Commodities Return During A Recession

IEOR4150

Valter Levander, Javier A Arevalo, Lorenzo Rega

Table of Content

Overview

For our project, we selected five futures contracts tracking the price of different commodities during a recession; specifically the dot-com bubble. We hope to gain insight into the correlation of different commodities during times of recession, with our main hypothesis that people reallocate their investment into commodities or cash during times of uncertainty and poor economic conditions.

First, we observed and analyzed the log returns of each ticker. We use the results to compare log-return and correlation between two chosen tickers.

The project is written using Python3. The code is in a Jupyter Notebook and can easily be modified to test different date ranges and tickers.

Data Set

Our data set consists of the daily close prices of five different Futures contracts that track different commodities, starting from the first day of the year that the dot-com bubble occurred (2001-01-01) until the end of the year (2001-12-31). Therefore, a total of 246 data points were collected for each ticker, corresponding to the number of trading days during the 2001 calendar year.

The data comes from Yahoo Finance, accessible with the Python3 library *yFinance*. The tickers for the different Futures commodities we used are

1. GC=F – CME Future tracking the price of gold
2. SI =F – CME Future tracking the price of silver
3. HG=F – CME Future tracking the price of copper
4. NG=F – CME Future tracking the price of natural gas
5. CL =F – CME Future tracking the price of crude oil

The commodities were chosen to track specific markets in the metal and oil sectors. Since we are using Futures contracts, we have taken into consideration the different contracts and the rollover dates. The data switches to the upcoming contract once the total trading volume, for a single day, of the new contract is larger than the previous one. This is automatically done by Yahoo Finance by adding “=F” at the end of each ticker.

Project Goals

We chose to focus on the commodities oil, natural gas, and precious metals during a somewhat recent recession. We observed that during the Covid-19 recession, large capital was allocated towards the tech sector and precious metals. Precious metals such as gold increased in value by about 23.8%, and silver was up by around 44.6%.

Currently, there are speculations and strong indications that the US is heading toward a new recession. We can see that a lot of capital is being allocated toward oil and natural gas. Natural gas peaked at a 150% gain for the year, crude oil reached a peak of 60% gain for the year.

Despite a lot of different advice regarding how to allocate your money during times of financial uncertainty, we believe that a correlation between different commodities can indicate the coming direction of the markets. If investments are being made into commodities, could this be interpreted as people attempting to protect their capital against uncertain future returns? Can we find a correlation between different commodities that can help indicate a reversal of a bear market? Is there a correlation between different commodities during a recession that we can expand and derive short-term trading strategies from?

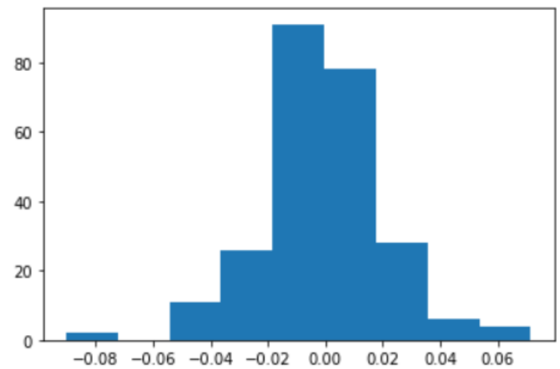
To investigate further, we check the log returns of the closing prices and compare their correlation, calculate confidence intervals, interpret probability distributions, and perform linear regression to find a simple trend that can somewhat indicate a relationship between recessions and investments in commodities.

Later on, we perform comparisons between two tickers to find correlations between their returns during the recession. Not specifically if they increase or decrease in price over time, but the correlation between the commodities' log return. Interpreting the confidence intervals for the variance and mean of each.

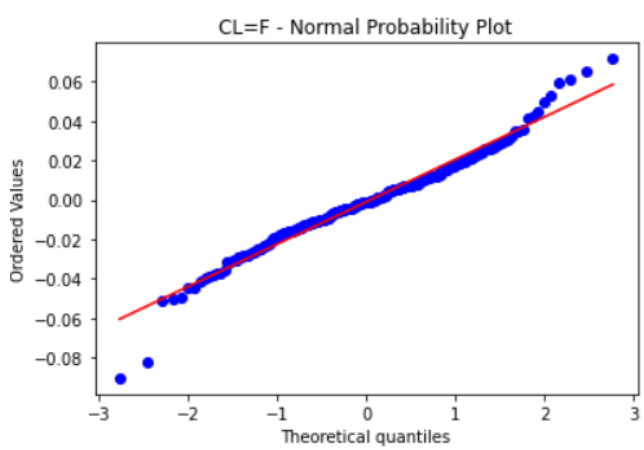
One Stock Analysis

In this section, we are going to show the capabilities of our Jupyter notebook. To do that, we are going to use Crude Oil as an example (Ticker “CL=F”).

First of all, we are able to plot a histogram of the log-returns. The outcome looks pretty much like a normal distribution.



This is not enough to prove it, and for this reason, we used a normal probability plot as well. In this kind of graph, if the points are on the red line, it means they do fit a normal distribution.



In the case of crude oil, it looks like the actual distribution is something similar to a normal one, but with fatter tails. We can see that from the last dots that are very far from the line.

We then computed the confidence intervals for the mean and the variance.

```
CI mean        (-0.003662656368559891, 0.0017745760002297615)
CI variance    (0.0003955765437776227, 0.0005640957115420063)
```

We had to estimate the unknown variance of the population with s^2. Then we standardized the mean. Since we estimated the variance, the standardized mean had a t-distribution with $n-1$ degrees of freedom (where n is the number of data points).

Crude oil, just like the other commodities, has a 95% confidence interval for its mean log-return that is very close to zero.

For the confidence interval of the variance, similarly, we get a 95% confidence interval close to zero. This shows a low variability of these commodities during the recession period.

In the end, we performed a regression of the log-return on time. The result was that the coefficients were all close to zeros, as well as the R^2. Suggesting that there is no linear relation between time and log-returns.

In particular, Crude Oil had a negative R^2. That means the model was so bad that its predictions are worse than a constant function that always predicts the mean of the data.

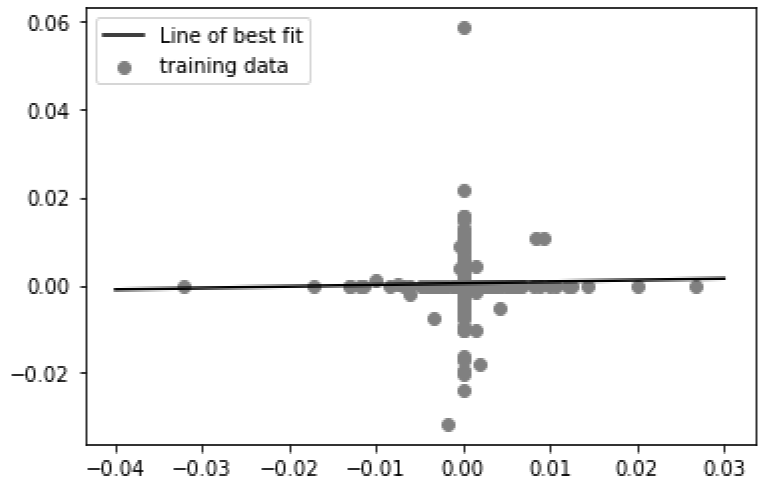
```
Intercept:      0.0029451008257134585
Coefficent:     -3.5631788674094434e-05
R-squared:      -0.010510488941555707
```

Two Stock Analysis

In the Two stock analyses we were evaluating whether any two stocks from our chosen set had any meaningful correlation or linear relationship. From our results, there was no clear evidence of a linear relationship or correlation between any two stocks. As evidently seen in the line of best fit between the two stocks it is evident that the line of best fit (the result of linear regression) is really similar to a horizontal line, $R^2 \approx 0.01$. What a horizontal line of best fit implies in statistics is that there is no correlation or linear relationship between the two variables (one stock's log-normal price vs another stock's log-normal price).

```
Line of best fit

Intercept:       0.00040474463534913443
Coefficent:      0.036163261744247735
R-squared:       0.01116262369194565
```



Given the fact that we used log-normal prices, this is not surprising. One of the drawbacks of using log-normal prices is that the prices end up appearing much closer to each other, due to the logarithmic transformation, making it difficult to find a linear relationship between them. Additionally, the team concluded that more data points needed to be used in the regression (using 2-5 years of data instead of less than 1 year of data) to be able to determine if the data is uncorrelated or if a linear relationship between any stock does exist. Furthermore, we believe that leaving out important factors in the pricing of these instruments, such as volatility, is a part of the reason why there was no evidence of a clear linear relationship between the variables we chose to analyze.

Conclusion

Based on our statistical analysis performed on our chosen financial derivatives, the team concluded that linear regression is not the most appropriate method to analyze a potential correlation. Based both on our individual stock analysis, as well as our two stock analyses, there was no evidence of a correlation or linear relationship between the date and close prices as well as no clear evidence of a linear relationship between any two given instruments. In both of our linear regression analyses, the line of best fit had various characteristics of a horizontal line implying that the two variables did not have a linear relationship.

The team was not surprised by the result since we believe significantly more data is needed to predict the closing prices of advanced financial derivatives such as futures. Additionally, since our analysis was limited to only using data under less than a year and during a recession period, and we were using log daily return, the data points were in the same order of magnitude, and relatively close to each other, making it significantly harder to find evidence of a linear relationship.