**Due Date Tuesday April 9th at 3:59 PM. Late Homeworks will not be accepted.**

On the Canvas website you will find the Excel Spreadsheet SP500Raw.xls

There is also an readme.txt file that tells what you each column represents.

The synapsis is that this spreadsheet contains all the constituents of the S&P 500 for each month as of the end of the month, along with each of the stocks end-of-month (EOM) stock price, its shares-outstanding, its total return of the prior month (for the EOM period), and its market capitalization. The period covered is 2011-01 thru 2023-12.

The purpose of this homework is to create a variety of “momentum” signals (or factors) and see if they can explain the cross-section of returns. Or to say it in a less technical manner, do stocks that have higher returns in some past period earn higher returns, on average, in the next period.

This homework will be considerably easier for you if you program. Python, Matlab, or S (or R) should all work well for you in this homework.

Questions:

1. Why does the dataset use PERMNO and not TICKERS, which is how stocks are identified on the major exchanges in the US?
2. For many of the Companies, the one month change percent change in price is not always equal to the PRC column.
   1. Is this a data error or is this correct? Explain.
   2. Why for some companies, the columns are always equal. Why?

Note when you calculate this for each company on each date, save this variable as we will be using it frequently. We be referring to it as Price\_Ret(T1)

1. For each date count the number of companies that exist. Why is not always 500? Is this a mistake? Explain.
2. How many of the companies are present over the entire sample? How many unique companies are in the sample?
3. Plot the time series of the max, 95th percentile, 75th percentile, median, 25th percentile, 5th percentile and min market cap of all stocks for each month, over the entire sample period.
   1. Calculate the above percentile ranges for the market cap of the stocks the month prior to their leaving sample.
   2. Repeat the above for companies on the month they enter the sample.

Note this set of percentiles as something that you will be frequently asked to use to see the behavior of your data. I will refer to the “Percentile Sets” going forward

We are now going to calculate a variety of *trailing* returns for each company. Specifically, for each company, on each date, calculate its prior:

1. Trailing twelve month return based on prc (we will refer to this PRC\_Ret(T12))
2. Trailing twelve month return based on prices (we will refer to this Prices\_Ret(T12))
3. Trailing twelve month return excluding the most recent trailing month based on prc. In other words, this eleven months of return starting from 12 months ago excluding the most recent month (we will to this as PRC\_Ret(T12M1))
4. Trailing twelve month return excluding the most recent trailing month based on prices. In other words, this eleven months of return starting from 12 months ago excluding the most recent month (we will to this as Prices\_Ret(T12M1))
5. The trailing one month return from exactly 12 months ago. In other words, if the EOM period is 2019-03 (March 2019), we want the return for the stock for 2018-03 (March 2018). Do this based both on prc column and price based return. We will refer to this as PRC\_Ret(T12\_1M) and Prices\_Ret(T12\_1M)
6. Calculate the standard deviation of the monthly price\_based returns used in calculating Prices\_Ret(T12M1); we call this Vol\_Prices\_Ret(T12M1). Divide Prices\_Ret(T12M1) / Vol\_Prices\_Ret(T12M1). We will call this SR\_Prices\_Ret(T12M1)

Now calculate the *forward* return for each stock at each date (point-in-time) using PRC. We will want the \*\* \*\*, forward 3 months returns and forward 6 months (e.g. over those total time periods). We will refer to these as PRC\_Ret(F1M), PRC\_Ret(F3M), PRC\_Ret(F6M).

Note: all returns that are calculated should be geometric returns not arithmetic returns.

1. For all of the variables you have calculated above, produce one times series graph for each variable, with each graph showing the Percentile Sets. For each one of these also calculate the median value of the time series of each of the Percentile Sets values. For example, what is the median value of the 95th percentile of PRC\_Ret(T12M) over the time series (sample)? What is the median value of the 75th percentile of PRC\_Ret(121M) over the times (sample)?

Now we are going to see the predictive power of each of the variables to explain the cross-section

1. Run a Fama-McBeth cross-sectional regression:
   1. For the period is November 2019 where the dependent variable is the one month forward return (e.g. the return for October 31st 2019 to November 30th 2019) and the independent variables is PRC\_Ret(T12M1). Please report and interpret the coefficient.
   2. Now run the same cross-sectional regressions as delinated above for the entire sample. Please plot the time series of the coefficients. Please calculate the statistical significance of the sample regression.
   3. Interpret your result, both economically and statistically. What conclusion can you draw about using Momentum [-12, -1] as investment strategy? How would you explain this to an educated layperson with relatively little quantitative knowledge. Are you results consistent across time and months? Are there any periods that stand out to you and are there any you would highlight and comment on?
   4. Review the various performance measures discussed in Peterson “Efficiently Inefficiently” in Chapter 2. Please comment on any measures that he describes that you think might be notable such hit rates, draw-downs, max draw-down, slugging rate, sortino ratio, etc. Note: you do not have to comment on each or all of these; focus on the one’s you to believe to be most salient. [Hint: what is the interpretation of the single regression coefficient? How might you view the time-series of the regression coefficients?]
2. Run Fama-McBeth Cross Sectional Regressions as outlined in 7(b) and report your results as per 7(b). Additionally as to the commentary above be sure to discuss economically if you would expect the sign of all the variables to be the same as well any particularities / issues you might see with the regression in 8 (b)
   1. Univariate regressions, where the dependent variable is one month forward returns and the independent variables are: Price\_Ret(T1); PRC; PRC\_Ret(T12); Prices\_Ret(T12); PRC\_Ret(T12M1); Prices\_Ret(T12M1); PRC\_Ret(T12\_1M); Prices\_Ret(T12\_1M); and SR\_Prices\_Ret(T12M1)
   2. Multivariate Regressions with same dependent variable as 8(a) with the following independent variables:
      1. Price\_Ret(T1) and Prices\_Ret(T12M1);
      2. PRC and PRC\_Ret(T12M1)
      3. PRC and Prices\_Ret(T12)
      4. PRC and SR\_Prices\_Ret(T12M1)
3. Repeat 7 (b) but now using the dependent variables: PRC\_Ret(F3M) and PRC\_Ret(F6M). Compare and contrast your answer with 7(b). Do you see any potential econometric / statistical issue with the regressions you ran just now? Please elaborate. Please comment on how you might correct any such issues.
4. Please now take a step back from the details of the homework. Please explain from a first principles standpoint why I might be having you do this homework. What might this have taught you about various investment strategies (as a quantitative researcher), what you have learned from the different forms of the independent and dependent variable(s) you created.