Midterm

High-Frequency Financial Econometrics

Guilherme Salomé

October 1, 2018

The purpose of this exam is to test your knowledge of the contents discussed in class and in the projects, and to test your ability to work with new concepts.

Read carefully the instructions below.

The exam is Due on October 4th by 11:59 PM. You must push your local repository with all the required files (all Matlab files and finished report in pdf) to GitHub before the deadline. The repository must contain all functions and scripts used in solving the exam. It should also contain a `main.m` file that generates all required plots for the exam when run. Finally, it should have a `report.pdf` containing your answers to the exam questions.

All results must be interpreted. Half of the work in the exam is doing the computations. The other half of the work is interpreting the results. You must interpret results regardless of whether the exercise explicitly 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.

The exam 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.

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.

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

The exam is open book. You can use all texts, lecture notes and previous Matlab scripts you developed. Note that if your previous Matlab scripts were not completely correct and you use them when solving the Midterm, leading to results that are not completely correct, then you may loose points in your solution.

Exam is strictly individual. You must not talk or communicate with each other or anyone else about the exam or course material. This restriction includes e-mail, voice, phone, text, or any other means of communication. The answers you turn in must be your own. Failure to do so will lead to a zero grade.

By turning in the exam you agree to the conditions above.

Good luck!

Exercise 1 - Local Variance

A.

- Load your stock data and plot prices and returns.
- Separate diffusive from jump returns using $\alpha = 4.5$.
- Plot continuous and jump returns.
- How many jumps occur each year?

B.

The local variance process c_t is a process defined over $[0, T]$. While we cannot estimate the value of c_t for all times $t \in [0, T]$, we can estimate the value of c_t at intervals containing t. Let $c_{t,i} \equiv c_{i\Delta_n + (t-1)}$ denote the value of the local variance process at day t and at the i-th interval.

Write a Matlab function that computes the local variance estimator for $c_{\mathrm{t,i}}, which we will denote by $\hat{c}_{t,i}$. Assume that the variance process is continuous. Hints:

1. The function should take as arguments a vector of continuous returns r^c, the window size parameter k_n, the day and time-interval for which to estimate c (that is, t and i), and the number of returns in a day (n);
2. Write a function to estimate the local variance assuming that you are in the middle of the day;
3. Write a function to estimate the local variance assuming that you are in the beginning or end of the day;
4. Combine both and use the information from i, k_n, n to decide whether you are in the middle or in the beginning/end of the day.

Select a day in your sample (fix some t) and compute the local variance estimator for intervals (for $i = 1, 2, \ldots, n$) using $k_n = 6$. Plot the results and comment.

C.

Use the function you created before to compute the local variance estimator $\hat{c}_{t,i}$ for your entire sample. Then, compute the average local variance across time for each interval:

$$\bar{\hat{c}}_i \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{c}_{t,i}$$

for $i = 1, 2, \ldots, n$.

Plot the values of $\bar{\hat{c}}_i$. Do you observe any pattern in the plot? Interpret. How does it relate to the time of day factor?

Exercise 2 - Confidence Interval for Jump Size

An useful application of local variance is to measure the precision with which the observed return across a jump interval measures the actual jump itself. Suppose a jump occurs at a time τ identified to lie somewhere in the interval $\tau \in ((i_\tau - 1)\Delta_n, i_\tau \Delta_n]$. The econometrics jump detector (threshold) can identify the interval labels (i_τ) of the jumps, but we do not know precisely when within that interval the jump occurred. Indeed, the return over that interval is:

$$r_{t,i_\tau} = \int_{(i_\tau - 1)\Delta_n}^{i_\tau \Delta_n} \sqrt{c_s} ds + \Delta J_\tau$$

That is, over the interval $((i_\tau - 1)\Delta_n, i_\tau \Delta_n]$ the return is due to the diffusive part plus a jump part that occurred at some point within that interval.

If Δ_n is small and the variance process is continuous across the jump interval, then the integral above is approximately normally distributed:

$$\int_{(i_\tau - 1)\Delta_n}^{i_\tau \Delta_n} \sqrt{c_s} ds \approx Y_\tau, \text{ where } Y_\tau \overset{d}{\sim} \mathcal{N}(0, c_\tau \Delta_n)$$

Therefore, the return is approximately:

$$r_{t,i_\tau} \approx Y_\tau + \Delta J_\tau$$

If we have an estimate of c_τ, then we can create a 95% confidence interval for the jump size ΔJ_τ:

$$CI(\Delta J_\tau) = [r_{t,i_\tau} - 1.96\sqrt{\hat{c}_\tau \Delta_n}, r_{t,i_\tau} + 1.96\sqrt{\hat{c}_\tau \Delta_n}]$$

We can use the local variance estimator to give us the $\hat{c}_\tau$ needed to construct the confidence interval above.

A.

- Select the three largest jumps in magnitude for each of your stocks.
- If a jump is located within the first or last half hour of the day, then choose the next largest jump in magnitude (to avoid dealing with end effects).
- Estimate c_τ with $k_n = 6$ for each of the selected jumps intervals.
- Compute the 95% confidence interval for each of the jumps.
- Create a table reporting the jump magnitude, the local variance estimator and the confidence interval.
- Comment on the accuracy with which the returns across the jump interval estimates the actual jump.

B.

If the local variance function is discontinuous at τ, that is, the local variance jumps when the price jumps, then the way to compute the confidence intervals needs to be changed. In this case, the return over the interval where there is a jump is approximately:

$$r_{t,i_\tau} \approx \Delta J_\tau + \sqrt{\Delta_n}\zeta_\tau, \text{ where } \zeta_\tau = \sqrt{c_\tau^- \rho} Z_\tau^- + \sqrt{c_\tau^+ (1-\rho)} Z_\tau^+$$

where $Z_\tau^- \overset{d}{\sim} \mathcal{N}(0,1)$, $Z_\tau^+ \overset{d}{\sim} \mathcal{N}(0,1)$, $\rho \overset{d}{\sim} \text{Unif}[0,1]$ and c_τ^-, c_τ^+ are the left and right limits of the local variance process.

The confidence interval in this case can be determined via a bootstrap-style simulationn. Simulate:

$$\begin{aligned} \tilde{Z}_\tau^- &\overset{d}{\sim} \mathcal{N}(0,1) \\ \tilde{Z}_\tau^+ &\overset{d}{\sim} \mathcal{N}(0,1) \\ \tilde{\rho} &\overset{d}{\sim} \text{Unif}[0,1] \end{aligned}$$

And then compute:

$$\tilde{\zeta}_\tau = \sqrt{\hat{c}_\tau^- \tilde{\rho}} \tilde{Z}_\tau^- + \sqrt{\hat{c}_\tau^+ \tilde{\rho}} \tilde{Z}_\tau^+$$

Repeat the procedure 1000 times to generate $\left\{\tilde{\zeta}_\tau^{(k)}\right\}_{k=1}^{1000}$, and then find the 2.5% and 97.5% quantiles. Multiply the quantiles by $\sqrt{\Delta_n}$ and then add/subtract the return to create the 95% confidence intervals for the jump magnitude.

- Create the confidence intervals for each of the jumps selected in the previous part.
- Comment on the importance of accounting for the variance jump for the confidence intervals.
- There is a lot of empirical evidence that price and variance co-jump, but are the possible co-jumps important for your particular data sets?

Exercise 3 - Realized Betas

For this exercise you will use the 5 minute data for the SPY (market index) in addition to your two stocks. If one of your stocks is already the SPY, then choose a third different stock from the data folder.

A.

We want to explore the dependency of moves in a stock to moves in the market index. To do so, create a Matlab function that computes the realized beta estimator for two different stocks. Hint: refer back to the realized beta lecture, what should be X_1 and what should be X_2?

Use this function to compute the realized beta between your stocks and the market index, for every day of the sample.

- Plot the realized betas.

- How do you interpret the realized beta?
- Does the realized beta vary over the years?
- Does it seem plausible to assume a fixed beta like in the usual CAPM?

B.

Use the bootstrap with $M = 7$ (or $k_n = 11$) to compute confidence intervals for the realized betas.

- Plot the realized beta alongside the confidence intervals and comment on the accuracy of intervals.
- Zoom in on the plot for a month of interest.

C.

A stock with $\beta = 1$ is called a unit beta stock, meaning it moves one-for-one with the overall market. Many of the large-cap stocks like those in the DOW are unit beta stocks. We can use the confidence intervals for the realized betas to test whether that is true. If the confidence interval contains 1, then we say there is evidence that the stock is unit beta, at least on a given day. For each stock, compute the number of days where:

- The confidence interval contains 1;
- The confidence interval is below 1 (we say the interval $[a, b]$ is below 1 if $a < b < 1$);
- The confidence interval is above 1 (we say the interval $[a, b]$ is above 1 if $1 < a < b$);

Report the numbers above. Would you say that the stock is generally as risky as the market? Or more so, or less so? Justify.