

Contents

1	Introduction	1
2	Data	2
2.1	Terminology	2
2.1.1	Autocorrelation Functions	2
2.1.2	Stationary Process	3
2.1.3	Related Distributions	3
2.2	Descriptive Statistics of Time Series	4
3	Methodology	6
3.1	ARMA-GARCH Models	6
3.1.1	The Autoregressive Moving Average Model	6
3.1.2	Generalized Autoregressive Conditionally Heteroscedastic Model	7
3.1.3	ARMA-GARCH Model	8
3.1.4	Estimation of ARMA-GARCH Model: Maximum Likelihood Estimator (MLE)	9
3.1.5	Identifying the Order of ARMA(p,q)-GARCH(P,Q) By Akaike Information Criterion (AIC)	10
3.1.6	Probabilistic Forecast for ARMA-GARCH Model	10
3.2	Multivariate modelling with copulas	11
3.2.1	Skalar's Theorem	11
3.2.2	Copula Densities	12
3.2.3	Independence copula	12
3.2.4	Normal copula	13
3.2.5	Student-t copula	13
3.2.6	Gumbel and Clayton copula	13
3.2.7	Independence Measure of copulas	14
3.2.8	Estimation of Copula-GARCH	16
3.3	Evaluation of Probabilistic Forecast	18
3.3.1	The Continuous Ranked Probability Score	18
3.3.2	Energy Score	19
4	Empirical Results	19
4.1	Computation- Packages in R software	19
4.2	Univariate Probabilistic Forecast with ARMA-GARCH Models	20
4.3	Multivariate Probabilistic Forecast with Copulas	25
5	Conclusion	29
Appendix		31

1 Introduction

Commodity prices play a crucial role in shaping global economies and influencing various sectors, including energy, manufacturing, agriculture, and finance. Forecasting and managing the volatility of commodity prices is important for investors and risk managers in the energy sector. Understanding the dynamics of natural gas, oil, and coal prices is crucial for making informed decisions and effectively managing investment risk.

In this report, we investigate the modeling and forecasting first differences of commodity price time series using the copula-GARCH models. We use the daily frequencies of the first differences commodity prices time series that starts from 2010-03-16 and ending in 2020-10-27. To make the probabilistic forecasts for the first differences of natural gas, oil and coal prices, we use the combination of two models namely the Autoregressive Moving Average (ARMA)[2] and the Generalized Autoregressive Conditional Heteroskedastic (GARCH)[2] in univariate case, then we use copula-GARCH[8] models to check the dependency between each time series natural gas, oil and coal, then we make forecast accordingly. We assume that the innovations are distributed with respect to the standard normal distribution. In order to find the optimal orders of the ARMA-GARCH process, we use the Akaike's Information Criterion[15] (AIC). We estimate the unknown parameters using the Maximum Likelihood Estimator (MLE)[2]. We start the fitting the model first 2500 observations for each commodity price differences and make the probabilistic forecast from 2501st to 2700th. We find that volatility clustering exists in first differences of the commodity prices time series. We compare the performance of the ARMA-GARCH model with Autoregressive (AR(1))[1] model at lag 1. The probabilistic forecasts are evaluated using Continuous Ranked Probability Score[15]. We find that ARMA-GARCH models are captured the volatility phases in the commodity time series and CRPS is obtained very close the CRPS of AR(1) forecast. Furthermore, we evaluate the investment risk by examining the Value at Risk (VaR)[6] at 0.05% quantile of the probabilistic forecasts. We find that the oil has less risk than other commodities.

Further, we use copula-GARCH models to find the dependency between time series natural gas, oil ,and coal. We implement the ARMA-GARCH models for the marginal distribution and the copula for the joint distributions. To obtain copula, we apply Probability Integral Transformation (PIT)[6] to the standardized residuals of the ARMA-GARCH models eliminating all the marginal information. Then we check the dependence structures of the PIT-transformed standardized residuals using rank correlation and tail dependence measures. We find that the PIT-transformed standardized residuals exhibit Student-t and Gaussian copula behaviours in bivariate setting. To find the appropriate Copula models, we fit the normal, the Student-t, the Gumbel, the Clayton and the independence copulas[8] to the PIT-transformed standardized residuals by maximum likelihood estimation. Using the Akaike's Information Criterion (AIC), we find the t-copula most suitable copula. Then we create one day-ahead multivariate probabilistic forecasts for the multivariate time series with the normal, the Student-t and the independence copula simulating the observations starts from 2501st to 2700th 1000 times 3-dimensional realizations from respective copula. Then we apply quantile transformation to obtain obtain probabilistic forecast. We evaluate the probabilistic forecast of each copula models using Energy Score[15]. We find that energy scores of the copula models are very close to each other but t-copula has the lowest energy score among of them.

The remaining paper is structured as follows. The following section presents a data description. Section 3, describes the overall methodology. Section 4 provides empirical

findings with discussion and an evaluation of the methods. Section 5 concludes the paper.

2 Data

The energy sector plays a pivotal role in global economic development, with natural gas, coal, and oil serving as major sources of energy worldwide. Understanding the dynamics of these energy commodities and their interrelationships is crucial for energy market participants, policymakers, and researchers. These resources are not only vital for meeting energy demands but also significantly impact environmental sustainability, geopolitical relations, and economic stability. In recent years, the prices of natural gas, coal, and oil have experienced considerable fluctuations, influenced by various factors such as geopolitical events, technological advancements, climate policies, and shifts in global energy demand. Analyzing the relationships between these commodities can provide valuable insights into the dynamics of the energy market and its potential implications.

2.1 Terminology

To understand the descriptive statistics of the time series for natural gas, oil, and coal prices, it is essential to familiarize ourselves with some key terminology.

2.1.1 Autocorrelation Functions

Autocorrelation function (ACF) [1] is a statistical tool used to measure the correlation between a time series and its lagged values. It quantifies the relationship between a time series and its lagged values. The ACF is calculated as :

$$\rho_s = \text{Corr}(X_t, X_{t-s}) = \frac{\text{Cov}(X_t, X_{t-s})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-s})}}$$

where ρ_s is the s^{th} autocorrelation of a series X_t ¹

Partial Autocorrelation function

The Partial Auto Correlation Function (PACF)[1] measures the correlation between two observations in a time series while accounting for the influence of the lagged observations. The formula for the PACF of a time series observation at lag h , denoted as $\text{PACF}(s)$, can be obtained through the following equation:

$$\text{PACF}(s) = \text{Corr}(X_t - \hat{X}_t, X_{t-s} - \hat{X}_{t-s})$$

where X_t and X_{t-s} are the observed values at time t and time $t - s$, respectively, and \hat{X}_t and \hat{X}_{t-s} are the corresponding predicted values based on a regression model.

In time series analysis, both PACF and the Autocorrelation Function (ACF) are crucial tools. The ACF measures the correlation between observations at different lags, providing insights into the presence of underlying patterns and dependencies in the data. It helps identify the seasonality and trend components in a time series. On the other hand, the PACF is particularly useful in identifying the specific lag orders that contribute significantly to the correlation structure of a time series.

¹see Appendix 1 for the detailed explanation to autocorrelation function and autocovariance function.

2.1.2 Stationary Process

A time series X_t is stationary [1] if its probability distribution does not change over time, that is, if the joint distribution does not depend on time t , regardless of the value of T ; otherwise, X_t is said to be non-stationary, Stock et al. (2020). If a time series is non-stationary, its statistical properties can change over time, making it difficult to make accurate forecasts. The stationary of a time series can be checked by looking at the plot or by using some statistical methods like Augmented Dickey-Fuller Test (ADF)[1] test. The augmented Dickey-Fuller (ADF) test is a commonly used statistical test to determine if a time series is stationary or exhibits unit roots. The ADF statistic is based on the augmented Dickey-Fuller regression model, which incorporates lagged differences of the dependent variable to capture the presence of unit roots. The null hypothesis of the test is that the time series contains a unit root, indicating non-stationarity, while the alternative hypothesis is that the time series is stationary.²

Strict Stationarity

The time series $(X_t, t \in N)$ is said to be strictly stationary[1] if the joint distribution of $(X_1, \dots, X_k)'$ is the same as that of $(X_{1+h}, \dots, X_{k+h})$, for any $k \in N$ any $h \in Z$. In other words, strict stationarity means that the joint distribution only depends on the “difference” h , not the time $t = 1, \dots, k$.

Second-order (Weak) Stationarity

The time series $(X_t, t \in Z)$ is said to be second-order (weak) stationary[2] if

- (i) $E(X_t^2) < \infty$, all $t \in Z$
- (ii) $E X_t = \mu$, all $t \in Z$
- (iii) $\text{Cov}(X_t, X_{t+h}) = \gamma_X(h)$, all $h \in Z$.

2.1.3 Related Distributions

Normal distribution

A normal (Gaussian) distribution[1] is a probability distribution that is widely used in statistics and probability theory. It is characterized by its bell-shaped curve, which is symmetric and centered around its mean. The shape of the curve is determined by two parameters: the mean (μ) and the standard deviation (σ). The probability density function of a normally distributed random variable (the normal probability density function (p.d.f.)) is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where e is the exponential function of x , the factor $\frac{1}{\sigma\sqrt{2\pi}}$ ensures that $P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$. The cumulative distribution function (cdf) of X is defined as:

$$F_{\mu,\sigma^2}(x) := P(X \leq x) = \int_{-\infty}^x f_{\mu,\sigma^2}(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Suppose X and Z are jointly normally distributed random variables. Then conditional distribution of Z given X is defined as $N(\mu_{Z|X}, \sigma_{Z|X}^2)$.

Chi-squared distribution

Let X_1, \dots, X_n be i.i.d standard normal random variables, then the random variable Y is

²see Appendix 2 for detailed explanation for the ADF test)

defined as:

$$Y = \sum_{i=1}^n X_i^2$$

Y has a chi-squared distribution[1] with n degrees of freedom. Because $E(X_i^2) = 1$ and $E(X_i^4) = 3$, $E(Y) = n$ and $Var(Y) = 2n$. The distribution is denoted as χ_m^2 .

Student t-distribution

Let X have a standard normal distribution and Y have a χ_m^2 distribution, X and Y are independently distributed, then random variable t

$$t = \frac{X}{\sqrt{W/m}}$$

has a Student t distribution[1] with m degrees of freedom. The distribution is denoted as t_m . If the degrees of freedom is ∞ , then t_∞ is standard normal distribution. Because, as the degrees of freedom increases, the t-distribution will get closer and closer to matching the standard normal distribution, until they are almost identical. The t-distribution gives more probability to observations in the tails of the distribution than the standard normal distribution, which is presented in figure 11. ³

Uniform distribution

The Uniform distribution[11] is a fundamental probability distribution that holds significant importance in statistics, as it is widely used for modelling random variables. It is a continuous probability distribution that deals with events that have an equally likely to occur. When a continuous random variable X follows a Uniform distribution, it is referred to as being uniformly distributed or having a rectangular distribution on the interval $[a, b]$. We write $X \sim U(a, b)$, if its probability density function equals $f(x) = \frac{1}{b-a}$, $x \in [a, b]$ and 0 elsewhere (Lovric 2011).

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & x < a \text{ or } x > b \end{cases}$$

Skewness: The skewness[1] describes how much a distribution deviates from symmetry. If the skewness is non-zero then the distribution is not symmetric.

Kurtosis : The kurtosis[1] of a distribution is a measure of how much mass is in its tails and therefore is a measure of how much of the variance of Y arises from extreme values (Stock, Watson et al., 2020). The kurtosis of a normally distributed random variable is 3, If it is more than 3 than we say that it has more mass in the tails than normal random variable. Then, it's called leptokurtic.

$$Skewness = \frac{E[(X_t - \mu_X)^3]}{\sigma_X^3}, \quad Kurtosis = \frac{E[(X_t - \mu_X)^4]}{\sigma_X^4}$$

2.2 Descriptive Statistics of Time Series

In this section, we examine the descriptive statistics of the time series. Prior to that, we conducted an assessment of the presence of missing values (NA) in each first-differenced asset. No missing values were found during this analysis. Table 1 presents the descriptive statistics for the time series of natural gas (NGas), Oil, and Coal prices, which consist of a total of 2700 observations in each time series. Since the time series comes with first differences, we do not expect non-stationarity in our time series. We confirm that by

³See Appendix 3 for the figure

applying the ADF test. The Augmented Dickey-Fuller (ADF) test results with a p-value of less than 0.01 suggest that all three time series are stationary, indicating they do not possess a unit root. The mean of the time series almost zero. Standard deviation of natural gas reported as 0.40. Since the the time series is scaled between -2 and 2 this number is not small it tells us there is some volatility exists, we confirm that by checking the left side of the figure 1 . We can see that natural gas time series has more volatility than other time series. All assets distribution have positive skewness[12], which indicates non-symmetric right skew of assets. Regarding kurtosis[12], all of the assets have positive values. This indicates the assets distributions have heavier tails than Normal distribution.

Table 1: Descriptive Statistic of N.Gas, oil and coal time series

Variable	Mean	SD	Skewness	Kurtosis	ADF
N.Gas	0.001211111	0.4049075	0.007792449	10.98039	< 0.01
Oil	-0.014096296	1.1983207	0.023061693	10.87502	< 0.01
Coal	-0.006185185	0.9346700	0.017987732	22.39216	< 0.01

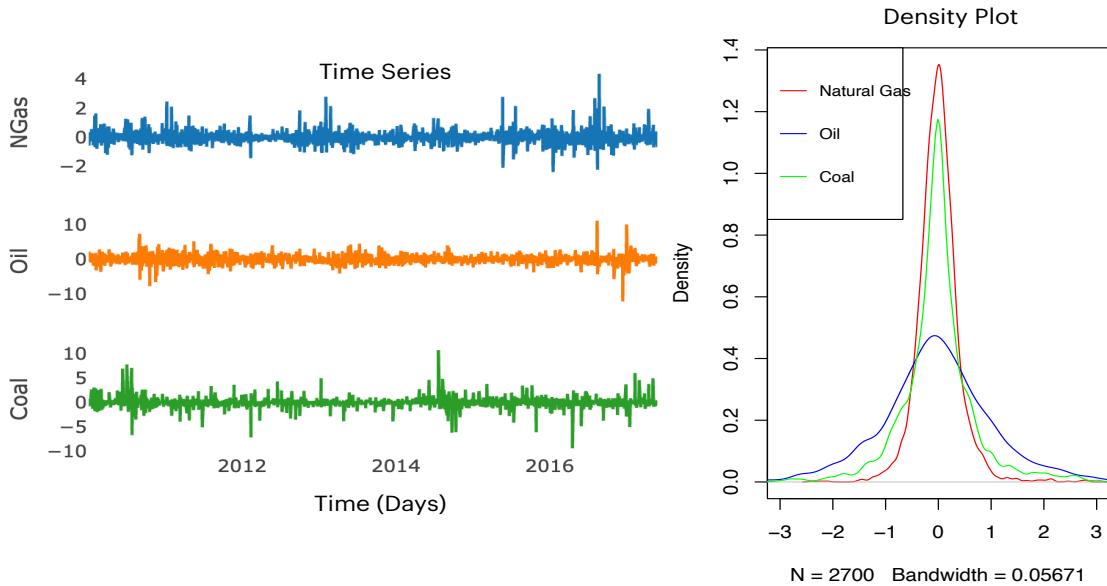


Figure 1: Plot of the time series(Right) and Density plots of the time series (Left) N.Gas, Oil and Coal.

As observed in Figure 2 and Figure 3, there are significant autocorrelations and partial autocorrelations present in both the first and second moments of the assets. Consequently, it is crucial to employ an appropriate model that can mitigate autocorrelations in both the first and second orders. To determine the suitable model orders, we analyse the ACF and PACF of the first moment to identify the order for the conditional mean, and the same approach applies to the squared assets to determine the order of the GARCH model, which addresses the conditional variance. However, it is important to note that optimising the order or finding the appropriate parameters for the model is not always straightforward by solely examining the ACF and PACF plots. In this report, we utilise the Akaike Information Criterion (AIC) to identify the optimal orders for ARMA and GARCH parts, as discussed in Section 3.

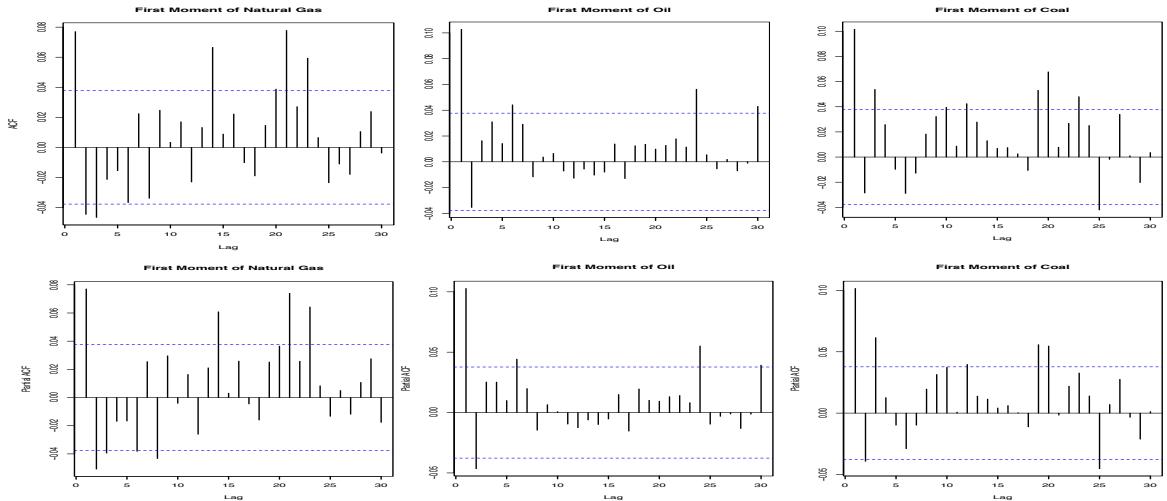


Figure 2: ACF and PACF plots of N.Gas, Oil and Coal time series. Upper plots are ACF and lower plots are PACF of first moment.

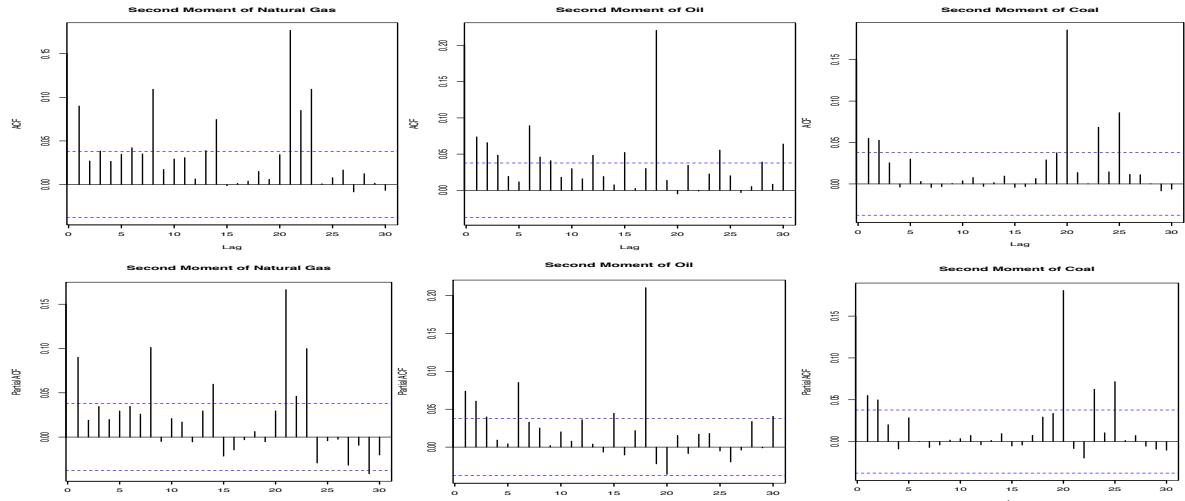


Figure 3: ACF and PACF plots of N.Gas, Oil and Coal time series. Upper plots are ACF and lower plots are PACF of second moment.

3 Methodology

In this section, we begin by introducing the Autoregressive Moving Average model (ARMA) and then proceed to the Generalized Autoregressive Conditional Heteroskedasticity model (GARCH). Following that, we introduce Dependency modelling with Copula. For each subsection, we provide a comprehensive explanation of the estimation and forecasting processes. Lastly, we present evaluation methods for the forecast in each of the models.

3.1 ARMA-GARCH Models

3.1.1 The Autoregressive Moving Average Model

The Autoregressive Moving Average (ARMA)[2] model is a combination of two time series models. These models are Autoregressive (AR) model which predicts the next value in the series based on past values, and the Moving Average (MA) model which predicts the next value based on previous residuals. We say X_t is ARMA(p, q) process if it is stationary

and there exist real coefficients c , a_1, \dots, a_p , b_1, \dots, b_q such that,

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j} \quad (3.1)$$

where p is the order of AR component and q is the order of MA component, ϵ_t is the white noise (Gaussian) of that $\epsilon_t \sim W.N(0, \sigma^2)$, $a_p \neq 0$ and $b_q \neq 0$ and the polynomials $1 - a_1 z - \dots - a_p z^p$ and $1 + b_1 z + \dots + b_q z^q$ have no common factors. The main attraction of this model, and the representations obtained by successively inverting the polynomials $a(\cdot)$ and $b(\cdot)$, is that it provides a framework for deriving the optimal linear predictions of the process.(Zakoian and Francq et al.2010).

ARMA(1, 0) or AR(1) Model

ARMA(1, 0) model or AR(1) model is a simple model that captures the linear relationship between an observation and its lagged value. It makes sense us to talk about AR(1) model since we are going to use in this report later on. AR(1) model equation is :

$$X_t = c + a_1 X_{t-1} + \epsilon_t$$

where a is the autoregressive parameter, which measures the impact of the lagged value, c is the constant mean and ϵ_t is white noise process. For a AR(1) stationary time series the coefficient has to hold that $a < 1$. This is because that $a = 1$ the variance $\text{VAR}(X_t) = \frac{\sigma_\epsilon^2}{1-a^2} = \infty$ which leads non-stationary time series.

3.1.2 Generalized Autoregressive Conditionally Heteroscedastic Model

The ARMA model can remove autocorrelation in the residuals . However, it can not remove the autocorrelation in the squared residuals caused by volatility clustering (Box et al. 2015). To deal the autocorrelation in the squared residuals we use Autoregressive Conditionally Heteroscedastic Model (ARCH) [2]which is introduced by Engle in 1982. The model is based on conditional variances which are dependent on the previous error values. The ARCH model equation is

$$\sigma^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2$$

where σ^2 is a function of constant term and previous squared errors, $\omega > 0$, $\alpha_i \geq i \in \{1, \dots, p\}$. The ARCH model is commonly used to address the issue of correlated squared residuals in time series analysis. This issue, the correlation in the squared residuals, is called ARCH effects and it can be checked using Lagrange multiplier (LM) test[2].To perform LM test, we first estimate the mean equation in (3.1), then extract the squared residuals $\hat{\epsilon}_t^2$ and regress $\hat{\epsilon}_t^2$ on lagged squared residuals :

$$\hat{\epsilon}_t^2 = a_0 + a_1 \hat{\epsilon}_{t-1}^2 + \dots + a_m \hat{\epsilon}_{t-m}^2 + \varepsilon_t$$

where ε_t is random error term. The null and alternative hypotheses are

$$H_0 : a_1 = \dots = a_m = 0 \text{ (No ARCH effects)}$$

$$H_1 : a_1 \neq \dots \neq a_m \neq 0 \text{ (there exists ARCH effects.)}$$

The test statistic for LM test is defined as $LM = n \times R^2$ where n is the sample size and R_ϵ^2 is the determination coefficient⁴. Under the null hypothesis, LM asymptotically follows chi-squared distribution with degrees of freedom $q \chi_q^2$. If $LM \geq \chi_q^2$ then we reject the null hypothesis and conclude that ARCH effects are present.

However, ARCH model has a limitation in that its lag structure tends to decline quickly. ARCH(p) models require a large p to account for the observed volatility clustering, and in order to capture the longer lag structure typically observed in empirical data, Bollerslev (1986) introduced the Generalised Autoregressive Conditional Heteroskedasticity (GARCH)[2] model. The GARCH model incorporates both the previous errors and previous variances in its conditional variance formulation. GARCH(p, q) model equation is defined as;

$$\begin{aligned}\sigma^2 &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \\ \epsilon_t &= \sqrt{\sigma_t^2} h_t \quad \text{and } h_t \stackrel{\text{iid}}{\sim} \mathcal{W.N}(0, 1)\end{aligned}$$

where ω is the long-run volatility with condition $\omega > 0$, $\alpha_i \geq 0$ which to ensure the conditional variance is non-negative, and $\beta_j \geq 0$ and $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ ensures that the unconditional variance of the error terms infinite and non-negative. The GARCH models have longer memory with smaller order. It can model the volatility clustering with less parameters.

3.1.3 ARMA-GARCH Model

By combining the two concepts, ARMA and GARCH, we obtain ARMA-GARCH[2] model that can be used for forecasting returns. The ARMA-GARCH model allows us to fit the conditional expectation of a time series with an ARMA model and the conditional variance with a GARCH model. ARMA(p, q)-GARCH(P, Q) model equation is defined as

$$\begin{aligned}X_t &= c + \sum_{i=1}^p a_i X_{t-i} + \sum_{j=1}^q b_j \epsilon_{t-j} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{W.N}(0, 1) \\ \epsilon_t &= \sqrt{\sigma_t^2} h_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^P \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2\end{aligned}$$

where the GARCH error parameter, α , measures the reaction of the conditional volatility. When α is relatively large, above 0.1, the volatility is very sensitive (see Figure 12⁵). The GARCH lag parameter, β , measures the persistence in conditional volatility, irrespective of anything happening in the market. When β is relatively large, above 0.9, the volatility takes a long time to die out as presented in Figure 12. $(\sum_{i=1}^P \alpha_i + \sum_{j=1}^Q \beta_j) < 1$ is necessary condition for stationary (strictly) solution and it determines the rate of convergence of the conditional volatility to the long-term average level. One assumption of the model is that, after fitting the model, there should be no serial correlation in the residuals. We assess this assumption by conducting the Ljung-Box test. The Ljung-Box test is presented in the Appendix.

⁴It means here the non-centered determination coefficient

⁵It presented in Appendix 4

3.1.4 Estimation of ARMA-GARCH Model: Maximum Likelihood Estimator (MLE)

The likelihood function represents the joint probability distribution of the observed data, with the unknown coefficients acting as parameters. By maximizing the likelihood function, we obtain the maximum likelihood estimator (MLE)[2], which provides the parameter values that maximize the probability of obtaining the observed data. In essence, the MLE selects the parameter values that are most likely to have generated the observed data. Let's assume X_1, \dots, X_t are generated by stationary ARMA(p,q)-GARCH(P;Q) process, then the model equations for MLE are defined as

$$\begin{aligned} X_t - c_0 &= \sum_{i=1}^p a_{0i}(X_{t-i} - c_0) - \sum_{j=1}^q b_{0j}\epsilon_{t-j} + \epsilon_t \\ \epsilon_t &= \sqrt{\sigma_t} h_t \\ \sigma_t^2 &= \omega_0 + \sum_{i=1}^O \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^P \beta_{0j}\sigma_{t-j}^2 \end{aligned}$$

where h_t is a sequence of independent and identically distributed random variables with $E(h_t^2 = 1)$, $\omega_0 > 0$, $\alpha_{0i} \geq 0$ and $\beta_{0j} \geq 0$.

$$\begin{aligned} \theta &= (\theta_1, \dots, \theta_{P+Q+1})^T = (\omega, \alpha_1, \dots, \alpha_Q, \beta_1, \dots, \beta_P)^T \\ \vartheta &= (\vartheta_1, \dots, \vartheta_{p+q+1})^T = (c, a_1, \dots, a_p, b_1, \dots, b_q)^T \end{aligned}$$

where θ is the parameter vector for GARCH(P, Q) model and ϑ is the parameter vector from ARMA(p, q) model. The parameter vector for ARMA(p, q)-GARCH(P, Q) is denoted as $\varphi = (\vartheta^T, \theta^T)^T$ and the true value is $\varphi_0 = (\vartheta_0^T, \theta_0^T)^T$ where,

$$\varphi_0 = (\vartheta_0^T, \theta_0^T)^T = (c_0, a_{01}, \dots, a_{0p}, b_{01}, \dots, b_{0q}, \omega_0, \alpha_{01}, \dots, \alpha_{0Q}, \beta_{01}, \dots, \beta_{0P})^T$$

The parameter space is $\Phi \subset \mathbb{R}^{p+q+1} \times (0, +\infty) \times [0, \infty)^{P+Q}$. If $P \geq q$, then the initial values are

$$X_0, \dots, X_{1-(P-q)}, \tilde{\epsilon}_{-P+q}, \dots, \tilde{\epsilon}_{1-P}, \tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{1-Q}^2$$

if $P < q$, then the initial values are

$$X_0, \dots, X_{1-(P-q)}, \tilde{\epsilon}_0, \dots, \epsilon_{1-P}, \tilde{\sigma}_0^2, \dots, \tilde{\sigma}_{1-Q}^2$$

where $\tilde{\epsilon}_t$ and $\tilde{\sigma}_t^2$ can be computed from

$$\begin{aligned} \tilde{\epsilon}_t &= X_t - c - \sum_{i=1}^p a_i(X_{t-i} - c) + \sum_{j=1}^q b_j\tilde{\epsilon}_{t-j} \\ \tilde{\sigma}_t^2 &= \omega + \sum_{i=1}^P \alpha_i\tilde{\epsilon}_{t-i}^2 + \sum_{j=1}^Q \beta_j\tilde{\sigma}_{t-j}^2 \end{aligned}$$

Then the normal log-likelihood is:

$$\tilde{I}_n(\varphi) = n^{-1} \sum_{t=1}^n \tilde{l}_t \quad \text{where} \quad \tilde{l}_t = \tilde{l}_t(\phi) = \frac{\tilde{\epsilon}_t^2(\vartheta)}{\tilde{\sigma}_t^2(\varphi)} + \log \tilde{\sigma}_t^2(\varphi)$$

We can insert the equations $\tilde{\epsilon}_t$ and $\tilde{\sigma}_t^2$ in $\tilde{I}_n(\varphi)$, then normal likelihood equation is

$$\tilde{I}_n(\varphi) = n^{-1} \sum_{t=1}^n \frac{(X_t - c - \sum_{i=1}^P a_i(X_{t-i} - c) + \sum_{j=1}^Q b_j \tilde{\epsilon}_{t-j})^2}{\omega + \sum_{i=1}^P \alpha_i \tilde{\epsilon}_{t-i}^2 + \sum_{j=1}^Q \beta_j \tilde{\sigma}_{t-j}^2} + \log(\omega + \sum_{i=1}^P \alpha_i \tilde{\epsilon}_{t-i}^2 + \sum_{j=1}^Q \beta_j \tilde{\sigma}_{t-j}^2)$$

$$\omega_i > 0, \alpha_i \text{ & } \beta_j \geq 0 \text{ and } \sum_{i=1}^P \alpha_i + \sum_{j=1}^Q \beta_j < 1$$

The MLE solution for φ is

$$\hat{\varphi}_n = \arg_{\varphi \in \Phi} \min \tilde{I}_n(\varphi).$$

This equation means that we take the first derivative of the equation for all unknown parameters.

$$\hat{\varphi}_n = \frac{\partial \tilde{I}_n(\varphi)}{\partial \varphi}$$

where φ is the parameter vector that we defined before. By solving the equation above we obtain the estimated parameters that can be used for the forecasting process. We see the detailed explanation in copula fitting process.

Estimation of AR(1) model

Same as ARMA-GARCH model we also use MLE to estimate the unknown parameters in AR(1) model. If we have normality assumption then we can write the likelihood as :

$$L(\varphi) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_t - \mu)^2}{2\sigma^2}\right)$$

where $\varphi = (c, a_1, \sigma^2)$ is parameter vector. We obtain the MLE $\hat{\varphi}$ by following the same procedure in ARMA-GARCH process.

3.1.5 Identifying the Order of ARMA(p,q)-GARCH(P,Q) By Akaike Information Criterion (AIC)

To find the optimal parameters for the model we use Akaike Information Criterion (AIC)[15]. The Akaike Information Criterion (AIC) is criteria used in model selection to discourage overfitting. As more parameters are added to a model its likelihood will improve, however if too many parameters are added the model will begin to fit noise. AIC penalise the increase in log likelihood by the number of parameters added. The formula is:

$$AIC = 2k - 2\tilde{l}_t \quad (3.2)$$

where k is the number of fitted parameters in the model and \tilde{l}_t is the log-likelihood of the fitted model(see section 3.1.4 for the definition of \tilde{l}_t). When fitting using AIC, the model with the lowest AIC is chosen. In this report we used the formula above to find the AIC scores of each parameter combination.

3.1.6 Probabilistic Forecast for ARMA-GARCH Model

Probabilistic forecasting is a valuable approach that incorporates uncertainty into forecasts by providing a probability distribution of possible future values. Instead of relying

solely on point estimates, probabilistic forecasting offers a range of potential outcomes along with associated probabilities or confidence levels. This enables decision-makers to make more informed choices, assess risks, and develop robust strategies. It gives a measure of the range of possible outcomes, capturing the inherent uncertainty in the forecast. To get the probabilistic forecast, we use the fitted ARMA-GARCH model which is helping us to obtain the forecasts for the mean \hat{X}_t and the volatility $\hat{\sigma}_t^2$. One-step-ahead probabilistic forecast for ARMA(p,q)-GARCH(P,Q)[2] model is given by

$$\begin{aligned}\hat{X}_{T+1|T} &= \hat{c} + \hat{a}_1 X_T + \dots + \hat{a}_p X_{T-p} + \hat{b}_1 \epsilon_T + \hat{b}_q \epsilon_{T-q} \\ \hat{\sigma}_{T+1|T}^2 &= \hat{\omega} + \alpha_1 \epsilon_T^2 + \dots + \alpha_P \epsilon_{T-P}^2 + \hat{\beta}_1 \sigma_T^2 + \dots + \hat{\beta}_Q \sigma_{T-Q}^2\end{aligned}$$

where the estimates ϵ_T and σ_T^2 are calculated iteratively with the initial values that are described in section 3.1.6.

Value at Risk

Value at Risk[6] (VaR) is a risk measure used in finance to estimate the potential maximum loss that a portfolio or investment may experience within a specified time horizon and at a specific confidence level under normal market conditions. It represents the amount of loss that is not expected to be exceeded with a certain probability. In simpler terms, VaR quantifies the worst-case scenario, in terms of potential financial loss, that an investor or portfolio manager should be prepared for within a given level of confidence and time frame, assuming normal market conditions. In the context of ARMA-GARCH and copula (section 3.2) modeling, we can utilize the concept of Value at Risk (VaR) to assess the level of risk associated with investing in an asset. To do that, we first obtain the probabilistic forecast from the ARMA-GARCH (or copula) model, then we can calculate the quantile corresponding to a desired confidence level to evaluate the potential downside risk. Which means that what is the amount or percentage of potential loss that an investment could experience. An effective model for this metric should have the 5% quantile closely aligned with actual outcomes. Once the VaR is determined, it allows us to estimate the expected loss in a portfolio once it falls below the VaR threshold. However, accurately modelling the tail of the distribution, where extreme events occur, is challenging due to the limited data available in that region. Despite this challenge, obtaining a reliable approximation of the distribution's tail shape is crucial for estimating portfolio losses beyond the VaR level.

3.2 Multivariate modelling with copulas

Copulas[8] are tools for modelling dependence of several random variables. The main purpose of copulas is to describe the interrelation of several random variables (Schmidt et al. 2007). The term comes from the latin word *copulare* which means "to connect" or "to join". It was first used by *Skalar* (1959). Copulas allow us to construct the dependency structure in a multivariate distribution.

3.2.1 Skalar's Theorem

According to Skalar's theorem[8], if $G(X_1, \dots, X_n)$ is the joint distribution function of random variables X_1, \dots, X_n with marginal distribution functions F_{X_1}, \dots, F_{X_n} , then there exist some copula C that for all (x_1, \dots, x_n) in Cartesian product $R_1 \times \dots \times R_n$

$$G(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (3.3)$$

if F_i is continuous for all $i = 1, \dots, d$ then C (Copula) is unique. Let $F_{X_1}^{-1}, \dots, F_{X_n}^{-1}$ be the inverse functions of F_{X_1}, \dots, F_{X_n} and $F_{X_i}(X_i) \sim U_i$ for $i = 1, \dots, n$ then U_i gives the quantiles of X_i and is uniformly distributed on $[0, 1]$, transforming a distribution function to quantiles is called **Probability Integral Transformation**. And transforming those quantiles to another continuous distribution we obtain a copula.

$$C(u) = G(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad (3.4)$$

The equation 3.4 allows us to extract a copula from multivariate distribution function. Let U_1, \dots, U_n be n random variables on $[0, 1]^n$. An n -dimensional copula, C , is a continuous, non-decreasing function that maps $[0, 1]^n$ to $[0, 1]$, with the properties:

- If any u_i ($i = 1, \dots, n$) is zero, then $C(u_1, \dots, u_n) = 0$
- $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for any $j = 1, \dots, n$
- $C(u, v) = u \times v$ for independent case.

Lets u_1, \dots, u_n are the transformed quantiles from marginal distributions. Then we can transform them into standart normal densities. We assume the joint model is multivariate normal with covariance matrix Σ . Then, using formula 3.4, we obtain a normal copula that holds tha joint dependency for model in Σ , so :

$$C(u_1, \dots, u_n) = \Phi(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n); \Sigma)$$

3.2.2 Copula Densities

A copula is a cumulative distribution function which is monotonically increasing. If the copula is sufficiently differentiable then the density[8] is :

$$c(u) := \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d}$$

We obtain the copula density in terms of the joint density with marginal cdfs if the copula is in form of equation 3.4. Denoting the joint density by f and marginal densities by f_i we obtain the copula density as :

$$c(u) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \dots f_d(F_d^{-1}(u_d))}$$

3.2.3 Independence copula

According the Sklar's theorem the random variables are independent if and only if their copula is the independence copula. Therefore, the copula density is a constant. Then the independence copula[8] is :

$$C(u) = \prod_{i=1}^d u_i$$

3.2.4 Normal copula

Lets assume that X_1 and X_2 are normal distributed random variables which are also jointly normal. Then their correlation

$$Corr(X_1, X_2) := \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} \quad (3.5)$$

fully describes the dependence structure. By following the equation 3.4, we obtain two-dimensional Normal (Gaussian) copula[8] which is defined as

$$C_{\rho}^{Ga}(u_1, u_2) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$

where Σ is the 2 x 2 correlation matrix with 1 on the diagonal and ρ otherwise, Φ is cdf of a standart normal distribution and Φ_{Σ} is the cdf of a bivariate normal distribution. This is equivalent to

$$C(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) ds dt$$

3.2.5 Student-t copula

The multivariate Student's t copula[8] is defined based on a symmetric, positive definite matrix Σ with diagonal elements equal to 1. It is used to model the joint distribution of correlated variables with a multivariate Student's t distribution. The multivariate Student's t distribution is a generalization of the multivariate normal distribution, allowing for heavier tails and capturing more extreme observations. It is characterized by the degrees of freedom parameter v and the correlation matrix Σ . t-copula is defined as

$$C_{v,\Sigma}^t(u) = t_{v,\Sigma}(t_{v,\Sigma}^{-1}(u_1), \dots, t_{v,\Sigma}^{-1}(u_d))$$

In the bivariate case, the copula expression can be written as:

$$C(u, v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 - 2\rho st + t^2}{v(1-\rho^2)}\right)^{-(v+2)/2} ds dt$$

From figure XX, we are able to find that both copulas are symmetric, so the dependence structure modelled by them is symmetric which means that extremes in both directions will happen with the same probability. However, being similar to t distribution, the density of t copula in both tails is heavier than gaussian copula, so variables with t copula are easier to fall into extremes than gaussian copula

3.2.6 Gumbel and Clayton copula

The bivariate Gumbel copula[8] has the following form:

$$C_{\theta}^{Gu}(u_1, u_2) = \exp[-((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta})^{\frac{1}{\theta}}]$$

where $\theta \in [1, \infty)$. For $\theta = 1$ we have the independence copula. The Gumbel copula has tail dependence in one corner.

The bivariate Clayton copula[8]has the following form:

$$C_{\theta}^{Cl}(u_1, u_2) = (\max(u_1^{-\theta} + u_2^{-\theta} - 1, 0))^{-\frac{1}{\theta}}$$

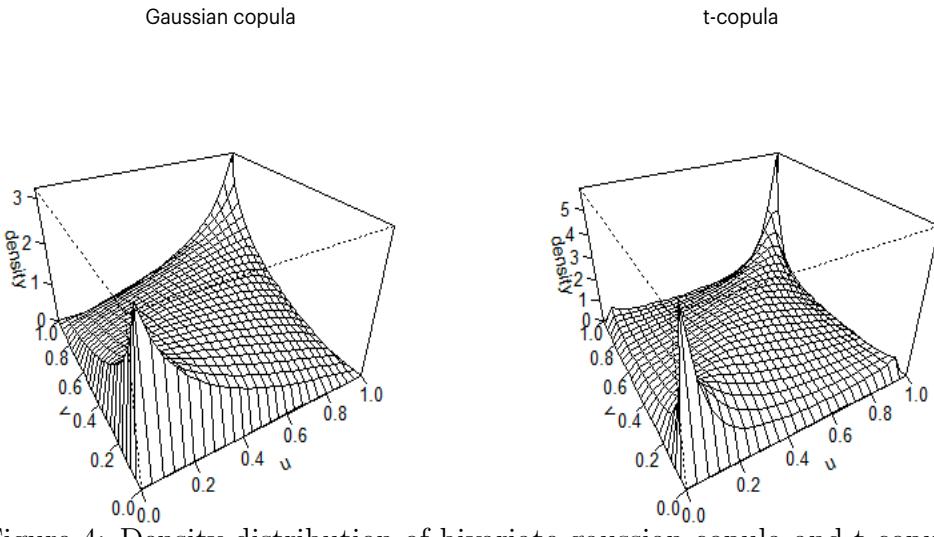


Figure 4: Density distribution of bivariate gaussian copula and t-copula

where $\theta \in [-1, \infty)/\{0\}$. For the limits θ goes zero that we obtain independence copula. Figure 5 shows the density distribution of Gumbel, Clayton copula respectively. We can see that Gumbel copula has heavier upper tail, while Clayton copula has heavier lower tail, so they can be used in modelling asymmetric dependence structure, thus they and their extensions are extremely useful in characterising tail dependence.

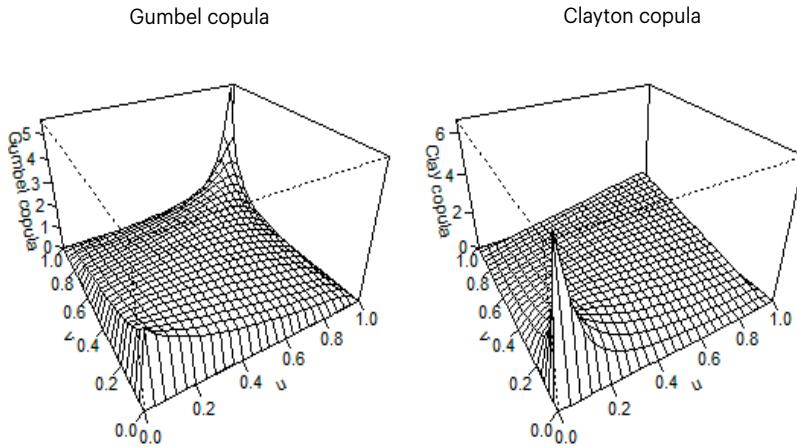


Figure 5: Density distribution of bivariate Gumbel copula and Clayton copula

3.2.7 Independence Measure of copulas

In this section, we talk about two important measure of copulas. First, we will explain the rank correlation, then the tail dependence. We skip the linear correlation since the linear correlation is explained in section 2 and the linear correlation is only suitable measure in a special class of distribution, i.e. elliptical distribution (Schmidt et al. 2006), where the normal distribution is also a family of the elliptical distribution. As an example, we can say that, where there is 0 correlation between random variables, then we say that there is independence. But this is only valid for normal distribution. Because for

Student t-distributed random variables, 0 correlation does not mean independence. As we introduced in definition of the Student t-distribution in section 2, that W has some dependence even the correlation of X 's are zero.

Rank correlation of copulas

The rank correlation[8] has paved the way for the development of prominent correlation measures, namely Kendall's tau and Spearman's rho, which exhibit a direct connection to copulas. By adopting a rank-based approach, these correlation measures provide estimates that are invariant to changes in scale. This quality proves to be highly advantageous when dealing with copulas. Consequently, employing rank correlations offers a promising avenue for fitting copulas to data, as it enables the consideration of inter-variable relationships without being influenced by the specific scales or distributions of the individual variables. Spearman's ρ is defined as: $\rho_s := \text{Corr}(F_1(X_1), \dots, F_d(X_d))$ where F_1 and F_2 are marginals of X_1 and X_2 random variables. In multivariate case Corr refers to correlation matrix, and in bivariate case it is the correlation of the uniform distribution $F_1(X_1)$ and $F_2(X_2)$ which gives us the correlation of the ranks itself.

Kendall's τ is defined as: $\rho_\tau(X) := \text{Cov}[\text{sign}(X - \tilde{X})]$ where X is d-dimensional random variables and \tilde{X} is independent copy⁶ of X . For bivariate case we can define ρ_τ as :

$$\begin{aligned}\rho_\tau &= E[\text{sign}((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2))] \text{ which can be written as} \\ &= P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0) - P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0)\end{aligned}$$

If the both probabilities are the same (which means $\rho_\tau = 0$) then we expect upward slopes with the same probability as downward slopes.

Tail dependence of copulas

In this section, we explain the tail dependence[8] of copula, distinguishing between upper and lower tail dependence. We examine the behaviour of copula densities, such as those of the Gaussian and Student t-copula, in the lower left and upper right corners, as depicted in figure 4. Upper tail dependence refers to the scenario where large values of one random variable (U_1) tend to be accompanied by large values of another variable (U_2). Specifically, we analyse the probability that U_1 exceeds a given threshold (q) when U_2 has already surpassed the same threshold (q), as q approaches to one. If this conditional probability decreases at a rate slower than q , it indicates that there is no tail dependence. Let X_1 and X_2 are random variables with cumulative distribution functions (cdf) $F_i, i = 1, 2$. Then the coefficients of lower and upper tail dependence are defined as follows.

$$\lambda_u := \lim_{q \rightarrow 1^-} P(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q))$$

$$\lambda_l := \lim_{q \rightarrow 0^+} P(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q))$$

where F_2^{-1} is the quantile function of X_2 , $\lambda_u \in [0, 1]$ is the coefficient for upper tail dependence, $\lambda_l \in [0, 1]$ is the coefficient for lower tail dependence. If $\lambda_u > 0$, then X_1 and X_2 have upper tail dependence, while for $\lambda = 0$, then X_1 and X_2 are asymptotically independent in the upper tail and analogously for λ_l . For continuous cumulative distribution functions:

$$\begin{aligned}\lambda_l &:= \lim_{q \rightarrow 0^+} \frac{P(X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q))}{P(X_1 \leq F_1^{-1}(q))} \\ &= \lim_{q \rightarrow 0^+} \frac{C(q, q)}{q} \quad \text{and} \quad \lambda_u = 2 + \lim_{q \rightarrow 0^+} \frac{C(1-q, 1-q) - 1}{q}\end{aligned}$$

⁶It means that it has same distribution as X , but being independent of X .

The tail dependence helps us to gain a deeper understanding of the relationship between extreme events within copulas. It provides valuable insights into the behavior of variables in extreme scenarios, which is crucial for risk analysis and modeling.

3.2.8 Estimation of Copula-GARCH

The inference function for margins method (IFM)

The process of fitting a copula to data involves estimating both the dependence parameters and the parameters for the univariate marginal distributions. Estimating copula parameters can be challenging as they often do not have simple analytic formulas for their estimators, requiring optimisation procedures for fitting. The Inference Function for Margins (IFM)[6] method, proposed by Joe and Xu (1996), provides an alternative approach for fitting copulas. It is particularly well-suited for copulas because it allows for the estimation of joint dependency parameters while fixing the parameters in the univariate margins. The IFM method proceeds by first fitting the univariate marginal distributions, which determines the parameters in the marginal distributions. Then, the joint likelihood is fitted, providing estimators for the joint dependency parameters. Under certain regularity conditions, the IFM estimators are asymptotically normal and unbiased. Xu's simulation studies have shown that the IFM estimators can be almost as efficient as maximum likelihood estimators in many cases. The advantage of the IFM method is that it offers computational tractability by fitting a reduced number of parameters in stages, making it feasible to solve complex problems. In this thesis, the IFM method is followed by fitting the marginal distributions using maximum likelihood and then fitting various copulas with different dependence structures to these fixed margins. Without the IFM method, it would be necessary to re-fit the marginal distributions for each copula model. In summary, the IFM method provides a practical and computationally efficient approach for fitting copulas by separately estimating the univariate marginal distributions and the joint dependency parameters. It strikes a balance between computational tractability and estimation efficiency, making it a valuable tool in copula modelling. Now we can define the model equations and use IFM (2-step Maximum likelihood method)[8] in bivariate case.

Let $F_{XY|W}(x_t, y_t; \theta|w_{t-1})$ be the conditional joint distribution of rvs X and Y, the conditional marginal distributions are $F_{X|W}(x_t; \theta|w_{t-1})$ and $F_{Y|W}(y_t; \gamma|w_{t-1})$, and $C(u_t, v_t; \kappa|w_{t-1})$ is the copula. According to Skalar's theorem :

$$F_{XY|W}(x_t, y_t; \theta|w_{t-1}) = C(F_{X|W}(x_t; \theta|w_{t-1}), F_{Y|W}(y_t; \gamma|w_{t-1}); \kappa|w_{t-1})$$

where θ is the set of all parameters of both marginal distributions and copula to be estimated, which is $[\varphi', \gamma', \kappa']$ and w_{t-1} is the information set until time $t - 1$. Since the joint bivariate density $f_{XY|W(x,y|w)}$ is equivalent to twice differentiation the conditional joint distribution , then we can obtain copula density as follows:

$$\begin{aligned} f_{XY|W(x,y|w)} &\equiv \frac{\partial F_{XY|W}(x, y|w)}{\partial x \partial y} \\ &= \frac{\partial F_{X|W}(x, |w)}{\partial x} \times \frac{\partial F_{Y|W}(y|w)}{\partial y} \times \frac{\partial^2 C(F_{X|W}(x|w), F_{Y|W}(y|w))}{\partial u \partial v} \\ &= f_{X|W}(x|w) \times f_{Y|W}(y|w) \times c(u, v|w) \end{aligned} \tag{3.6}$$

where $u = F_{X|W}(x|w)$ and $v = F_{Y|W}(y|w)$. Then, we obtain copula density as

$$c(u, v|w) = \frac{f_{XY|W}(x, y|w)}{f_{X|W}(x|w) \times f_{Y|W}(y|w)}$$

By taking a logarithm in equation 3.6, we can implement the maximum likelihood estimation, and obtain the log-likelihood function as $L_{XY} = L_X + L_Y + L_C$ which is

$$\begin{aligned} L_{XY}(\theta) &= L_X(\varphi) + L_Y(\gamma) + L_C(\kappa) \\ &= \sum_{t=1}^T \log f_X(x_t; \varphi) + \sum_{t=1}^T \log f_Y(y_t; \gamma) + \sum_{t=1}^T \log c(F_X(x_t; \varphi), F_Y(y_t; \gamma); \kappa) \end{aligned}$$

where, $\varphi \equiv \text{int}(\varphi) \subseteq \Re, \gamma \in \text{int}(K) \subseteq \Re^r$. Then we can estimate the unknown parameters vector $\theta \equiv [\varphi', \gamma', \kappa']' \in \text{int}(\Phi) \times \text{int}(\Gamma) \times \text{int}(K) \subseteq \Re^{p+q+r} \equiv \Re^s$ ⁷, using IMF method as:

- Parameters φ and γ of univariate marginal distributions are estimated in the first stage as :

$$\begin{aligned} \hat{\varphi} &= \arg_{\varphi \in \Phi} \max \sum_{t=1}^T \log f_X(x_t; \varphi) \\ \hat{\gamma} &= \arg_{\gamma \in \Gamma} \max \sum_{t=1}^T \log f_Y(y_t; \gamma) \end{aligned}$$

- Then, copula parameters κ can be estimated as :

$$\hat{\kappa} = \arg_{\kappa \in K} \max \sum_{t=1}^T \log c(F_X(x_t; \hat{\varphi}), F_Y(y_t; \hat{\kappa}))$$

The IFM estimator is defiend as :

$$\hat{\theta} = [\hat{\varphi}', \hat{\gamma}', \hat{\kappa}]'$$

Copula selection

The choice of copula is very important. To find the suitable copula we use Akaike's Information Criterion (AIC)[6], which is defined as :

$$AIC(M) = -\frac{2LL}{n} + \frac{2M}{n}$$

where M is the number of parameters being estimated and LL is the value of maximum likelihood function. This approach is helpful to find the appropriate copula, but it is not able to provide any understanding about the power of the decision rule employed.

Copula simulation

Copula simulation aims to produce deterministic outcomes with increased precision, while maintaining reproducibility. The following algorithm describes the process of simulating ARMA-GARCH model with desired copula C:

- a) Fit ARMA-GARCH model and extract the standardised residuals. (See section 3.1.3)

⁷int(.) presents the interior of a set

- b) Apply PIT to standardised residuals.
- c) Fit the desired copula to PIT residuals from step b).
- d) Run the simulation for n times and using the copula coefficients from step c).
- e) Obtain simulated standardised residuals by using the inverse function(quantile) of estimated marginals in step d.
- f) Use the standardised residuals from step e) , and the forecasted means and variances from step a) to get the simulated asset.
- g) Repeat the steps d)-f) for N times.

This process can be extended to get VaR score to analyse the investment risk for each commodity.

3.3 Evaluation of Probabilistic Forecast

The probabilistic forecast can be evaluated by applying scoring rules. Scoring Rules provide a measure for the evaluation of probabilistic forecasts by assigning a numerical score based on the predictive distribution. We will be using two scoring rules to evaluate the probabilistic forecast in this report.

3.3.1 The Continuous Ranked Probability Score

In certain forecasting scenarios, such as quantitative precipitation forecasts, the distributions involved may have point masses at zero or may be expressed as samples originating from techniques like Markov chain Monte Carlo. In such cases, defining scoring rules directly in terms of predictive Cumulative distribution functions is more practical. The continuous ranked probability score (CRPS)[12] is a scoring rule that evaluates the accuracy of probabilistic forecasts. The CRPS takes into account the entire predictive distribution and penalizes forecasts that assign low probabilities to observed values or assign high probabilities to values near but not identical to the observed outcome. Let X be a random variable, F is the cumulative distribution function (CDF) of X . CRPS is defined as

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(y) - I(y \geq x))^2 dy$$

Closed form expression for the integral is defined as

$$\frac{1}{2} \mathbb{E}_P |X - X'| - \mathbb{E}_P |X - x|$$

Since we assume the normal distribution as a distribution of the forecast of ARMA-GARCH model, then CRPS is

$$\text{CRPS}(N(\mu, \sigma^2), x) = \sigma \left[\frac{1}{\sqrt{\pi}} - 2\vartheta\left(\frac{x - \mu}{\sigma}\right) - \frac{x - \mu}{\sigma} (2\Phi\left(\frac{x - \mu}{\sigma}\right) - 1) \right]$$

where ϑ and Φ denote the probability density function and the cumulative distribution function of a standard normal variable. In summary, the use of predictive cumulative distribution functions and the continuous ranked probability score addresses the limitations of predictive densities in certain forecasting scenarios. These approaches provide a more practical and flexible framework for evaluating probabilistic forecasts and assessing their accuracy forecasts.

3.3.2 Energy Score

Energy score (ES)[12] is a generalization of CRPS (Gneiting and Raftery, 2007). It evaluates the multivariate probabilistic forecasts. Energy Score is defined as

$$ES(P, \mathbf{x}) = \frac{1}{2} \mathbb{E}_P \|\mathbf{X} - \mathbf{x}\|^\beta - \mathbb{E}_P \|\mathbf{X} - \mathbf{X}'\|^\beta$$

where \mathbf{x} and \mathbf{X}' are independent copies of a random vectors with d-dimensional distribution P , $\|\cdot\|$ is Euclidean norm. In negative orientation the ES is defined as

$$ES^*(P, \mathbf{x}) = \mathbb{E}_P \|\mathbf{X} - \mathbf{x}\|^\beta - \frac{1}{2} \mathbb{E}_P \|\mathbf{X} - \mathbf{X}'\|^\beta$$

A lower value of ES^* means a better probabilistic forecast.

As we said before Energy Score is generalization of CRPS. To show that lets assume that we have a one-dimensinal distribution , then $d = 1$ and $\beta = 1$, then Energy Score is

$$\begin{aligned} ES^*(P, x) &= \mathbb{E}_P |X - x| - \frac{1}{2} \mathbb{E}_P |X - X'| \\ &= \mathbb{E}_P \sqrt{(X - x)^2} - \frac{1}{2} \mathbb{E}_P \sqrt{(X - X')^2} \\ &= \mathbb{E}_P |X - x| - \frac{1}{2} \mathbb{E}_P |X - X'| \\ &= CRPS^*(P, x)^8 \end{aligned}$$

4 Empirical Results

In this chapter, we utilize the models introduced in Section 4. Initially, we fit the assets using the ARMA-GARCH model and compute the probabilistic forecast. Then, we explore the dependency structure between each asset and employ copula models to determine the best fit. Finally, we evaluate the models using the methods introduced in probabilistic forecast evaluation

4.1 Computation- Packages in R software

tseries package

To check the stationarity of time series (augmented Dickey-Fuller test) we use function adf.test (augmented Dickey-Fuller test) from the tseries[14] package (Trapletti and Hornik 2023).

rugarch package

The R package rugarch[15] provides an environment for simulating, fitting and testing ARMA-GARCH models (Ghalanos 2020). The package allow us to fit model and a variety of conditional distributions for the error terms.

scoringRules

The scoringRules[16] package for R provides functionality for comparative evaluation of probabilistic models based on proper scoring rules, covering a wide range of situations in applied work (Jordan 2018). The package allow us to evaluate the probabilistic forecast

⁸(CRPS* represent negative orientation of the closed expression of CRPS)

from ARMA-GARCH and Copula-GARCH , by using the functions *crps()* (Continuous Ranked Probability Score) and *es.sample()* (Energy Score).

copula package

The *copula*[17] package provides (S4) classes of commonly used elliptical, (nested) Archimedean, extreme value and other copula families; methods for density, distribution, random number generation, and plots (Hofert 2023).

4.2 Univariate Probabilistic Forecast with ARMA-GARCH Models

As we mentioned in section 2.2, there are significant autocorrelations and partial autocorrelations in the first and second moments of the assets. Therefore, an appropriate model is required which can reduce both autocorrelations in the first and second orders. Hence, in this report, we first consider the ARMA-GARCH model to analyze the data. The selection of the ARMA-GARCH order is based on Akaike's information criterion (AIC). We start by determining the order of the ARMA part for each asset, with the parameter restrictions of $0 \leq q \leq 2$ and $0 \leq p \leq 2$. The AIC scores are computed, and the order with the lowest AIC score determines the values of p and q for each asset. Once the ARMA orders are determined, we fit the time series data with the ARMA model using maximum likelihood estimation (MLE) and obtain the residuals. It is observed that the squared residuals of each asset from the ARMA estimations exhibit a conditional heteroskedasticity issue known as the ARCH effect. The ARCH effect is further confirmed using the LM test. The results of the LM test are presented in Table 2.

After confirming the presence of the ARCH effect, we proceed to find the optimal orders for the GARCH(P,Q) models. Based on the orders obtained from the ARMA part, we determine the GARCH part orders using the AIC scores. The selected orders for the models and their corresponding AIC scores are presented in Table 3. Consequently, the parameters of the ARMA-GARCH models for each asset are estimated using MLE, and the estimated parameters are reported also in Table 3. It is worth noting that for all models, the GARCH order is chosen as (1,1) for the GARCH part. This choice is a common selection for GARCH models, as Hansen and Lunde [7] have shown through extensive empirical studies that there is no model better than the GARCH(1,1) specification.

In Table 3, it is evident that the estimated parameters satisfy the model restrictions, suggesting that the assumptions of the model hold for the estimated parameters. Furthermore, we assess the residuals of each model to verify if they exhibit any serial correlation. According to the model assumptions, the residuals should be free from serial correlation. To confirm that, we conduct the Ljung-Box test. Ljung-Box test is a statistical test used to examine the presence of serial autocorrelation in the residuals. The test results indicate no significant serial correlation in the residuals. Detailed results of the Ljung-Box test for serial autocorrelation can be found in the Appendix. Furthermore, we conducted an investigation into the distribution of the residuals and observed that they do not follow a normal distribution. This is evident from the deviation of the residuals from the theoretical linear line, as depicted in Figure 13⁹. The QQ plot illustrates that the residuals could not be adequately captured by the expected normal distribution. This reinforces our belief that the true data-generating process may deviate from normality, indicating the need for alternative distributional assumptions or modeling approaches that better accommodate the observed residual distribution.

⁹Figure presented in Appendix

Table 2: LM test for ARMA models

Commodities	ARMA Orders	Test Statistics's p-value	H_0 : No ARCH effects
N.Gas	ARMA(2,1)	0.00	H_0 Reject
Oil	ARMA(2,0)	0.00	H_0 Reject
Coal	ARMA(2,2)	0.00	H_0 Reject

Table 3: Orders of the ARMA-GARCH models with AIC scores and estimated parameters

Assets :	Natural Gas	Oil	Coal
Orders :	ARMA(2,1)-GARCH(1,1)	ARMA(2,0)-GARCH(1,1)	ARMA(2,2)-GARCH(1,1)
AIC	0.001211111	0.4049075	0.007792449
Estimated Parameter Vector	$\hat{\varphi}$	$\hat{\varphi}$	$\hat{\varphi}$
\hat{c}	-0.000465884()	0.02018657 ()	-0.02364725 (*)
$\hat{\alpha}_1$	0.677700031(***)	0.12413556(***)	0.65473747 (**)
$\hat{\alpha}_2$	-0.094188163(***)	-0.03834471(***)	0.29983135 (**)
$\hat{\beta}_1$	-0.609029045(***)	-	-0.54435993(***)
$\hat{\beta}_2$	-	-	-0.37919840(***)
$\hat{\omega}$	0.002449823(***)	0.02937485 (***)	0.01497727(***)
$\hat{\alpha}_1$	0.125227906(***)	0.05604234(***)	0.01209159 (***)
$\hat{\beta}_1$	0.873772011(***)	0.9237526(***)	0.96956205(***)
$\alpha, \beta > 0$	True	True	True
$\sum \alpha + \sum \beta < 1$	True	True	True
$\omega > 0$	True	True	True
Ljung-Box Test	No Serial Correlation	No Serial Correlation	No Serial Correlation

Figure 6 showcases the comparison between the estimated conditional variances and the actual time series. The red dashed lines represent the presence of volatility clustering in the time series, while the green lines depict the conditional variances estimated from the ARMA-GARCH model. Notably, in each subfigure, the region between the red dashed lines illustrate the occurrence of volatility clustering in the time series. This indicates that periods of high volatility tend to cluster together, leading to periods of relative calmness in between. The presence of volatility clustering suggests that the volatility of the time series is not constant over time. On the other hand, the green lines represent the conditional variances estimated from the ARMA-GARCH model. These lines demonstrate how the model captures the volatility clustering observed in the data.

Next, we proceed to one-day-ahead probabilistic forecasts using the estimated coefficients from the ARMA-GARCH models. We perform iterative forecast from the 2501st observation until the 2700th observation for each asset. Then we obtained the mean and the variance forecasts for each commodities. One effective way to represent these forecasts is by utilizing the mean and variance forecasts to generate simulations from a normal distribution. This approach allows us to create density plots for each forecast based on the simulated values. For that, we generated 1000 observations from a normal distribution using the forecasted mean and standard deviation (sigma). In Figure 7, we present the forecast densities and the actual time series for each asset. In the upper left side, we have the natural gas series along with the density plot of the probabilistic forecast. We can observe that when there is volatility clustering, the density plot exhibits fat tails. On the other hand, if there is no clustering but a peak in the actual time series, the density plot also shows a peak. Although the peaks may not be easily discernible in this density plot, we can identify that the middle of the density is curved for each peak, indicating a shock in the time series. Similar patterns can be observed in the density plots of the other assets as well. After obtaining the probabilistic forecasts for each asset using the ARMA-GARCH models, we also applied a probabilistic forecast using an AR(1) model for each time series. We compared the results in Figure 8, and we can see that the CRPS scores for each model are almost the same. This is because the AR(1) model assumes only the mean forecast and does not account for volatility. Since our time series exhibits few volatility phases, the forecast of the AR(1) model closely aligns with the mean of

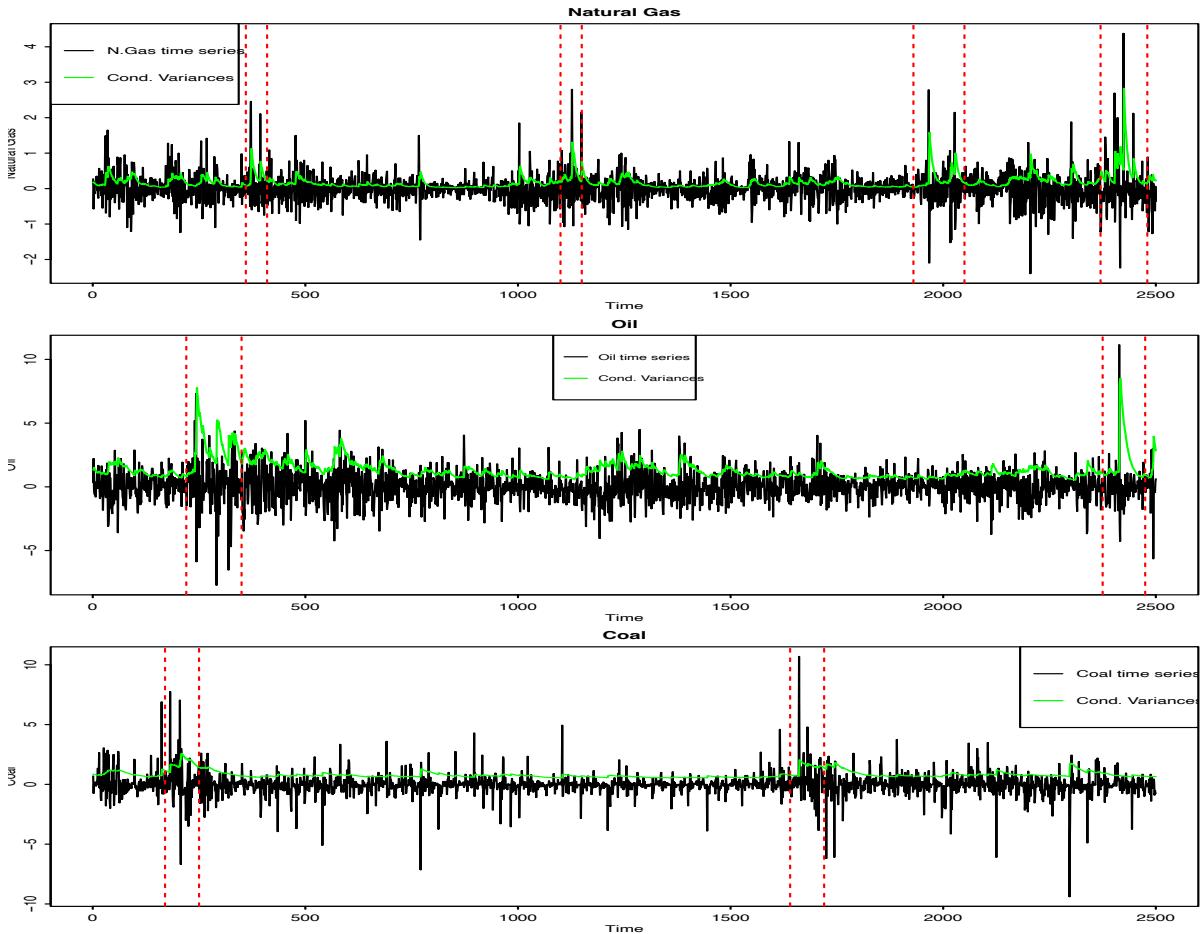


Figure 6: Natural Gas, Oil and Coal time series with estimated conditional variances

the model at each point, resulting in very similar CRPS scores. According to the CRPS scores in the table 4, that the natural gas has the lowest CRPS score among all the commodities. This indicates that the probabilistic forecasts from the ARMA-GARCH model for the natural gas perform better in terms of CRPS. It is important to note that the ARMA-GARCH models capture the volatility clustering, which is not accounted for in the AR(1) model's assumptions.

A proper scoring rule is a function that compares a probabilistic forecast with a realization of the variable, such that it is maximized when the forecast corresponds to the true distribution generating the data (Iacobini, et al., 2023). Since we have doubts about the true distribution of the commodities and suspect that it may not follow a normal distribution, we should be cautious in relying solely on the CRPS (Continuous Ranked Probability Score) for model evaluation. However, if the true distribution deviates from normality, it is important to consider alternative evaluation measures that are more appropriate for non-normal distributions. This could include measures that account for tail behavior, skewness and kurtosis.

Furthermore, we also check the investment risk in each commodities by employing 0.05 quantile (VaR). In Figure 9, the red line represents the 0.05 quantile of each probabilistic forecast, while the actual time series for natural gas, oil, and coal are shown. The purpose of examining the 0.05 quantile is to assess the downside risk and determine the safety of investing in each commodity. However, it is important to note that this analysis assumes a normal distribution, which may not accurately capture extreme tail events as normal distributions tend to ignore the tails. To further evaluate the safety of each commodity,

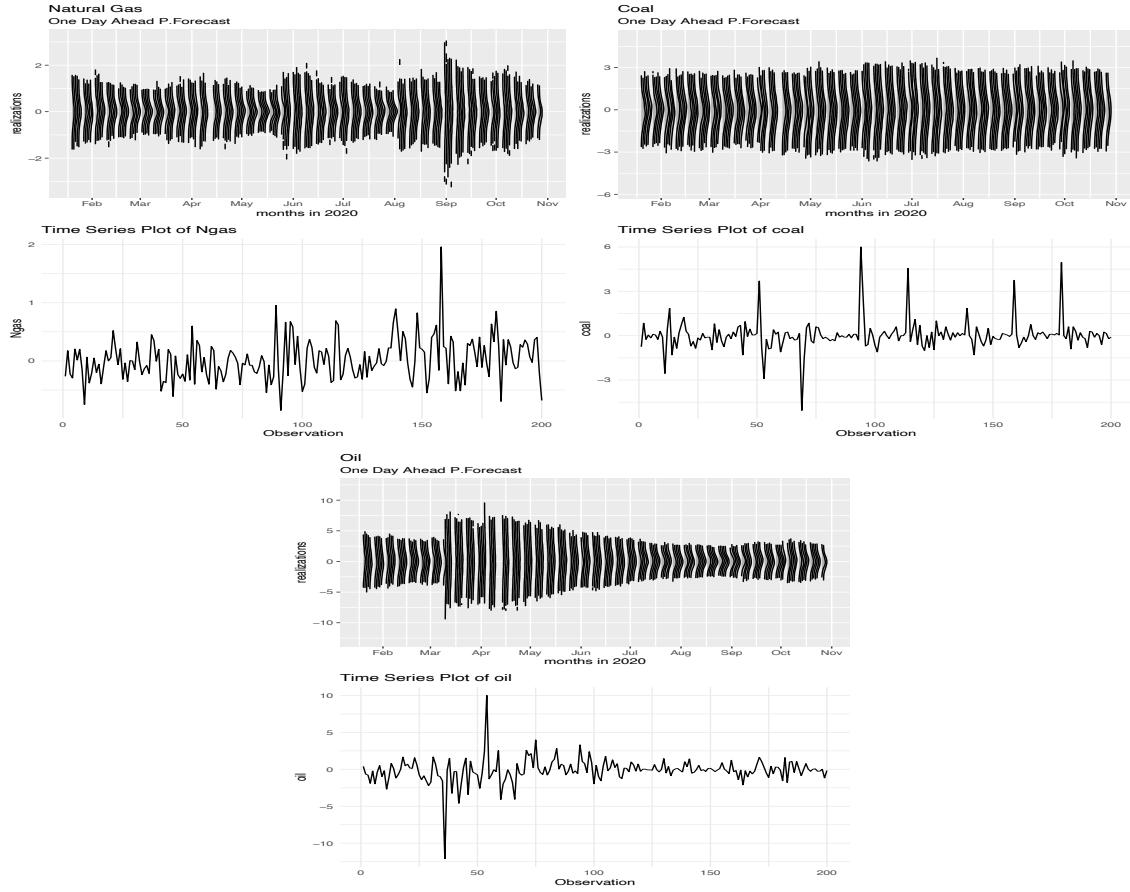


Figure 7: Actual observations of Natural Gas, Oil and Coal time series and the density of the probabilistic forecast from ARMA-GARCH model

Table 4: CRPS scores of ARMA-GARCH and AR(1)

Commodities	Natural Gas	Oil	Coal
Models :	ARMA(2,1)-GARCH(1,1)	ARMA(2,0)-GARCH(1,1)	ARMA(2,2)-GARCH(1,1)
Mean.CRPS	0.1948112	0.739014	0.4668022
Models :	AR(1)	AR(1)	AR(1)
Mean.CRPS	0.1943891	0.7455772	0.4620877

we calculated the percentage of observations that fall below the 0.05 quantile. We found that 3% of the natural gas observations, 4% of the oil observations, and 1.5% of the coal observations fall below the 5% quantile. However, it is crucial to exercise caution in relying solely on these numbers, as they may be influenced by overfitting the model during the forecasting period. In the left side of Figure 1, we can observe that the oil and coal series exhibit fat tails. This means that there are observations in the data that fall outside the range predicted by a normal distribution. By assuming a normal distribution, we ignore these extreme observations and fail to capture the true extent of risk associated with natural gas and coal. To account for the fat tails and better capture the risk associated with natural gas, oil and coal, the alternative distributions that are more robust to extreme events, such as heavy-tailed distributions should be considered. These models can provide a more accurate representation of the tail behaviour and improve the assessment of investment risk. We believe that investing in oil would be a safer option based on our analysis. When examining the absolute losses in each model, we observed

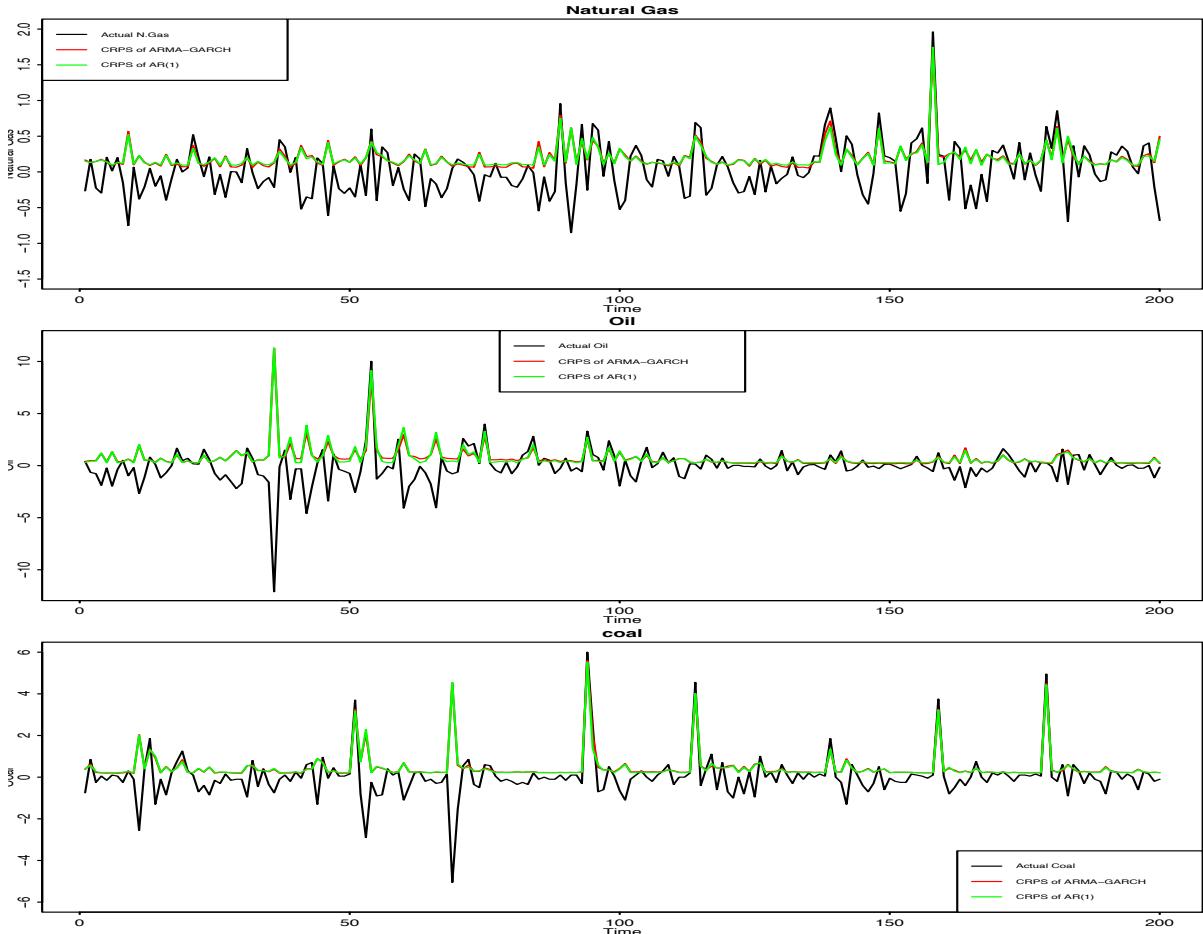


Figure 8: Actual observations of Natural Gas, Oil and Coal time series with CRPS scores from ARMA-GARCH and AR(1) probabilistic forecasts

that both natural gas and coal series had fewer points below the 0.05 quantiles. However, the losses associated with those points were considerably larger compared to oil. In this context, we prefer to minimize the risk of losing a substantial amount and prioritize smaller losses. Considering both the quantile analysis and absolute losses, oil appears to be a safer investment choice.

In summary, our analysis highlights several areas for improvement in univariate probabilistic forecasting. Taking these factors into account can enhance the accuracy and reliability of the forecasts:

- Distribution Assumptions: Instead of relying solely on the normal distribution assumption, exploring alternative distributions that capture fat tails, such as the t-distribution or skew-t distribution, can provide a more realistic representation of the underlying data.
- Leverage Effect: It is important to consider the leverage effect when improving univariate probabilistic forecasting. The leverage effect refers to the asymmetric relationship between asset returns and volatility. Specifically, it suggests that negative shocks or downturns in asset prices tend to have a stronger impact on future volatility compared to positive shocks or upturns. To account for the leverage effect, one can consider incorporating leverage effects in the volatility modeling process. This can be achieved through the use of models such as GARCH (Exponential GARCH) and Asymmetries-GARCH (Treshold).

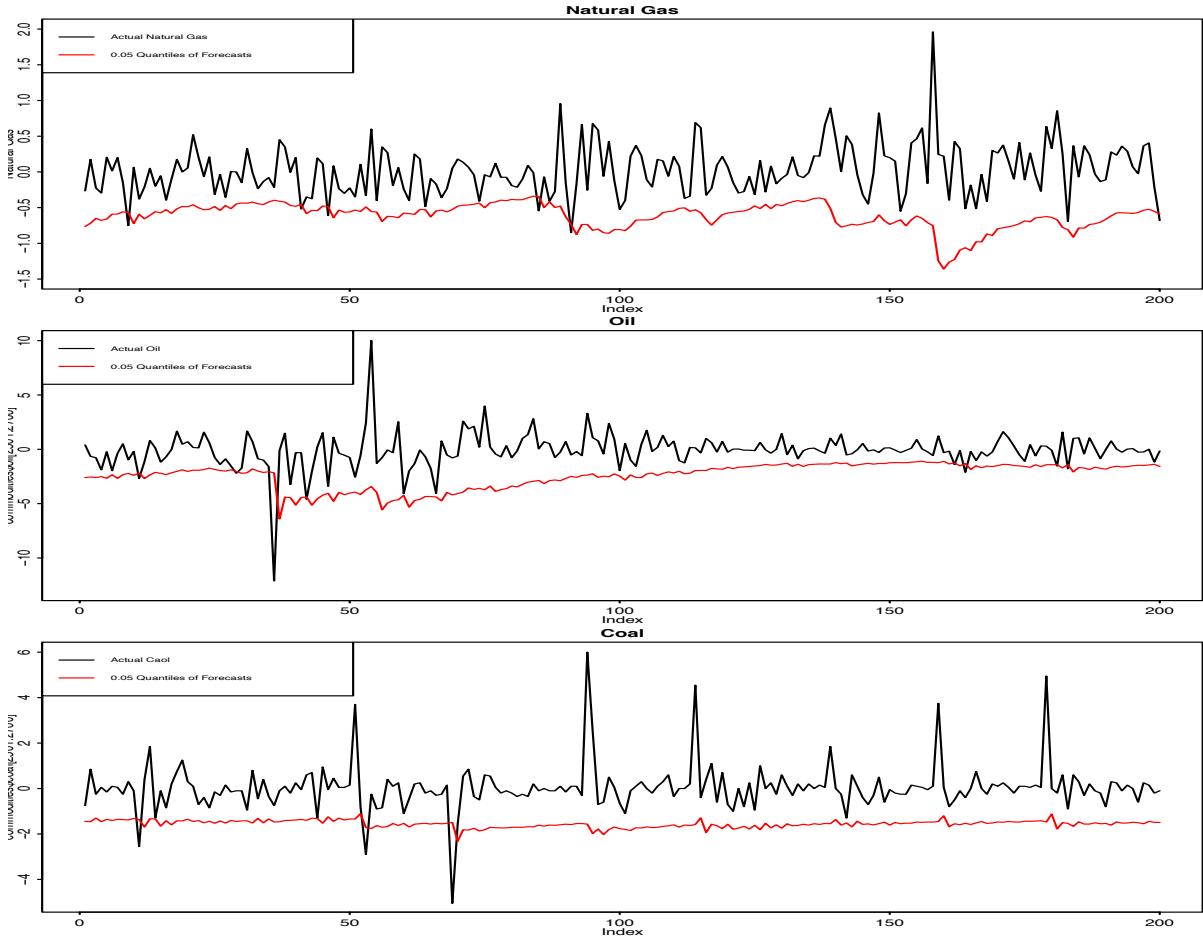


Figure 9: 0.05% Quantiles of the probabilistic forecast and actual time series natural gas (top), oil (middle) and coal (bottom)

- Seasonal Effects: Incorporating the influence of weekly or daily seasonal patterns can improve the accuracy of forecasts. Accounting for these patterns can help capture the specific behavior of the time series on different days or at different times of the year.
- Multivariate Dependency: Exploring the dependencies among multiple assets can provide a more comprehensive understanding of the overall market dynamics. Consideration of multivariate models, such as copulas can capture the joint distribution and dependency structure among the assets, leading to more accurate forecasts.

4.3 Multivariate Probabilistic Forecast with Copulas

To account for the dependency structure between the commodities, we extracted the standardized residuals from the ARMA-GARCH estimation and applied the Probability Integral Transformation (PIT). The PIT-transformed standardized residuals are then used to assess the dependence between the commodities. In Figure 10, we presented the plot of PITs standardized residuals. As we explained in the methodology part that the linear correlation only measures the degree of dependence but does not clearly explain the structure of dependence. That's why we have employed the Spearman rank correlation coefficients to interpret the plots, which were found as 0.22 for natural gas - coal, 0.16 for natural gas - oil, and 0.11 for oil - coal. These coefficients suggest very weak positive

relationships between the commodities. Analyzing the individual plots, starting with the top plot of natural gas and oil, we observed relatively random points with some concentration. This indicates a potential Gaussian copula between the two commodities, supported by the low Spearman correlation coefficient. Moving to the middle plot, we noticed aggregation or concentration in both the upper and lower tails. This behavior suggests the presence of tail dependence, which aligns with the characteristics of a t copula. However, the strength of the tail dependence is relatively weak, as indicated by the low Spearman rank correlation coefficient of 0.22 between natural gas and coal. Similarly, in the last plot, the points resembled a t copula, but with a very low coefficient of tail dependence. This indicates that there might be little to no tail dependence. We were uncertain about the specific copula type, leaning towards either an gaussian copula or a t copula with weak dependence. To further explore the tail dependencies, we calculated the upper and lower tail coefficients for various copula families such as Gaussian, t copula, Clayton, and Gumbel copula. Table 5 presents the upper and lower tail coefficients with the Kendall coefficient for the copula model with the lowest AIC scores from bivariate copula fit. Analysing the table, we observed that the PIT plot of the coal and the natural gas series showed upper and lower tail dependency with the coefficients of 0.03, this is suggesting some symmetrical tail dependency, which aligns with a t copula. Similarly, the coal and oil series exhibited a tail dependency with the coefficient of 0.001, which is close to 0, indicating potential t-copula (or Gaussian copula) characteristics. As expected, the last plot, which showed no tail dependency, it is also indicating to a Gaussian copula. Also, the Kendall tau coefficients were relatively low, further supporting the finding of low or almost no dependency between the commodities. Since we aim to employ multivariate

Table 5: Upper and lower bound coefficients with Kendall's Tau coefficients

Assets	Natural Gas-Coal	Coal-Oil	Natural Gas -Oil
Copula :	t-copula	t-copula	Gaussian copula
Upper Tail	0.03	0.001	0
Lower Tail	0.03	0.001	0
Kendall's tau	0.15	0.08	0.1

modelling with copulas, we now switch to multivariate copula modelling. Initially, we fit various copula models to the PITs standardized residuals. To determine the most suitable copula, we compare the Akaike Information Criterion (AIC) scores of different copula models. The copula model with the lowest AIC score would indicate the best order for the model. By following the steps outlined in Section 3.2.8, we calculated the likelihood score for each copula model and subsequently computed the Akaike Information Criterion (AIC) scores. Table 6 presents the AIC scores for five different copula models. Upon examining the table, we observe that the t-copula has the lowest AIC score among the considered copula models. This result aligns with our previous findings, indicating that the t-copula provides the best fit for the data in terms of both dependence structure and tail behaviour. We see that the AIC scores of each copula models are very close each other, that's why we examine 3 copula models which are Student-t, normal and independent copula. According to the procedures outlined in Section 3.2.9, we carried out the simulation and forecasting process as follows:

- Extract the standardized residuals from the ARMA-GARCH estimation for natural gas, oil, and coal.

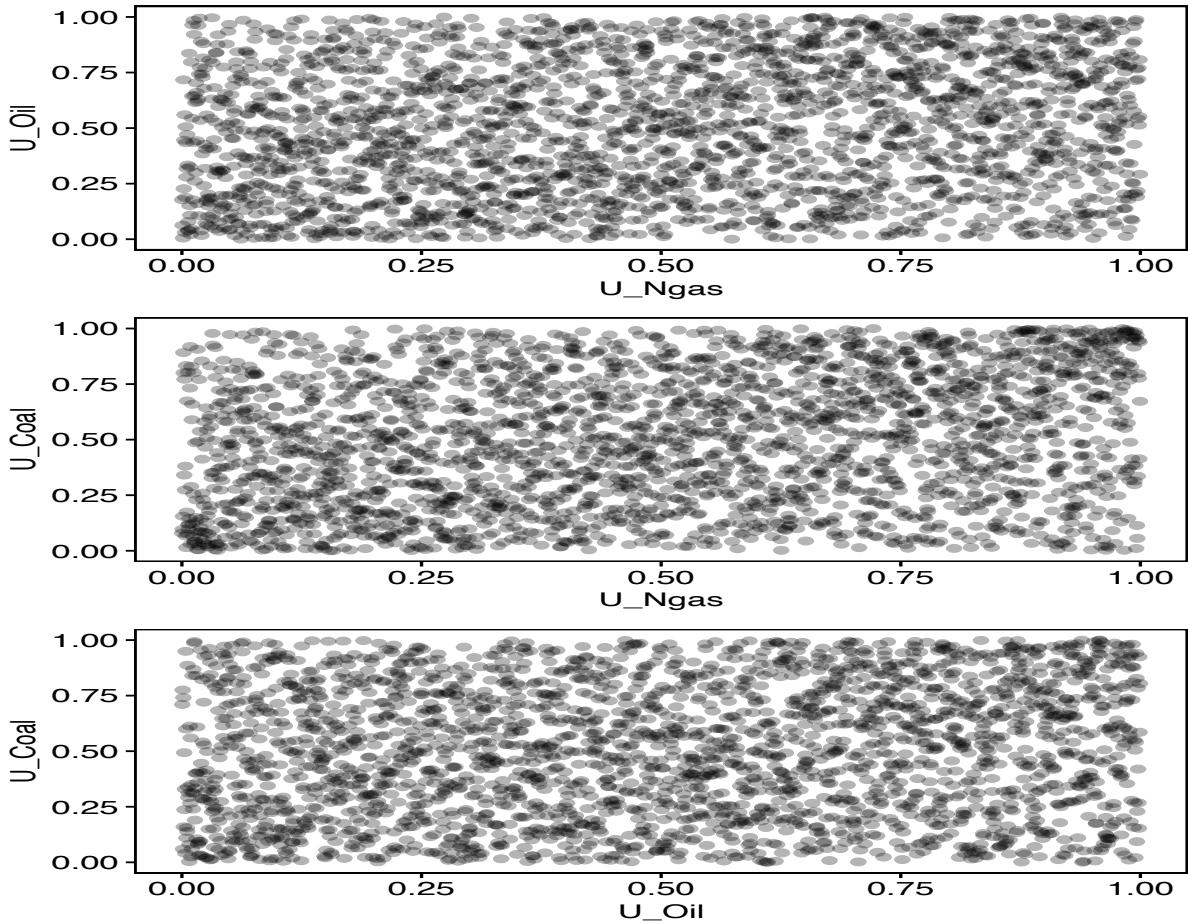


Figure 10:

Table 6: AIC scores of normal, independent, Clayton, Gumbel and t copulas.

Copulas :	Normal copula	t-copula	Independent copula	Clayton copula	Gumbel copula
AIC	-0.08	-0.09	-0.055	-0.058	-0.058

- Apply the PIT transformation to the standardized residuals.
- Fit a normal copula to the PIT residuals.
- Simulate the PIT-transformed residuals 1000 times using the estimated copula coefficients.
- For the 2501st observation, we performed a quantile transformation using the mean and variance forecasts from the ARMA-GARCH model to obtain standardized residuals.

The process described above is repeated for observations 2501 to 2700 for each copula model, resulting in 200 lists. In the end, we have 200 lists for each copula, each containing 1000 rows and 3 columns. In Figure 14 (see Appendix), we present the density of the returns obtained from Copula simulation alongside the actual density plots of each commodity. It is evident that the distribution of the simulations differs from the actual distribution of the commodities. Which proves that misspecified distribution on innovation results in lead unreliable forecast. Next, we evaluate the simulation results using the energy score. A lower energy score indicates a better model fit. Figure 7 displays the

energy scores of each forecast using the Student-t, normal, and independent copula models. We can see that the energy scores of the copula-based simulations are relatively close to each other. Among them, the Student-t copula exhibits the lowest energy score, indicating better performance compared to the normal copula model. The rationale behind this analysis is that by using copulas, we do not change the marginal distributions but rather modify the dependency between assets. According to the research on multivariate scoring rules [3], changing the dependency structure of the model may still yield the same energy scores for the copula models. This emphasizes the importance of distributional assumptions since we altered the dependency structure using copulas and obtained same energy scores for each copula model. However, if we modify the distributional assumptions for the innovation differently, we may obtain different energy scores. Hence, we extracted the simulation results from the independent copula model and computed the energy score for each asset individually. The results are not surprising because we expect

Table 7: Evaluation of the Probabilistic Forecast of Copula Models with energy Score

Models:	Independence Copula	Student t-copula	Normal copula
Mean of Energy Scores	1.018	1.017	1.019

that the CRPS score would be close to the energy scores of independent copula, since the correlation coefficient is zero between each commodities. This finding strongly reinforces that the we might have independency between each commodities, since we obtain the energy scores very close to each other. In summary, our analysis highlights several areas

Table 8: Energy Score of Independent Copula Model and CRPS of ARMA-GARCH Model

Commodities :	Natural Gas	Oil	Coal
Energy Scores :	0.221248	0.6588646	0.5230257
CRPS	0.1948112	0.739014	0.4668022

for improvement in multivariate probabilistic forecasting.

- Firstly, as mentioned before, the assumption of normality is not realistic for financial analysis. Other distributions such as the t-distribution or skewed t-distribution may provide a better fit. The t-distribution accounts for tail dependence, while the skewed t-distribution captures high kurtosis.
- Traditional copula methods may not be flexible enough to capture complex or non-linear dependencies. Alternative approaches, such as vine copulas, offer improved modelling capabilities. Vine copulas decompose the joint distribution into a series of bivariate copula functions. Each bivariate copula models the dependence between a pair of variables, and the overall dependence structure is constructed by combining these bivariate copulas. Vine copula methods enable more flexible and comprehensive modeling of complex dependencies compared to traditional copula methods.
- Lastly, we did not consider that correlations may vary over time. In reality, the dependency or correlation structure between variables is time-varying.

5 Conclusion

In this report, we first consider the ARMA-GARCH model to analyze the data. The optimal orders for the ARMA-GARCH models are determined based on Akaike's information criterion (AIC) and the ARCH effect is further confirmed using the Lagrange Multiplier (LM) test. Then we estimate the parameters of the ARMA-GARCH models for each asset using Maximum likelihood Estimator (MLE). The estimated parameters satisfy the model restrictions and we assess the residuals of each model to verify if they exhibit any serial correlation. The Ljung-Box test indicates no significant serial correlation in the residuals. We conducted an investigation into the distribution of the residuals and observed that they do not follow a normal distribution. This indicates that the true data-generating process may deviate from normality, and that alternative distributional assumptions are needed to accommodate the observed residual distribution. Then we used the estimated coefficients from the ARMA-GARCH models to perform one-day-ahead probabilistic forecasts for each asset. We demonstrated the effectiveness of the ARMA-GARCH model for capturing autocorrelations and volatility clustering in the time series data of natural gas, oil, and coal. The models provided accurate one-day-ahead probabilistic forecasts. Then we use AR(1) model to compare the probabilistic forecast of ARMA-GARCH models using Continuous Ranked Probability Score (CRPS). We find that the CRPS of AR(1) and ARMA-GARCH models very close to each other. However, the AR(1) model only accounts for the conditional mean and does not consider the conditional variance. Therefore, comparing these two models may not provide valuable insights in terms of CRPS. Furthermore, we highlighted the importance of assessing investment risk beyond the assumptions of normal distributions and 0.05 quantiles. We find that the oil has less risk than other commodities.

Lastly, we incorporated copulas to account for the dependency structure between the commodities. The t-copula was identified as the best-fit copula model, considering the AIC score. We then performed probabilistic forecasts using simulations for the normal, t, and independence copula models. The t copula exhibits the lowest energy score, indicating better performance compared to the other copula model.

In conclusion, the assumption of normality is unrealistic in financial markets due to the presence of fat tails. This highlights the need to explore alternative distributions that can accurately capture the extreme events observed in these markets. Secondly, we emphasize that the correlation structure of assets is not constant over time. This implies that relying on static correlation measures may lead to inadequate risk assessments and portfolio allocations. Incorporating time-varying correlation models can provide a more accurate representation of the dynamic relationships between assets.

References

- [1] Stock, J. H., Watson, M. W. (2020). Introduction to Econometrics, Fourth Edition. Pearson. 11,15,406-408, 565 - 567,586 - 589, 710-725cite1
- [2] Francq, C., Zakoian, J. M. (2019). GARCH models: structure, statistical inference and financial applications. John Wiley Sons.
- [3] Alexander, C., Coulon, M., Han, Y., Evaluating the discrimination ability of proper multivariate scoring rules, (2022)
- [4] Mittnik. S., Marc S. Paoletta and Svetlozar T. Rachev, (2000), Diagnosing and treating the fat tails in financial returns data, Journal of Empirical Finance, 7, (3-4), 389-416
- [5] Hyndman, R. J. and Fan, Y. (1996) Sample quantiles in statistical packages, American Statistician 50, 361–365
- [6] Lu, X. F., Lai, K. K., Liang, L. (2014). Portfolio value-at-risk estimation in energy futures markets with time-varying copula-GARCH model. Annals of operations research, 219(1), 333-357.
- [7] Hansen, P.R. and Lunde, A., "A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)?", Journal of Applied Econometrics, Vol. 20, No. 7, (2005), 873-889.
- [8] Schmidt, T., Coping with copulas (2007), Research Gate
- [9] Box G E P and Jenkins G M 1976 Time Series Analysis: Forecasting and Control ed Second (San Francisco: Holden- Day)
- [10] Brockwell P., Davis R., (2016). Introduction to Time Series and Forecasting, Third Edition, Springer, 73-94,
- [11] Durrett, R., (2010) Probability: Theory and Examples, Fourth Ddition, Cambridge University Press, 8 , 8
- [12] Stock, J. H., Watson, M. W. (2020). Introduction to Econometrics, Fourth Edition. Pearson. 2, 63-64
- [13] Gneiting, T., Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association, 102(477), 359-378.
- [14] Trapletti, A., tseries: Time Series Analysis and Computational Finance, (2023)
- [15] Galanos, A., rugarch: Univariate GARCH Models (2022) <https://github.com/alexisog/rugarch>
- [16] Jordan, A., scoringRules: Scoring Rules for Parametric and Simulated Distribution Forecasts, (2023) <https://github.com/FK83/scoringRules>
- [17] Hofert, M., copula: Multivariate Dependence with Copulas, (2023), <https://copula.r-forge.r-project.org/>
- [18] Iacopini, M., Proper Scoring Rules for Evaluating Density Forecasts with Asymmetric Loss Functions. (2023)

Appendix

Autocorrelation Function

The autocorrelation function (ACF)[1] is a statistical tool used to measure the correlation between a time series and its lagged values. It quantifies the relationship between observations at different time points within the same series. The ACF is calculated by computing the correlation coefficient between the original time series and its lagged versions. The correlation coefficient measures the strength and direction of the linear relationship between two variables. The ACF is typically plotted as a function of the lag, which represents the time interval between the observations being compared. The lag is denoted by the symbol "k" and can take on different values, such as 1, 2, 3, and so on. The ACF provides insights into the presence of autocorrelation in a time series. It helps identify patterns or dependencies within the data, such as seasonality or trends. A positive autocorrelation at lag k indicates that there is a similarity between observations separated by k time units, while a negative autocorrelation indicates dissimilarity. The ACF is calculated as :

$$\rho_s = \text{Corr}(X_t, X_{t-s}) = \frac{\text{Cov}(X_t, X_{t-s})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-s})}}$$

where $\text{Cov}(X_t, X_{t-s}) = \frac{1}{T} \sum_{t=1}^T (X_t - \mu_{X_t})(X_{t-s} - \mu_{X_{t-s}})$

The Augmented Dickey-Fuller (ADF) Test

The augmented Dickey-Fuller (ADF) test is a commonly used statistical test to determine if a time series is stationary or exhibits unit roots. The ADF statistic is based on the augmented Dickey-Fuller regression model, which incorporates lagged differences of the dependent variable to capture the presence of unit roots. The null hypothesis of the test is that the time series contains a unit root, indicating non-stationarity, while the alternative hypothesis is that the time series is stationary. The ADF regression model can be expressed as follows:

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_k \Delta y_{t-k} + \epsilon_t$$

where Δ is difference operator. We test the hypothesis $H_0 vs H_1$ which are defined as:

$$H_0 : \gamma = 0 \quad \text{vs} \quad H_1 : \gamma < 0$$

Test statistic is calculated as follows:

$$ADF_\tau = \frac{\hat{\gamma}}{\sqrt{\text{Var}(\hat{\gamma})}}$$

If the test statistics is less than the critical value with a significance level, we reject H_0 and conclude that the time series is stationary.

Ljung-Box test

Absence of autocorrelation in the ARMA-GARCH residuals is one of the assumptions of the model. To check this we introduce Ljung-Box test. The Ljung-Box test is a statistical test used to examine the presence of serial autocorrelation in the residuals or errors of a time series model. It is a way to assess whether the model adequately captures the underlying patterns in the data or if there is additional structure in the residuals. The test

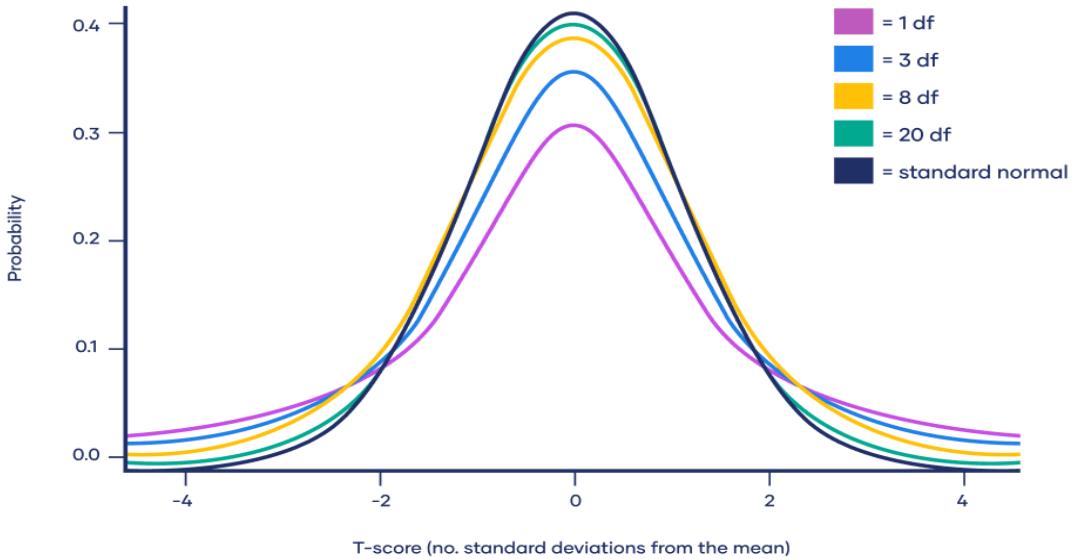


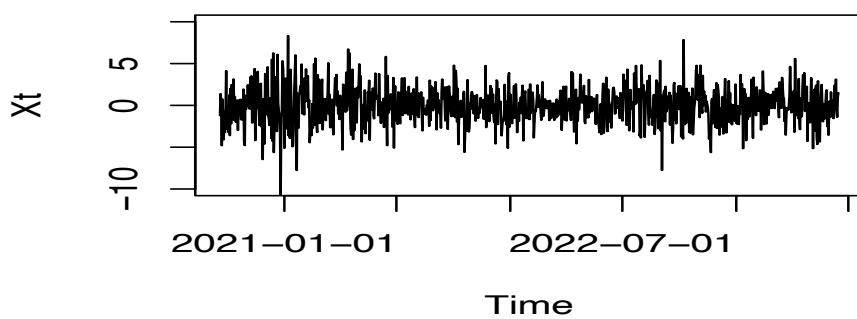
Figure 11: T-distribution and the standard normal distribution from <https://www.scribbr.com>

evaluates whether the autocorrelations of the residuals are statistically different from zero up to a specified lag k . The null hypothesis (H_0) of the Ljung-Box test is that the model does not exhibit a lack of fit, implying that the residuals are independent and identically distributed (iid) or resemble white noise. The alternative hypothesis (H_1) suggests that there is significant autocorrelation present in the residuals, indicating a lack of fit. The test statistics for Ljung-Box test is defined as:

$$Q(m) = n(n + 2) \sum_{j=1}^m \frac{\rho_j^2}{n - j}$$

where, ρ = the accumulated sample autocorrelations, m = the time lag. We reject the null hypothesis and say that the model shows lack of fit if the test statistic Q is bigger than $\chi^2_{1-\alpha,h}$, where the value found on the chi-square distribution table for significance level α and h degrees of freedom.

Alpha=0.09 and Beta=0.9



Alpha=0.6 and Beta=0.3

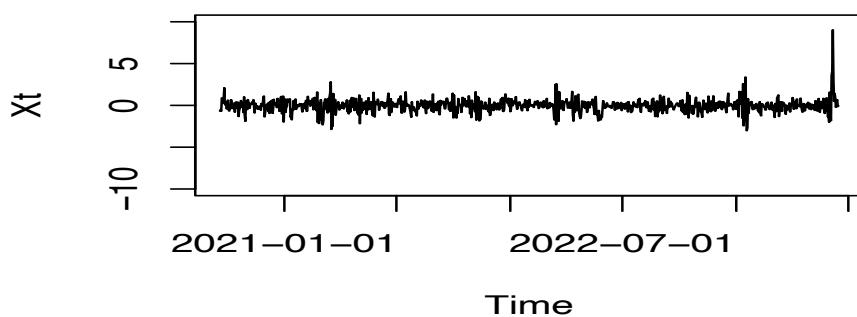


Figure 12: GARCH(1,1) model simulation with different alpha and beta values

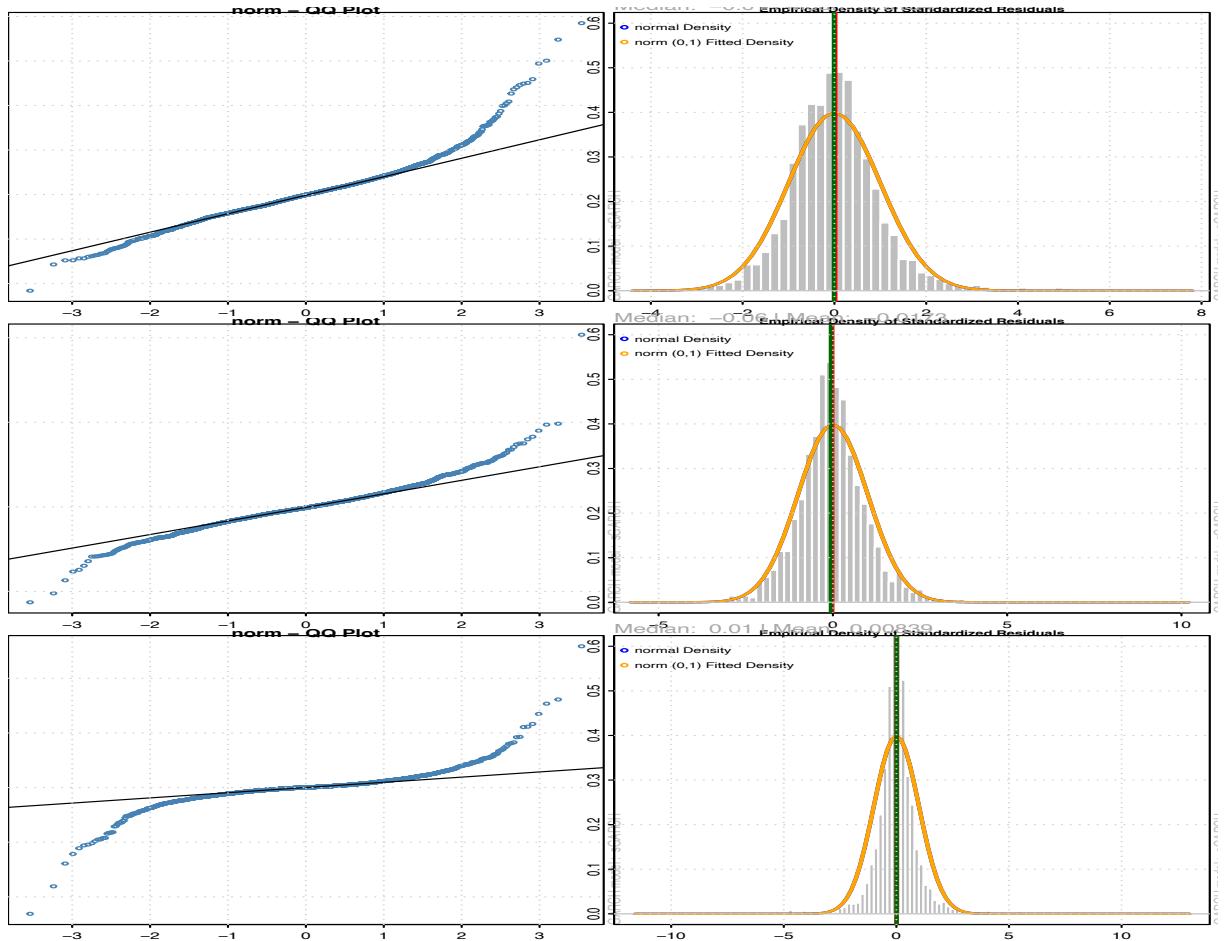


Figure 13: QQ-plot of standardized residuals from ARMA-GARCH with Normal distributional assumption. First row is Natural gas, second row is oil and last row is coil that each row present the standardized residuals of ARMA-GARCH process.

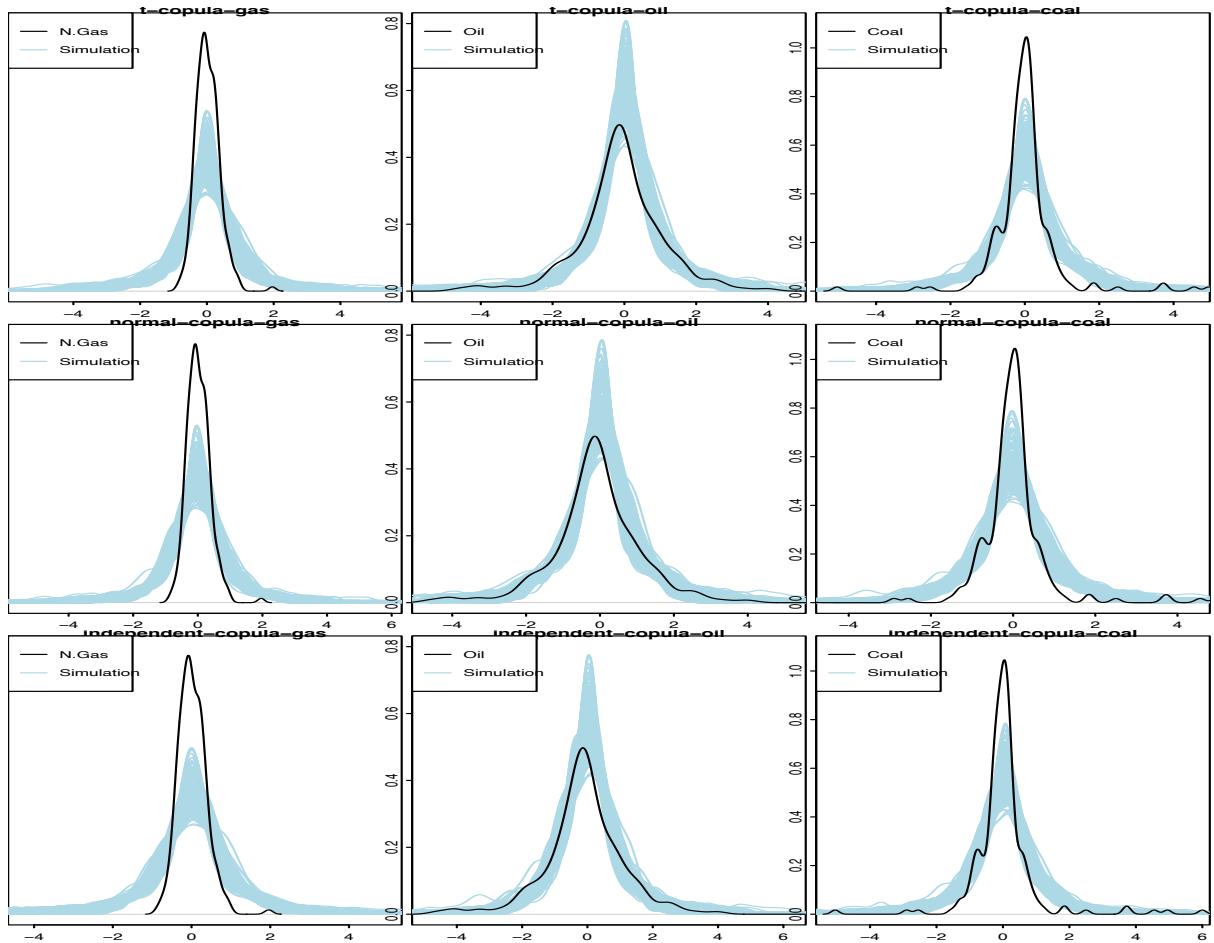


Figure 14: Density plot of the mean forecast of the commodities from normal (first row), t (second row) and independent (last row) copula