Project 3

High-Frequency Financial Econometrics

Guilherme Salomé

September 16, 2018

The purpose of this project is to understand how to separate diffusive returns from jump returns, use the diffusive returns to compute a better integrated variance estimator, and conduct a Monte-Carlo analysis of the estimator.

This project is Due on September 27th by 11:59 PM. You must push your local repository with the `report.pdf` file back to GitHub before the deadline.

All results must be interpreted. The objective of the project is understanding the theory through its implementation, and also learning how to explain your results. Half of the work is making the computations. The other half of the work that is equally or more important is interpreting the results. You must interpret results regardless of whether the exercise asked for it or not.

All plots in the report must be self-contained. Self-contained means that a reader who only sees your figure (image and caption, but not the surrounding text) can understand what you are plotting. This translates to all plots having axis titles, correct units on the axis, and a caption that summarizes what is plotted.

This project makes use of stock data. Refer to this page to get access to the data (requires Duke login). You must complete the exercises below for both of your stocks using 5 min. data, unless stated otherwise.

Students must uphold the Duke Community Standard. Projects with excessive overlap with other student's answers will receive a zero grade.

The data files follow the .csv format and contain the prices of different assets. The name of the file represents the ticker symbol for a given stock. For example, the file `AAPL.csv` contains data for Apple's stock. Each file has 3 columns (no headers): date, time and price. The first column of a file contains the date of a given price in the `YYYYMMDD` format. For example, a date of `20070103` means January 3rd of 2007. The second column contains the time of a given price in the `HHMM` or `HHMMSS` format. For example, a time of `935` means that the price in the 3rd column was recorded at 9:35 am. If the value is `93500`, it means the price was recorded at 9:35:00 am (this is only for 5 seconds data). The last column contains the price in dollars of the stock at the given date and time.

Exercise 1

The theory for separating diffusive from jump returns uses the threshold defined below:

$$cut_{t,i} \equiv \alpha \Delta_n^{0.49} \sqrt{\tau_i BV_t}$$

To compute the threshold, first you compute the bipower variance for each day $t = 1, 2, \ldots, T$. Then, you compute the time-of-day factor, τ_i, for each time interval in a day $i = 1, 2, \ldots, n$.

The time-of-day factor is used to account for the intraday pattern of the volatility. To compute τ_i we first estimate the average (across days) bipower factors at each time interval:

$$
\begin{aligned}
b_i &\equiv \frac{1}{T} \sum_{t=1}^{T} |r_{t,i} r_{t,i-1}| \text{ for } i = 2, 3, \ldots, n \\
b_1 &\equiv b_2
\end{aligned}
$$

Then, the time-of-day factor is defined as:

$$
\tau_i \equiv \frac{b_i}{\frac{1}{n} \sum_{j=1}^{n} b_j}
$$

That is, the time-of-day factor is just a re-scaled version of b_i so that the mean of τ_i over a day is 1.

A.

Compute the time-of-day factor and plot it.

B.

Let $\alpha = 4$ and compute the jump separation threshold, $cut_{t,i}$, for your entire sample. Then compute the diffusive and jump returns:

$$
\begin{aligned}
r_{t,i}^{c} &\equiv r_{t,i} 1_{\{|r_{t,i} \leq cut_{t,i}|\}} \\
r_{t,i}^{d} &\equiv r_{t,i} 1_{\{|r_{t,i} > cut_{t,i}|\}}
\end{aligned}
$$

You may need to re-adjust α, since we usually expect about 10 to 15 jumps per year.

Plot the diffusive and jump returns.

C.

- How many jumps occur in each year?
- Are there more jumps during periods of crisis?
- Are jumps approximately evenly distributed throughout the years?

D.

- What is a density function?
- If you could estimate the density of the diffusive returns what would you expect it to look like? Explain.
- If you could estimate the density of the jump returns what would you expect it to look like? Explain.

E.

Use the ksdensity function from Matlab to estimate the density of the diffusive and jump returns. The function takes a few optional arguments which we will make use of. Specify the 'Kernel' to be 'epanechnikov', and adjust the 'Bandwidth' argument so that the density is reasonably smooth, but not over-smoothed.

- Plot the densities and interpret.
- Does the jump density estimate look strange? Explain.

F.

Many people believe that stock crashes (negative jump returns) are larger and occur much more often than rallies (positive jump returns). What do your density plots suggest about this belief?

Exercise 2

The truncated variance is defined as:

$$TV_t = \sum_{i=1}^{n} (r^c_{t,i})^2$$

A.

Compute the truncated variance for all days in the sample and make plots. Remember to annualize the values for plotting.

B.

Plot the truncated variance and the realized variance and compare. Remember to annualized the values for plotting.

C.

Compute the 95% confidence intervals for the integrated variance based on the asymptotic distribution of the truncated variance. Plot the truncated variance alongside the confidence interval.

D.

The plot in the previous exercise is quite condensed. Zoom in on a month of interest (say October 2008) and plot TV and the confidence intervals. Discuss.

E.

Answer the following questions:

- Were the asymptotic distribution results developed for the truncated variance (TV_t) or for the annualized truncated variance ($\sqrt{252TV_t}$)?
- When plotting the truncated variance and the confidence intervals did you annualize TV and the confidence interval?
- If the answer to the previous question was yes, explain what was the transformation you applied to the confidence interval. If not, explain why you did not annualize the values.

F.

It turns out that annualizing the confidence interval does not lead to the correct confidence interval for the annualized TV. Remember, the asymptotic distribution for the TV is:

$$\Delta_n^{-\frac{1}{2}}(TV_t - IV_t) \xrightarrow{d} \mathcal{N}\left(0, 2\int_{t-1}^{t} c_s^2 ds\right)$$

But we want to find the asymptotic distribution for the annualized TV:

$$\Delta_n^{-\frac{1}{2}}\left(\sqrt{252TV_t} - \sqrt{252IV_t}\right) \xrightarrow{d} \mathcal{N}(0, ?)$$

To do so, we will use the Delta Method theorem.

- Read the univariate case of the delta method theorem.
- What does the Delta Method theorem allow us to do?
- Why do we need it?
- What should be g defined as in our case?
- What is g'?
- What is the asymptotic distribution of $\Delta_n^{-\frac{1}{2}}\left(\sqrt{252TV_t} - \sqrt{252IV_t}\right)$?

G.

Given the asymptotic distribution of:

$$\Delta_n^{-\frac{1}{2}}\left(\sqrt{252TV_t} - \sqrt{252IV_t}\right)$$

What is the confidence intervals for the annualized integrated variance:

$$CI(\sqrt{252 \times IV_t}, \alpha) = \; ?$$

Compare it to the way you annualized (if you did) the confidence interval for IV in the previous exercises.

H.

Plot the annualized TV and the correct 95% confidence intervals.

I.

- How would you construct the confidence intervals for IV_t based on the bootstrap?
- How would you construct the confidence intervals for $\sqrt{252 IV_t}$ based on the bootstrap?
- Do you need to use the Delta Method as in the previous exercises? Explain.

J.

Compute the 95% confidence intervals for the integrated variance based on the bootstrap of the truncated variance. Use 1000 bootstrap repetitions for each day when testing your code, and when your code works change it to 10000 repetitions. You need to be careful on how you do the bootstrap, otherwise your code will take incredibly long to run. You can read more about being efficient with bootstrap here.

Plot the truncated variance and the confidence intervals. Remember to annualize the values.

K.

Zoom in on the same month used in part D. Compare the confidence intervals based on asymptotic distribution and on the bootstrap.

Exercise 3

In this exercise you will do a Monte Carlo analysis to evaluate the accuracy of the IV estimators (and the asymptotic theory) we have developed so far.

The idea is to simulate a jump-diffusion model, and use the simulated values of the c_t process to compute the actual integrated variance. Then, we will put ourselves in the place of the practitioner, who only has access to the stock returns, and uses these returns to estimate IV via the estimators we have studied.

A.

Simulate the following model:

$$\begin{aligned} dX_t &= \sqrt{c_t} dW_t \\ dc_t &= \rho(\mu_c - c_t)dt + \sigma_c \sqrt{c_t} dW_t^c \end{aligned}$$

Assume that:

$$
\begin{aligned}
N &= 79 \\
T &= 1.25 \times 252 \\
N_E &= 20 \times N \\
\rho &= 0.03 \\
\mu_c &= 0.000121 \\
\sigma_c &= 0.0005 \\
c_0 &= \mu_c \\
X_0 &= \log 100
\end{aligned}
$$

Use the following additional rule. If the simulated c becomes smaller than $\mu_c/2$ change the value to $\mu_c/2$ instead.

Plot the simulated variance process and the simulated prices.

B.

Throw away the first price, so that we have exactly $N = 79$ prices for each of the 315 days. And throw away the first value of c ($c_0 = \mu_c$) so that we have exactly $N_{euler} = 1580$ values for the variance process per day.

Because we "observe" the values of the variance process c we can compute the actual IV:

$$IV_t \equiv \int_{t-1}^{t} c_s ds \text{ for } t = 1, 2, \dots, T$$

- What does the numerical value of IV_t represent? Hint: start with the case where the variance is constant, and then think about the variance being constant into small intervals.
- We do not have a continuum of values for c_t, but only as many as we simulated. How can we use those to compute the actual IV_t?

C.

Compute the actual IV for each of the simulated days:

$$IV_t = \frac{1}{N_{euler}} \sum_{j=(t-1)N_{euler}+1}^{tN_{euler}} \tilde{c}_j \text{ for } t = 1, 2, \dots, T$$

Plot the values, remember to annualize.

D.

Now, use the prices from the simulation to compute the intraday returns. Remember, we use the Euler scheme to reduce the discretization error, but we are simulating 5-minute prices, not 15-second prices. Use the intraday returns to compute RV_t for all simulated days.

Plot the realized variance alongside the actual integrated variance. Discuss.

E.

Remember that for RV the asymptotic distribution is (we are ignoring jumps for now):

$$\Delta_n^{-\frac{1}{2}}(RV_t - IV_t) \xrightarrow{d} \mathcal{N}\left(0, 2\int_{t-1}^{t} c_s^2 ds\right)$$

Use the Delta-Method to compute the asymptotic distribution for the annualized RV:

$$\Delta_n^{-\frac{1}{2}}(\sqrt{252RV_t} - \sqrt{252IV_t}) \xrightarrow{d} \mathcal{N}(0, ?)$$

What is the confidence interval for the annualized IV?

F.

Compute the 95% confidence interval for IV using the simulated returns. Plot the IV and the confidence intervals.

The theory says that, for n big enough, the probability of the actual integrated variance be within the confidence interval should be 95%.

G.

- For how many days is IV_t within the confidence interval?
- What is the average coverage rate?
- Is it close to the expected 95%?

H.

We know that when the stock price can jump the RV_t estimator becomes biased. Let's see this effect in practice.

First, let's add jumps (compound Poisson) to the model

$$\begin{aligned} dX_t &= \sqrt{c_t}dW_t + dJ_t \\ dc_t &= \rho(\mu_c - c_t)dt + \sigma_c\sqrt{c_t}dW_t^c \end{aligned}$$

Use the following parameters to simulate the jumps:

$$\begin{aligned} \lambda &= 15/252 \\ \sigma_{jump} &= 25 * \sqrt{\mu_c/N} \end{aligned}$$

Add the jumps to the simulated prices.

Plot the simulated prices with jumps.

I.

Use the simulated prices from the model with jumps to compute the intraday returns. Then compute RV for each day of the sample, and compute the confidence intervals for the annualized IV based on the asymptotic theory.

- What is the coverage rate of the confidence intervals?
- How does it compare to when there are no jumps?
- Is the value of the coverage rate what you expected? Explain.
- How did the failure to account for jump returns affect the inference?

J.

Change the jump intensity to a higher value, say $\lambda = 1$, and redo the previous exercise.