MLT-05 Summary Document

Overview

This document summarises the MLT-05 lecture on Statistical Arbitrage using PCA. This lecture talks about building PCA from scratch and presents examples of using PCA to construct a stock index and doing statistical arbitrage using 2 stocks as well as using a portfolio of stocks.

Kommentar [1]: That seems to be no-connected with the text

Kommentar [2]: Fixed

The lecture covers the following topics:

- Building PCA from scratch
  - Covariance matrix
  - Eigenvalue and eigenvectors
- Example 1 - Volatility curve modelling using PCA
  - Level of the volatility curve
  - The curvature of the volatility curve
  - The tilt of the volatility curve
- Example 2 - Building a stock index using PCA
  - Running PCA with sklearn PCA function
- Cointegration vs correlation
- Test for cointegration
- Example 3 - Statistical arbitrage using PCA
- Example 4 - Portfolio statistical arbitrage using PCA

Kommentar [3]: Is it a volume curve or a volatility curve? Likewise in the next two sentences?

Kommentar [4]: Yes, my bad. It should be volatility. Fixed it.

Building PCA from scratch

Principal Component Analysis (PCA) is a dimensionality reduction technique. Simply put, it is used to reduce the number of independent variables in a dataset so that the remaining variables capture most of the information.

Eigenvectors
The vectors whose direction doesn't change while applying the linear transformation.

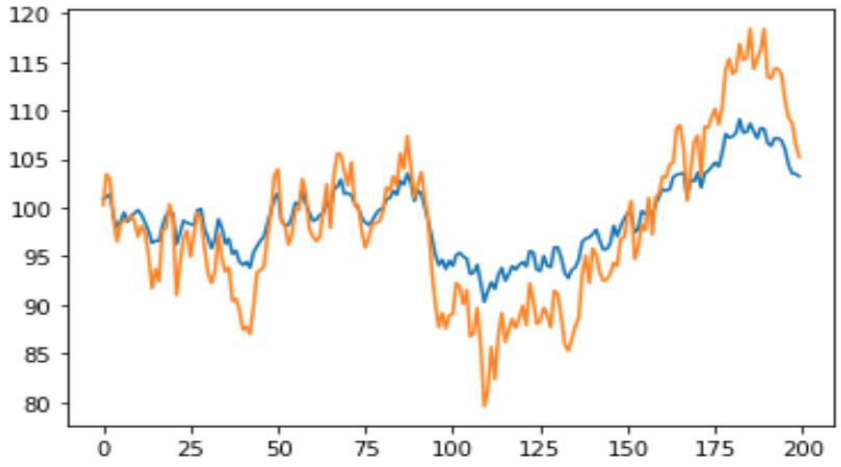
Eigenvalue
The scalar value that is used to transform eigenvectors.

They help in reducing the noise from the data by eliminating strongly correlated features. They form the basis of Principal Component Analysis (PCA). Let's look at an example.

The following are the steps to build PCA:

- Generate two series with different SD using random() function.

The following snapshot shows two randomly generated series with different standard deviations.

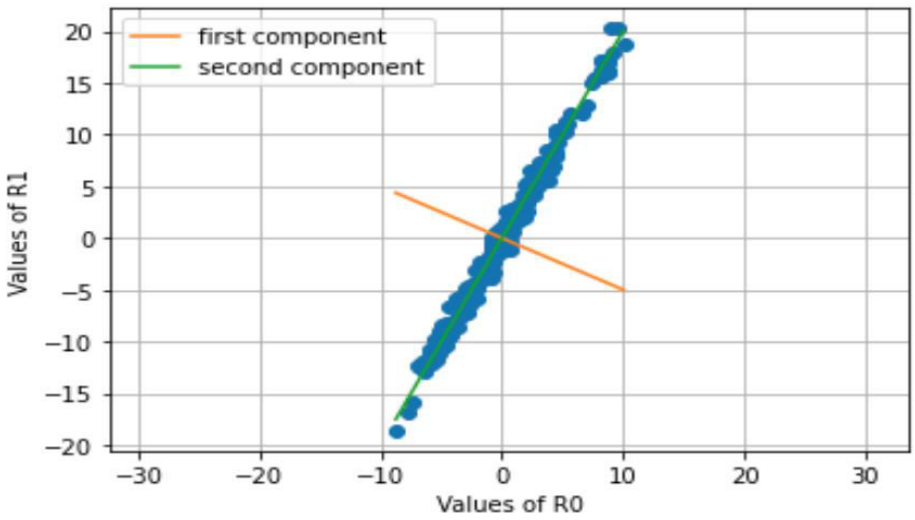


- Compute the covariance matrix using np.cov()
- Calculate the eigenvalue and eigenvectors of the covariance matrix.

The following snapshot shows the two calculated eigenvectors.

Kommentar [5]: maybe a word on eigen vectors/values could be interesting

Kommentar [6]: Agreed. Added.



Observation - Eigenvectors are orthogonal.

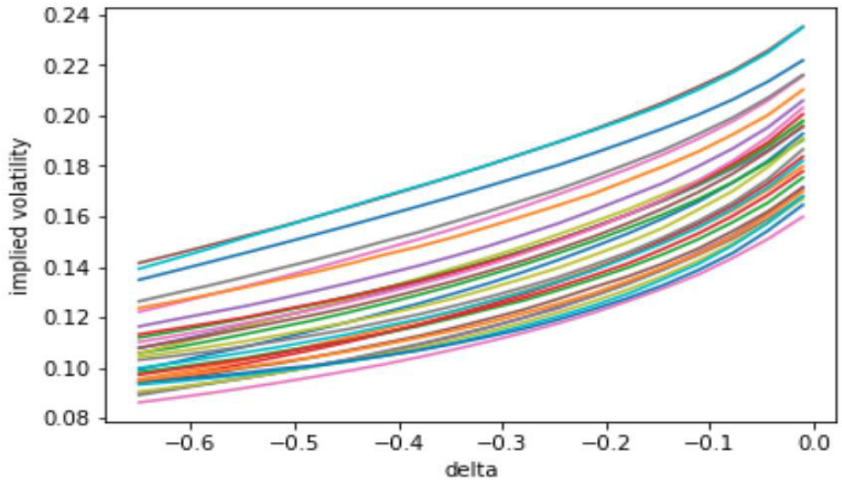
Example -1 Volatility curve modelling using PCA

- A widespread use case for PCA is modelling yield curves for bonds or implied volatility curves for options. Here, we look at the implied volatility (IV).
- The IV of an option tells you whether the option is relatively cheap or expensive.
- An ATM or at-the-money option (where the strike is closest to the price of the underlying stock) usually has a delta close to 0.5 while ITM or in-the-money options have deltas > 0.5 and OTM (out-of-the-money) options have deltas less than 0.5.

Here are the steps for volatility modelling using PCA -

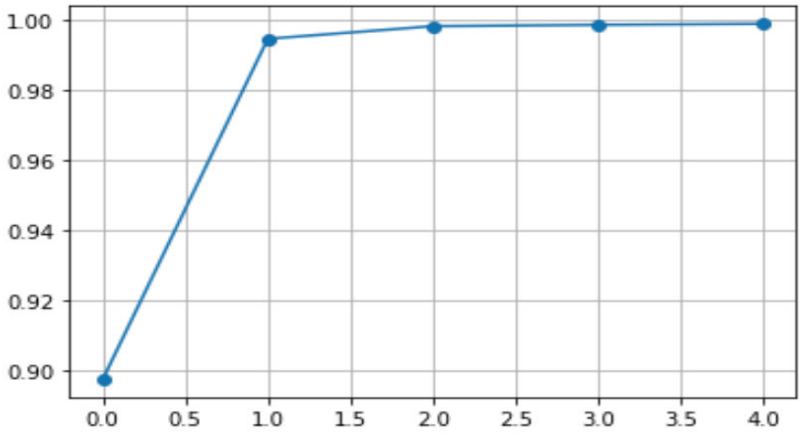
- Get volatility curve data.

The following snapshot shows the volatility curve for OTM put options. The x-axis is expressed in deltas, which is a way to normalise the strike prices of options.



- Perform the PCA and plot the most significant principal components.
- The cumulative sum of them is also called the "explained variance".

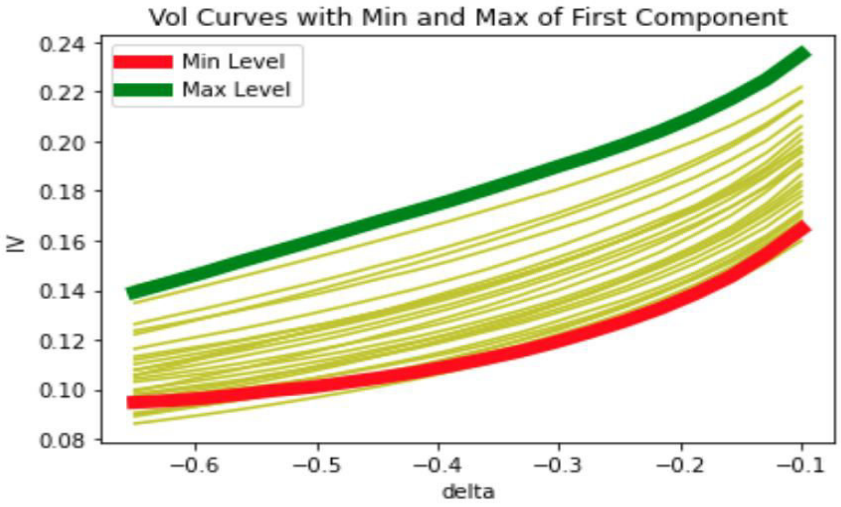
The following snapshot shows the significance of the principal components generated using the PCA.



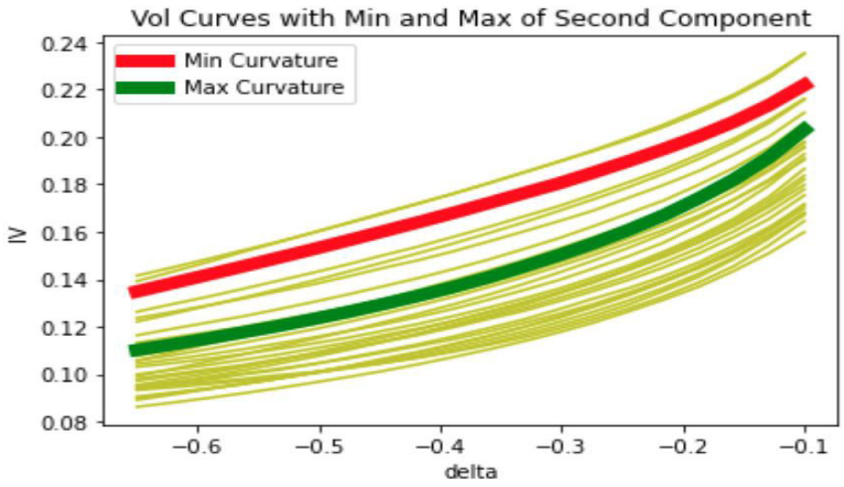
Observation - We can see that out of the 30 components, the four most significant ones can explain almost 100% of the variance.

- In some cases, such as in curve modelling, principal components have a meaning.
- The first three components in vol-curves denote the level, slope, and tilt curves.
- We can reconstruct our vol-curves quite accurately using these components.
- Here's a demonstration of the three components -

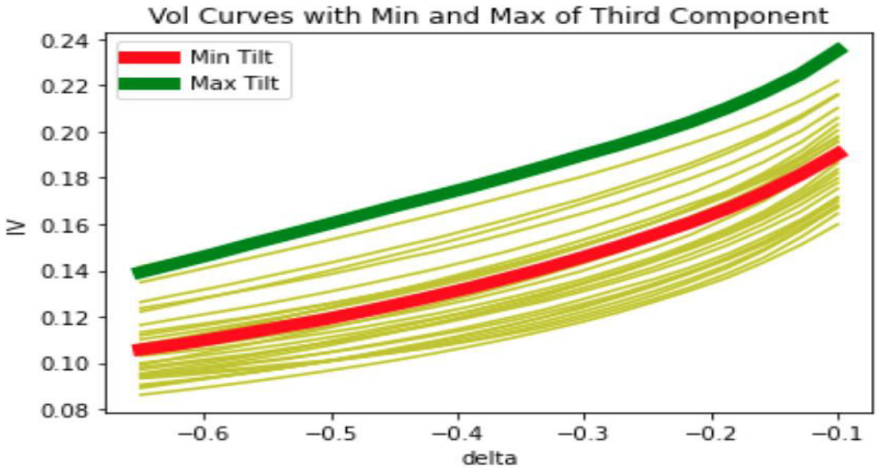
Level of the volatility curve



The curvature of the volatility curve



The tilt of the volatility curve



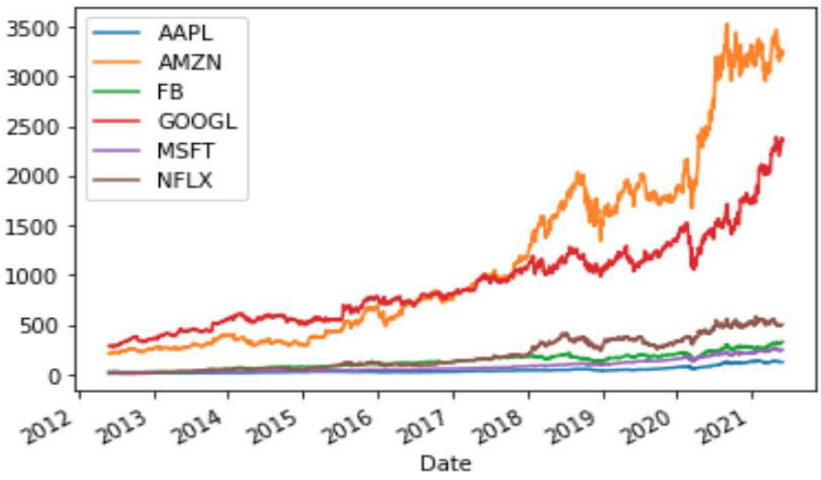
Note - Here, we have built a three-component model for the volatility curve.

Example 2 - Building a stock index

Here's an example of building a FAANG index. FAANG stands for Facebook, Amazon, Apple, Netflix and Google. Here are the steps to create a stock index using these securities -

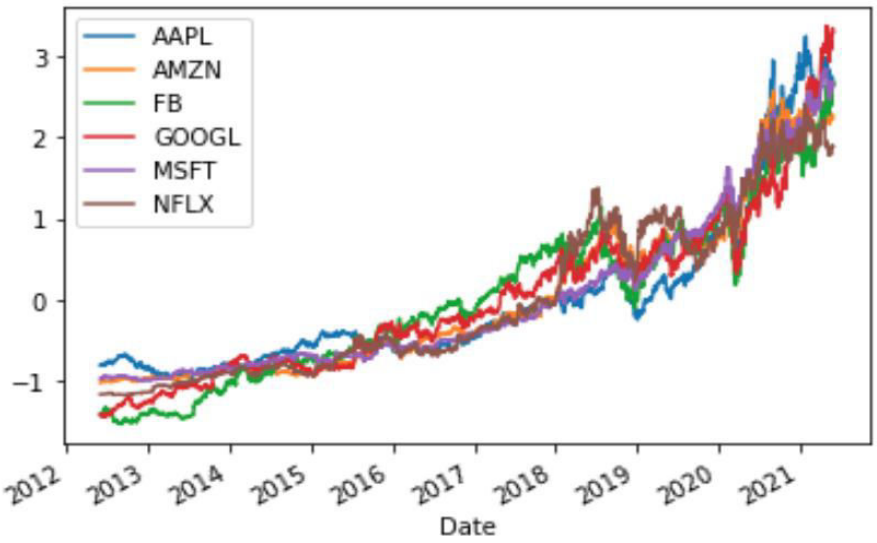
- Download the data from Yahoo finance.

The following snapshot shows the adjusted close price of FAANG stocks.



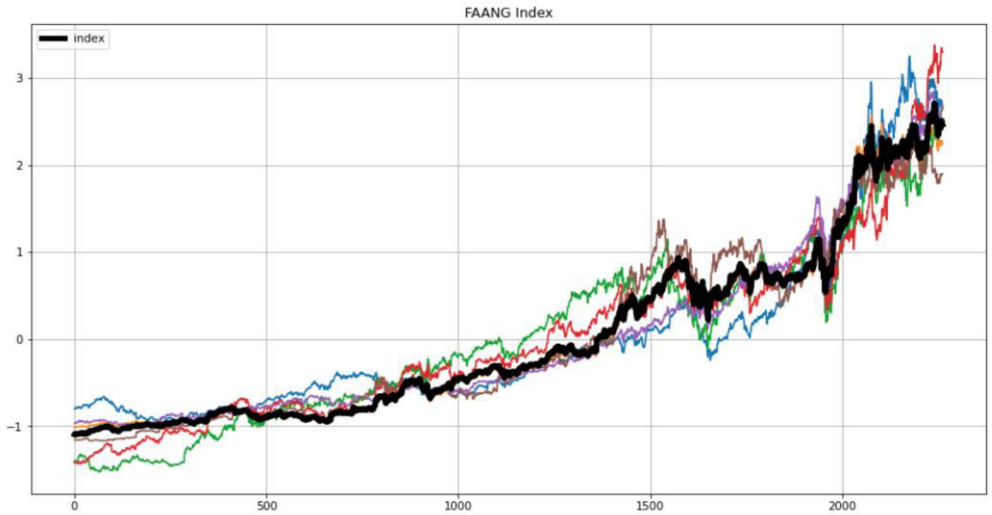
- Normalise the adjusted close prices

The following snapshot shows the normalised prices for FAANG stocks.



- Run PCA with sklearn library's PCA function.
- Generate the index from the principal components

The following snapshot shows the FAANG index created using the principal components.

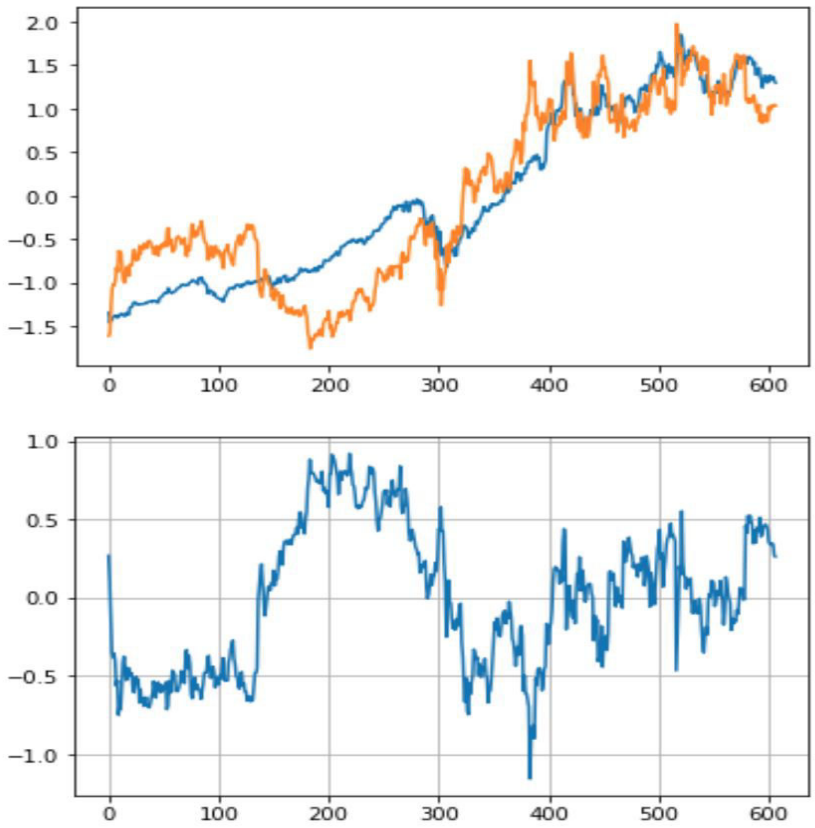


Example 3 - Statistical arbitrage using PCA

Here are the steps to build a statistical arbitrage model using PCA -

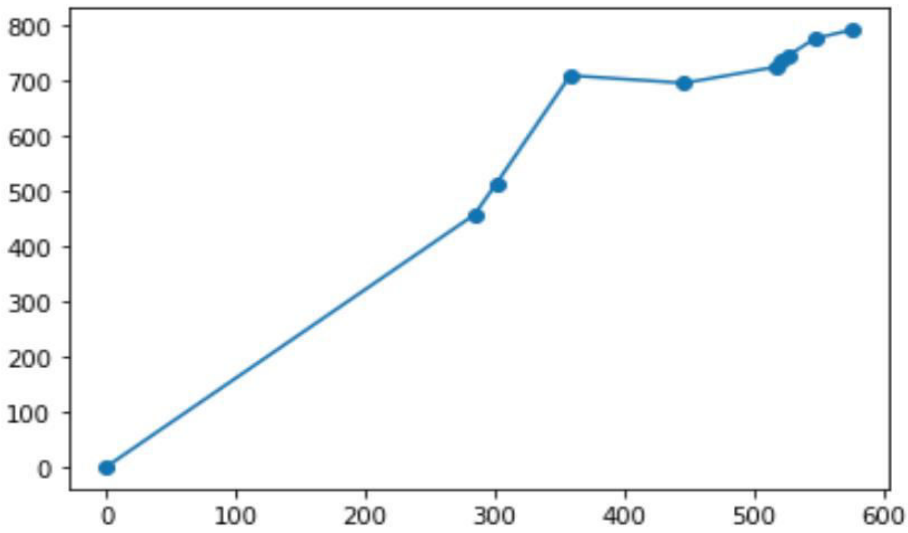
- Download the data from Yahoo finance for two stocks.

The snapshot shows the z-score of Apple and Netflix. The second plot shows the difference in the z-scores.



- Generate the backtested result

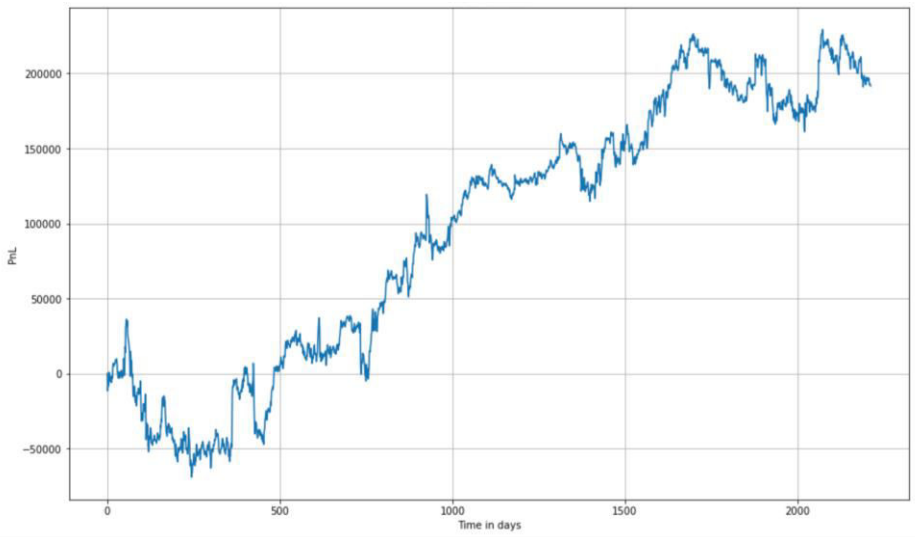
The following snapshot shows the PnL graph using the backtest for Apple and Netflix stocks.



Example 4 - Portfolio statistical arbitrage using PCA

- In this example, we apply statistical arbitrage to a portfolio.
- It is, in some sense, a generalisation of pairs trading.
- Using PCA, stocks are ranked and a long-short portfolio is created.
- The no. of long and short positions are decided by the maximum position defined in the beginning.

The following snapshot shows the PnL of statistical arbitrage implemented on a portfolio (FAANG stocks) using PCA.


200000
150000
100000
50000
0
−50000
PnL
0
500
1000
1500
2000
Time in days