Realized Volatility Estimation with two-scale estimator

Juncheng Li

As the sample interval becomes smaller, each time span contains more sample points. If we plot log returns of some prices' series, they fluctuate more violently. Such fluctuation stems from the behaviour of the trading price. It often bounces, representing different opinions on fair price of the market.

The challenge of sampling densely is the distorted volatility estimate. Such distortion is called microstructure noise. Here are the questions we seek answers:

Problem1: Does microstructure noise really exists?
Problem2: Does the Two-Scale estimator alleviate microstructure noise?

Firstly, we plot the time series of log return calculated from mid-quote price, defined as the average of bid1 quote and ask1 quote at the same timing point.

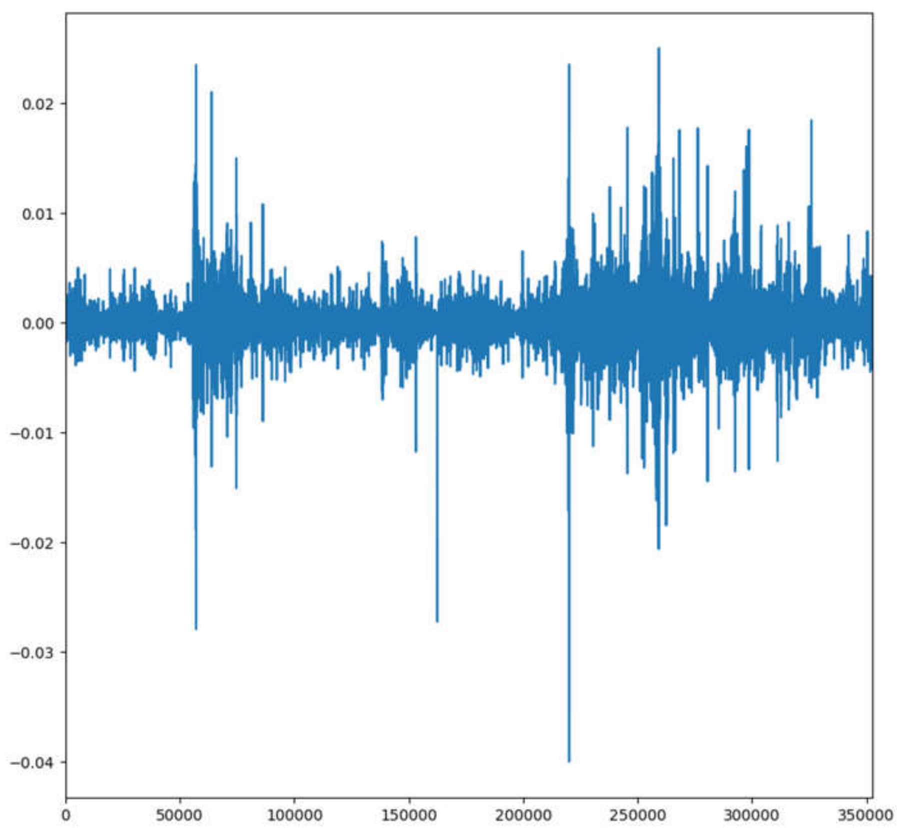

Fig1. Log Return Mid-Quote Series of the BTC_CNY, from 8^{th} Sep. 2017 to 18^{th} Sep. 2017.

Next, we derive the daily estimation of realized variance. The connection of realized variance and the volatility can be written as follows: $RV_{[0,T]} = \int_0^T \sigma_s{}^2 ds$

Without loss of generality, all the estimated variances are annualized with time-scale dependent coefficient.

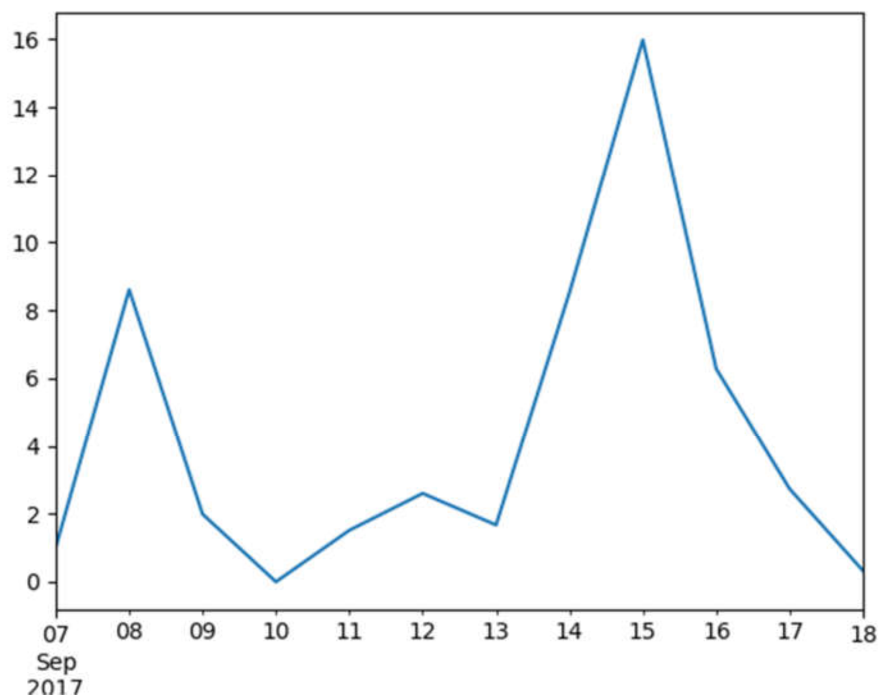

Fig2. Annualized Daily Realized Variance Estimates of the BTC_CNY.

While we take a closer look on the hourly realized variance annualized by the scaling factor $8760 = 365(days) \times 24(hour / day)$, mid-quote return series become more volatile.

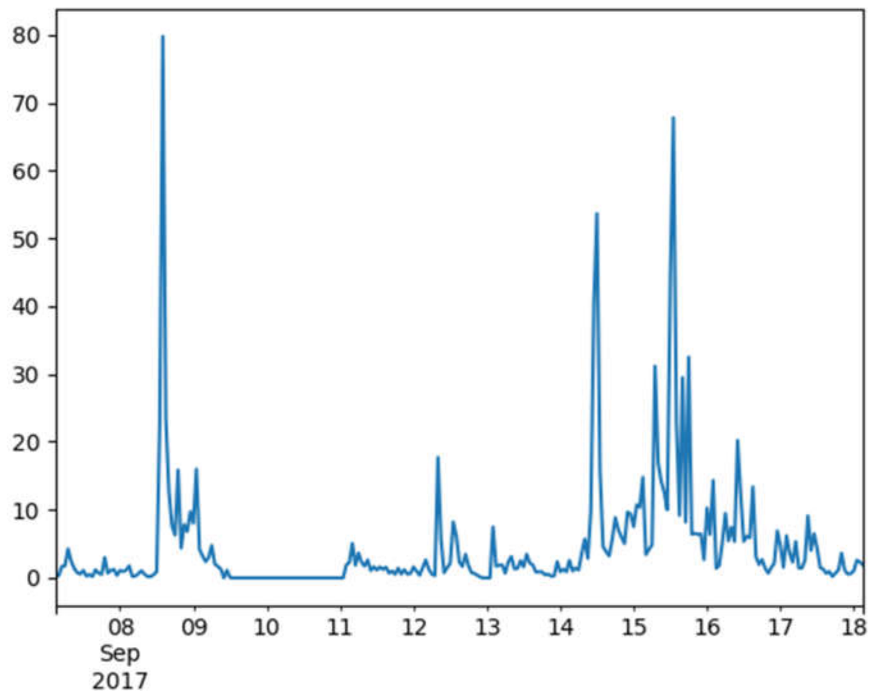

Fig3. Annualized Hourly Variance Estimates of the BTC_CNY.

In addition, the annualized realized variance by minute skyrockets, with scaling factor $525600 = 365(days) \times 24(hour/day) \times 60(minutes/hour)$,

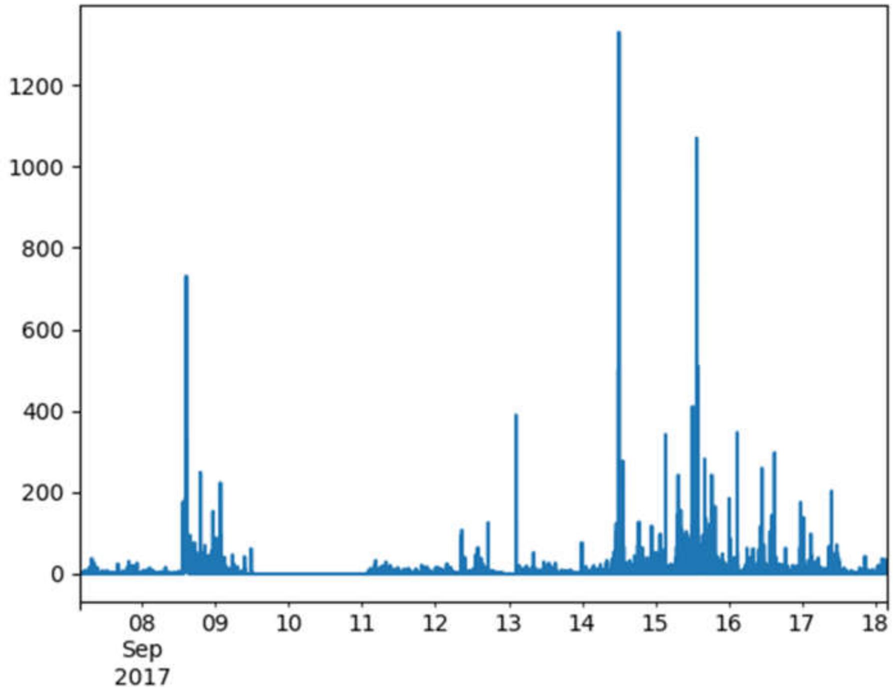

Fig4. Annualized Minutely Variance Estimates of the BTC_CNY.

A good starting point may be estimate the realized Variance every hour with two-scale estimator[1]. The two-scale estimator on realized hourly annualized Variance is defined as follows:

$$RV_{Two\_Scale,hour(j)} = \left(\frac{1}{K}\sum_{i=1}^{K} RV^{(i,n_i)} - \frac{\overline{n}}{n} RV^{(all)}\right) \times (365 \times 24) \quad j = 1,2,\ldots N,$$

N represents the number of hours occurred over the dataset. K represents the number of bins in the tick dataset and $\overline{n}$ represents the average number of the ticks fall into each bin. $\overline{n} = \frac{1}{K}\sum_{i=1}^{K} n_i$. In addition, n_i represents the number of ticks included in one of the K bins and n is the total number of ticks within the sample period.

We remind readers that the mid-quote price generates from the depth table, in which multiple levels of bids and ask are included, and in which sampling interval is prefixed. But the trade price is extracted from the order book, in which every row corresponds to a tick.

In Zhang and Mykland's paper, authors assumed that the tick data is evenly sampled with pre-assigned time interval. However, since the arrival of the transaction is random, it may be

[1] We also estimated the volatility on daily basis. But the two-scale estimator does not smooth out the estimated volatility. Interested readers may refer to the Appendix Section for the results.

more reasonable to estimate with pre-assigned tick number. But sampling by tick makes it more difficult to choose the appropriate time scaling factor for estimating the Variance. To reduce complexity, we can assign the number of ticks in each bin (denoted as)as the average of tick number every hour.

To check the sensitivity of the estimated Variance over the change of $\bar{n}$, we can tune $\bar{n}$ by multiplying it to several appropriate scaling factors and observe changes.

Firstly, we present the histogram of number of ticks in each hour.

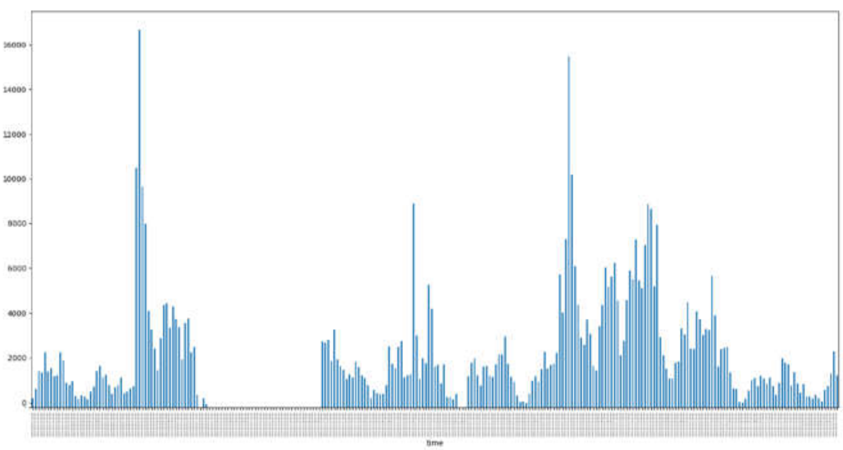

Fig 6. Number of Ticks Each Hour Over the Trading Period.

Size of the ticks fluctuates without regular patterns. To observe the distribution of tick numbers, we draw a boxplot.

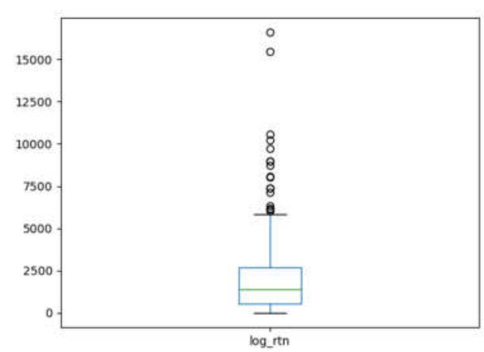

Fig7. Boxplot of Each Hour's Ticks Size

On average, size of ticks is about 1,500, with about a dozen of outliners, implying that long tail of trading ticks exists. By taking the number of ticks in each period as the mean, we can conduct the estimation on integrated Variance with two-scale estimator. The results are displayed in Fig 8.

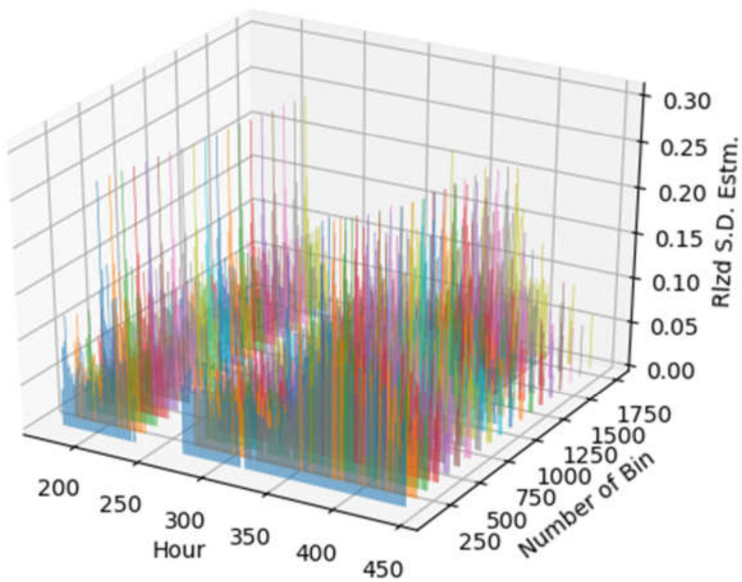

Fig 8 Bar plot of Estimated Two-Scale Realized Variance by Hour with different K

The smooth-out effect is also not evident, despite increasing the number of bin resulting smaller n bar. Consequently, some spikes in the Variance may smooth out.

To derive further analysis, we begin with minute estimate.

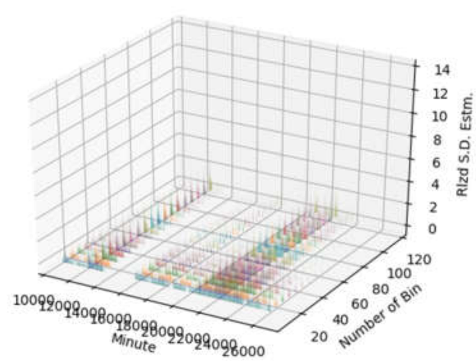

Fig 9 Bar plot of Estimated Two-Scale Realized Variance by Minute with different K

Negative values and extreme positive values (up to 12) appeared in the estimation of the Variance may imply that minute interval is too small. Consequently, we continue our experiment in the following manner, by setting the sample interval as follows:

1. 30minutes

2. 45minutes
3. 15minutes

Now we show the 30 minutes results:

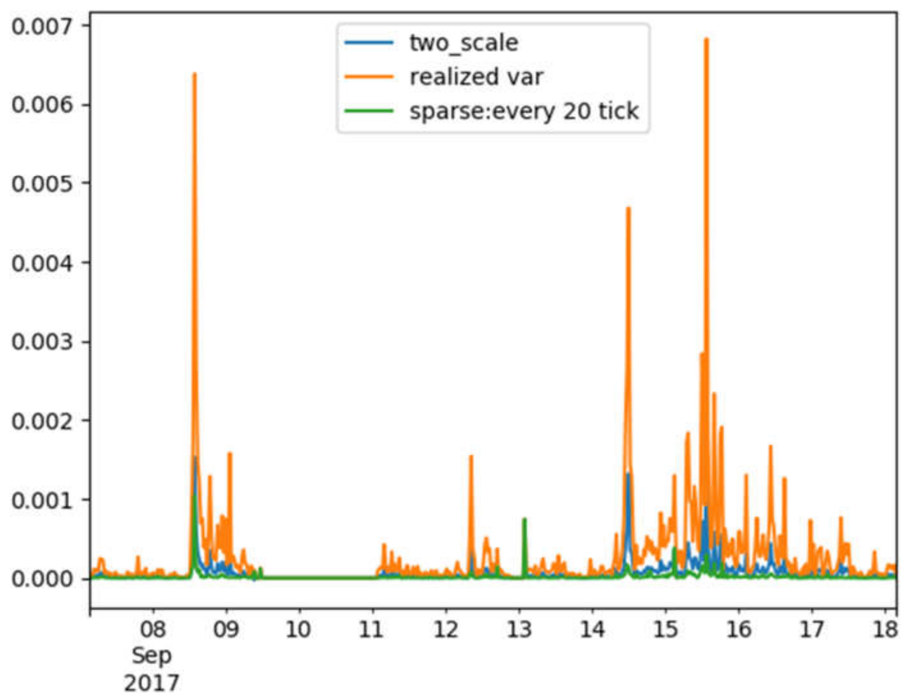

Fig 10. Line Graph of Two-Scale Variance and Realized Variance(Mid-Quote), 30 minutes

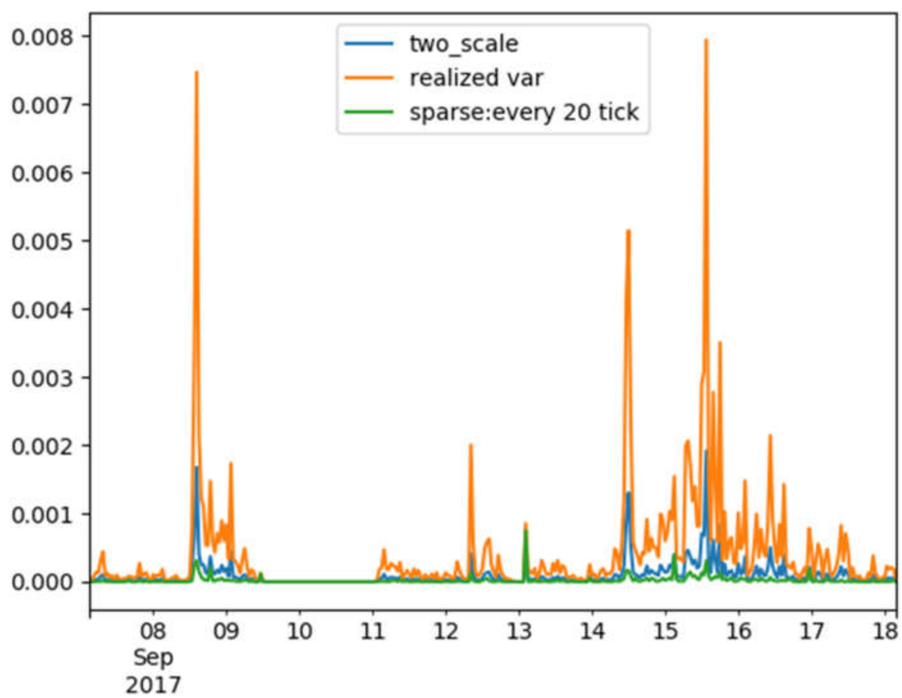

Fig 11. Line Graph of Two-Scale Variance and Realized Variance(Mid-Quote), 45 minutes

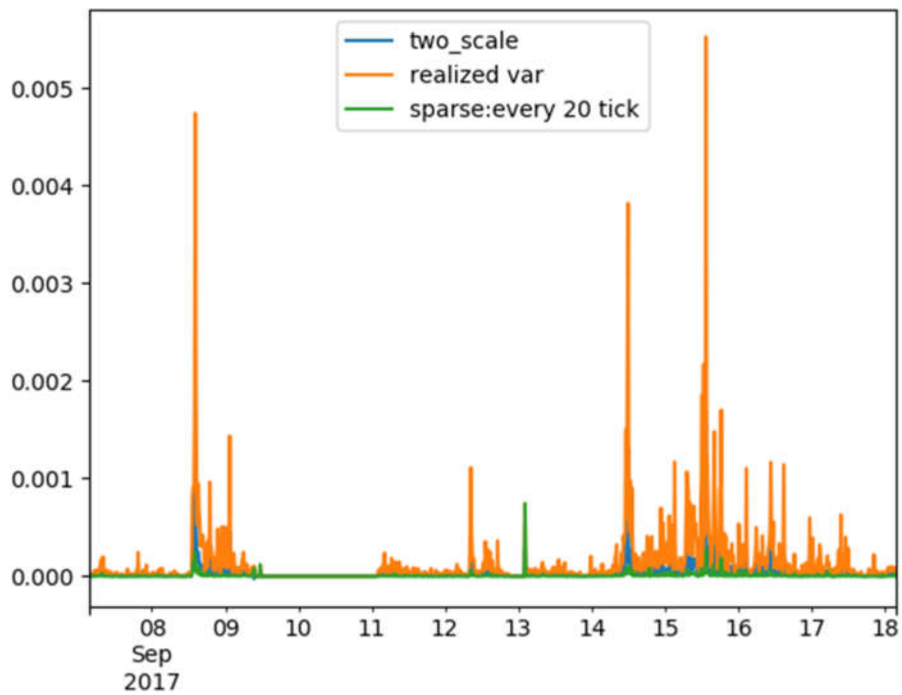

Fig 12. Line Graph of Two-Scale Variance and Realized Variance(Mid-Quote), 15 minutes