STATA Homework Assignment

1. Go to <https://wrds-www.wharton.upenn.edu/login/> CRSP, annual update, security files, monthly stock file. Select dates from Jan 1965 through Dec 2019. Select entire database and the following variables and download all as STATA dataset:

* Company name
* Share code
* CRSP Permanent Company Number (permco, unique firm identifier in CRSP)
* New CRSP permno (unique security identifier in CRSP)
* Exchange code
* Price
* Share volume
* Holding period return
* Number of shares outstanding
* Value-weighted return
* Equal-weighted return

1. Remove non-common stocks from the sample (common stocks have share codes of 10 or 11)
2. For each month, calculate equal-weighted and value-weighted market return, as average return across all common stocks. In the case of equal weighted market return, you can use simple average. In the case of value weighted market return, your weight should be market capitalization of the firm in the prior month. Market capitalization should be calculated as a product of the absolute value of price and number of shares outstanding; and then lagged by one month. The mathematical formula of the value weighted average return is:

So, overall we can also write:

These formulas may be useful to you as you consider a programming solution for the value-weighted market return.

1. Once you have your value and equal weighted returns, compare to the equal- and value-weighted returns provided in the CRSP database. Produce a table that will report average, median, min, max, and standard deviation for each of the following four versions of market returns: equal- and value-weighted returns that you created and equal- and value-weighted returns from CRSP. Provide a 3-4 sentence description of how the returns compare?
2. Go to WRDS again and download Fama and French factors, monthly frequency data. Select all variables and the same time period (Jan 1965 – Dec 2019) as before.
3. Merge Fama and French factors into your original data based on date.
4. Calculate full sample summary statistics (mean, median, stdev, key percentiles) for MKTRF, SMB, HML, RF, and your equal- and value-weighted returns.
5. Calculate the same summary statistics for various subsamples: 2-equal length subperiods; 4 equal-length subperiods (over the 55-year period). Discuss persistence in the values of the variables over time. How do factors compare? Does any factor behave like risk?
6. Create lag of each variable (MKTRF, SMB, HML, RF, and your equal-weighted return). Regress each variable on its own lagged value and lagged values of the remaining variables (exception, when your dependent variable is MKTRF or equal-weighted return, please do not use lagged value of the other of the two variables because MKTRF and equal-weighted return essentially measure the same thing, so there will be collinearity). Does any of the variables have any predictive power? Discuss results.