Systematic Investing
Professor Dhar
Spring 2025

Assignment VOL
(Data available on class website)

Using the S&P500 daily cash data provided, **determine whether there has been any relationship between volatility and direction in equity markets**. The relationship could be *concurrent* or *lead/lag*. The raw data you have consists of a date and a price of the S&P Cash Index. The first three rows of data are shown below along with column headers showing the returns and volatilities that you need to calculate for each day.

Date	Close	daily % ch	20day vol	previous 20d return	next 20 day return	zvol20	zfret20	zret20
1/4/1960	59.91							
1/5/1960	60.39							
1/6/1960	60.13							

Part A
Setup Data and Indicators:

Start by computing the daily returns, and the 20-day volatility of these daily returns. Call the latter **vol20**. Then for each day, compute historical 20-day returns until that day (i.e. using 21 prices) – call this **ret20**. Similarly, for each day, calculate and 20 day FUTURE return from that day, that is, return for the *next 20 days* – call this **fret20**.

Now, "**normalize**" the volatility and returns. Specifically, do the following

- Convert **vol20** into **zvol20** by computing the z-score of volatility:
  Z(i) = [X(i)-AVERAGE(X(i-1),X(i-2)…X(i-N))]/STDEV(X(i-1),(X(i-2),…X(i-N))
  Use the last 250 days as the historical window for normalization, i.e. N=250. To normalize volatility, subtract from today's value, the average of the last 250 values prior to today, and divide this by the standard deviation of the last 250 days prior to today. This gives you the daily "z-score" of volatility, namely, **zvol20**.
- Similarly, calculate **zret20** and **zfret20**.

IMPORTANT: Note that there will be blanks in some parts of the spreadsheet. For example, the daily%ch series will start in row 3 (since it requires two prices). Similarly, the 20day vol of the daily returns will require 20 values for the daily%ch series, so it will start further down in row 22 of the spreadsheet. Similarly, the "zvol20" column will start even further down (row 272) since it requires 250 trailing values of vol20 to calculate its first "normalized" value in row 272. Start calculating zret20 and zfret20 also from that row.

Equally importantly, note that your "forward" indicators such as **fret20** and **zfret20** will not have values for the last 20 days of the series since the required forward data do not exist for those dates. When you are doing your analysis, delete these days.

Questions:
After you have composed the data, bucket the volatility, zvol20, into quintiles. You can do this easily by sorting the data by zvol20 (lowest to highest) and then segmenting the sorted data into 5 equal sized segments (quintiles). These represent the lowest to highest volatility quintiles. If the number isn't divisible by 5, approximate. For example, if you have 10,053 records, you could have 2,010 records in the first 4 quintiles and 2.013 records in the fifth quintile.

For the two analyses below, I would suggest copying the relevant data from your "data" worksheet into a new worksheet and doing the analysis in the new worksheet. Leave the original data worksheet uncluttered.

1. **Is there a (concurrent) relationship** between the normalized volatility of the market and its normalized returns? I.e. Between **zvol20** and **zret20**? What is it?

   For this question, calculate the average normalized volatility **zvol20** for each quintile (1 to 5 or lowest to highest) and the average **zret20** within each quintile. Plot the five averages, with **zvol20** on the X-axis and **zret20** on the Y-axis.

2. **Is there a (lead-lag) relationship** between the normalized volatility of the market and its FUTURE normalized returns? I.e. Between **zvol20** and **zfret20**? Plot the five averages, with **zvol20** on the X-axis and **zfret20** on the Y-axis.

What are the relationships in the two cases above? Can you hypothesize the possible reasons behind your results? If you want, you can try experimenting with different durations of volatility (say 5 and 10 days) and returns to test whether the relationship holds up more generally.

Part B (Extra Credit)
Do your results suggest a potential trading strategy on either the long side or the short side? What would this be? How could you "test" it? What is its Information Ratio?

Some Important Details

- You need N+1 prices to calculate N returns.
- To calculate your first value of vol20 you need 20 values of daily returns which means you will "burn" up the first 20 values of daily returns to calculate your first vol20 value. For these days vol20 will be blank.
- Similarly, you will burn up 250 values of the vol20 values in order to calculate the first normalized value of volatility (zvol20).
- Similarly, at the end of the data, there will be (20) days for which you are not able to calculate the 20 day future return since you won't have all 20 future return values for these days. For these days zfret20 and fret20 will be blank.

In your analysis, eliminate the days for which you don't have data for ALL the required data.