Data Assignment 1

Financial Data Analytics

It's understood that all assignments are individual work. Failure to observe this may result in immediate failure of the course.

This assignment accounts for 5% of the course mark.

1. (5 pts) Access WRDS and download monthly stock information for Amazon (ticker symbol AMZN, permno 84788) for the period of 2000-2021 from CRSP.

Variable	Label
PERMNO	PERMNO
DATE	Names Date
CUSIP	CUSIP Header
COMNAM	Company Name
TICKER	Ticker Symbol
PRC	Price or Bid/Ask Average
RET	Returns
RETX	Holding period return without dividends
SHROUT	Shares Outstanding
VWRETD	WRDS value-weighted market returns
SPRTRN	Return on the S&P 500 Index

Calculate the mean and std deviation of Amazon's (cum-dividend) returns. What is annualized returns and volatility (proxied by standard deviation) of Amazon?

2. (10 pts) Download the Fama-French riskfree rate and Fama-French factors from WRDS ("Fama-French Portfolios and Factors"). Ensure that your factors and rate are at the monthly frequency to match that in question 1.

a. Using WRDS value-weighted market returns (VWRETD) as the market return, estimate the CAPM (with an intercept) for AMZN for the entire period of 2000-2021. Interpret the results.

b. Estimate the model repetitively, for every month starting from January 2005, using a rolling period of every five years (e.g., for January 2005 estimate your model using the past 60 months of returns from Feb 2000 to Jan 2005). Report your estimates of alpha and beta into a single table (with output columns of coefficient estimate and *t*-statistics), and graph the time series of alpha and beta estimates out. Interpret the results.

3. (10 pts) Repeat the same exercise of Question 2, but with the Fama-French three-factor model. Interpret the results, focusing on the slope and the beta's. You can pool questions 2 and 3 into one single block of codes.

   a. To aid with grading, please put questions 2b and 3b time-series outputs into a separate excel file, and name it as "2b3b.xlsx".

4. (15 pts) Take the following snapshot as Berkshire Hathaway's (i.e., Warren Buffet's) portfolio composition: https://www.cnbc.com/berkshire-hathaway-portfolio/. To make this exercise trackable, I've already cleaned the data and got rid of some foreign stocks.

   a. Using stock symbols in the "BH_cleaned" sheet (not the "BH_original" sheet) in the excel file "BHHoldings_202209.xlsx", download the monthly returns of the stocks there from WRDS for the period of 2005-2021.
      i. You can use the permno list from the sheet "BH_cleaned_permno" as the stock unique identifier for downloading.

   b. Assuming that from January 2005 and on, Berkshire Hathaway held the stocks in weights proportionate to the column "portfolio_weight" for the stocks available, estimate Berkshire Hathaway's monthly returns and standard deviations for the period.
      Two data cleaning tips:

   i. In the process, if for any given month you have fewer than 45 stocks, the summed weights for the month will be less than 100%. Make sure you make proportionate adjustment so that the weight sum to 100% for that month.

   ii. There may be multiple classes of shares/multiple stocks corresponding to a ticker symbol. Ensure that you only download for the list of permnos. The best practice is to use permno instead of ticker to download data.

 c. What's Berkshire Hathaway's alpha? Is it significant? Does Berkshire Hathaway beat the benchmark(s)?

Please submit to Dropbox "WRDS Data Assignment 1", <mark>by 5pm, Jan. 23.</mark>

1. Original datasets. To aid with grading, please output all your datasets in one excel file. Label your sheet clearly.

2. Your codes (can be Python, SAS, Stata, R, SPSS, Matlab, etc.) in Dropbox.

3. A final output report. Please make sure your output report is easy to read. Coefficient estimates do not exceed 4 decimal places, and *t*-statistics do not exceed 2 decimal places. Any submitted work with output that is *only* embedded into codes will automatically get at least 25% off the entire mark. We grade your work on your final output "report" and only recourse to your codes and data if needed.

4. Any notes if you wish to identify problems and any thoughts in the entire process. As it goes, the key to data analytics rests on good data cleaning work (called "data curation" if you want a fancy big-data word). A good note that has good understanding of data issues may have 5 bonus points.