COVID-19 IMPACT ON INDIAN STOCK MARKET USING PCA ON BASIS OF CAPM MODEL

2020-21

JANUARY 10

Authored by: Lubdhak Mondal
Collaborated with: Sahil Mulewar

IISER PUNE

What is CAPM Model?

The Capital Asset Pricing Model (CAPM) is a model that describes the relationship between the **expected return** and risk of investing in a security. It shows that the expected return on a security is equal to the risk-free return plus a **risk premium**, which is based on the **beta** of that security. Below is an illustration of the CAPM concept.

CAPM Formula:

$$E(R_i) = R_f + \beta_i * (E(Rm) - R_f) + €_i$$

Where,

$E(R_i)$ = capital asset expected return

R_f = risk-free rate of interest

β_i = sensitivity

E(Rm)= expected return of the market

$€_i$ = idiosyncratic return

$E(Rm) - R_f$ = Risk Premium

The CAPM formula is used for calculating the expected returns of an asset. It is based on the idea of systematic risk (otherwise known as non-diversifiable risk) that investors need to be compensated for in the form of a risk premium. A risk premium is a rate of return greater than the risk-free rate. When investing, investors desire a higher risk premium when taking on more risky investments.

The intuition behind this equation is that:

(1) the return of a stock should be at least equal to the return of the risk-free asset (otherwise why take the extra risk in the first place?)

(2) the return of the asset is also explained by the market factor, which is captured by the term $(E(R_m) - R_f)$ (measures the excess return of the market with respect to the risk-free asset) and β_i (measures the degree to which the asset is affected by the market factor).

(3) the return of a stock is also affected by idiosyncratic factors, which are stock specific factors (e.g. the earnings release of a stock affects that individual stock only, but not the overall market).

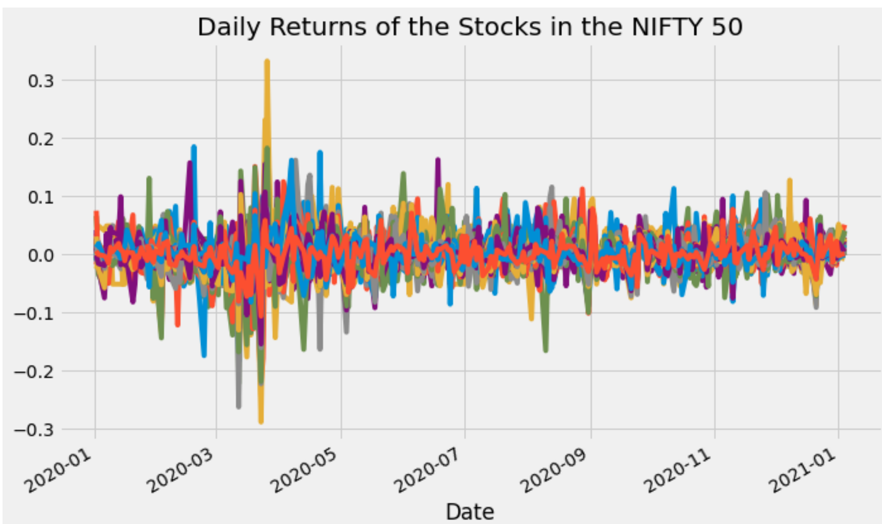
Empirically speaking, the market factor is the primary driver of the stock market returns, as it tends to explain most of the returns of any given stock in any given day.

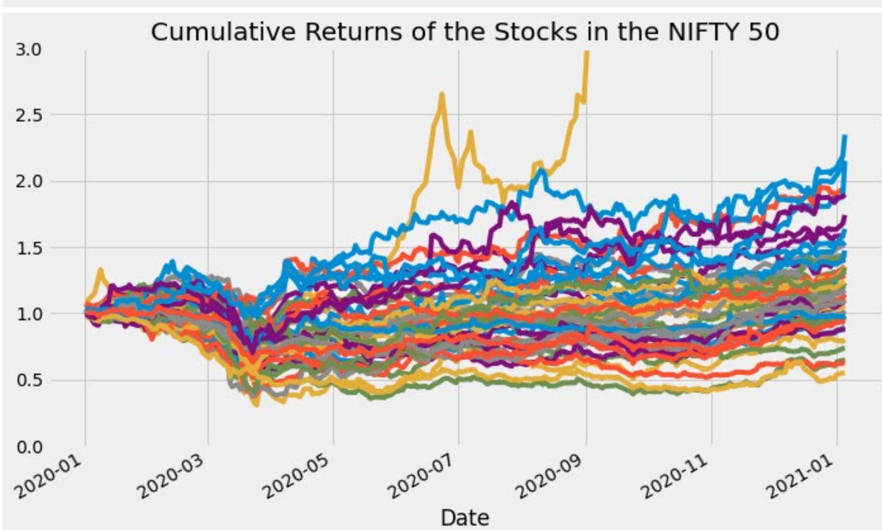
About Data:

We have extracted data from *Yahoo Finance* and we used the data of each individual stock of NIFTY50 which is the flagship index on the National Stock Exchange of India Ltd. (NSE). The Index tracks the behavior of a portfolio of blue-chip companies, the largest and most liquid Indian securities. It includes 50 of the approximately 1600 companies listed on the NSE, captures approximately 65% of its float-adjusted market capitalization and is a true reflection of the Indian stock market. We have individually extracted **Adj. Close** column of each stock available in NIFTY50.

Basic Plotting and Filtering:

Considering 50 stocks of NIFTY50 we plotted their daily return and Cumulative returns starting from 01/01/2020 to current date.





Getting some conclusion from the raw data and plotted graphs is tough at first glance, so we've to filter the data in the first hand to draw some insights from this. Here we're going to use PCA to filter the raw data.

What is PCA?

Principal component analysis (PCA) is a procedure for reducing the dimensionality of the variable space by representing it with a few orthogonal (uncorrelated) variables that capture most of its variability. Reducing the dimensions of the feature space is called dimensionality reduction. Reduction of dimensions is needed when there are far too many features in a dataset, making it hard to distinguish between the important ones that are relevant to the output and the redundant or not-so important ones.

PCA Algorithm:

1. Scale the data by subtracting the mean and dividing by std. deviation.

$$z = \frac{x - \mu}{\sigma}$$

2. Compute the covariance matrix.

$$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n} (Xi - \bar{x})(Yi - \bar{y})$$

3. Compute eigenvectors and the corresponding eigenvalues. $A\vec{v} = \lambda\vec{v}$ .....[1] (Lambda is eigenvalue)

$$A\vec{v} - \lambda\vec{v} = 0$$
$$\Rightarrow \vec{v}(A - \lambda I) = 0,$$ (I is the Identity matrix)

4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues, these becoming the principal components.
5. Derive the new axes by re-orientation of data points according to the principal components.

Single Value Decomposition (SVD):

We can find the eigenvalues and eigenvectors using this method. This decomposes a matrix into three other matrices each with some specific properties. If A is a matrix (here the covariance matrix), we can think of that as a transformation that is acting on vector v.
Now, from equation…[1] we can decompose A as:

A=CΣC'

Where **Σ** is a matrix whose diagonal elements are eigenvalues of the corresponding eigenvectors in C, which is the matrix containing all the eigenvectors along each column.

Generalized version of SVD is as follows:

$$A=U\Sigma V^T$$

Where A is a m×n matrix, U is a m×n orthogonal matrix, Σ is n×n diagonal matrix, V is n×n orthogonal matrix.

U is an orthogonal matrix with orthonormal eigenvectors chosen from $\Sigma\Sigma^T$.

V is an orthogonal matrix with orthonormal eigenvectors chosen from $\Sigma^T\Sigma$.

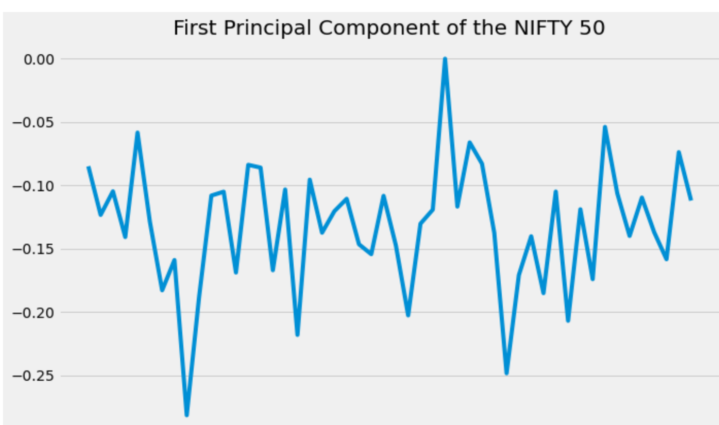
How PCA is used in this particular problem:

First Principle component (PC1) approximate the market factor when PCA is applied to daily stock returns.

Why so is very hard to explain in this short document or presentation and also requires copious amount of math and finance derivations.

The weights of the first eigenvector of a covariance matrix represent the market factor and also the largest source of systematic risk (variation of returns). PCA simply identifies the eigenvector that maximally explains the variance of the system. It turns out that this is the "market factor" - i.e. the tendency of securities to rise and fall together as an asset class. Why is this the market factor? If you examine the weights (factor loadings) of the first eigenvector in a histogram you will find they are generally all of the same sign whereas this is not the case for any of the subsequent eigenvectors (which represent sectors or style factors, that is to say other sources of systematic risk). In other words, it is empirically the case that there is a dominant systematic factor called the equity risk premium explaining the variance of returns.

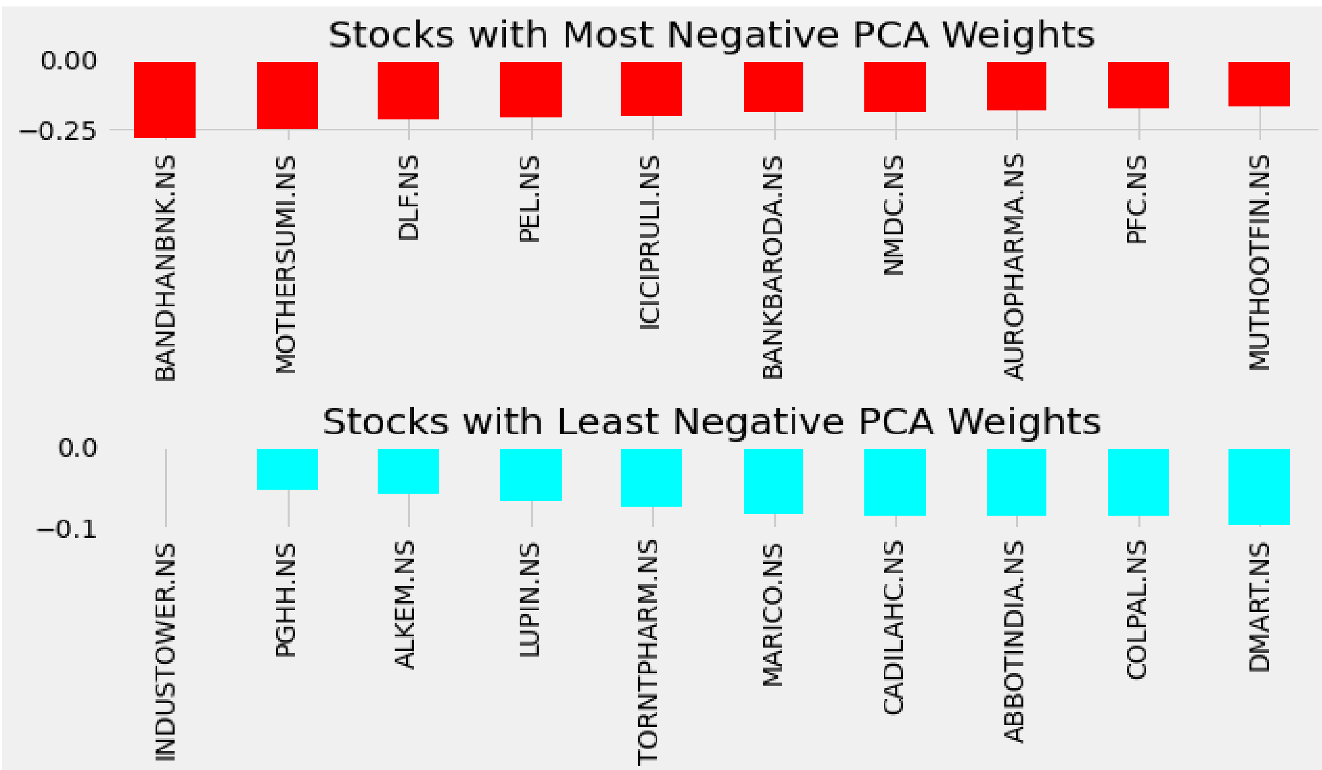
We have processed the raw data of 50 columns of Adj. Close of each stock via PCA by computing the 1st principal component of the daily returns. The figure below shows the values of the 1st principal component, which is essentially a vector of dimension 50 that contains a value for each of the 50 stocks. Here the first principal component represents the linear combination of the 50 columns of input data extracted from yahoo-finance, that explains most of the **variance**, and the primary driver of stock returns is the overall market factor.



Its very easy to calculate 1st principal component of PCA in Sci-Kit library in Python (details in code).

Analysis of Impact of Covid-19 using this model:

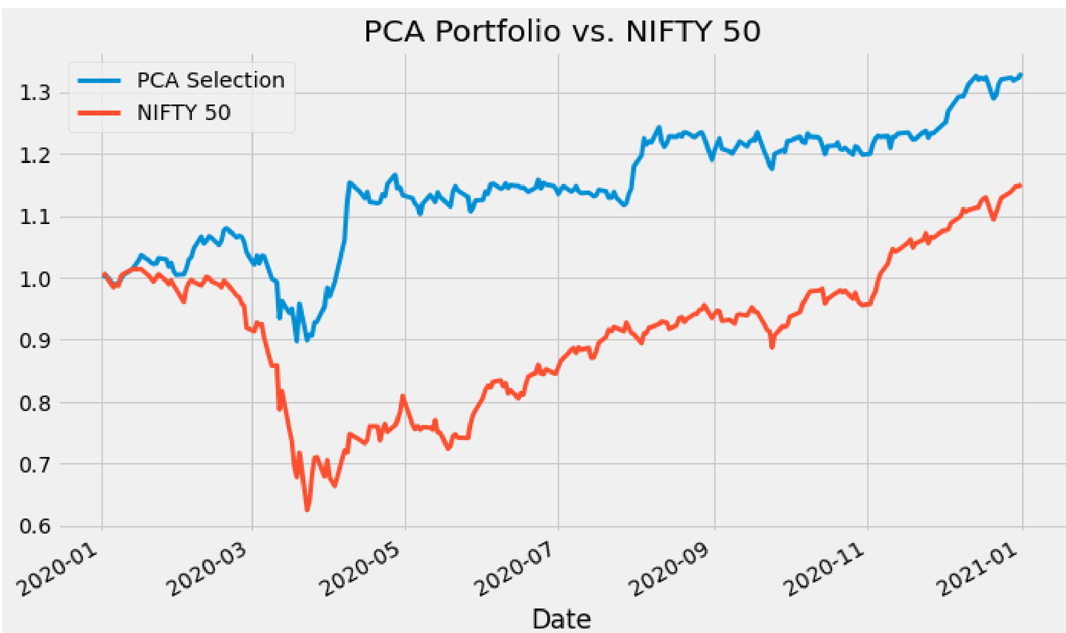
When Covid-19 pandemic hits hard, all normal daily activities ceased and lockdown was announced all over the world, the stock markets also gone through many many ups and down and Indian Stock market wasn't no different. Now using PCA, we can cluster most and least affected stocks in Indian Market without having any prior knowledge about them and can get lot of interesting inferences about the situation of Indian economy. Below it is given the plot of the stocks that have the most and the least negative PCA weights. Stocks with most negative weights (RED) were devastatingly affected in this year and vice versa for the Positive weights (CYAN).



The results are quite interesting from the latter plot. Lets first see the most **badly affected** stocks and those stock are from Banking sector or Government owned stocks. There are few reasons behind why financial and banking sector stocks (e.g. Bandhanbank, BankofBaroda, MuthootFinance, Icici Prudential Life Insurance) have suffered such a massive blow during Covid-19. One of the key reasons was government exempted companies from insolvency process for one year under the Insolvency and Bankruptcy Code. The move aims to give relief to companies defaulting on loans due to the COVID-19 stress. And at that point of time very few people were investing in new stocks so naturally these stocks crashed. Few other sector's stocks have performed pretty bad and one of the key common factors we have observed is these badly performed stocks are owned by Government of India (e.g. National Mineral Development Corporation (NMDC), Power Finance Corp (PFC)). During lockdown almost every public sector's work was ceased and so does the new investment in these government stocks and that aids to crash these stocks.

Now let's talk about the stocks those have performed well significantly during this period. These are either fast moving consumer goods stocks (e.g MARICO, DMART) or Pharmaceutical/Toilet Preparation (e.g. Colgate-Palmolive (COLPAL), Lupin, MARICO, PGHH, Abbot India) stocks. This observation is matching our intuition since these sectors benefited from the boost in sales of consumer goods due to the quarantine measures. And also, a lot of people invested in various biomedical companies predicting a boom in these companies once vaccine comes out for real. These create a "*Bubble*" in the biomed markets, which can be seen from the graph.
Though its evident from the first graph that market is recovering from the massive blow after COVID-19 and so does Indian economy.

If we could go back to time and then we can create a winning portfolio with the help of PCA, consisting top 10 well performed stock instead of the top 10 poorly performed stocks. This allocation would be based on the PCA weights. And the figure below shows how well our PCA portfolio performed instead of the actual market if we have invested in the PCA winning portfolio.



Some Insights:

1. PCA can be used as a powerful tool in these kinds of financial problems but one has to ensure two things. a) The data is normalized which I've did in the first place and b) Data should be clean and comparable. I tried once with NIFTY 500 thinking about getting more accurate result but that data wasn't cleaned properly and there were too many loopholes so we sticked to NIFTY 50
2. There will be some amount of **Loss of Information**. Although principal components try to cover the maximum variance among the features in a dataset, if we don't select the number of principal components with care, it may miss some information as compared to the original list

of features. Also, Principal components are the linear combination of your original features. Principal components are not as readable and interpretable as original features.

Further Study:

For the limited time we could not extended the project for further. I've planned to study further based on this and explore the market situation of Asian Stock market while SARS hit in 2001-2002 and compare it with COVID-19 situation and then if any potential clues to relate them and any viable solution for future if again humanity have to face these kinds of dreadful pandemic.

Acknowledgements:

1. I truly appreciate our course instructor Prof. Joy Merwin Monteiro for motivating us to take up new problems and inspiring us throughout the sem.
2. Sahil Mulewar
3. My mentor Anubhab Chatarjee, pursuing MS in Data Science at CMI

References:

1. https://towardsdatascience.com/
2. https://towardsdatascience.com/the-mathematics-behind-principal-component-analysis-fff2d7f4b643
3. https://blog.statsbot.co/singular-value-decomposition-tutorial-52c695315254
4. https://youtu.be/FgakZw6K1QQ
5. http://store.ectap.ro/articole/1066.pdf