Selection of Value at Risk Models for Energy Commodities

Ilkin Namazli

Zahid Emin Aker

Rexina Bushi

Gianmarco Sperati

30 April , 2022

**Keywords**: GARCH, RUGARCH, Midas, Quantile models, Energy commodities, Value at Risk

**Abstract**

This project investigates the empirical properties of WTI crude oil, Propane, Gasoline and Kerosene. However, only the findings of WTI are discussed in the paper but the R codes of other variables are shared in appendix section. We investigate the time-varying correlation between the energy commodities returns – high frequency data – and the covid deaths curve, mixing them with the rumidas package in R. We eventually make forecasts about the volatility and the returns of the variables through GARCH models. Then, we conduct the VaR analysis followed by a backtesting procedure. Finally we measure forecasting performance and conditional correlation dynamics, as well as evaluate and compare the performance of univariate and multivariate models using a variety of diagnostic and forecast performance tests.

**Introduction**

The covid phenomenon hit the world economics with unprecedent results. During the outbreak of the pandemic, in early 2020, the price of energy commodities experienced an above-average volatility because of the huge fall in demand that occurred as consequence of the many lockdowns imposed by national governments.

For the first time in history, energy prices fell below zero in April 2020, highlighting that the demand was so low that suppliers were willing to pay buyers – in that particular moment – since they could not store oil barrels anymore. In the long run, the evolution of prices sheds light on the health state of the world economy, revealing possible bubbles and stagnation periods. Many research papers before investigated the effects of energy commodity fluctuations, highlighting the importance of the price these raw materials for inflation and industrial production.

The period considered starts from July until the midst of the vaccination campaign in June 2021, corresponding to the beginning of the economic upturn and the consequent recovery of energy commodities’ prices.

The empirical results show that this statistical approach represent a valuable methodology that can be exploited by risk managers, investors and policy makers to assess the effect of the pandemic on spillover effects in energy markets.

**Fundamental Analysis**

We start our analysis with WTI variable from the “energydata”. The following chart shows that the data exhibits an increasing trend, together with a seasonality component.

Chart, line chart, scatter chart

Description automatically generated

Before we attempt to build any model using our variable, we should firstly determine whether it is stationary. Additionally we can't use statistical methods when the process isn't stationary. Stationarity is the long-term equilibrium in respect to a mean level of the process. Moreover we must eliminate everything that causes a process to be non-stationary. We can check it by building the autocorrelation function (ACF) and the partial autocorrelation function (PACF). ACF provides information on the process's memory, quantifies the correlation (linear dependence) between process values at different lags, and it also indicated the amplitude and length of the memory of the process.

Chart, histogram

Description automatically generated

Based on this decaying ACF of WTI, we are likely dealing with an autoregressive process, but we can’t say for sure since a given ACF doesn’t fully characterize a process. Since our data is daily, the distance between our lags is very small, indicating that the process has a strong memory. Also, the ACF of the series exhibits a trend component. Our time series is not stationary.

PACF is the partial autocorrelation function that reveals the partial correlation between the series and their lags. Both, from the graph and ADF results of the series, we can say that its non-stationary (due to non-constant mean and sd, and an obvious trend component).

Chart

Description automatically generated

In order for us to turn our times series into a stationary one we differentiated it.

> d\_WTI <- diff(WTI)

Now as can be seen from the chart below the series displays a stationary behavior. Moreover, this variable will be used in the further analyses.

Graphical user interface

Description automatically generated

Taking the first difference of WTI we built again the ACF and the PACF. In this way we have eliminated the seasonality and trend of the WTI time series.

Graphical user interface

Description automatically generated

In the differenced ACF we observe one significant value indicating a correlation at lag 1. All the other values lie inside the blue dashed lines. So, this can also be sign for a White Noise.

Graphical user interface

Description automatically generated

Additionally, to test the stationarity we used the Augmented Dickey Fuller Test, which uses the following null and alternative hypotheses:

**H0:** The time series is non-stationary. In other words, it has some time-dependent structure and does not have constant variance over time.

**HA:** The time series is stationary.

If the p-value from the test is less than some significance level (e.g. α = .05), then we can reject the null hypothesis and conclude that the time series is stationary.

>adf.test(WTI) -> For our non-differentiated data the p-value is bigger than 0.05, so we do not reject the null hypothesis meaning that the time series of WTI is not stationary.

>adf.test(d\_WTI) -> As for our differentiated data, the p-value is 0.01, smaller than the significance level of 0.05, so we reject the null hypothesis meaning that our time series is stationary.

Next on our analysis of WTI we have also included the skewness and kurtosis.

Chart, histogram

Description automatically generated

As we can see, the histogram of the of the returns seems to be more skewed than the normal distribution, meaning that considering the normal distribution for the returns is not a good choice. We will see further what model is confirmed by the model estimation.

The skewness is a metric of symmetry. Moreover, negative skewness generally means that the mean of the data values is smaller than the median and that the data distribution is skewed to the left. On the other hand, positive skewness indicates that the mean of the data values is greater than the median, and the data distribution is skewed to the right.

>skewness(WTI) -> -0.6658455 Our data are moderately skewed .

The kurtosis measure describes the tail of a distribution. Firstly, a negative value for kurtosis indicates a thin tailed distribution; the values of the sample are distributed closer to the median than we would expect for a standard normal distribution. In addition to a positive kurtosis number suggests that we are dealing with a fat-tailed distribution, in which extreme outcomes are more prevalent than a typical normal distribution would estimate.>kurtosis(WTI) -> 5.254791

We have a positive kurtosis value, which means that we are dealing with a fat tailed distribution.

To check the goodness of fit test we consider the Jarque-Bera Test .

The Jarque-Bera test is a goodness-of-fit test that measures if sample data has skewness and kurtosis that are similar to a normal distribution. The Jarque-Bera test statistic is always positive, and if it is not close to zero, it shows that the sample data do not have a normal distribution.

>jarque.bera.test(WTI) -> with the results being: X-squared = 90.799, df = 2, p-value < 2.2e-16

This indicates that the test statistic is 90.799, with a p-value of 2.2e-16. We would reject the null hypothesis that the data is normally distributed in this circumstance.

To check if the time series contains correlation, we use The Ljung-Box test that is a hypothesis test. The null Hypothesis H0 is that the residuals are independently distributed. The alternative hypothesis is that the residuals are not independently distributed and exhibit a serial correlation.

>Box.test(WTI) -> with the results being: X-squared = 285.85, df = 1, p-value < 2.2e-16

Here we see a p-value much smaller than 0.01, thus we can reject the null hypothesis, indicating the time series does contain an autocorrelation.

**GARCH**

The generalized autoregressive conditional heteroskedasticity (GARCH) process is an econometric term developed in 1982 by Robert F. Engle, who won the Nobel Memorial Prize in Economics in 2003. The term GARCH refers to a method for estimating volatility in financial markets. GARCH processes are widely used in finance due to their effectiveness in modelling asset returns and inflation.

There are several types of GARCH modeling. Financial professionals frequently prefer the GARCH process because it provides a more real-world context than other models when trying to forecast the prices and rates of financial instruments. GARCH models explain financial markets in which volatility may shift, becoming more volatile during periods of financial crises or world events and less volatile during periods of relative stable and steady economic growth. Because they are autoregressive, GARCH processes rely on previous squared observations and variances to model current variance.

Main purpose of GARCH is to minimize number of errors in estimating by accounting for errors in prior forecasting and enhancing the accuracy of ongoing predictions. The GARCH process is based on the assumptions of Normal, Student t, and Generalized Error Distributions (GED). The Normal distribution is the usual assumption in any time series estimation, but due to the fact that the distribution of GARCH process is leptokurtic (distributions with kurtosis greater than three, it can be described as having a wider or flatter shape with fatter tails resulting in a greater chance of extreme positive or negative events). Normal distribution was shown to be ineffective in capturing the tail behavior of the series.

**GARCH Model Specifications**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Optimal Parameters | | | | | | | |
|  | **Normal GARCH** | | **Skew Normal GARCH** | | **Student-t GARCH** | | **Skew Student-t GARCH** | |
|  | estimate | p-value | estimate | p-value | estimate | p-value | estimate | p-value |
| **mu** | 0.140178 | 0.027565 | 0.128672 | 0.040522 | 0.15946 | 0.008641 | 0.12529 | 0.046088 |
| **omega** | 0.065229 | 0.104945 | 0.063102 | 0.095350 | 0.12910 | 0.283890 | 0.10537 | 0.261614 |
| **alpha1** | 0.124098 | 0.003790 | 0.136751 | 0.002074 | 0.17118 | 0.050763 | 0.17820 | 0.028364 |
| **beta1** | 0.830129 | 0.000000 | 0.819499 | 0.000000 | 0.73662 | 0.000001 | 0.75215 | 0.000000 |
| **skew** | - | - | 0.824006 | 0.000000 | - | - | 0.83929 | 0.000000 |
| **shape** | - | - | - | - | 6.63129 | 0.025869 | 6.98290 | 0.035762 |

The above table summarizes the R output for various GARCH specifications. It can be seen from the table that the associated p-values are generally significant in every distribution. However, only the omega parameter tends to be insignificant in all the cases. Skew Normal GARCH seems to be the best selection here. Because the parameters in this distribution are more significant than the parameters of other distributions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Information Criteria | | | |  |
|  | **Normal GARCH** | **Skew Normal GARCH** | **Student-t GARCH** | **Skew Student-t GARCH** |  |
| **Akaike** | 2.9805 | 2.9673 | 2.9523 | 2.9467 |  |
| **Bayes** | 3.0403 | 3.0421 | 3.0271 | 3.0364 |  |
| **Shibata** | 2.9799 | 2.9664 | 2.9514 | 2.9454 |  |
| **Hannan-Quinn** | 3.0046 | 2.9975 | 2.9825 | 2.9829 |  |

This table presents the information criteria. It displays the Akaike (AIC), Bayes (BIC), Hannan-Quinn and Shibata criteria for the model estimation. The lower these values, the better the model is in terms of fitting. From the table, one can understand that the distributions Student-t and Skew Student-t are the ones with lower values. Therefore, they seem potentially acceptable.

The below table presents the Ljung-Box test for testing the serial correlation of the error terms. The null hypothesis is that there is no serial correlation of the error terms. As can be seen, since the associated p-values are higher than 5%, meaning that there is not enough evidence to reject the null hypothesis. Therefore, all the distribution models passed the test.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Weighted Ljung-Box Test on Standardized Residuals | | | | | | | | |
|  |  |  | **Normal GARCH** | | **Skew Normal GARCH** | | **Student-t GARCH** | | **Skew Student-t GARCH** | |
|  |  |  | **statistic** | **p-value** | **statistic** | **p-value** | **statistic** | **p-value** | **statistic** | **p-value** |
| **Lag [1]** | | | 0.1561 | 0.6928 | 0.1469 | 0.7015 | 0.2229 | 0.6368 | 0.1981 | 0.6562 |
| **Lag [2\*(p+q)+(p+q)-1][2]** | | | 0.4851 | 0.7006 | 0.4478 | 0.7186 | 0.5176 | 0.6854 | 0.4624 | 0.7115 |
| **Lag [4\*(p+q)+(p+q)-1][5]** | | | 1.2622 | 0.7980 | 1.1869 | 0.8161 | 1.2071 | 0.8113 | 1.133 | 0.8283 |
| **d.o.f=0** | | |  | | | | | | | |
| **H0 : No serial correlation** | | |  | | | | | | | |

The below table shows the output of Adjusted Pearson Goodness of Fit Test for each distributions. This table is useful to check if the error term follows the normal distribution. The null hypothesis is that the conditional error term follows a normal distribution. If the p-value is lower than 5%, the null hypothesis is rejected. As we can see, the normal distribution is not rejected (as the p-values are higher than the 5% level).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Adjusted Pearson Goodness-of-Fit Test: | | | | | | | | | | | |
|  | **Normal GARCH** | | | **Skew Normal GARCH** | | | **Student-t GARCH** | | | **Skew Student-t GARCH** | | |
|  | **group** | **statistic** | **p-value (g-1)** | **group** | **statistic** | **p-value (g-1)** | **group** | **statistic** | **p-value (g-1)** | **group** | **statistic** | **p-value (g-1)** |
| **1** | 20 | 19.57 | 0.4212 | 20 | 15.91 | 0.6631 | 20 | 16.43 | 0.6281 | 20 | 11.74 | 0.8965 |
| **2** | 30 | 32.33 | 0.3011 | 30 | 22.78 | 0.7864 | 30 | 31.13 | 0.3593 | 30 | 24.61 | 0.6984 |
| **3** | 40 | 45.83 | 0.2100 | 40 | 31.22 | 0.8081 | 40 | 38.17 | 0.5074 | 40 | 29.83 | 0.8546 |
| **4** | 50 | 53.48 | 0.3064 | 50 | 50.00 | 0.4334 | 50 | 52.17 | 0.3516 | 50 | 37.39 | 0.8872 |

**VaR and Backtesting**

In this section, VaR and Backtesting results is mentioned. The below table highlights several components and their respective p-values. In this table, the outcome of two tests is included; Kupiec and Christoffersen. The variables in the first column are the ones for different distribution errors. For example, WTI\_95\_1 is for the GARCH model with constant mean (normal GARCH) with 95% level of significance. The decision rule is simple, if the associated p-value exceeds the benchmark threshold (0.05), then we do not reject the null hypothesis. Checking the first variable (WTI\_95\_1), we do not reject the null hypothesis. Indeed, this is true for all the variables. This means that VaR model doesn’t produce the correct number of exceedances at the 1% and 5% levels. Additionally, from all the variables, we observe that the LR.uc values are less than the LR.cc ones. Therefore, GARCH (1,1) VaR estimates are not much better than unconditional estimated based on statistical test results.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Kupiec (LR.uc, p-value) | Christoffersen (LR.cc, p-value) | Decision |
| WTI\_95\_1 | 0.266 | 0.38 | DNR (do not reject) |
| WTI\_95\_2 | 0.69 | 0.796 | DNR |
| WTI\_95\_3 | 0.266 | 0.38 | DNR |
| WTI\_95\_4 | 0.266 | 0.38 | DNR |
| WTI\_99\_1 | 0.311 | 0.578 | DNR |
| WTI\_95\_3 | 0.311 | 0.578 | DNR |

**MIDAS Procedure**

The below is the code to construct the matrix with MIDAS variable. Here K=3, meaning that the monthly lag is 4, as it’s defined as K+1. This matrix also takes into account that the high-frequency data, which is d\_WTI\_xts in our case, as converted into weekly.

mv\_m <- mv\_into\_mat (d\_WTI\_xts, weekly\_d\_covid, K =3 , "weekly")

Then we fit the model and estimated the coefficients.

fit\_gm\_norm <- summary (ugmfit(model = "GM", distribution = "norm",

mv\_m = mv\_m, K=3,

skew = "YES", daily\_ret = d\_WTI\_xts))

For the first fitting, we tried GARCH-MIDAS with normal distribution. The R output can be seen below.

|  |  |  |  |
| --- | --- | --- | --- |
| Fitting the model for GARCH\_MIDAS, with Normal Distribution | | | |
| **Coefficients:** |  |  |  |
|  | **estimate** | **p-value** | **sig.** |
| **alpha** | 0.0001 | 0.9995 |  |
| **gamma** | 0.2184 | 0.2245 |  |
| **beta** | 0.6078 | 0.0007 | \*\*\* |
| **m** | 0.1203 | 0.5977 |  |
| **theta** | -0.0001 | 0.1455 |  |
| **w2** | 1.0021 | 0.7695 |  |
| Signif. Codes: 0.01 '\*\*\*', 0.05 '\*\*', 0.1 '\*' | | |  |
|  | | |  |

These two tables report estimation results for GARCH-MIDAS models with a time-varying GARCH coefficient based on one MIDAS lag year of a daily macro variable, d\_WTI. As it can be seen from the first table, all the associated p-values are insignificant, apart from the beta coefficient. Therefore, it is not rational to take into account these results. On the other hand, the R output for the student-t distribution has 2 significant values. Therefore, this would be a better choice.

**Dynamic Quantile (DQ) Test**

The Dynamic Quantile test developed by Engle and Manganelli (2004). It includes putting certain linear constraints to the test in a linear model that links the violations to a collection of explanatory variables.

Consequently, it is based on a linear regression model of the hit variable on a collection of explanatory factors that includes a constant, the hit variable's lagged values, and any function of the previous information set suspected of being useful.

The following is the R code for the test and associated p-values:

> DQtest(y, VaR, VaR\_level)

[1,] 1.841899

[1,] 0.7648085

The output above suggests that the variables passed the Backtesting procedure since their related p-values are higher than the 0.05 benchmark level.

**CONCLUSION**

The below code is run for the ultimate selection of model through the Model Confidence Set (MCS) Test. The table attached at the end is the final results. According to the R output, the best fit model seems to be gjrGARCH-snorm. However, this finding does not match with the findings of this paper. In other words, gjrGARCH-snorm model has not been analyzed previously and further analysis must be done in order to be confident with this decision.

library(MCS)

VaR.comp=list()

for( s in specifications ) {

VaR.comp[[s]] <- as.data.frame(roll.comp[[s]], which = "VaR")[, 1]

}

Loss <- do.call(cbind,lapply(specifications,

function(s) LossVaR(tau=0.01, realized=tail(d\_WTI\_xts,150)/20,

evaluated=VaR.comp[[s]]/20)))

colnames(Loss) <- specifications

data(Loss)

SSM <- MCSprocedure(Loss = Loss, alpha = 0.2, B =1000 , statistic = "Tmax")

|  |  |  |  |
| --- | --- | --- | --- |
|  | Rank\_M | v\_M | MCS |
| gjrGARCH-snorm | 1 | - 2.735953838 | 1.000 |