Data Story

Even after applying portfolio optimization to the set of available stocks in the healthcare industry, MGNX, a top performing stock in the industry, is non stationary. In other words, finding top performing stocks that meet the criteria for normality and, screening for lesser volatility with the Sortino Ratio does not secure any hope of selecting a set of stocks that retain the properties of a stationary process.

To reiterate the entire statistical process from the beginning, we take our choice of stocks from any industry we wish on the yahoo finance webpage. You obtain a csv and proceed with the proper python imports to load price columns from the set of tickers obtained in the csv file. You could take the data provided in the csv but since it is limited to a short time interval, get_data_yahoo() is more appealing in terms of providing data over a larger time interval.

The approach to stock selection is quite simple. Because our ultimate goal is to model and forecast univariate prices, it is preferable to work with a normalized time series. In this case, portfolio analysis was reduced to checking conditions of normality through statistical properties such as kurtosis, and skew. Normal distributions would have us choose a skewness between -0.05 and 0.05, and a kurtosis less than or equal to 3. These were the only conditions required to fulfill the normality criterion. The next and last part of the selection process is reduced to the setup for the Sortino Ratio. Considering the assumed volatility of the healthcare industry, the Sortino Ratio was chosen as an alternative to the Sharpe Ratio because of its focus on the hedge against negative volatility in returns.

$$Sortino\ Ratio = \frac{R_p - R_t}{\sigma_{np}}$$

R_p = Expected portfolio/asset return

R_{rf} = Target rate of return defined by investor

σ_p = Downside of portfolio/asset standard deviation

Fig 1

Once again, the focus is the standard deviation of negative returns whereas the Sharpe Ratio has its focus on all returns. A crucial difference in our approach to the analysis was the decision to stop after the calculation of the Sortino Ratio. Instead of calculating over a set of stocks, SR was measured on each stock individually. Finally, the traditional way is to minimize variance of the portfolio using the Efficient Frontier approach but we did not use portfolio weights. These weights represent the proportion invested in a particular stock or asset. These are the weights that are optimized. They are necessary to do EF optimization. Below you'll find an image of the Efficient Frontier with tangents along the Efficient Frontier curve representing optimal portfolios with risk and reward tradeoffs on the x and y axis, respectively.

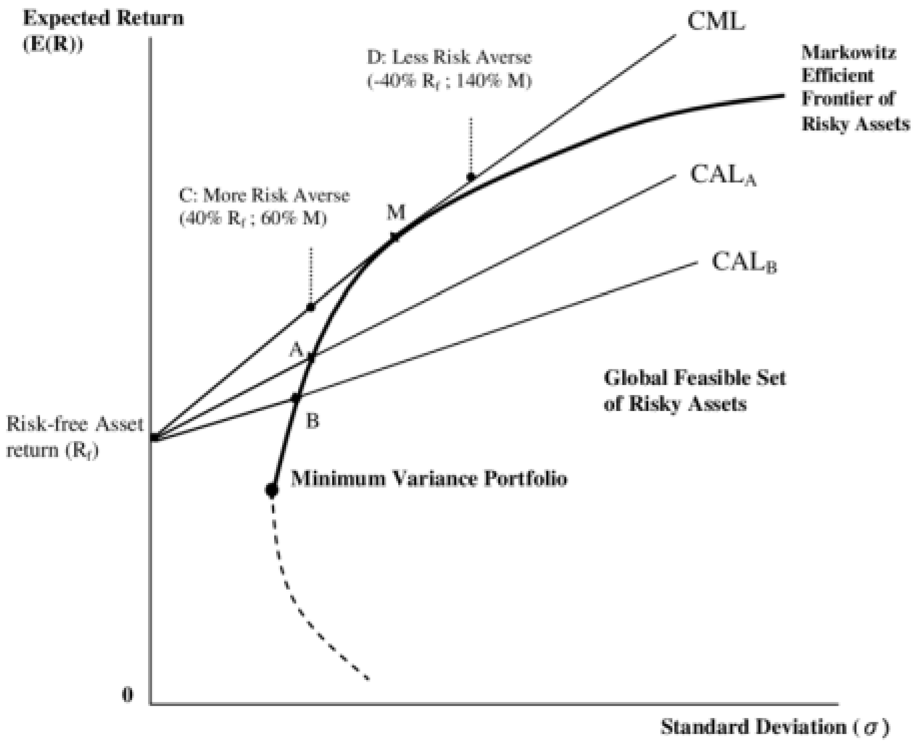

Fig 2

Consequently, portfolio weights were left out and portfolios in general. We simplified the process and the end goal to be selection based on the individual stock's Sortino Ratio.

After measuring the Sortino Ratio of all individual stocks, we selected the 5 most positive of the set. The next phase in the exploration would be the inspection of autocorrelation in the time series. Since modeling is univariate in this study, we focused on a single stock in our selection pool, MGNX. Our autocorrelation test gave us further confirmation of the volatility of the time series. The ACF demonstrated high correlation with 30 to 253 lags and the PACF gave us a set of two statistically significant lags. Further analysis using the Dickey Fuller test, confirmed the non stationarity of the time series. So we continued to demonstrate the process of detrending for the purpose of obtaining a stationary time series. Not once has the hypothesis regarding the state of the time series changed. These stocks are highly volatile and one could only take precautions in order to normalize and detrend the data before proceeding with the final stage of modeling.