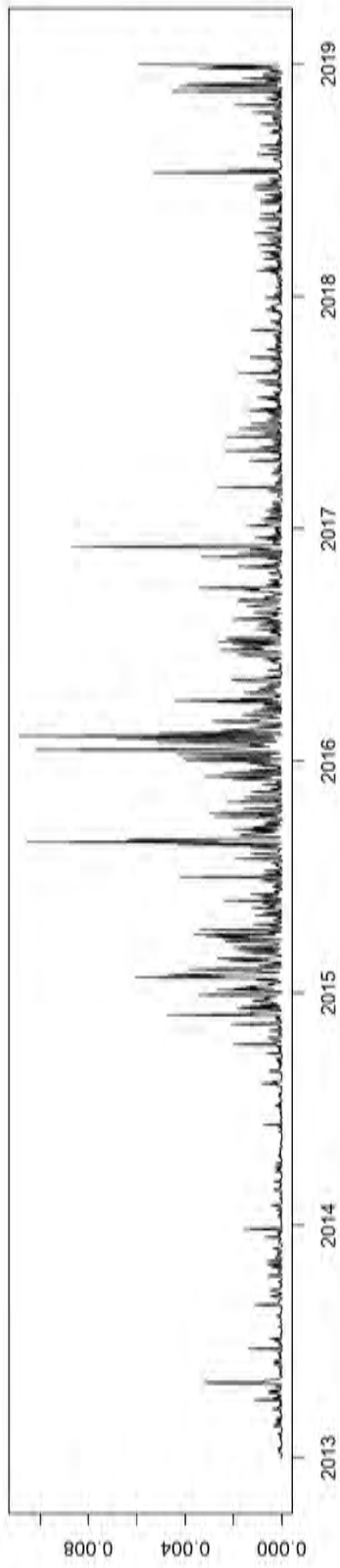
Margin sqaured log returns
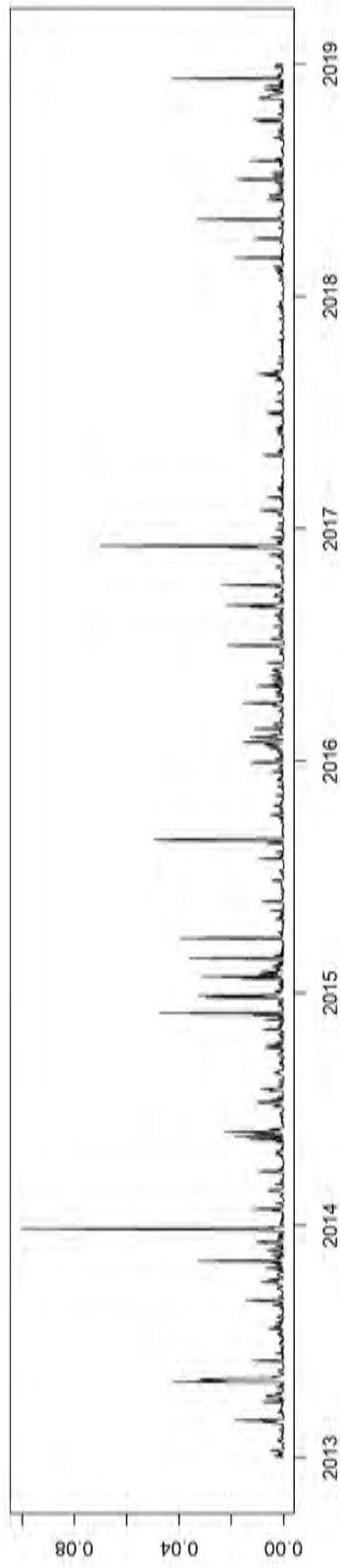
C   Appendix 3: r_t and r_t^2 ACF plots
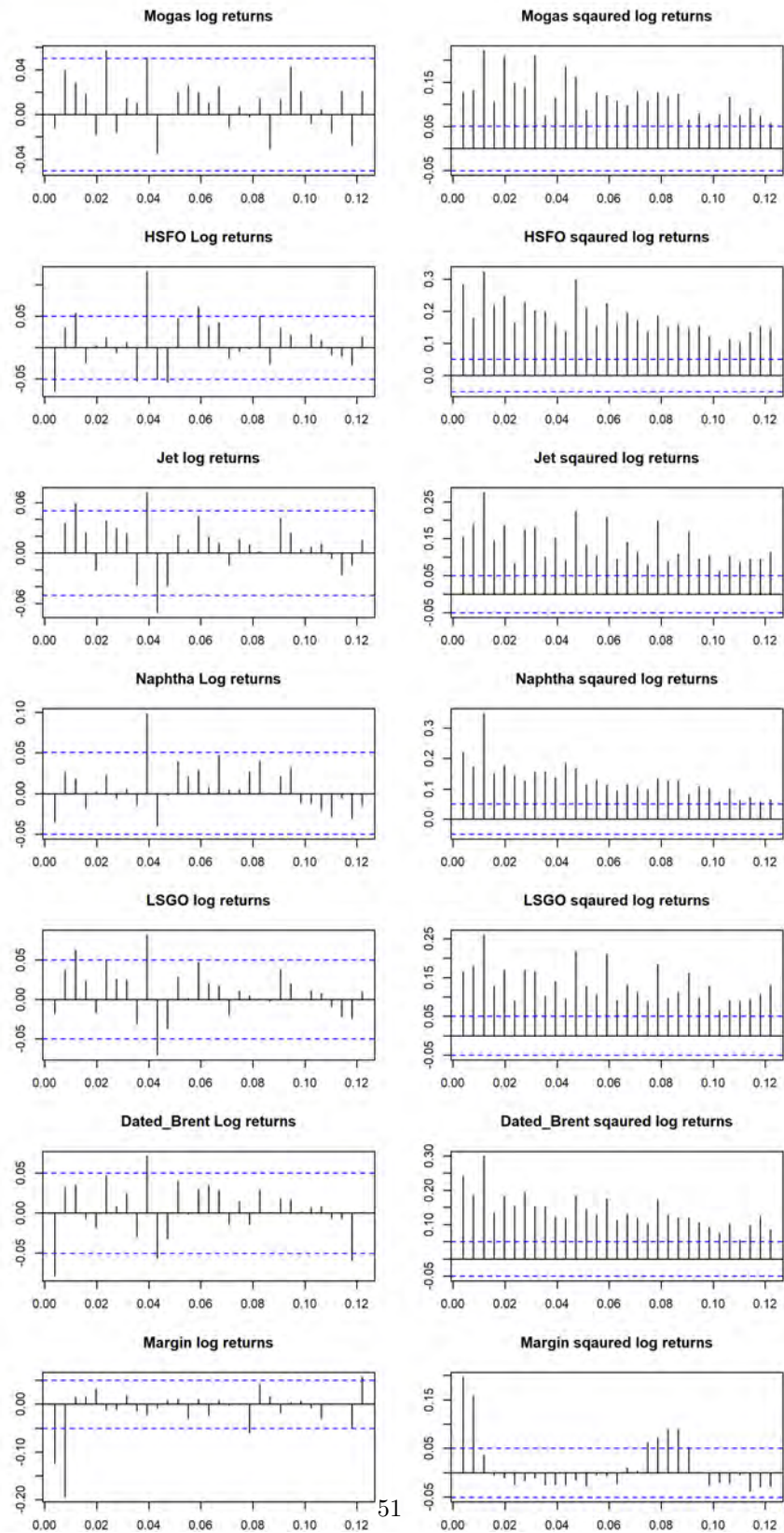
D   Appendix 4: Error distribution QQ-plots

D.1   Mogas
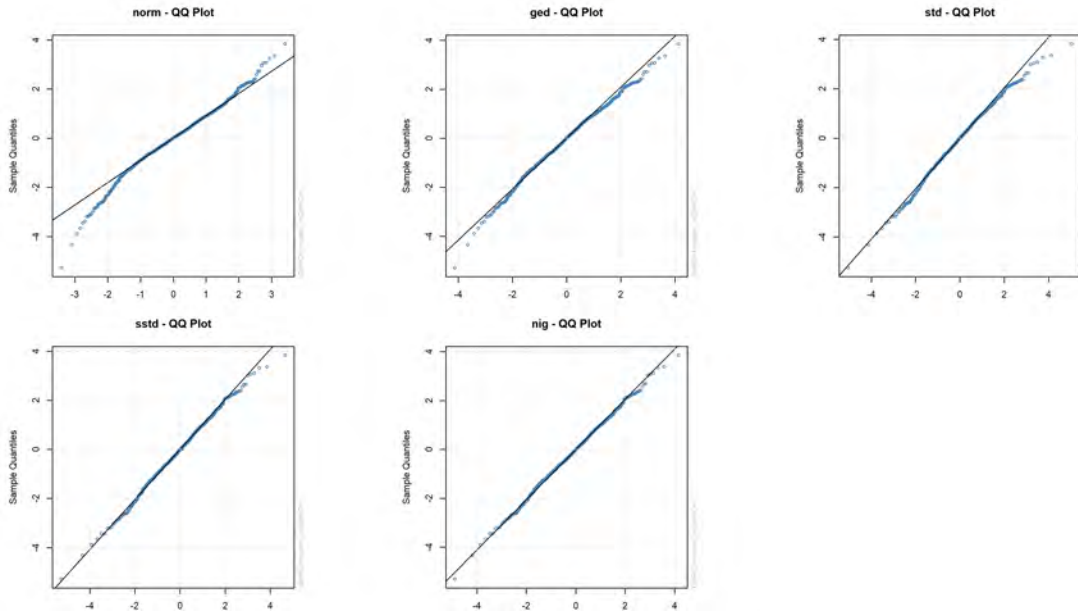
D.2   HSFO
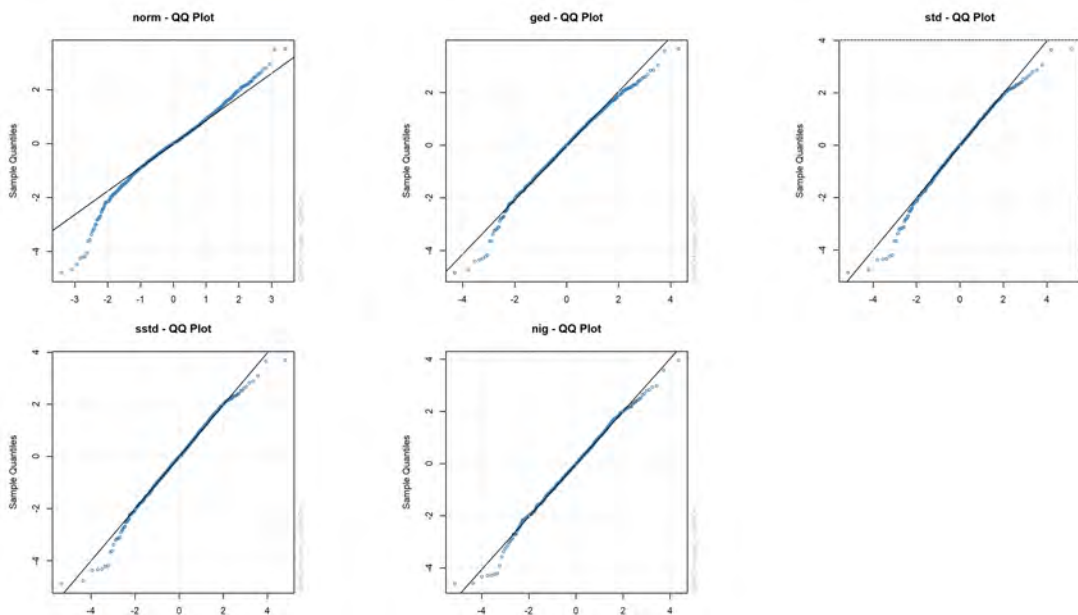
D.3 Jet
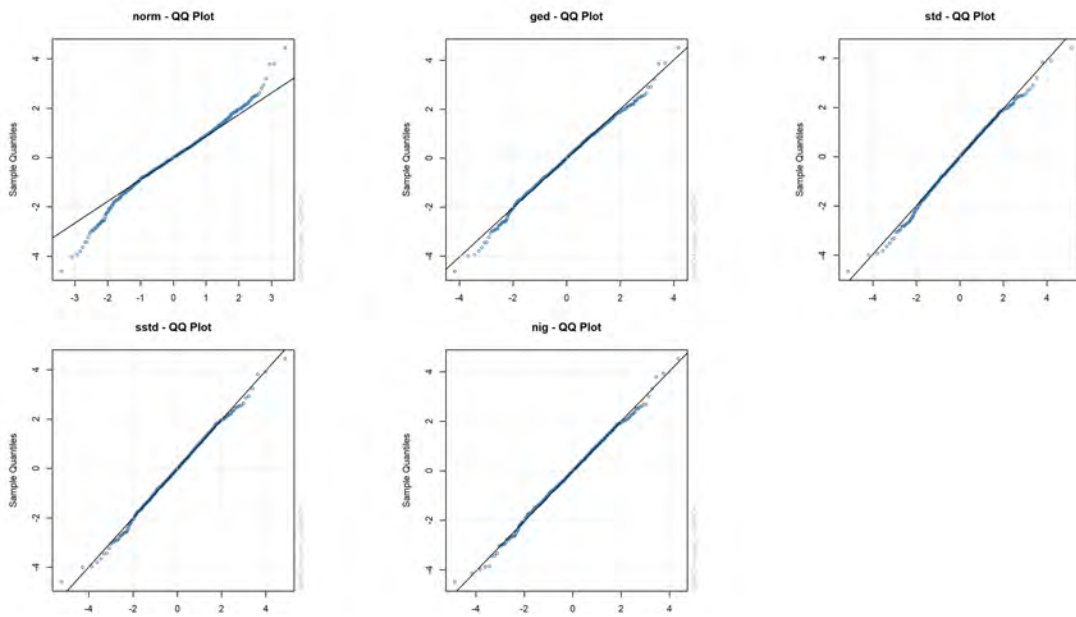
D.4 Naphtha
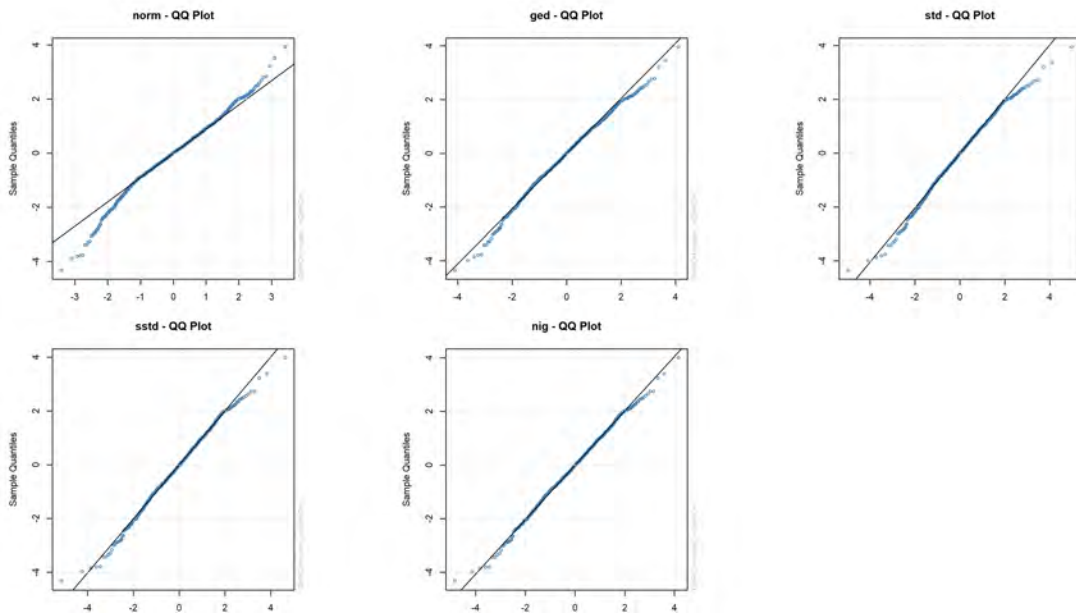
D.5 LSGO

D.6 Dated Brent

D.7 Margin

norm - QQ Plot

ged - QQ Plot

std - QQ Plot

sstd - QQ Plot

nig - QQ Plot
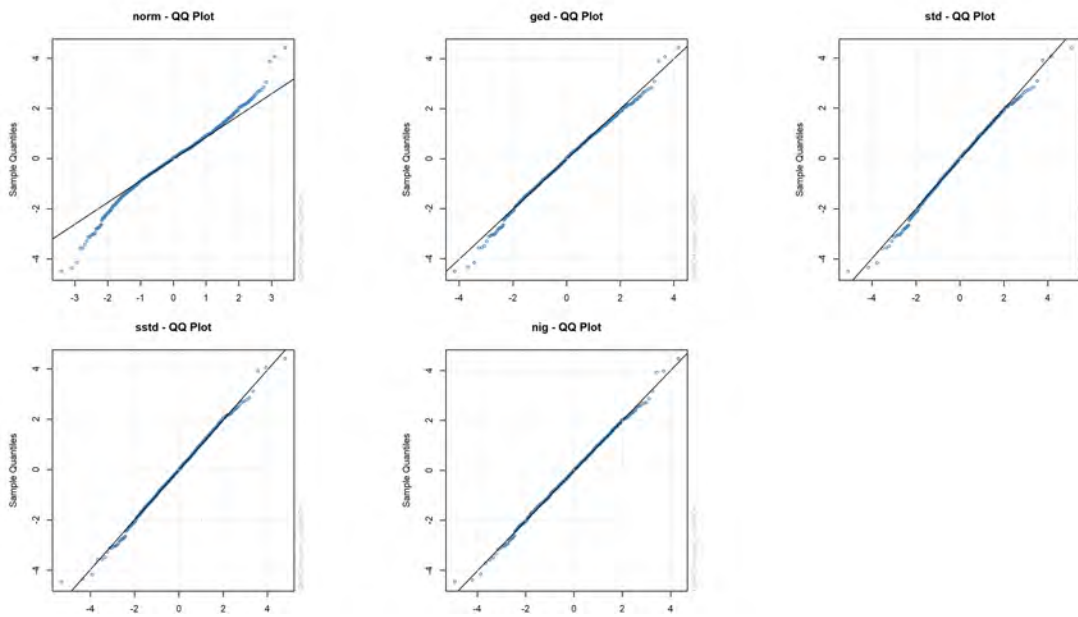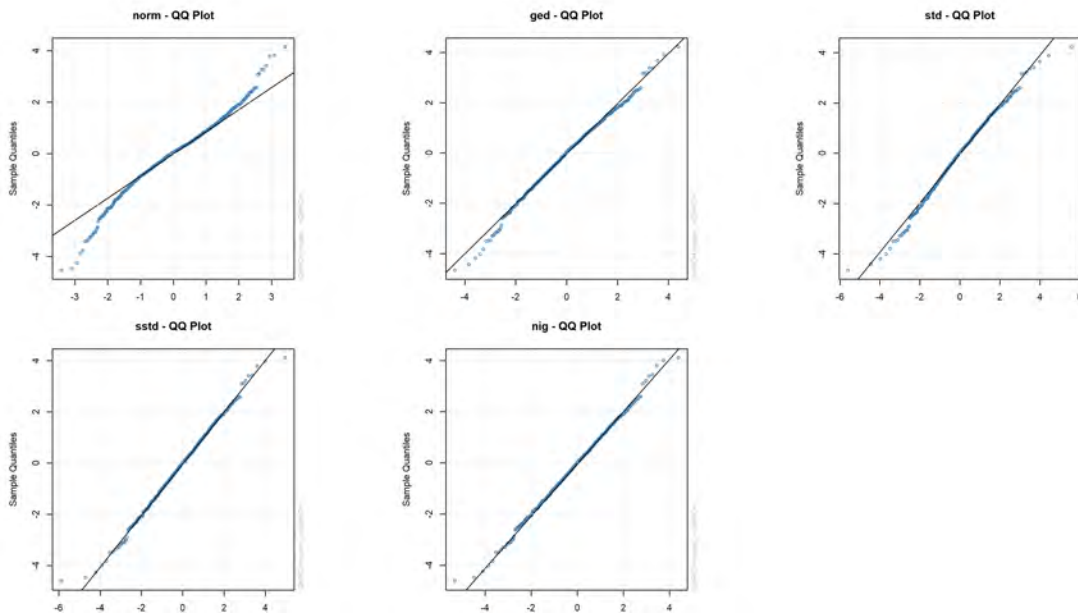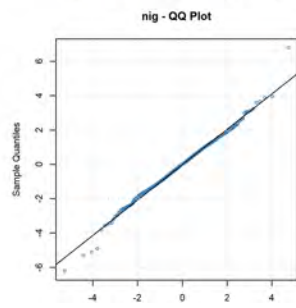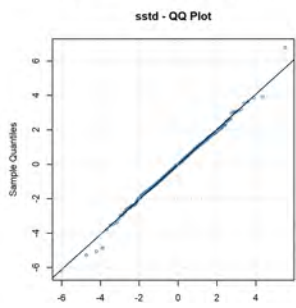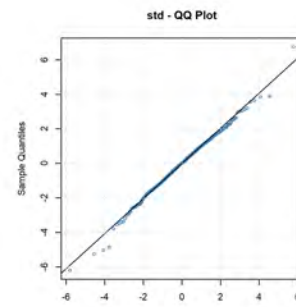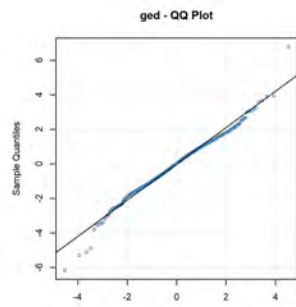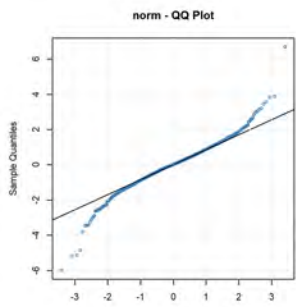
E   Appendix 5: Mean and volatility result plots

E.1   HSFO

E.2   Jet

58
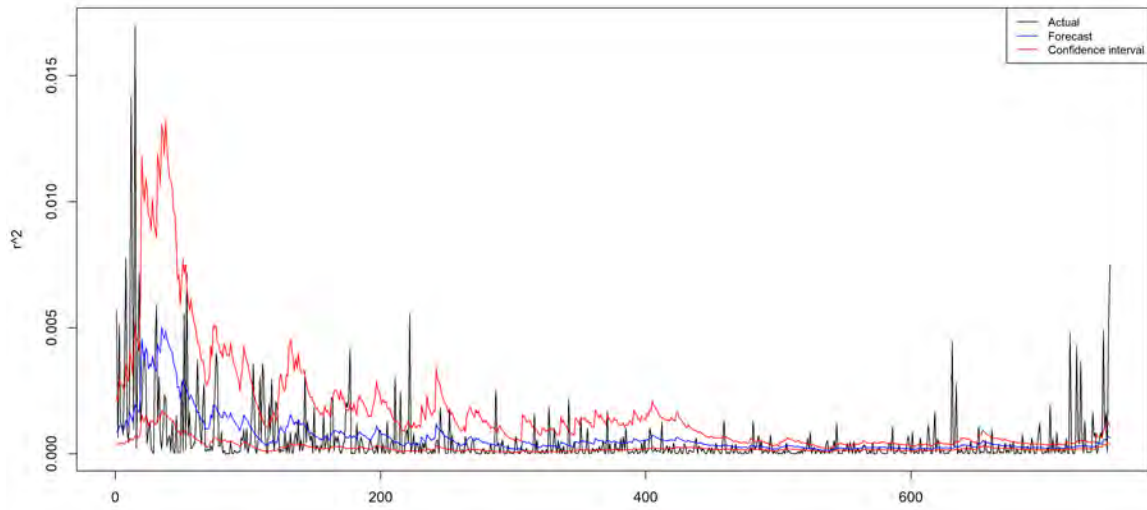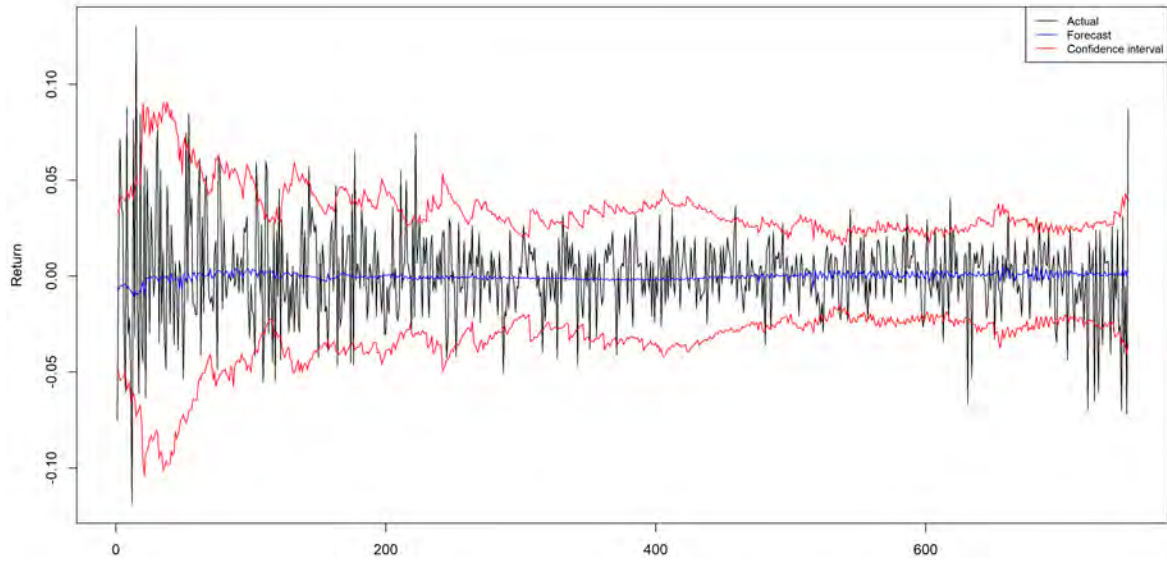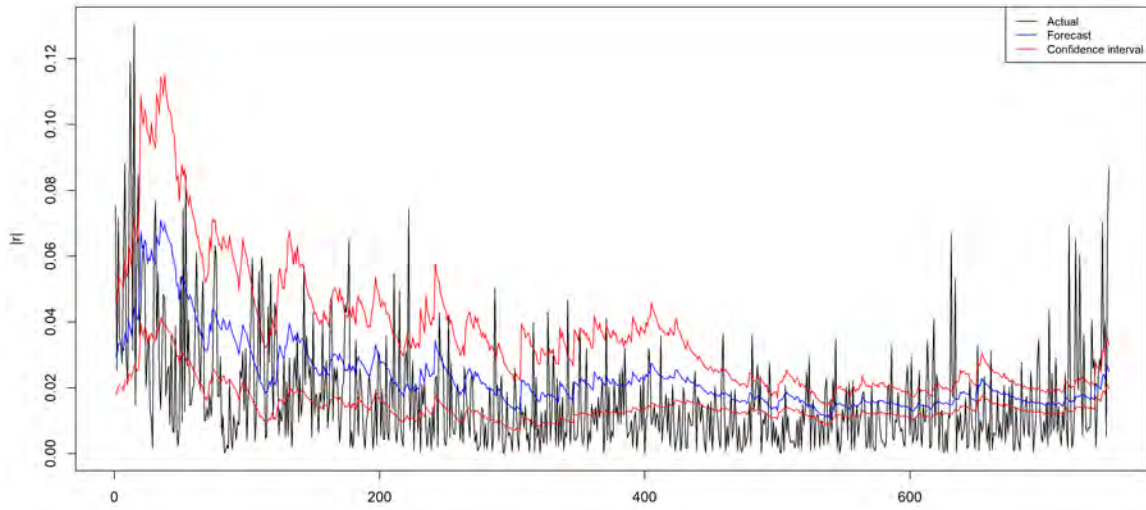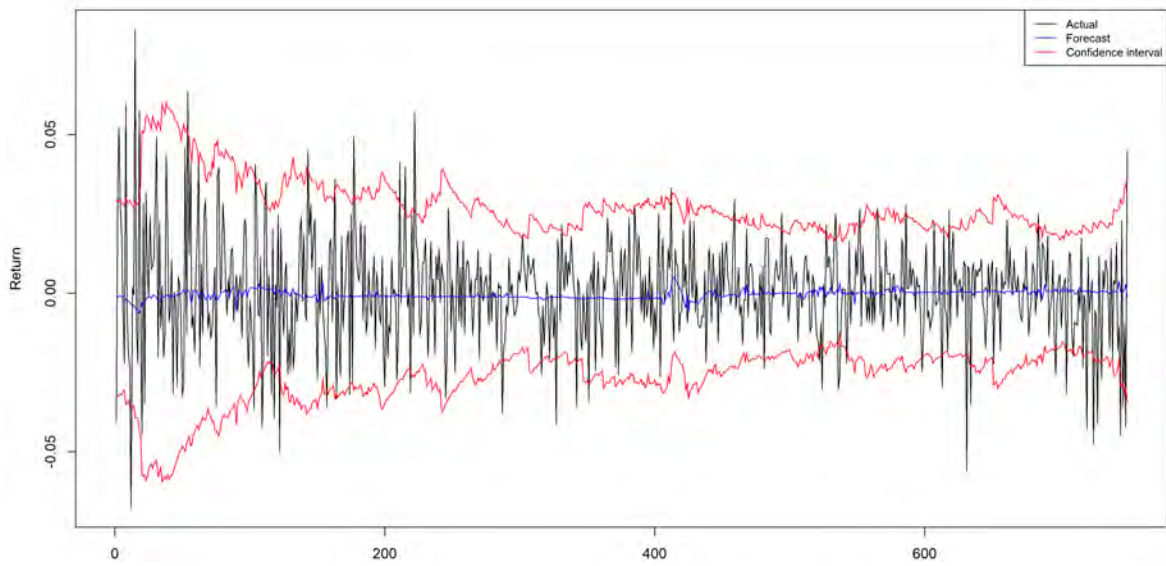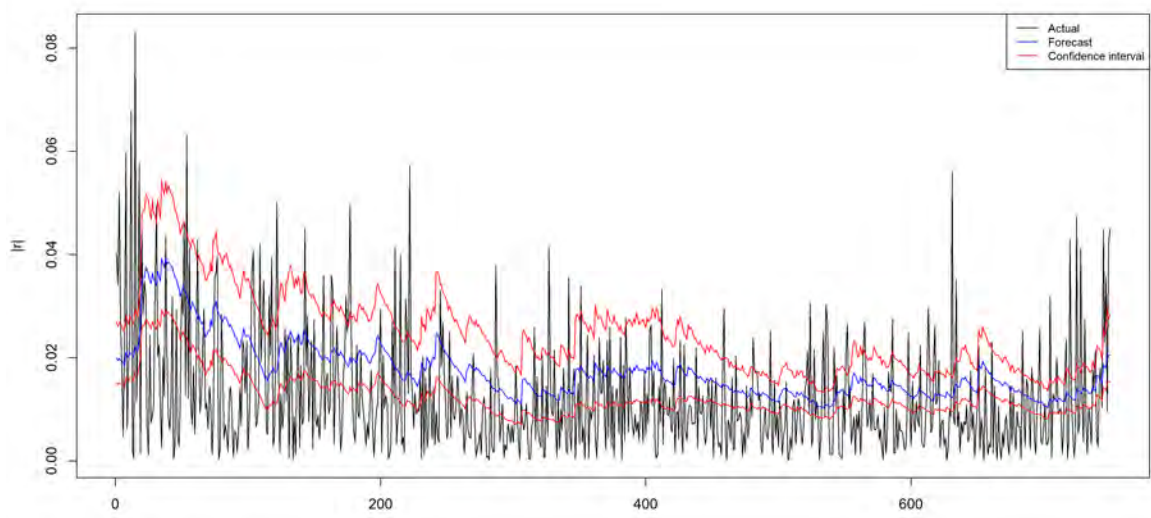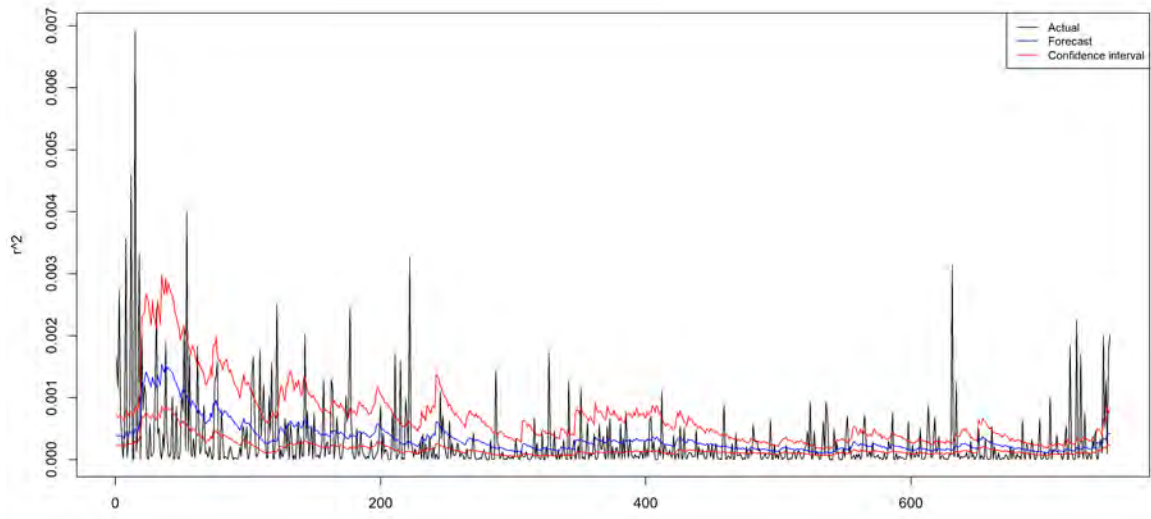
E.3   Naphtha
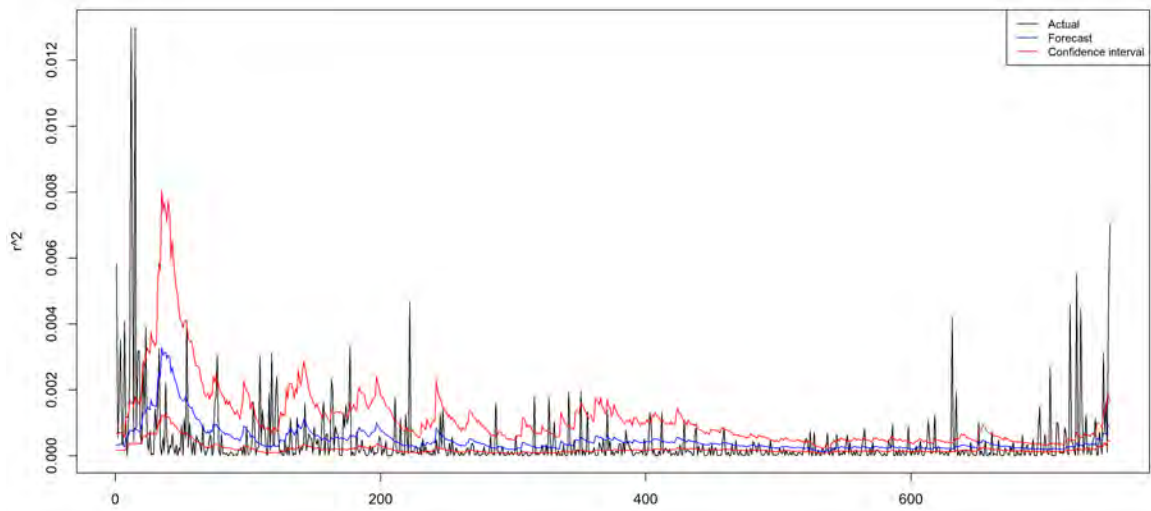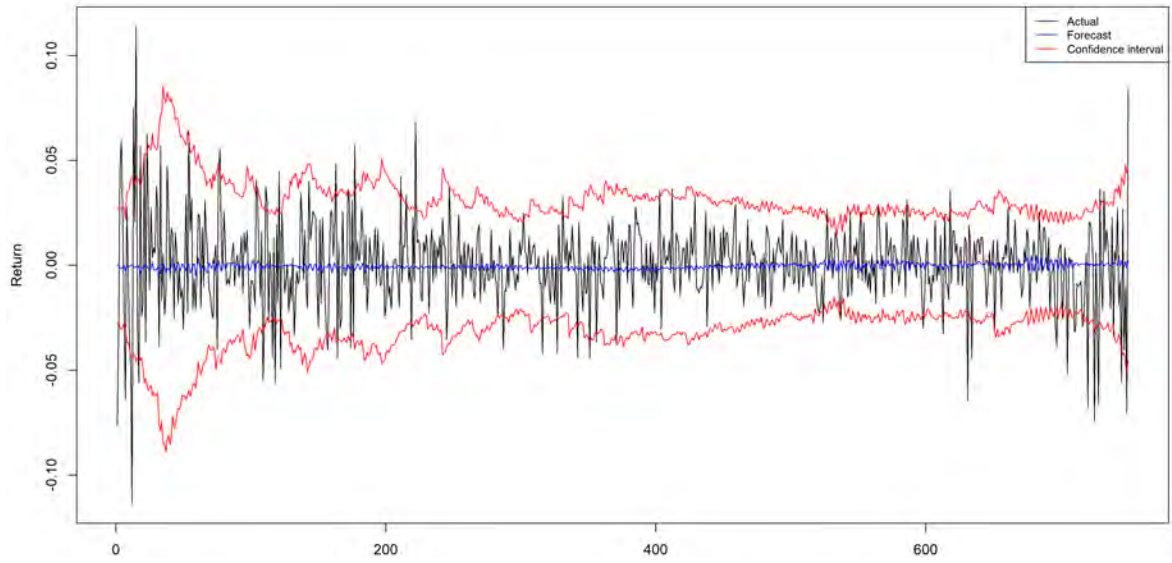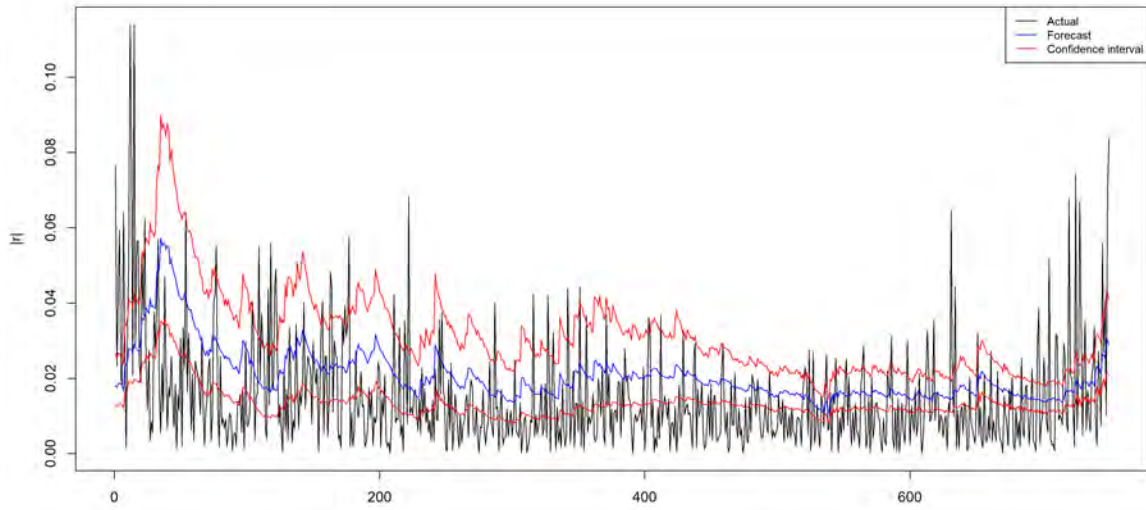
E.4 LSGO

E.5 Dated Brent

E.6 Margin