Homework 7 Realized Variance
Due Sunday, March 29, 2020
Total 10 points

Instructions. This is a group assignment. Groups may include up to 5 people. Please submit a Word or .pdf document with your solutions via the Compass site prior to 11:59 p.m. on Sunday, March 29. Please also upload your R or Python scripts.

0. Gather the data. (0 points, but necessary for the other questions) Login in to WRDS, go to the data vendor TAQ, and find the millisecond "Consolidated Trade" data:
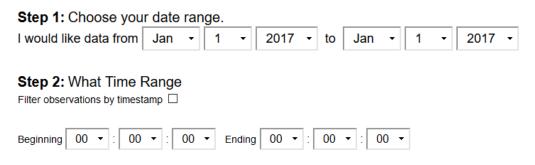

Click on "Consolidated Trades" and you will see:

NYSE Trade and Quote Millisecond Consolidated Trades

> **Note to TAQ users: Missing trades in the April 5, 2012 Consolidated Trades dataset.**
>
> There are roughly 65,000 missing trade records for tickers with initial letters M through R from 08:42 through 10:02:50 on April 5, 2012. NYSE has reported to WRDS that NASDAQ was not disseminating trade reports during that time period. The actual number of trade records (for "M" through "R") during that time from NASDAQ (exchange code "Q") was 2,232, but the average during the rest of the week was 68,274. The missing trades amount to about 12% of total trades in those equities.
>
> WRDS will provide an update as soon as the files are available.

Step 1: Choose your date range.
I would like data from [Jan ▾] [1 ▾] [2017 ▾] to [Jan ▾] [1 ▾] [2017 ▾]

Step 2: What Time Range
Filter observations by timestamp ☐

Beginning [00 ▾] : [00 ▾] : [00 ▾] Ending [00 ▾] : [00 ▾] : [00 ▾]

Step 3: Apply your company codes.

◉ SYMBOL (root portion)

You want the data from January 2017, the time range 9:30:00 – 16:00:00 (the regular trading day), root symbol JPM, and the fields "Exchange that issued the trade," "Trade Sale Condition," "Volume of trade," "Price of trade," and "Trade Correction Indicator." I recommend against downloading data outside the time range 9:30:00 – 16:00:00 or fields that you do not need. It might be convenient to download and process the data for one day at a time rather than try to download and process the data for the entire month at once.

If you download the data in .csv format, the columns will be headed DATE, TIME_M, EX (= exchange), SYM_ROOT, SYM_SUFFIX, TR_SCOND (= trade sale condition), SIZE, PRICE, and TR_CORR (= trade correction indicator). You should use the data for which SYM_SUFFIX is blank (these are trades in JPM common stock), TR_SCOND = @ (these are regular trades), and TR_CORR = 0.

For each date in January 2017, construct a sequence of trade prices at one-minute frequency, starting from 9:31 am and ending at 4:00 pm. To this, for each minute, set the price equal to the trade price of the last trade that occurred on or prior to the minute. For example, for the minute 9:37, set the price to be equal to the trade price of the last regular trade with trade time less than or equal to 9:37:00.000.

1. Realized variance when the market is open for a liquid stock. (3 points) For each of the 20 trade dates during January 2017, compute the 1-minute continuously compounded returns from 9:45 am to 4:00 pm (the first return interval should be from 9:45 am to 9:46 am, and the last should be from 3:59 pm to 4:00 pm.). Use these continuously compounded returns to compute the realized variance during the period when the market is open. (This is what Christoffersen calls RV_t^{OPEN}.) What is the average of the 20 estimates?

2. Average realized variance estimator (3 points) Compute 2-minute, 5-minute, 10-minute, and 15-minute continuously compounded returns. Use these returns and the average realized variance estimator to estimate the realized variance on each of the 20 days based on returns at 2-minute, 5-minute, 10-minute, and 15-minute frequency. For each return interval, what is the average (across days) of the realized variance estimators? (Your answer should consist of four numbers, one each for the 2, 5, 10, and 15-minute intervals.)

When the return interval is greater than one minute and the market is not open 24 hours, the average realized variance estimator when the market is open, RV_t^{OPEN}, might seem not quite right. For example, there are 25 15-minute returns based on the prices at 9:45, 10:00, 10:15, …, 15:45, 16:00, but only 24 based on the prices at 9:46, 10:01, 10:16, …, and 15:46. There are also only 24 returns in the other series of 15-minute returns. (The 2, 5, and 10-minute returns share the same problem.) Should you somehow adjust for this? If so, how?

Remark: This is not in the textbook or lecture notes. You have to think about it, and figure out what to do.

Do your estimates of realized variance increase or decrease as the return interval becomes longer? What is the ratio of the average (across days) average realized variance based on 15-minute returns to the average (across days) realized variance based on 1-minute returns? What is the ratio of the average (across days) average realized variance based on 15-minute returns to the average (across days) average realized variance based on 10-minute returns?

3. Autocorrelation of realized variance. (1 point) Use the 20 15-minute average realized variance estimates that you computed to answer Question 2 to compute the autocorrelation of realized variance. What is the autocorrelation of this time series of realized variance estimates?

4. 24-hour realized variance estimator. (2 points) Use the 15-minute average realized variance estimates from Question 2 for each of the 20 days and the two different approaches on slides 8 and 9 of the lecture notes IntradayVolModelling-04.pptx to compute estimates of the 24-hour variance of JPM for each of the 20 days. (Because there are two approaches and 20 days, you should compute a total of $2 \times 20 = 40$ estimates of the realized variance.) What are the averages (across days) of the variance estimates computed using the two approaches?

Remark. Your answer to this question should consist of two numbers. You will have to collect the JPM opening and closing prices. Yahoo Finance is a convenient source.

5. (1 point) The Chinese stock exchanges are open four hours per day, from 9:30 – 11:30 a.m. and then from 1:00-3:00 p.m. Suppose you use data from the period when the Shanghai Stock Exchange is open to compute the realized variance of the common stock of China Construction Bank during the period when the Shanghai Stock Exchange is open. Call the estimate RV^{Open}. Please briefly describe two ways to adjust RV^{Open} to obtain an estimate of the variance over a 24-hour period?