Intraday volatility estimation from high frequency data

Sarahnour Ghaith & Thomas Roiseux

2023-2024

Python has been utilized as the computational tool for performing all the required calculations and model simulations. To align with our practical methodology, please consult the attached Jupyter Notebook file.

1 Introduction

This project aims to estimate daily volatility from high-frequency time series data while assessing the impact of microstructure noise. By utilizing tick data from Ishare S&P (IVE), we will be:

1. Estimating and graphically representing realized volatility using observation frequencies ranging from 30 seconds to 15 minutes.
2. Comparing these estimates with a long-term estimation based on one month of daily data.
3. Assessing the size of microstructure noise using methods such as autocorrelation between returns at different scales.

With the data provided by KIBOT - Free historical intraday data, offering access to tick transaction data for Ishare S&P, we have a unique opportunity to explore the evolution of intraday volatility over the past year. This project is part of a comprehensive analytical approach to understand the underlying mechanisms of price behavior in a dynamic financial context.

2 Data Collection

Chosen data set: Ishare S&P values (IVE) tick data We've opted for tick data from Ishare S&P (IVE), providing transaction-specific details like time stamps and prices. IVE represents iShares S&P 500 Value ETF, enabling a detailed analysis of high-frequency market dynamics for large-cap U.S. value stocks. This dataset aligns with our focus on estimating daily volatility, forming a solid basis for our analytical pursuits.

Source of Data The data for our analysis is sourced from `http://www.kibot.com/free_historical_data.aspx`, a platform offering free tick intraday data. The information is stored in text format files, and for our study, we have selected the first file, **"Tick with bid/ask data (unadjusted)"**. This dataset encompasses tick data with bid/ask details, providing a comprehensive view of market dynamics. Notably, the data series initiates from September 2009, offering a substantial timeframe for our analytical investigations. The choice of this source aligns with our objective to access high-quality, free tick data for the Ishare S&P (IVE) values.

Data Structure The tick data obtained follows a structured format with six base columns: Date, Time, Price, Bid, Ask, and Size. The timestamp is recorded in seconds, allowing for precise temporal analysis. Notably, the data collection approach is not based on a predefined regular frequency but is rather conducted frequently enough to facilitate analyses with a minimum analysis frequency of 10 seconds.

3 Pre-processing

3.1 Data cleaning and filtering

After lading all data, we noticed that some price values could be considered as outliers (values lower than 10 \$) so we dropped these values. This could be done as we have more than 11 millions of samples and dropping 9 of them is not a big issue.

3.2 Organizing data into time series format

Our dataframe is now made of all timestamps and the associated prices. Plotting it result in figure 1, combining all prices and the daily opening prices.

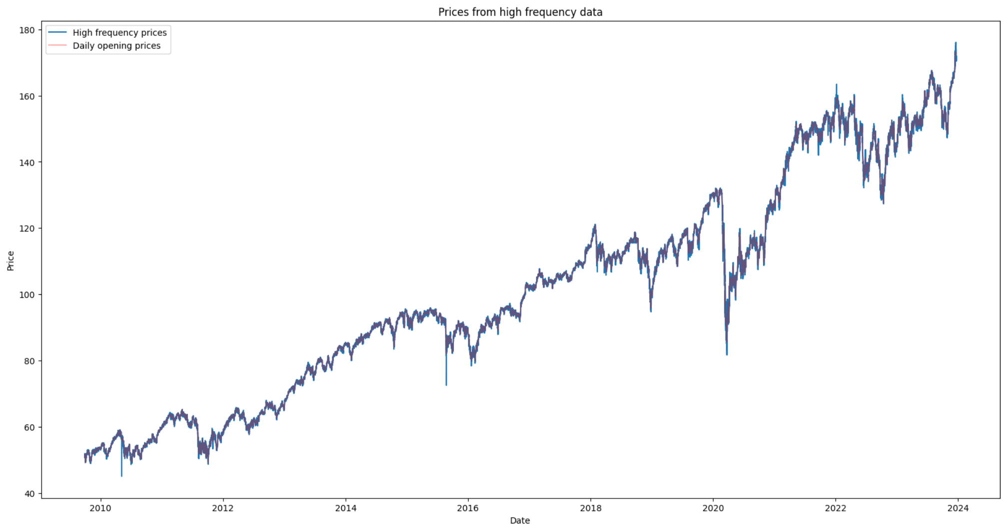

Figure 1: All prices since 30/09/2009

4 Intraday Volatility Estimation

4.1 Introduction to realized volatility

Realized volatility is the volatility computed using the past values of the index. It is defined by:

$$RV = \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i})^2 \tag{1}$$

4.2 Estimation method: rolling window approach

To compute this volatility, we chose the approach of using rolling windows. To do so, we selected several time intervals to parse data using these intervals. Then we computed the realized volatility with the remaining data. For example, for the time interval `30 seconds`, we chose points separated by at least 30 seconds between two consecutive ones.

We expect to notice that the realized volatility will decrease as we increase the time interval, as we will have fewer price variation among the day.

4.3 Visual representation: plot the estimated realized volatility over time

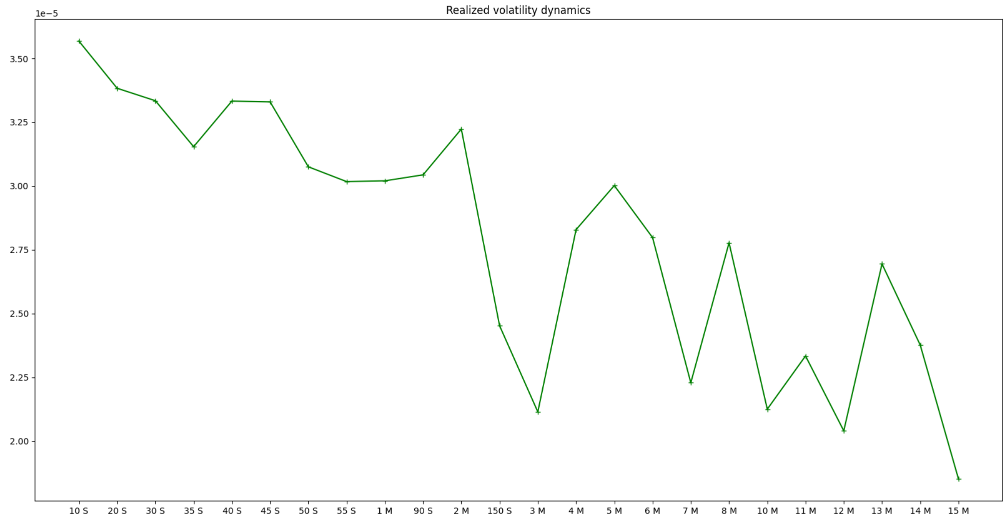

Figure 2: Realized volatility with variable time interval

After plotting all the realized volatility with all time intervals (figure 2), we indeed notice that volatility decreases when time interval grows. This can be explained as volatility in related to the price variations and by increasing the interval, we select fewer prices among the day.

5 Long-Range Volatility Estimation

5.1 Estimating volatility based on 1 month of daily data

To compare this value with the previous graph, we will compute the volatility using the same month. To do it using daily data, we are going to compute the daily realized volatility during the whole month and plot this curve.

5.2 Comparing intraday volatility estimations with long-range volatility

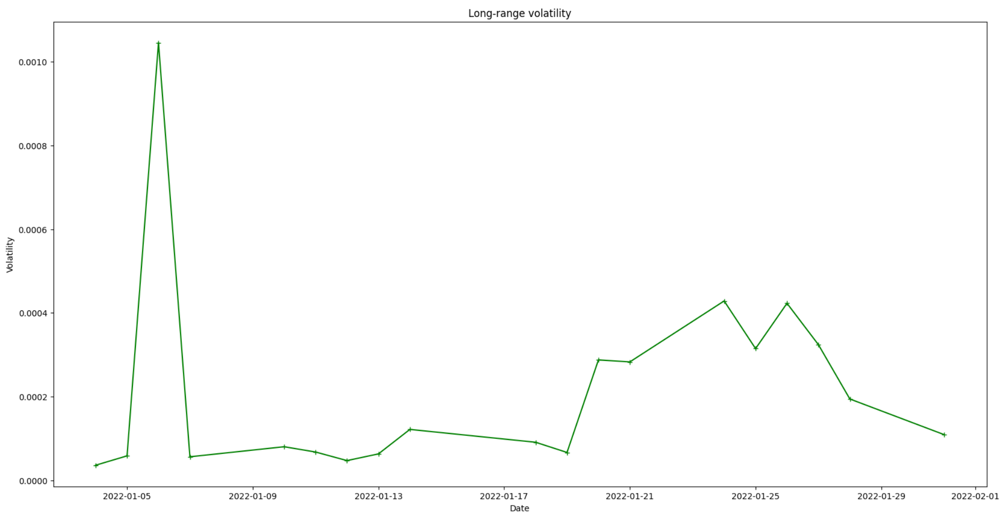

Figure 3: Long-range volatility

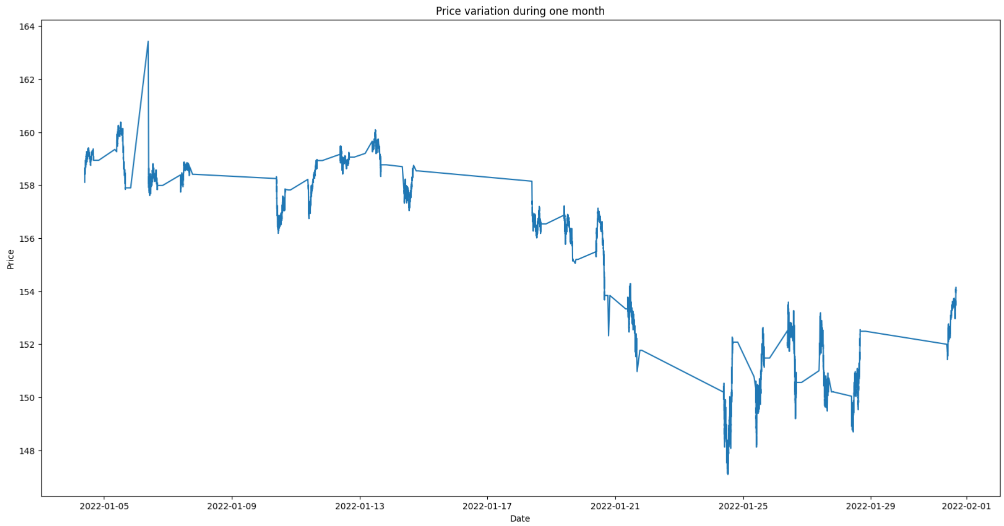

Figure 4: Prices over one month

Figure 3 shows the long range volatility during the same month. We therefore notice that the dynamics are quite different: the global trend for the intraday one is decreasing whereas this one doesn't have any outlying trend.

Also, for high frequency data, intraday volatility may be interesting but the long range one seems to be quite often near 0, which could be mean that this parameter is not really useful for high frequency data, as these data are related to small time intervals and not months.

Moreover, high frequency data, regarding figure 1 show prices that have thousands of small variation in a very time period. This long range volatility along one month tends to describe that these price don't vary that much, which seems strange regarding the previous mentioned figure.

Plotting the prices during the same month (figure 4) confirms this aspect but also reveals that some areas don't have values, are the markets are closed during night for example or week-ends. When computing and plotting the long-range volatility as a continuous function, we have some blank intervals that don't exist in reality, which could be an explanation of the bad result given by the long-range volatility.

6 Microstructure Noise Analysis

After studying the realized volatility and the long-range one, we are now going to focus on microstructure noise and its size. Our goal is to estimate the size of this noise.

6.1 Introduction to microstructure noise

Microstucture noise can be defined as the result of the ask & bid mechanism, which will interfere with the prices.

This noise will therefore modify the prices that could be predicted by a regular model, like the Brownian motion. Estimating it can help price prediction by taking it in account when computing the models or the estimated daily volatility.

The size of this noise is particularly useful to understand its importance and its impact in a model. Also, it can be easily estimated.

6.2 Methodology for estimating noise size

Let's denote η an estimator of the micros-structure noise size. Then, recall that:

$$\frac{RV}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 2\eta^2 \iff \sqrt{\frac{RV}{2n}} \xrightarrow[n\to\infty]{\mathbb{P}} \eta \tag{2}$$

Using the previous formula, we can now compute, with a large sample size n the estimator η of the micro-structure noise size.

6.3 Computation of the estimator

Using the previous formula and data sampled from last year, we computed the value of the estimator η between 01/01/2022 and today (figure 5).

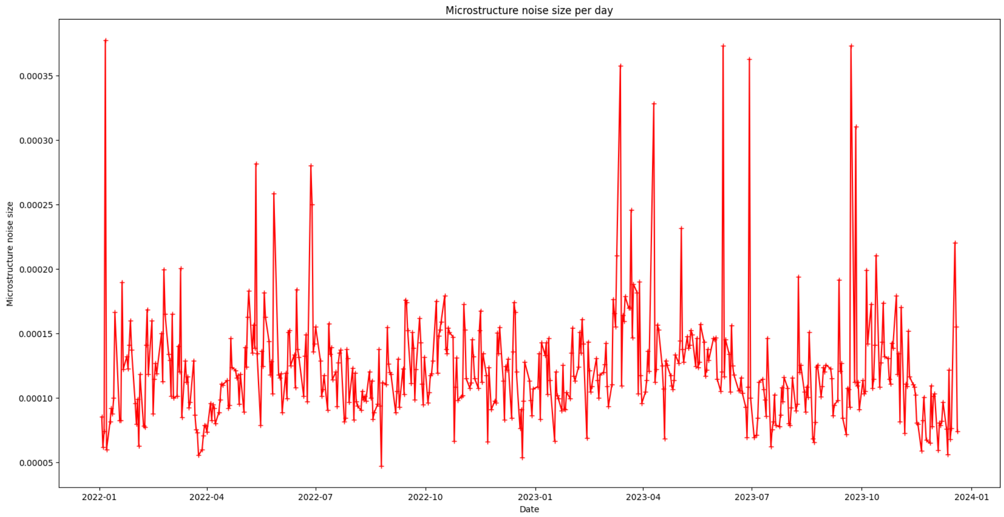

Figure 5: Daily microstructure noise

6.4 Autocorrelation

Another way to achieve the mcirostucture noise is by using the auto-correlation function on the log-prices.

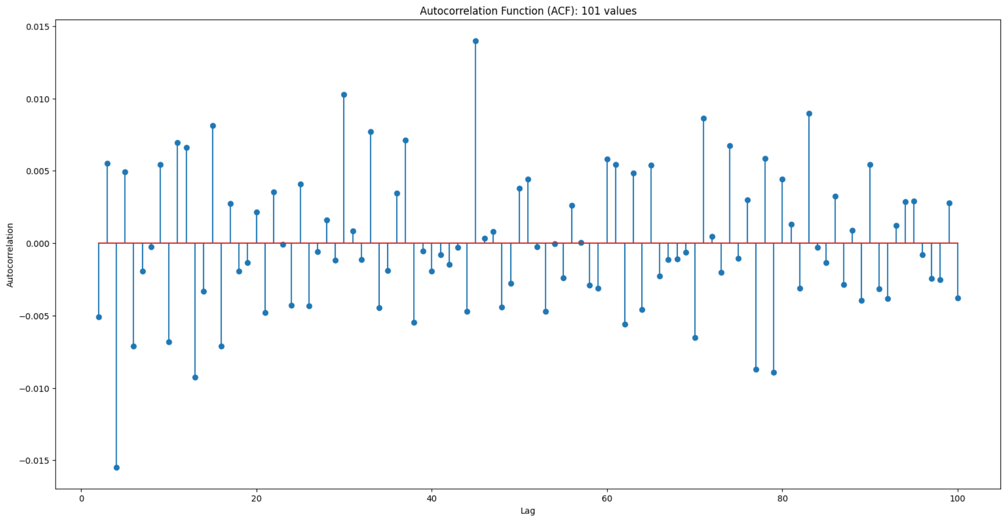

Figure 6: Microstructure noise with autocorrelation

Using this method shows all the impact of the microstructure noise on the prices, and the way it impacts it (positively or negatively).

7 Estimated daily volatility

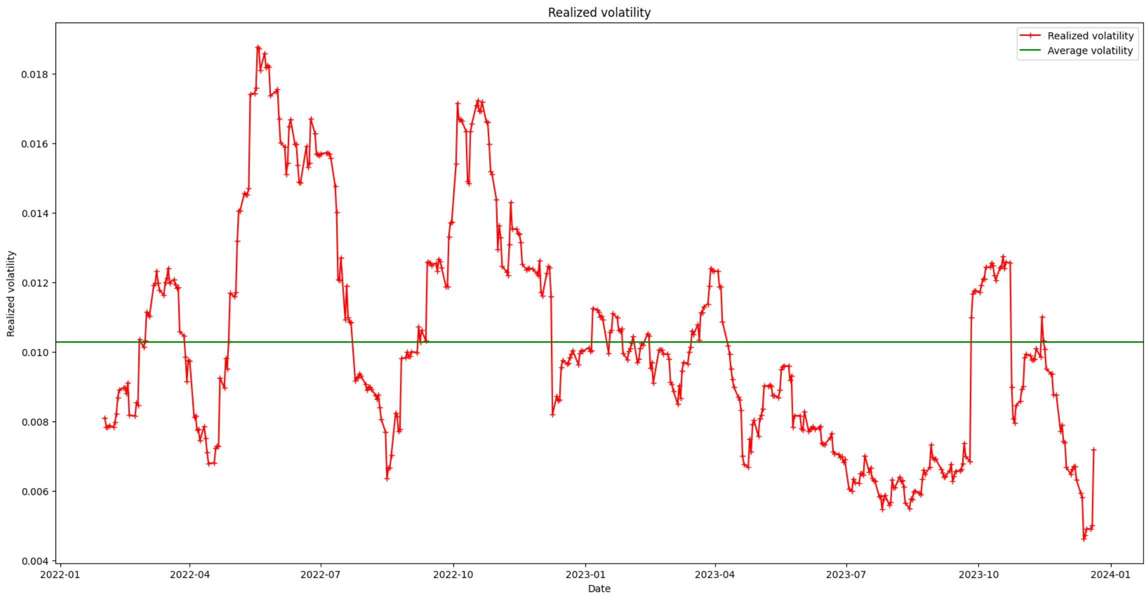

Figure 7: Estimated daily volatility over the last year

Figure 7 shows the estimated daily volatility from last year (red) and the average volatility (green).

8 Conclusion

In conclusion, the analysis of realized volatility, offers notable advantages such as simplicity and the utilization of intra-daily log returns. While the desire is to maximize information by choosing a large time interval (n), challenges arise as the time interval between returns diminishes. The effects of market microstructure, including bid-ask bounce, can introduce bias into volatility estimates

Addressing the second challenge, the substantial overnight return poses a concern for stock returns. Neglecting overnight returns can lead to a significant underestimation of volatility. However, our observations, particularly with foreign exchange returns or index returns like the considered asset (IVE), indicate that overnight returns tend to be minimal.

In response to these challenges, both industry and academic literature have developed numerous solutions over the past two decades. These solutions aim to mitigate the impact of market microstructure noise and address the influence of overnight returns [2]. The ongoing evolution of methodologies in this field underscores the importance of refining volatility estimation techniques for a more accurate understanding of financial market dynamics.

References

[1] Cristina Mabel Scherrer, Gustavo Fruet, and Dias Marcelo Fernandes. Price discovery and market microstructure noise. *SSRN*.

[2] Ruey Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley, 2005.