PHY408 Final Project: Correlations of different pairs of ETF prices

Darya Zanjanpour

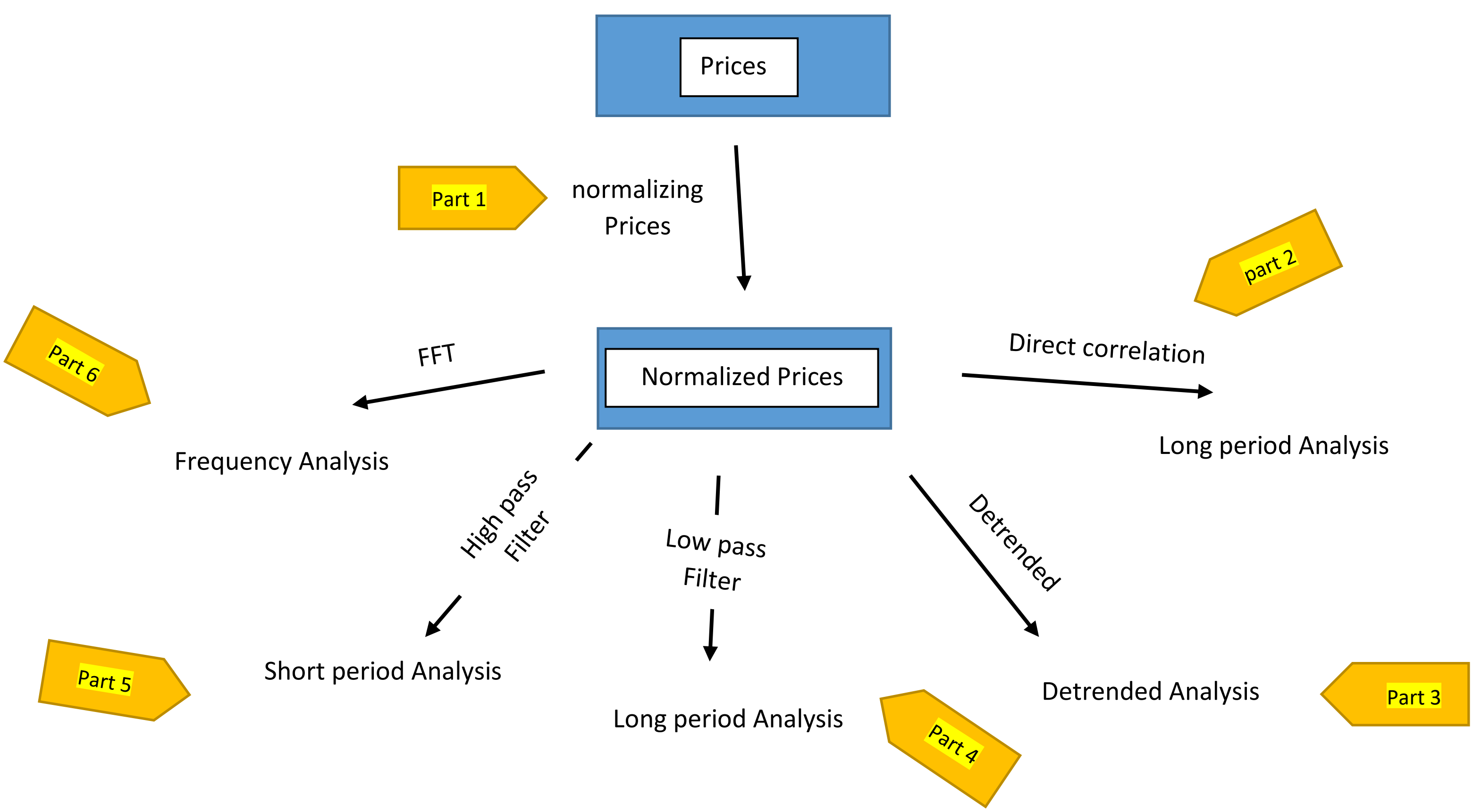
April 19th 2022

Abstract

In this report, the prices of exchange-traded funds investing in Gold, Silver, Oil, Gas and Bitcoin, over the past 9 years, from 2013 to 2022, have been used to investigate correlations between them. Four different approaches were used to investigate different aspects of similarity between prices: correlation was directly performed on prices and also performed after quadratic detrending, after low pass filtering, and after high pass filtering. The different correlation results reveal various things about the relationships between the investigated markets.

1 Introduction & Motivation

The question we are trying to solve using time series analysis techniques is: "How do trends in ETF prices correlate with each other in general, in long-term trends, and in short-term trends?". To answer this question, I have followed the following flowchart to perform different types of analysis for the data.



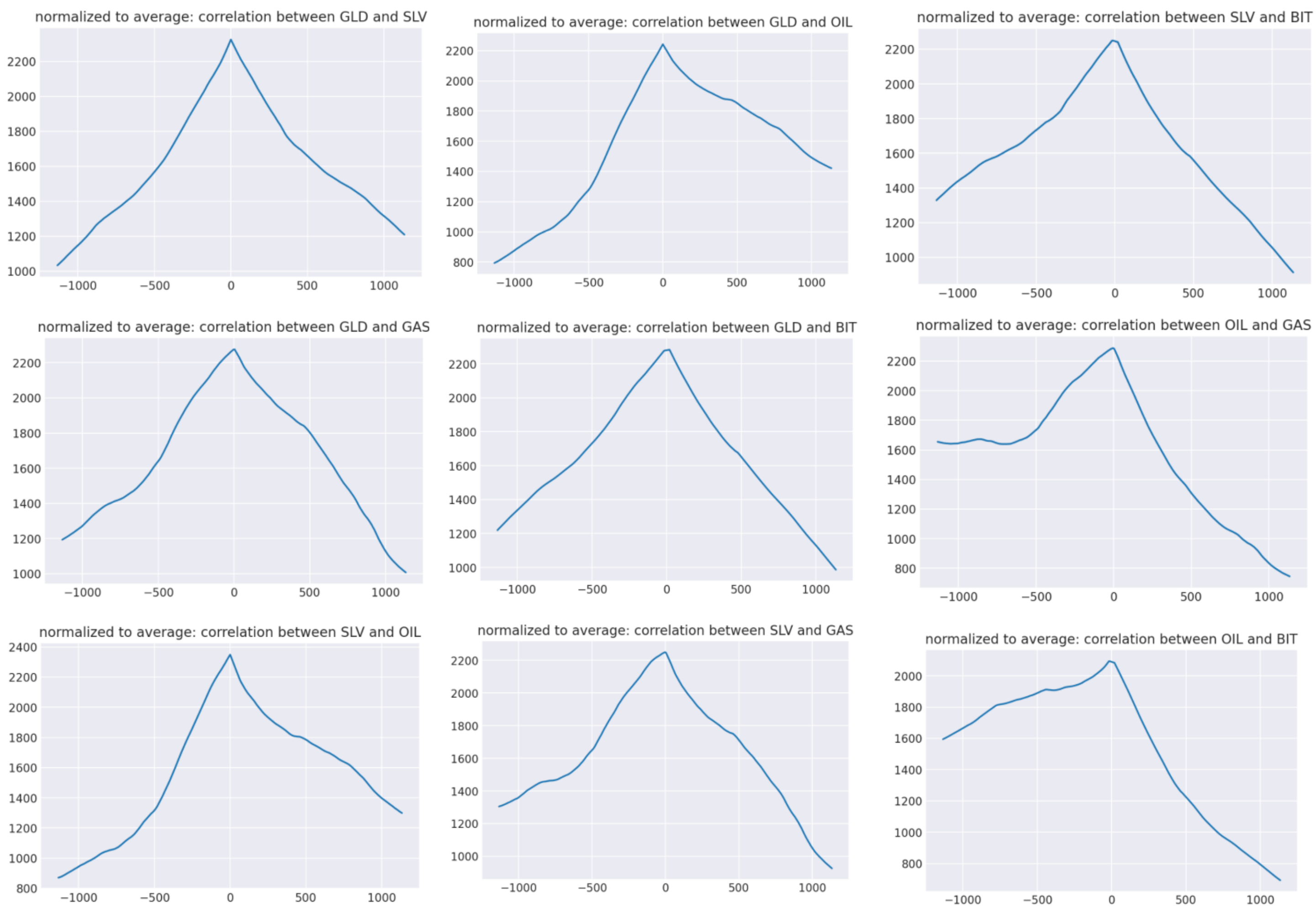
In this report, I have used gold, silver, oil, gas and Bitcoin market prices. This selection is based on the availability of data over recent years. I limited the data selection to the past 9 years, back to 2013, due to unavailability of some datasets before 2013. The aim was to have the same time interval for each dataset to simplify analysis. We have to consider that by having a larger data sets, the result will be more trustable and accurate. For instance, having all data sets for 50 years would be idealistic for the purpose of this analysis but due to unavailability of the data set, I limited myself to most recent data set which are available for all of the samples. All the analysis can be done on the larger sample size with more accurate results for the future investigation. I used Ishan shah's method [1] to import my data set to python. The final result is as the following:

Figure 1: imported data for ETF prices from 2013 to 2022. Note that the scales of the prices are quite different from each other. For instance, gas has the largest scale and Bitcoin has the smallest scale. Each dataset was therefore divided by its mean for normalization. The lower right graph shows the normalized data sets (Part 1) all in one plot.

2 Analysis and Results:

- Direct cross correlation

After normalizing each data set, we go back to the original question of this report. To answer this question, we introduced 4 general strategies for different types of analysis. The first strategy is cross correlating the datasets without applying any other filtering which shows how prices correlate to each other overall. (Part 2) Both long and short term trends are reflected in these correlations, but they may be dominated by longer term trends due to their relative size. These correlations produce the results shown in the plots below.



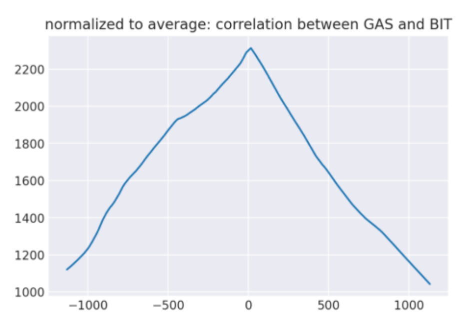
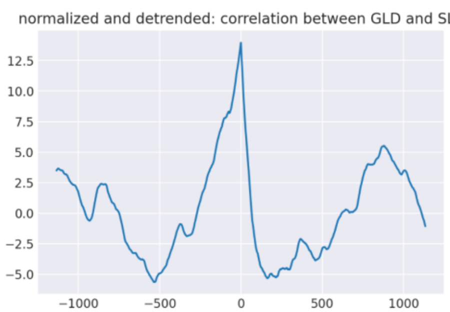


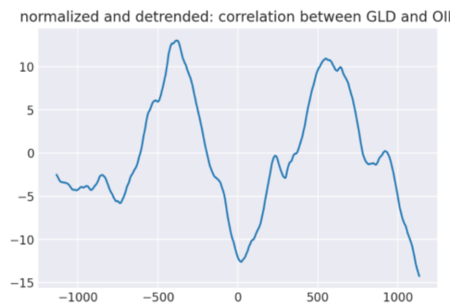
Figure 2: Cross correlated results of each pair of prices. Horizontal axes show time shift in business days while vertical axes are in a relative unit indicating the strength of the correlation.

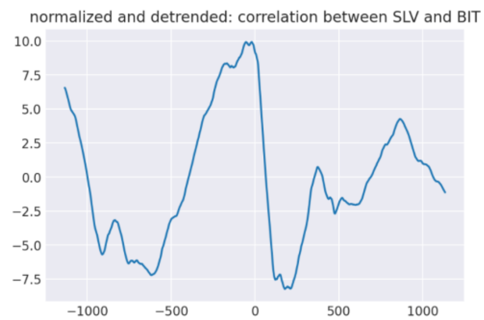
Generally, since all correlation results have a maximum at 0, it can be seen that all these prices correlate positively with each other in general.

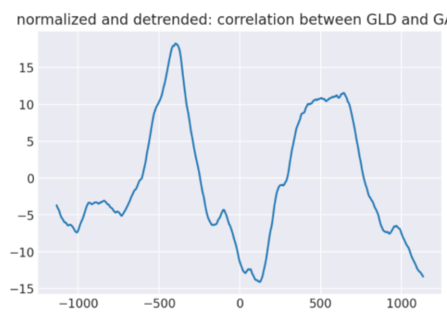
- Detrending and then cross correlation:

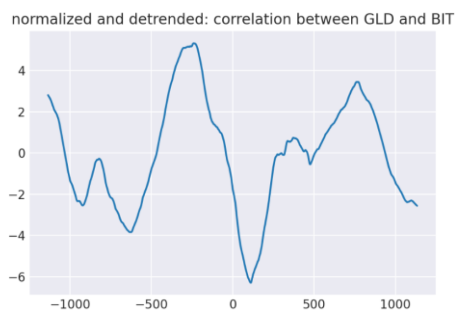
Moving to part 3 of analysis, in this section, all datasets are detrended before every pair is cross-correlated. By detrending the data, the overall trend in the data is eliminated. Cross-correlation results therefore better represent correlations between shorter-term trends. Instead of a first order linear fit, a second order quadratic fit was used for detrending because it visually appeared that a second order fit would be much more effective in removing long-term trends. **(Appendix 4.1)** I then performed cross-correlation on each pair and observed how the correlation results were changed by detrending.

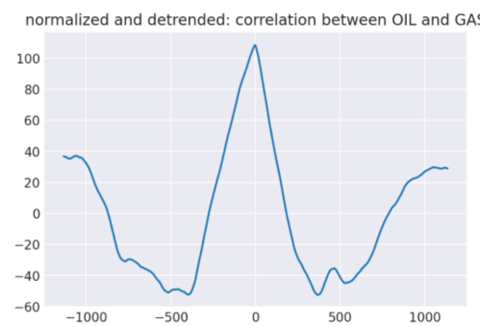


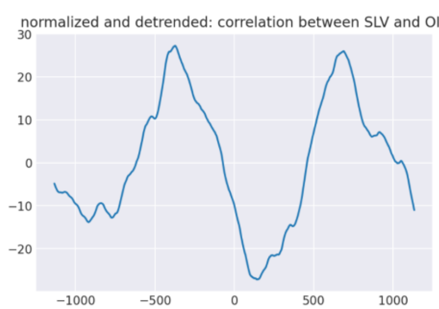


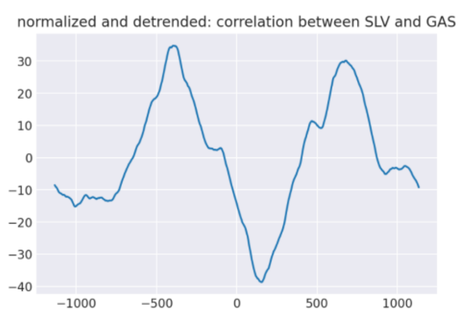


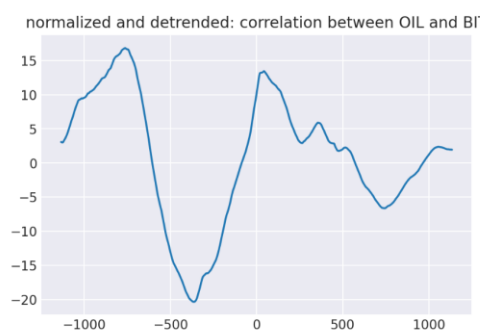












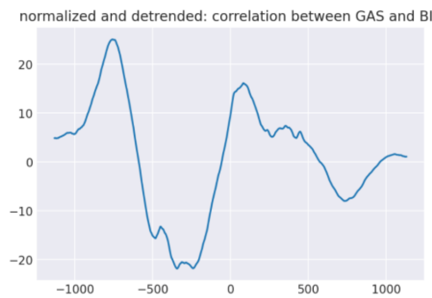


Figure 3: Cross correlated results of each pair of datasets after normalization and detrending. Horizontal axes show time shift in business days while vertical axes are in a relative unit indicating the strength of the correlation.

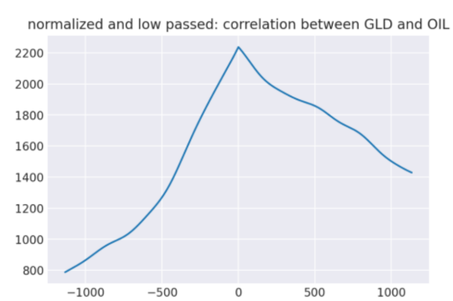
Compared with part 2 (cross correlation without detrending) most of the pairs lose their correlation. Correlation is lost when the overall pattern represented by a quadratic fit is removed. The only pairs for which correlations survive are: **[gold and silver], [gas and oil] and [silver and bitcoin]**. Note that the strength of the maximum at 0 shift for oil and gas is 10 times stronger than that of the [gold and silver] and [silver and bitcoin] pairs. The short term correlation between gold and silver is expected because they are both precious metals. The cross correlation between oil and gas is also reasonable for similar reasons. The correlation between silver and bitcoin in the short term, meanwhile, is unexpected and interesting!

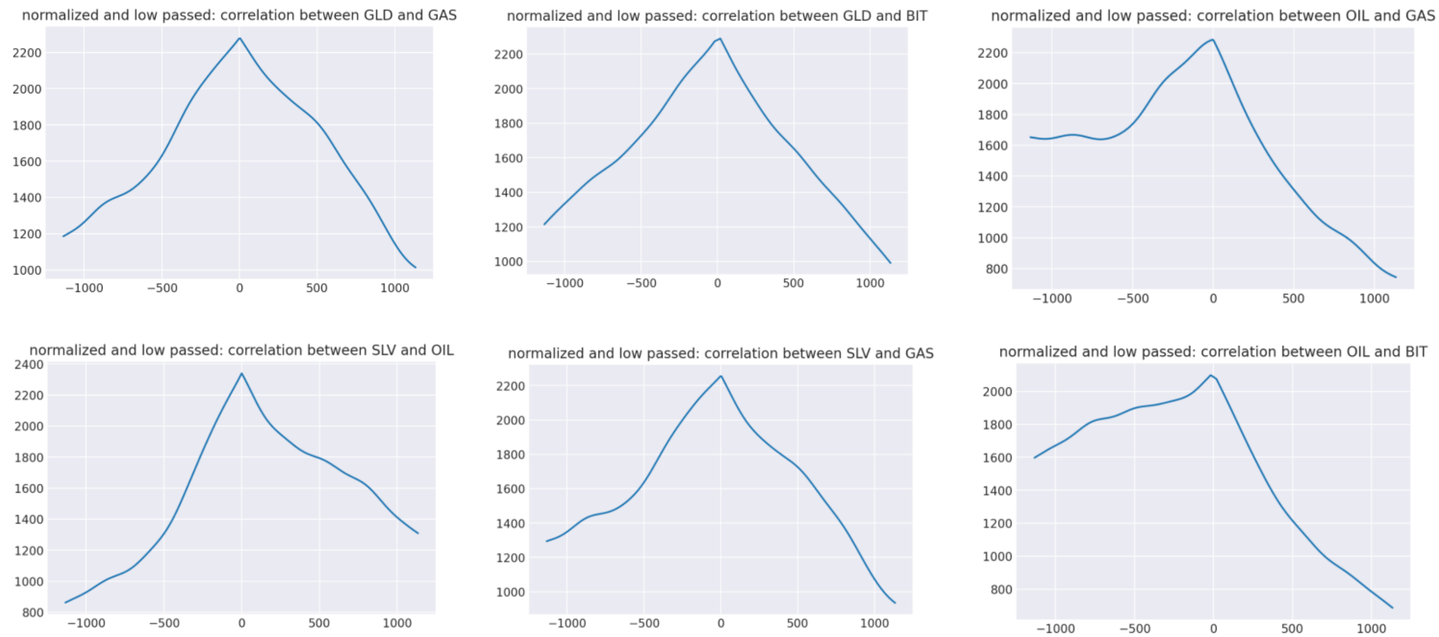
- Applying low pass filter and then cross correlation: (Using Fast Fourier transform)

In this section, part 4, after normalizing the datasets, a low pass filter is applied before correlation. In this filter, each dataset is first linearly detrended to limit edge effects. An FFT is then applied. Frequencies above a threshold are removed. The threshold has been manually set to 1/250, corresponding to a period of 250 business days, because that is the approximate number of business days per year. After removing unwanted frequencies, the data is transformed back to the time domain using the IFFT and the trend is added back. **(Appendix 4.2)**

After applying the low pass filter, cross-correlation is performed on each pair. In addition to plots of correlation results, plots of each pair together after low pass filtering are available because they help interpretation. **(Appendix 4.3)**

Cross-correlation results after low pass filtering are as follows:





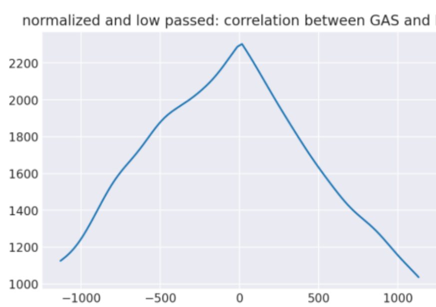
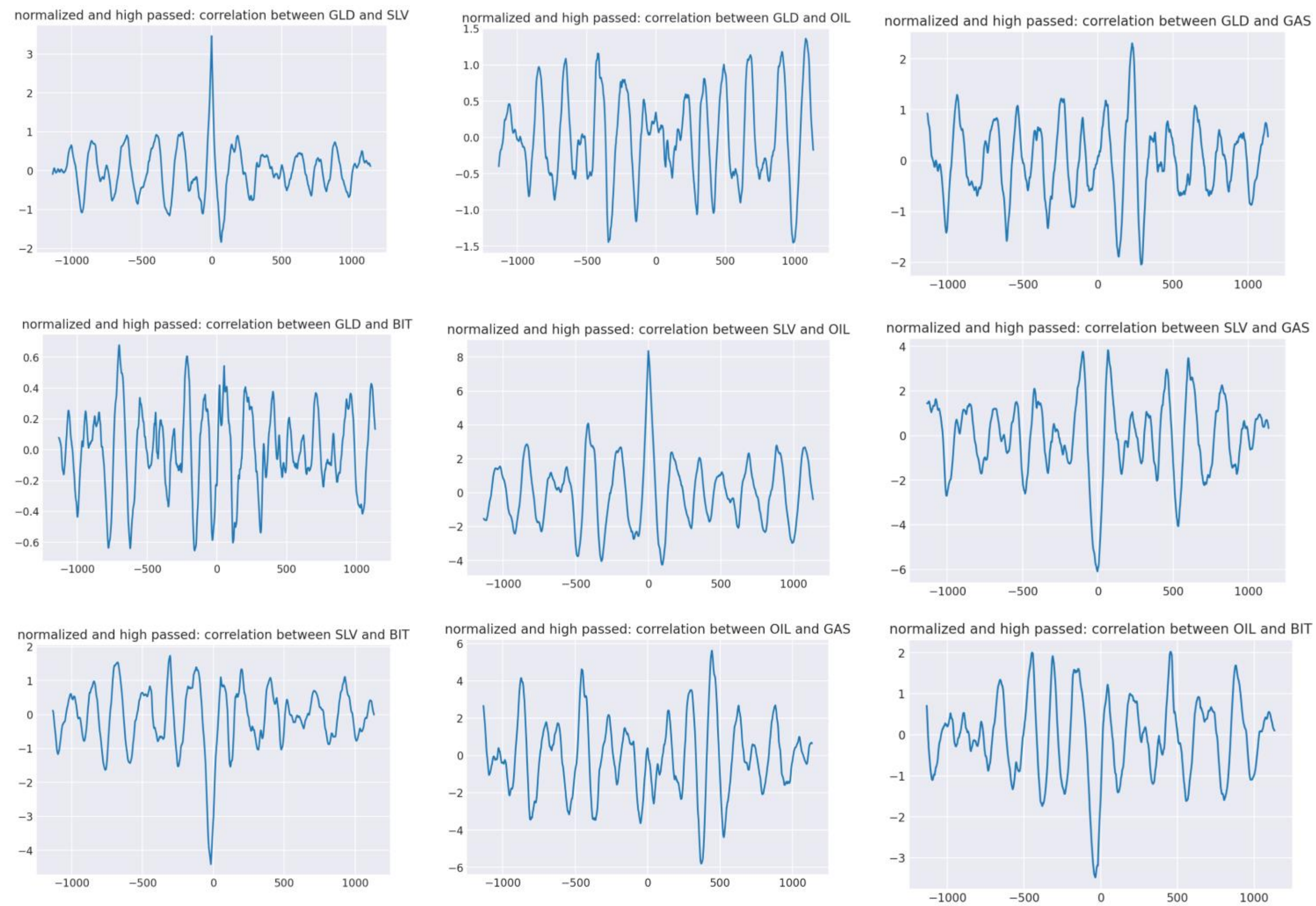


Figure 4: Cross-correlation results after performing low pass filtering. Units are the same as for previous cross-correlation results.

The cross correlation results for all pairs in figure 4 are very similar to the direct correlation results which shows that low pass filtering has little effect on the relationships between datasets. The plots of pairs of low pass filtered data (not correlation results) in **Appendix 4.3** are more useful for visually assessing relationships.

- Applying high pass filter and then cross correlation: (Using Fast Fourier transform)

The same strategy as part 4 has been applied in part 5 except that frequencies below the threshold are removed instead of frequencies above the threshold. **Appendix 4.4** shows high pass filtering results. Just like in part 4, we also have pair plots for high pass filtering in **appendix 4.5** which in this case are not as useful as the equivalent for low pass because these plots mostly contain noise. Here, cross-correlation results are more useful than visual comparison of results. The cross correlation results for the high pass filter are as follows:



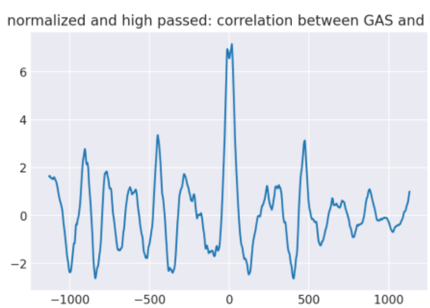


Figure 5: Cross-correlation results after performing high pass filtering. Units are the same as for previous cross-correlation results.

This time the result is quite different from previous cases. Few pairs have peaks at 0 shift. In some cases, there is a peak there but its height is no greater than the surrounding peaks. These may not represent significant true correlations. In the plots of pairs of filtered datasets in **appendix 3.5**, there is similarly little visible correlation. Correlations appear to exist for **[gold and silver] and [gas and bitcoin] and possibly for [silver and oil].** Interestingly, the correlation between [oil and gas] appears to be lost in high pass filtering: their short-term trends do not correlate.

- Frequency Analysis (Fast Fourier Transform) **(In Appendix 4.6)**

3 Discussion:

Analyzing procedure :	Result:
Direct correlation: Both long and short term trends are reflected, but they be **dominated by longer term**	All the cross-correlations are highly positive, indicating strong overall correlations between all pairs.
Detrended and then correlation: Tracking the correlation in semi long-range	The correlated pairs are **[gold and silver], [gas and oil] and [silver and bitcoin].** The strength in correlation for [oil and gas] is 10 times stronger than for [gold and silver] and [silver and bitcoin].
Applying low pass filter and then correlation: Tracking the correlation in semi-long range	Cross correlation panel: All the cross-correlation results are highly positive. This suggests that long-term trends are correlated between all markets. Low pass filter result panel**: [oil and gas], [gold and silver] and [Silver and bitcoin]** have clearly visible resemblances to each other.
Applying High pass filter and then correlation: Tracking the correlation in short range	No strong correlations. Correlations appear to exist for **[gold and silver]** and **[gas and bitcoin]** and possibly for **[silver and oil].** Interestingly, the correlation between ~~[oil and gas]~~ appears to be lost in high pass filtering.

According to the result summarized in the table above, in the long term, the behaviors of all ETF prices are quite similar to each other. After detrending or low pass filtering, highly correlated pairs are reduced to **[gold and silver], [gas and oil] and [silver and bitcoin].** The correlation between gold and silver is expected. The correlation between oil and gas is also reasonable. Having a correlation between silver and bitcoin, however, is very unexpected and interesting!

In short-term trends, according to the results after high pass filtering, few clear correlations exist. Small correlations appear to exist for **[gold and silver]** and **[gas and bitcoin]** and possibly for **[silver and oil].** Interestingly, the correlation between **[oil and gas]** appears to be lost in high pass filtering despite the fact that they are both fossil fuels. The correlation of short-term trends between **[silver and oil] and [gas and bitcoin]** is unexpected too.

Most of overall results were broadly as expected. One interesting and unexpected result is the correlation between **silver and bitcoin**.

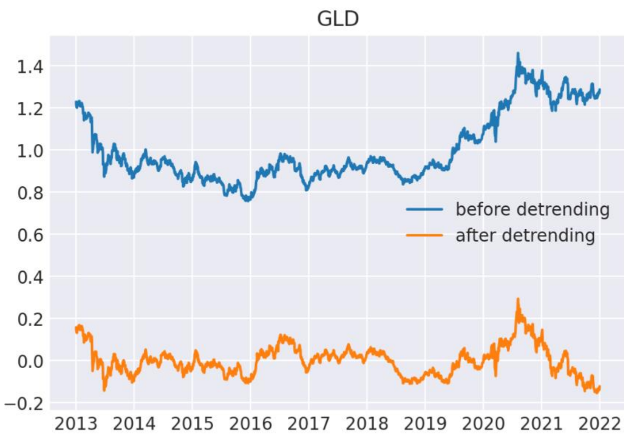
One option for improving results is to extend the time interval over which analysis is done. In our case, this could not be done due to limited availability of data from earlier years.

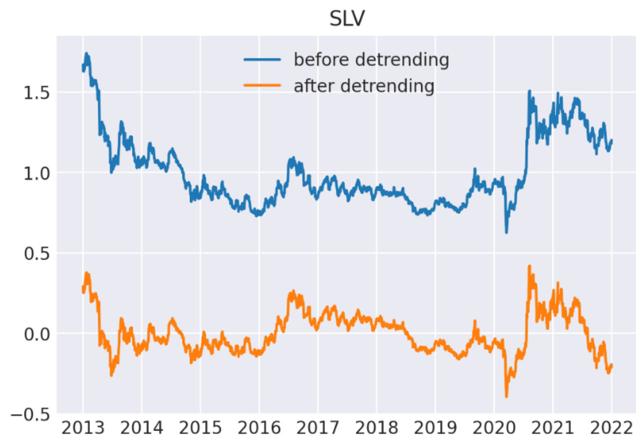
I consider cross-correlation after detrending to be the most useful type of correlation result. The result I consider the second most useful was the low pass filtering results plotted without cross-correlation, appendix 4.3. The filtering procedure contains detrending. I therefore believe that the most important and useful procedure in making results I find useful is **detrending**.

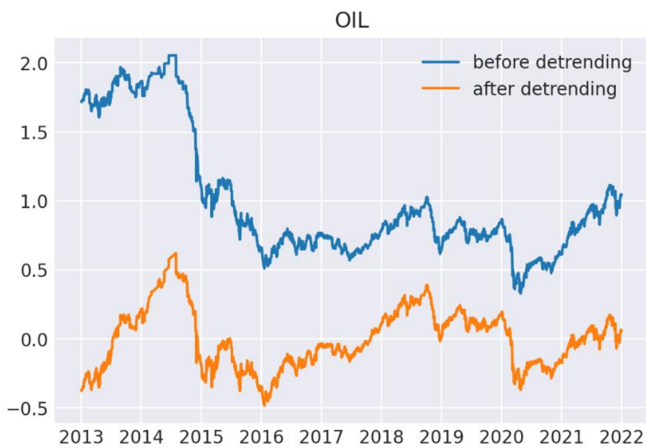
A direction for future work extending this project can be the construction of a predictive model. Based on the overall behavior of one market we are able to predict the other markets' behavior in the long term. This prediction can be useful for predicting the future using time shifts between markets. This report was focused on studying correlations without shifts. The next step would be analyzing the same data set to identify shifts.

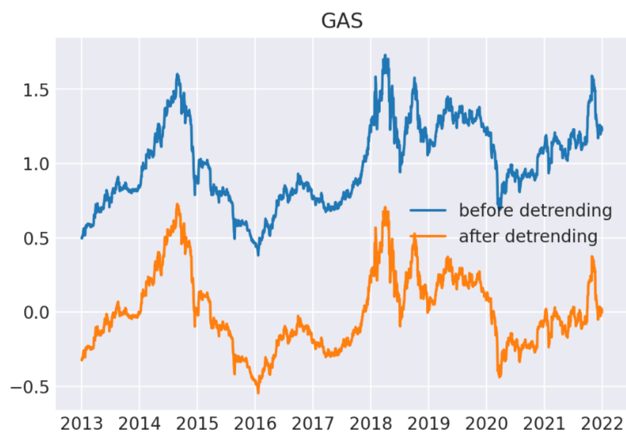
4 Appendix:

4.1 Detrending all data:









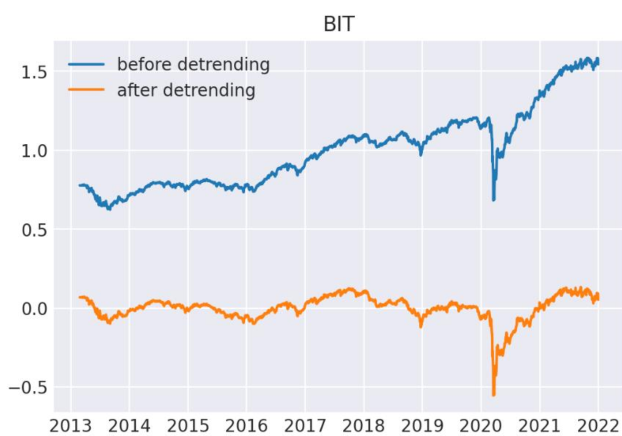
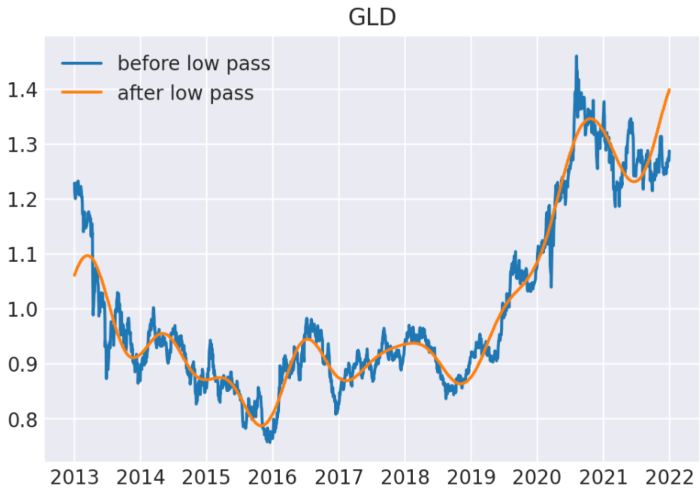
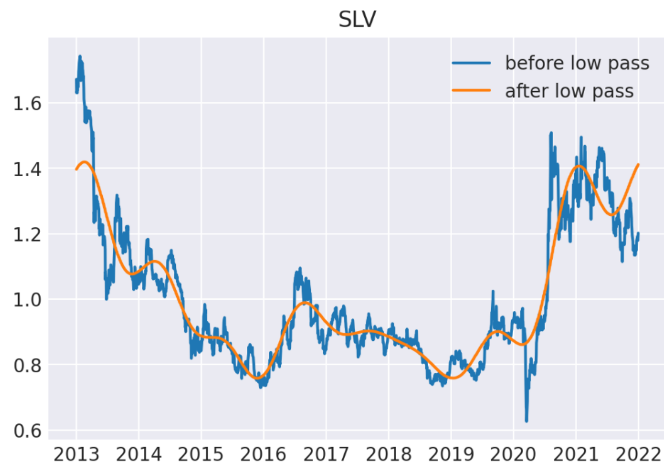
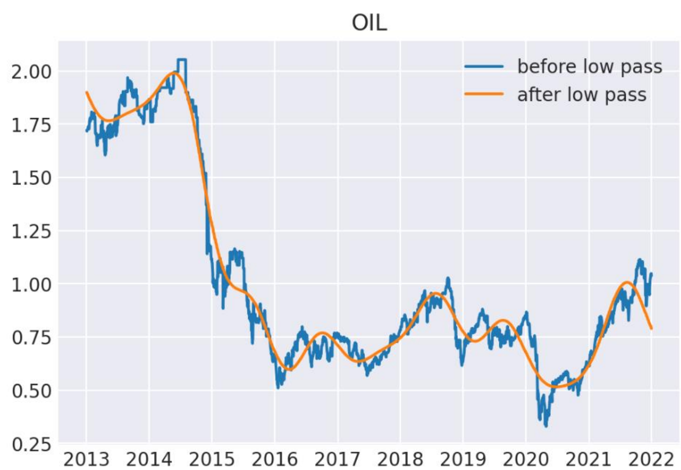


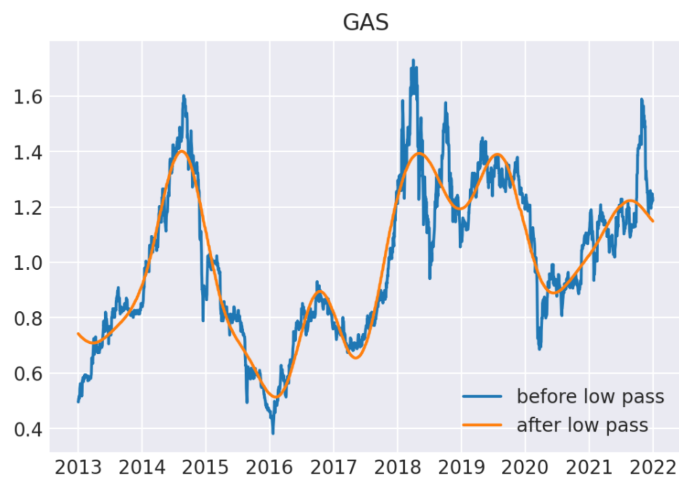
Figure: Detrending on each data set has been applied on the normalized data set. The vertical axis is relative price and the horizontal axis is in time unite.

4.2 Low pass filtering:









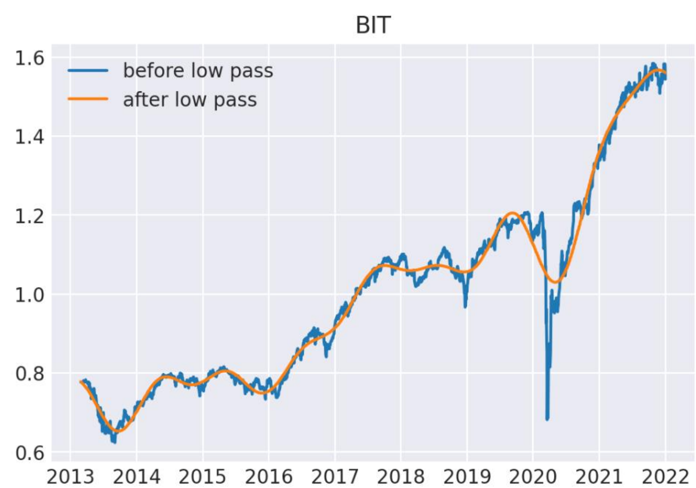
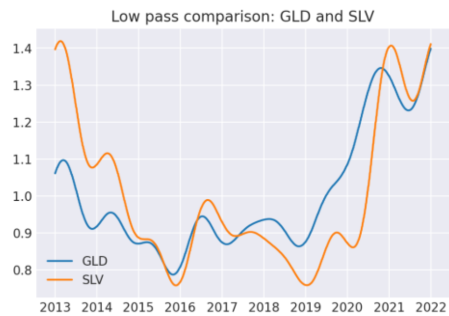
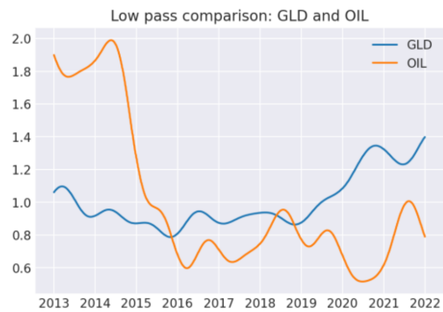
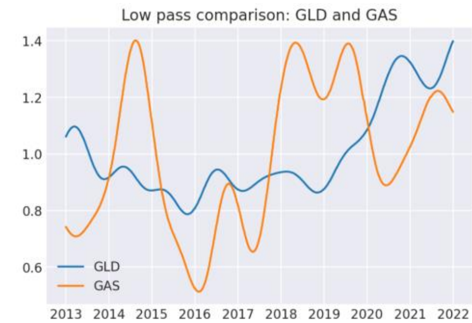


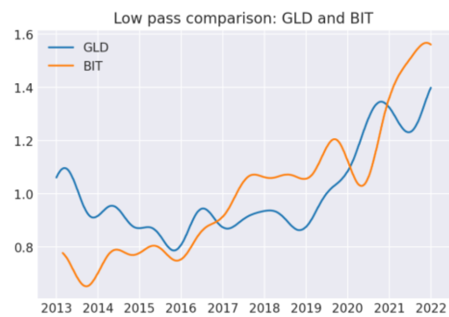
Figure: After normalizing and Detrending on each data set, low pass filter has been applied.

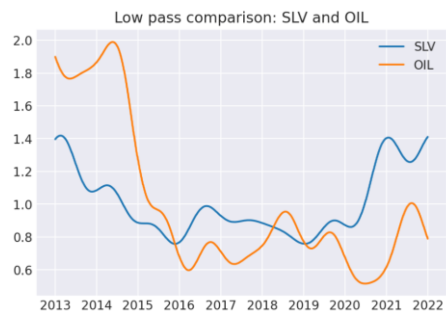
4.3 Low pass filtering. (Comparing each pairs based on the overall shape of the filter)

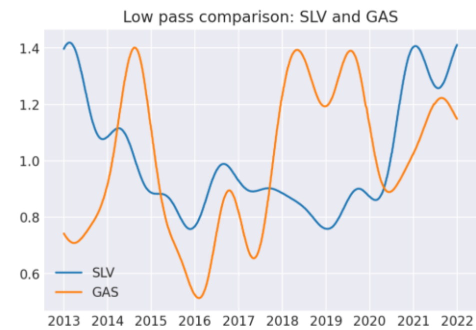


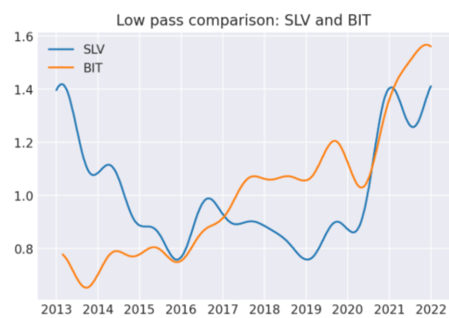


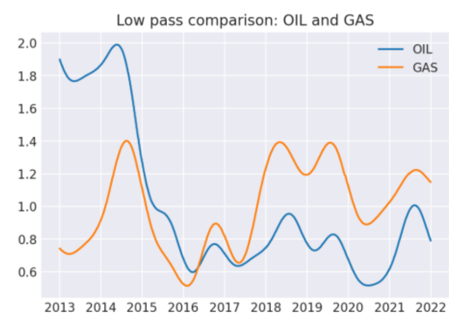


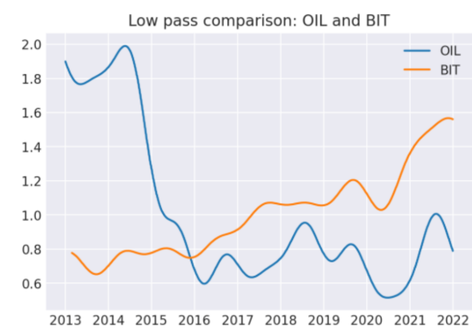












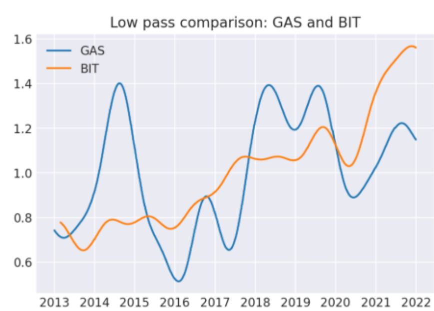
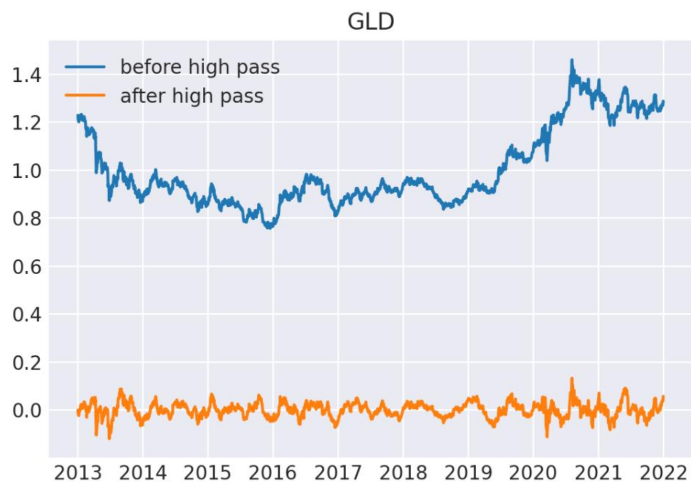
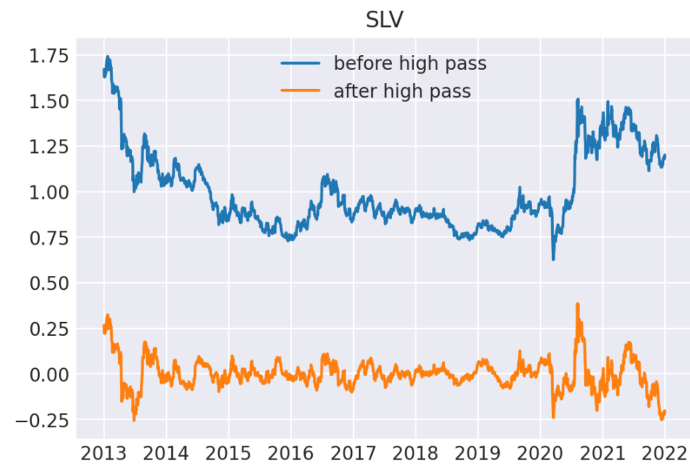
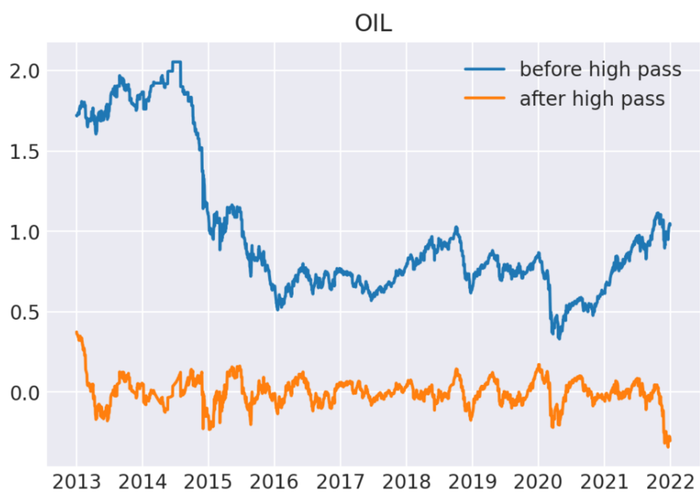


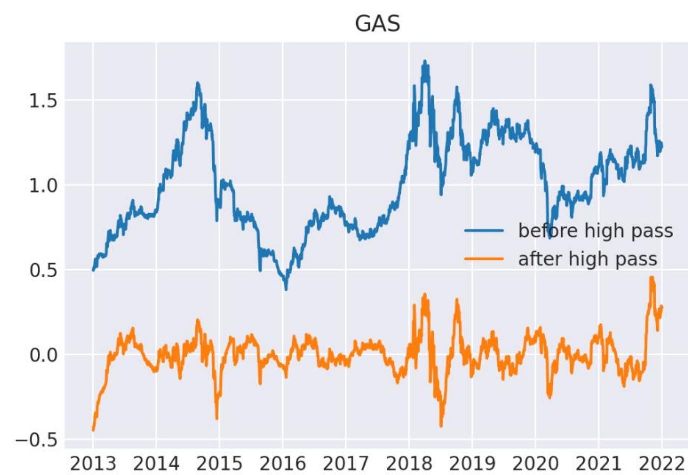
Figure: Low pass filter of each pair. The general pattern of each variable is an outcome of low pass filter. [oil and gas], [gold and silver] and [Silver and bitcoin] have the overall similar pattern which is consistent with the result in the trending section.

4.4 High pass filtering









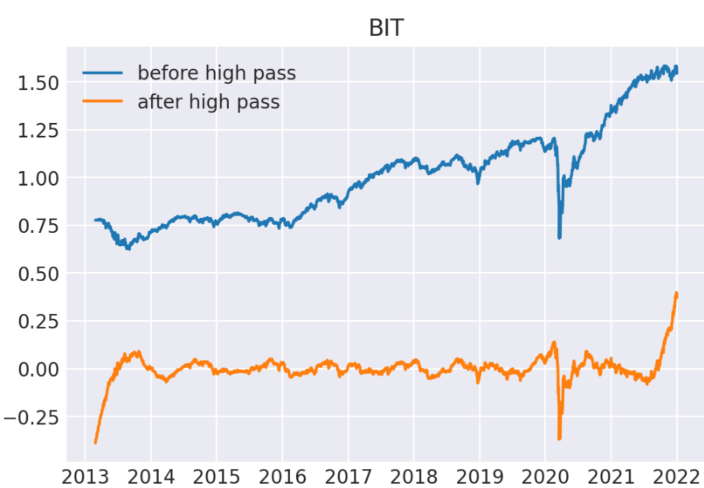
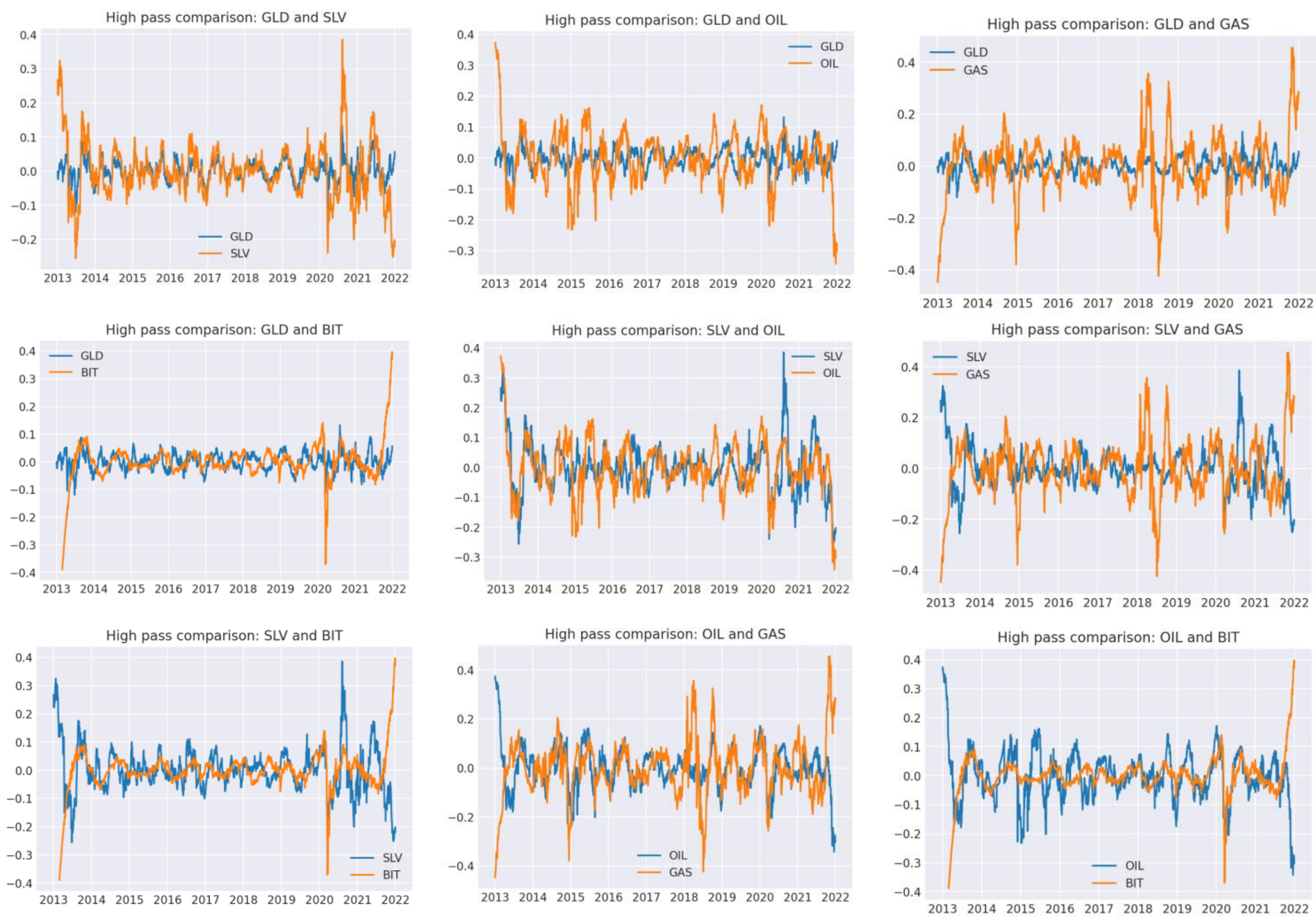


Figure: After normalizing and Detrending on each data set, high pass filter has been applied.

4.5 High pass filtering. (Comparing each pairs based on the overall shape of the filter)



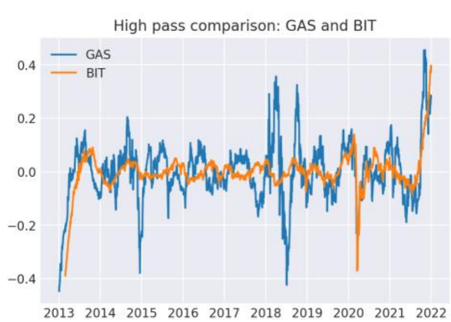


Figure: high pass filter of each pair. The general pattern of each variable is an outcome of high pass filter. In contrast with the same panel for low pass filtering, this panel does not contain useful information and the cross correlation panel works better for high pass filter case.

4.6 Frequency Analysis (Fast Fourier Transform)

In the final part, part 6, I used the FFT to find similarities in frequency spectra between detrended datasets. I limited myself to examining one of the best-correlated pairs, gold and silver, since in the rest there were no obvious frequency matches. Figure 6 is the FFT result for gold and silver. At very low frequencies, there is no obvious match, but at higher frequencies, in the range of 0.01 to 0.03, matches occur. Since higher frequency means lower period, this result can be interpreted as showing correlation of gold and silver in short-term trends. This is consistent with other parts.

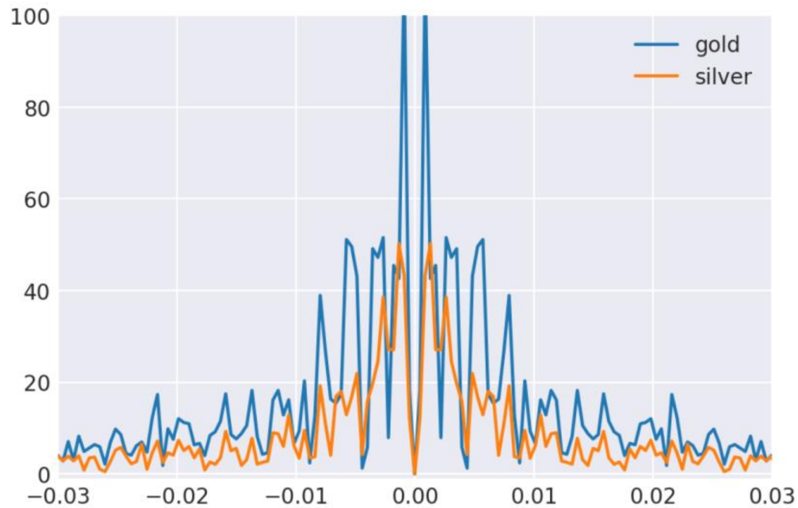


Figure 6: FFT transform of detrended gold and silver price data. The horizontal axis has unit of frequency = (1/day) and vertical axis is strength of the unit.

5 References:

[1] https://blog.quantinsti.com/gold-price-prediction-using-machine-learning-python/