FE670 Homework Assignment #1

Due Date: Sept. 28 (Thursday).

Problem 1: We can simulate 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 take two random price datasets: stock1_data.csv and stock2_data.csv. Now we will use ChatGPT3.5 to generate a Python program to test cointegration of these two random stocks. Please first create a ChatGPT account using the following website: https://chat.openai.com/chat.

Once you created the account and try to ask ChatGPT the following question: "Please write a Python code to test cointegration of two stocks using both ADF and Johansen test." Test the Python code generated by ChatGPT and answer the following questions:

(1) Does the code generated by ChatGPT work with the data you generated?

(2) Please provide ways to improve the Python code generated by ChatGPT. Please note we have price data not return data.

(3) Interpret the ADF test result to see whether the stock1 and stock 2 returns are cointegrated.

(4) Interpret the Johansen test result to see whether the stock1 and stock 2 returns are cointegrated.

(5) Discuss whether you can trust the results from the (3) and (4) questions, and explain the reason.

Problem 2: The generalized autoregressive conditional heteroscedasticity (GARCH) model of Bollerslev (1986) is an important type of time series model for heteroscedastic data. It explicitly models a time-varying conditional variance as a linear function of past squared residuals and of its past values. The GARCH process has been widely used to model economic and financial time-series data.

Many extensions of the simple GARCH model have been developed in the literature. This example illustrates estimation of variants of GARCH models using the AUTOREG and MODEL procedures, which include the GARCH-in-mean (GARCH-M)

model, Exponential GARCH (EGARCH) model, Quadratic GARCH (QGARCH) model, the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model, etc.

The the general form of GARCH(p,q) model (where p is the order of the GARCH terms σ^2 and q is the order of the ARCH terms ϵ^2), following the notation of the original paper, is given by

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t|\Psi_{t-1} &= \mathcal{N}(0, \sigma_{t-1}^2) \\ \epsilon_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + ... + \alpha_q\epsilon_{t-q}^2 + \beta_1\sigma_{t-1}^2 + ... + \beta_p\sigma_{t-p}^2 \\ &= \omega + \sum_{i=1}^{q} \alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i\sigma_{t-i}^2 \end{aligned}$$

where r_t is the return at time t, μ is the mean return, σ_t is the standard residual at time t. We assume the residual ϵ_t follows a normal distribution $\mathcal{N}(0, \sigma_t^2)$, ω, α_q and β_p are the model parameters to be estimated. σ_{t-1}^2 is the conditional variance at time $t-1$.

Generally, when testing for heteroskedasticity in econometric models, the best test is the White test. However, when dealing with time series data, this means to test for ARCH and GARCH errors. In practice GARCH(1,1) model is mostly used. In this assignment, you are to develop a GARCH(1,1) model and estimate the parameters with the assistant of ChatGPT.

From the general GARCH(p, q) mode, we can derive the

GARCH(1,1) model as the follows:

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \end{aligned}$$

Use Python "arch" package to fit a volatility model for Apple stock daily returns for the duration from 1/1/2020 to 12/31/2022. You may download the data from Yahoo finance or other sources. Try to ask ChatGPT the following question: "How to build a GARCH(1,1) volatility model for a stock return data?" Please answer the following questions:

(1) Is the procedure generated by ChatGPT correct? Please focus on the data preparation or pre-processing. Please describe what you have learned from the recommendation from ChatGPT.

(2) Use the "arch" Python package to fit a GARCH(1,1) model for the Apple stock and report the model parameters.

(3) Please ask ChatGPT to write a python code for you as well, and then compare your code with the Python code generated by ChatGPT.

(4) Please describe the differences between your code and the ChatGPT code. Can you learn from ChatGPT? Why?

Problem 3: 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.