# Financial Data Science and Computing, Project A
# Michele Marinucci

## 1. Question 1

After using Bloomberg to get OHLC time series data for the S&P 500 index, I answered the assignment's several questions here below. Questions are reported at the beginning of every poiint in italics.

### 1.1. Point a

*What date range did you obtain? Is your data set complete?*

After an initial deta pre-processing, Data range obtained goes from 01/03/1928 to 09/01/2020. The data set seems to be complete in terms of dates, but 281 of these dates are without data.

### 1.2. Point b

*List data integrity checks that can be performed on the data. Apply them. What errors, if any, did you find? Which errors, if any, could be corrected without using alternative data sources?*

Some possible data integrity checks include (1) no negative or zero values, (2) High is higher than or equal to Low, (3) no duplicates, (4) low is lowest and high is highest, (5) and finally count all of the consecutive dates that have the same Close, which is usually indicative of forward or backward filling given the likelihood of the same Close price happening on consecutive dates.

### 1.3. Point c

*Based on your data set, estimate the probabilities that the market's daily high value occurs at the open or at the close, respectively. Do the same for the market's daily low value. That is, you should estimate four probabilities, Prob(phigh = popen), Prob(phigh = pclose), etc. Explain your method and assumptions. Can these results be used to test the random walk hypothesis?*

Refer to Table 1 for the table of probabilities. Moreover, looking around the day in 1982 in which intraday data became more detailed, one can tell why this data may not be used for a random walk at a small, intraday level. However, there should be no problem using one day returns if the sample is large enough.

*Table 1.* Probability table

| | CLOSE | OPEN |
|---|---|---|
| HIGH | 0.603 | 0.624 |
| LOW | 0.585 | 0.636 |

*Table 2.* Top 10 Intraday Range

| DATE | INTRADAY RANGE |
|---|---|
| 1987-10-19 | 0.25739448 |
| 1987-10-20 | 0.13471311 |
| 2008-11-13 | 0.11520844 |
| 2008-10-10 | 0.11497976 |
| 2008-10-28 | 0.11267406 |
| 2008-10-09 | 0.10565448 |
| 2008-10-13 | 0.10318269 |
| 2008-10-15 | 0.10023341 |
| 2008-11-20 | 0.09727460 |
| 2010-05-06 | 0.09550662 |

### 1.4. Point d

*Compute the intraday range. Note that the range is insensitive to the time-ordering of the high and the low. From 1/1/1980 through 8/30/2011, find the top 20 intraday ranges. List them, ordered by size. How many occurred during the final three-year sub-period (i.e., 9/1/2008–8/30/2011)?*

A total of 15 of the top 20 happened in the last three years. Notice that top observation was Black Monday and that quite a few observations were in 2008.

### 1.5. Point e

*Compute the overnight return. List the top 20 positive overnight returns during the period 1/1/1980–8/30/2011 in reverse chronological order, and separately list the top 20 negative overnight returns. Which three-year period had the largest number of each?*

As Table 3 and Table 4 show, 1980-1982 period has the most instances in both cases; the reason is perhaps that there was no intraday before 1982 and that out time period under analysis starts in 1980.

*Table 3.* Top 10 Overnight

| DATE | OVERNIGHT RETURN |
|---|---|
| 1982-03-22 | 0.01952807 |
| 1982-02-24 | 0.01757690 |
| 1982-01-28 | 0.03282960 |
| 1981-11-02 | 0.01895151 |
| 1981-10-30 | 0.02376953 |
| 1981-10-02 | 0.01947386 |
| 1981-09-28 | 0.02447459 |
| 1981-03-25 | 0.01811836 |
| 1981-03-12 | 0.02493267 |
| 1980-12-17 | 0.01753446 |

*Table 4.* Bottom 10 Overnight

| DATE | OVERNIGHT RETURN |
|---|---|
| 1986-09-25 | -0.01883359 |
| 1982-02-08 | -0.02242879 |
| 1982-02-01 | -0.02176080 |
| 1982-01-11 | -0.02317022 |
| 1982-01-05 | -0.02191625 |
| 1981-09-25 | -0.01947657 |
| 1981-08-24 | -0.02886327 |
| 1981-02-02 | -0.02037823 |
| 1981-01-20 | -0.02024261 |
| 1981-01-07 | -0.02200985 |

### 1.6. Point f

*Market volatility is time-varying, yet persistent on short time scales. So let's evaluate the impact of a "sharp swing" by comparing it to the general level of volatility that was present beforehand.Define a one-day jump measure,*

$$j_t = r_t/_t$$

*where we look at the one-day (log) return rt = log(pt/pt1) relative to its standard deviation. For the denominator, , compute the standard deviation of returns using the three months (63 trading days) prior to the start of day t. That is, jt reflects the size of the return relative to a typical daily move, where "typical" is based on recent expectations, as of the prior day's close. Pay careful attention to the endpoints, to the scale, and to the units. List the top 20 jumps jt (ranked by absolute value) in the data set. How many occurred during the 3 years ending 8/30/2011?*

As show in Table 5, only 5 of these jumps happen during the final three years. This may be partly because of the implementation of circuit breakers and generally due to lower historical volatility.

*Table 5.* Top 10 Jumps

| DATE | JUMPS |
|---|---|
| 1987-10-19 | 17.69887 |
| 1989-10-13 | 9.291160 |
| 2007-02-27 | 7.626093 |
| 1997-10-27 | 6.785078 |
| 2011-08-08 | 6.131791 |
| 1982-08-17 | 5.920905 |
| 1998-08-31 | 5.818365 |
| 1994-02-04 | 5.739030 |
| 1986-09-11 | 5.297913 |
| 1993-02-16 | 5.271930 |

## 2. Question 2

*On October 6, 1982, equities in the U.S. soared on news of falling interest rates. "The stock market surprised experts..." reported the Washington Post in an article the following day. It apparently took data providers by surprise, too. Different data vendors report conflicting results for SP 500 index OHLC data for that date.† What discrepancies can you find? How economically significant are they? How could they be resolved? Explain what you think were the correct SP 500 index values on that day and why. Start with Bloomberg and Yahoo, and use any other sources that are helpful.*

Open and low are the same, however for high and close Yahoo shows 125.97 whereas Bloomberg shows 126.97. There could be two potential explanations in my opinion. First, Yahoo missed a high price at the end of the day, whereas Bloomberg did not miss it. Second and more likely, either data provided plugged the wrong unit digit for the price number. I would probably tend to say that Bloomberg is the right one due to the first possibility and due to the higher reliability of the platform.

## 3. Question 3

*On August 24, 2020, Dow Jones announced that three of the thirty companies making up the venerable Dow Jones Industrial Average would be replaced, effective August 31, 2020.*

### 3.1. Point a

*Determine the value the index divisor would have had if the change were made at the market close on the announcement date.* The new divisor is 0.159.

### 3.2. Point b

*What fraction of the total index value was made up of the departing companies? That is, what was the total index weight being replaced, as of the announcement date?* The dropped companies were 3.46 percent of the index.

### 3.3. Point c

*Did the 27 remaining companies (i.e., those not being changed) have greater or lesser total weight within the index after the change? That is, what was their total weight with and without the change, as of the announcement date?*

Their weight afterwards is 93.18% , while beforehand it was 96.54%.

### 3.4. Point d

*Suppose instead that one replacement member had been AMZN instead of AMGN. What would be the new estimate of the divisor as of the announcement date? Suppose the replacement had been Berkshire Hathaway (Class A shares). What would be the new estimate of the divisor as of the announcement date?*

Same exact calculation as above; however, since Berkshire's stock price is in the magnitude of houndreds of thousands and the dow is not market cap weighted, the dow would've basically become close to 100% Berkshire.

### 3.5. Point e

*What role did stock splits play in the timing of the replacement?*

They were basically a consequence of Apple's stock split!