FE670 Homework Assignment #1

Due Date: Sept. 29 (Thursday).

Problem 1: Russell 1000 Index includes the largest 1,000 securities (large-cap) in the Russell capitalization in the U.S. equity market. We apply a selection rule similar to that of the Russell 1000 to artificially generated random walks. We form 10,000 independent random walk price processes, each representing the price of a company's stock, over 1,000 periods using the recursive formulation:

$$
\begin{aligned}
P_i(2) &= 1 + 0.007 \times \epsilon_i(2) \\
P_i(3) &= (1 + 0.007 \times \epsilon_i(3) \times (1 + 0.007 \times \epsilon_i(2)) \\
&... \\
P_i(n) &= (1 + 0.007 \times \epsilon_i(n)) \times ... \times (1 + 0.007 \times \epsilon_i(3)) \times (1 + 0.007 \times \epsilon_i(2))
\end{aligned}
$$

where $P_i(n)$ represents the i-th stock's price at period n, $\epsilon_i(n)$ are normally distributed random processes, and we assume $P_i(1) = 1$. The level of volatility, 0.007, is compatible with realistic market values.

Let's make the simple assumption that each company has the same number of stocks outstanding say (100,000 shares). Every 50 periods, we select the 1,000 stocks with the largest market cap. Given our assumption, these are the stocks with the highest market prices. This selection rule is roughly corresponds to the Russell 1000 selection rules assuming that one period represents one year.

We will look at two different universes. The first is the universe where we consider only those processes selected at the latest selection period. For example, at any time between 500 and say 501, we include the first 1,000 processes selected at the period 500. The second universe includes all processes selected at any period for the entire simulation duration.

Expected returns can be estimated as moving averages of returns. We select a moving window of 100 periods. If one period represents one week, 100 periods correspond roughly to two years. Let's choose two moving windows that terminate at the periods 500 and 501, that is, immediately before and immediately after the selection applied at time 501. For each moving window, we plot the average return of all 10,000 price processes as well as the average of the processes that were in the Russell 100 index at the corresponding period. Therefore, for the time window that ends at time 500, we average the 1,000 processes selected at time 451, while for the time window that ends at time 501 we average the 1,000 price processes selected at time 501. Please do answer the following questions using Python as a tool:

(1) Plot simulated Russell 1000 Index (selected at the last selection period) return path for the 1,000 simulation periods.

(2) Apply ADF test to see whether the Russell 1000 Index return has a unit-root.

(3) Plot the average return of Russell 1000 Index in the window after the selection is applied.

(4) Plot the average return of the 10,000 stocks in the window after the selection rule is applied.

(5) Discuss the significant difference between the (2) and (3) plots, and explain causes of the difference. Hint: we will see 20 observations after the selection.

Problem 2: The S&P 500 stock market index, maintained by S&P Dow Jones Indices, comprises 500 large-cap American companies covering about 75 percent of the American equity market by capitalization. The index is weighted by market capitalization, so large companies account for relatively more of the index with the symbol SPY. The amount of change in the price of these stocks will be highly correlated, as they are all part of the larger market. Factor analysis will be used to reduce the dimensionality of the 500 stocks in the S&P 500 stock market index to reflect the influence of 11 major industry sectors.

Data: Using Python **pandas_datareader** package, use factor analysis on Vanguard Sector & specialty ETFs ('VOX', 'VCR', 'VDC', 'VDE', 'VFH', 'VHT', 'VIS', 'VGT', 'VAW', 'VNQ', 'VPU') which represents the following corresponding sectors: 'COMM', 'CONSUMER DISC', 'CONSUMER ST', 'ENERGY',

'FINANCIALS', 'HEALTH', 'INDUSTRIALS', 'TECHNOLOGY', 'MATIREALS', 'REAL ESTATE', 'UTILITIES'. Data consists of the high, low, opening, and closing price of each of the 11 ETFs for each day. This includes all data from January 01, 2010 to December 31, 2020. We also have the S&P500 index prices during the same period.

(1) Plot the return processes of the 11 selected sector ETFs and the S&P 500 index. Using a scree plot and by analyzing the eigenvalues of the correlation matrix, choose a sufficient number of factors.

(2) Upon finding the adequate number of factors, use the varimax method to find a final rotated factor solution. Build a factor model to predict the S&P 500 stock index returns.

(3) Apply the factor sorting method on the most important factor and design a long-short strategy for the 11 selected sectors where you put the neutral return stocks in the middle bucket and divide the positive and negative factor stocks evenly into other four buckets. In this, we assume a normal distribution of the factor value, and we long the best performing bucket and short the worst performing bucket. Compare this strategy with the benchmark return of the S&P 500 stock index return.

Homework Honor Policy: You are allowed to discuss the problems between yourselves, but once you begin writing up your solution, you must do so independently, and cannot show one another any parts of your written solutions.