

## Project A: Market Data, Volatility, and Indices

1. Read the article from the New York Times, “Market Swings Are Becoming New Standard” (9/12/2011, links available on Canvas) and then use the Bloomberg terminals to get OHLC time series data for the S&P 500 index to answer the questions below.

In the Bloomberg Excel “Spreadsheet Builder,” these data fields are found under “Market Activity” and designated as Open (PX\_OPEN), High (PX\_HIGH), Low (PX\_LOW) and Last (PX\_LAST), respectively.) Get the longest set of daily data available, ending with data from 8/31/2023.

- (a) What date range did you obtain? Is your data set complete?
- (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?
- (c) Based on your data set, estimate the probability that the market’s daily high value occurs at the open. That is, you should estimate the probability,  $\text{Prob}(p_{\text{high}} = p_{\text{open}})$ . Explain your method and any assumptions. Can these results be used to test the Random Walk Hypothesis intraday?
- (d) Define the intraday range by

$$H_t \equiv \frac{p_{\text{high},t} - p_{\text{low},t}}{p_{\text{low},t}}. \quad (1)$$

Note that the range is insensitive to the time-ordering of the high and the low. From 1/1/1980 through 8/31/2011, find the top 20 intraday ranges. List them, ordered by size. How many occurred during the final three-year sub-period (i.e., 9/1/2008–8/31/2011)?

- (e) Define the overnight return (also known as “close-to-open” return) as

$$R_{\text{overnight},t} = \frac{p_{\text{open},t} - p_{\text{close},t-1}}{p_{\text{close},t-1}}. \quad (2)$$

List the top 20 positive overnight returns during the period 1/1/1980–8/31/2011 in reverse chronological order, and separately list the top 20 negative overnight returns. Which three-year period had the largest number of each?

- (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,

$$j_t \equiv r_t / \sigma_t^*, \quad (3)$$

where we look at the one-day (log) return  $r_t = \log(p_t/p_{t-1})$  relative to its standard deviation. For the denominator,  $\sigma_t^*$ , compute the standard deviation of returns using the 30 trading days *prior to the start* of day  $t$ . That is,  $j_t$  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  $j_t$  (ranked by absolute value) in the data set. How many occurred during the 3 years ending 8/31/2011?

- (g) How does the volatility of the year pandemic year 2020 compare with periods discussed in the article? Use any metrics you think are appropriate.
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 S&P 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 S&P 500 index values on that day and why. Start with Bloomberg and Yahoo, and use any other sources that are helpful.
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.

- (a) Determine the value the index divisor would have had if the change were made at the market close on the announcement date.
- (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?
- (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?
- (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?
- (e) What role did stock splits play in the timing of the replacement?

---

\*Even worse, at least one vendor overwrote good data with bad data many years after the event.