Project 7: Microstructure Noise and the Two-Scale Realized Variance

Instructions

Project 7 is due on November 1st by 10:00 pm. This is a hard deadline, so no exceptions. You must push your local repository back to GitHub before the deadline. Your repository must contain:

- The Matlab code you used to complete the project;
- A script named `main.m` file that generates all required plots;
- A `report.pdf` file with your answers to the project questions. The report must also contain an Appendix with the code used to solve the project;
- 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 the Data page for instructions on how to download the data and which files to download (requires Duke login). You must complete all exercises for both of your stocks using the data at the 5-minutes sampling frequency, unless stated otherwise.

You can obtain the repository for this project by clicking **on this link**.

Questions

The purpose of this project is to implement to understand the effect of noise on the estimates for the integrated variance, and implement the two-scale realized vairance estimator, that takes the noise into account when estimating the integrated variance. In the data folder you can find data available at the 5 seconds frequency. You should use that data for this exercise.

A.

Load the data using the appropriate values for n and T. Create a function to compute the realized variance using coarse sampling. For now, assume the summation starts at the very first price. The function should take a parameter k_n that specifies how coarse the sample should be. For example, if we want to compute RV using a 5-minute sampling frequency, we would input $k_n = 60$ (since the raw data is sampled every 5 seconds).

B.

Use the function created in the previous item to make a volatility signature plot for your stock. For each day, compute RV using $k_n = 1, 2, 3, \ldots, 120$. Average the value of RV over the entire year for each k_n. Plot RV (in the appropriate units) against the sampling frequency (scale the axis to units of 1 minute). Using the theory discussed in class about noise interpret the results of the plot.

C.

Estimate the value of σ^2_χ using the data at the highest frequency. Compute the estimate for each day of the sample. Explain the idea behind the $\hat{\sigma}^2_\chi$ estimator?

D.

When we sample at rate k_n there are $\lfloor n/k_n \rfloor$ terms in the RV sum. Therefore, the relative contribution of the noise to the total return variation RV is:

$$contribution(k_n) \equiv \frac{2\frac{n}{k_n}\sigma^2_\chi}{RV}$$

Using the material from the last two lectures, explain why $contribution(k_n)$ above is a measure of:

$$\frac{\text{Noise}}{IV + \text{Noise}}$$

at frequency k_n.

E.

Compute the average (across days) $\widehat{contribution}(k_n)$ for each frequency $k_n = 1, 2, \ldots, 120$ by substituting σ^2_χ for $\hat{\sigma}^2_\chi$:

$$\text{average contribution}(k_n) = 100\frac{1}{T}\sum_{t=1}^{T}\text{contribution}(k_n, t)$$

Plot the average contribution against the frequency k_n.

Does the noise dominate RV at the very high frequencies in the sense of the contribution being close to 100%? The noise is considered unimportant at low frequency if it contributes less than 10% of the total variation at 5-min and 8-min. Is the noise unimportant at low frequencies in your data set?

F.

Modify the function that computes RV with coarse sampling so that it starts the summation at a price different from the first one. For example, in the case of coarse sampling from 5-seconds to 30-seconds, we could compute RV starting at 6 different values:

$$\begin{aligned}
RV_0 &= (Y^n_6 - Y^n_0)^2 + (Y^n_{12} - Y^n_6)^2 + \cdots + (Y^n_{L(0)} - Y_{L(0)-6})^2 \\
RV_1 &= (Y^n_7 - Y^n_1)^2 + (Y^n_{13} - Y^n_7)^2 + \cdots + (Y^n_{L(1)} - Y_{L(1)-6})^2 \\
&\vdots \\
RV_5 &= (Y^n_{11} - Y^n_5)^2 + (Y^n_{17} - Y^n_{11})^2 + \cdots + (Y^n_{L(5)} - Y_{L(5)-6})^2
\end{aligned}$$

Further modify the function so that it returns the average RV based on the coarse sampling. For example, if we consider the example above, then the function should return $\frac{1}{5}(\mathrm{RV}_0 + \cdots + \mathrm{RV}_5)$.

Use this function to compute $\mathrm{RV}^{\mathrm{subave}}$ for the sampling-frequency of 5-minutes. Plot $\mathrm{RV}^{\mathrm{subave}}$ and the usual RV based on 5-minutes data. Are there any differences between the estimates? Should there be differences? Explain.

G.

Compute the two-scale realized variance estimator for the highest frequency $k_n = 1$ (5-seconds) for each day. Plot TSRV and RV based on 5-min coarse sampling against time. Interpret the differences.

H.

Compute TSRV for the frequencies $k_n = 1, 2, 3, 4, \ldots, 120$. Compute the average value of TSRV over all days, for each of the sampling frequencies. Add the values of TSRV to the volatility signature plot. Does TSRV deal with the microstructure noise? Is the curve relatively flat even when the sampling frequency increases? How does it compare to the usual RV estimator?