Project 4

High-Frequency Financial Econometrics

Guilherme Salomé

October 11, 2018

The purpose of this project is to analyze how jumps in the market index affect the moves of individual stocks. You will analyze two stocks

This project is Due on October 18th 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.

Exercises

The main idea of jump regressions is just the opposite of what we saw in realized beta regressions. Instead of discarding the jump data and using only the diffusive returns, in the jump regression we keep only the data points at which the market index jumped and discard the rest.

Suppose the data are (X_1, X_2) where X_1 is the market index, as given by the S&P500 index (SPY), and X_2 is one of your stocks. Notice that if one of your original stocks is SPY then you need to download a third stock for this project.

The jump regression model is:

$$\begin{aligned} \Delta X_{2,t} &= \beta \Delta X_{1,t} + \Delta J_{2,t} \\ \Delta X_{1,t} \Delta J_{2,t} &= 0, \forall t \in [0, T] \end{aligned}$$

where T is fixed. This model is the CAPM from finance applied to the jump moves.

We do not observe the jumps directly, but we can identify the intervals over which the market jumps. The index of the market jump intervals are:

$$\mathcal{I}'_n \equiv \left\{ i : |\Delta_i^n X_1| > \alpha \Delta_n^{0.49} \sqrt{\tau_i BV_t} \right\}$$

The list of integers $\mathcal{I}'_n$ is just a list of the subset of the nT intervals where the market is detected to have jumped. There are P_n elements in $\mathcal{I}'_n$ (the detected market jumps), which we denote by i_p for $p = 1, 2, \ldots, P_n$.

A.

Run the jump detection on the market index to locate the jump indices i_p. Use $\alpha = 5$ or some other appropriate value. Report the number of jumps per year and the average magnitude of the jumps each year.

B.

Make a scatter plot of the stock returns at the jump times, that is $\Delta_{i_p}^n X_2$, against the market jump returns, $\Delta_{i_p}^n X_1$. Does a linear jump regression appear plausible?

C.

The OLS jump regression estimator is:

$$\hat{\beta} \equiv \frac{\sum_{p=1}^{P_n} \Delta_{i_p}^n X_1 \Delta_{i_p}^n X_2}{\sum_{p=1}^{P_n} (\Delta_{i_p}^n X_1)^2}$$

Compute the OLS jump beta for each of your stocks. Report the estimated values and interpret.

D.

Add to the scatter plot of part B. a regression line $y = x\hat{\beta}$ for $1.3 \times \min_p \Delta_{i_p}^n X_1 \leq x \leq 1.3 \times \max_p \Delta_{i_p}^n X_1$. Use a fine grid for the values of x. The 1.3 just extends the line over the domain so that the plot looks better.

E.

Given the jump beta estimate, $\hat{\beta}$, compute the local residuals: $\hat{E}_t = X_{2,t} - \hat{\beta} X_{1,t}$. Estimate the local variance of the residual process using the local variance estimator at the jump times. Use $k_n = 7$ to compute $\hat{c}_{e,t_p}$ for $p = 1, 2, \ldots, P_n$. Compute:

$$\widehat{V}_\beta = \frac{\sum_{p=1}^{P_n} (\Delta_{i_p}^n X_1)^2 \hat{c}_{e,t_p}}{\sum_{p=1}^{P_n} (\Delta_{i_p}^n X_1)^2}$$

Report the value of $\sqrt{\Delta_n \widehat{V}_\beta}$.

F.

Create a 95% confidence interval for the jump beta. Report the confidence interval.

G.

Imagine you work in a hedge fund and have an account holding \$100 million worth of a particular stock. You are concerned about a possible market jump in the near future. Suppose you can trade market index futures (SPY futures). How can you hedge your position using futures? Based on the confidence intervals what is the range of values you need to short sell in order to hedge your position?

H.

One of the underlying assumptions is that the jump beta is constant. Let's test whether that is plausible. Split your sample into two periods:

- Period 1: 2007-2011
- Period 2: 2012-2017

Estimate the jump beta for each of the periods: $\hat{\beta}_1$ and $\hat{\beta}_2$. Compute $\widehat{V}_\beta$ for both periods: $\widehat{V}_{\beta_1}$ and $\widehat{V}_{\beta_2}$.

I.

It is possible to create a 95% confidence interval for the difference of the jump betas:

$$CI(\beta_1 - \beta_2) = \hat{\beta}_1 - \hat{\beta}_2 \pm 1.96\sqrt{\Delta_n \hat{V}_{\beta_1 - \beta_2}}$$

We need to compute $\hat{V}_{\beta_1-\beta_2}$. The fluctuations in the jump betas are generated by the diffusive parts of the process, and so are independent across the two periods. Thus, we can compute $\hat{V}_{\beta_1-\beta_2}$:

$$\hat{V}_{\beta_1-\beta_2} = \hat{V}_{\beta_1} + \hat{V}_{\beta_2}$$

Compute the confidence interval above. Does the confidence interval contain the number 0? What does that mean?