Noise Reduction in High-Frequency Volatility Measurement

Optiver-07

April 6, 2025

Abstract

This document presents a comprehensive analysis of techniques to mitigate noise in high-frequency volatility measurements. Focusing on the microstructure noise arising from the bid-ask spread in Level 2 Limit Order Book (LOB) data, we compare the Adjusted Realised Volatility (Adjusted-RV) method and the Realised Kernel Estimator. In addition, we discuss the Lee-Ready algorithm's role in trade classification and data cleaning. The study details theoretical foundations, practical considerations, and applicability to our specific data structure, culminating in clear recommendations for further exploration.

1 Introduction

High-frequency financial data is prone to microstructure noise, particularly from the bid-ask spread, which may distort volatility measurements. This paper examines two primary techniques aimed at reducing such noise: the Adjusted Realised Volatility (Adjusted-RV) method and the Realised Kernel Estimator. While the Adjusted-RV method corrects for the bias introduced by the bid-ask spread, the Realised Kernel Estimator applies non-parametric smoothing to high-frequency returns. Furthermore, the Lee-Ready algorithm is reviewed for its potential role in enhancing data quality through accurate trade classification.

2 Background

2.1 Microstructure Noise and Volatility

Microstructure noise is introduced by market frictions such as transaction costs, bid-ask bounce, and other high-frequency effects. When volatility is computed from mid-point prices (derived from bid and ask prices), these noises can lead to biased realised volatility estimates. Reliable noise reduction is crucial for improving risk management and pricing models.

2.2 Level 2 Limit Order Book (LOB) Data

Level 2 LOB data provides detailed information on the top levels of bids and asks. Although it offers a granular view of market activity, the derived mid-point can be significantly affected by the bid-ask spread noise. This presents a challenge in obtaining accurate volatility measures from such data.

3 Methodologies

3.1 Adjusted Realised Volatility (Adjusted-RV)

The Adjusted-RV method aims to directly correct for the bias introduced by the bid-ask spread. Its formula is given by:

$$\text{Adjusted-RV} = RV - c \times (BAS)^2, \tag{1}$$

where:

- RV is the realised volatility computed from high-frequency returns.
- BAS denotes the bid-ask spread.
- c is a constant calibrated from market data.

This approach is theoretically sound as it compensates for the noise inherent in the squared returns. However, its practical effectiveness depends on the accurate calibration of the constant c and the availability of detailed bid-ask information.

3.2 Realised Kernel Estimator

The Realised Kernel Estimator utilizes a kernel function to smooth high-frequency return data, effectively filtering out microstructure noise. Key characteristics include:

- **Noise Filtering:** The kernel function is applied to high-frequency returns to reduce the impact of noise.
- **Robustness:** It yields heteroskedasticity and autocorrelation consistent (HAC) estimates of integrated volatility.
- **Empirical Validation:** Studies (e.g., Barndorff-Nielsen et al., 2008) have shown that the estimator performs significantly better than conventional methods, especially in noisy environments.

This method is highly promising due to its consistency, robustness, and empirical support in reducing noise in high-frequency datasets.

3.3 Lee-Ready Algorithm for Trade Classification

The Lee-Ready algorithm is primarily used for classifying trades by comparing trade prices to the quote mid-point, thereby distinguishing between buyer-initiated and seller-initiated trades. Although it is not a direct noise reduction tool, accurate trade classification can improve data quality. Important observations include:

- **Theoretical Accuracy:** In clean, well-matched datasets, the algorithm classifies trades effectively.
- **Practical Limitations:** Empirical studies have shown misclassification issues in high-frequency, tick-level data, especially under market conditions with high noise.

Thus, while beneficial for data cleaning, caution is advised when employing the Lee-Ready algorithm for noise reduction in high-frequency contexts.

4 Literature Review and Comparative Analysis

A review of academic papers and industry reports provides the following insights:

- **Adjusted-RV:** Research suggests that while the adjusted realised volatility method corrects for noise by adjusting squared returns, practical studies offer limited evidence of its superiority in real-world scenarios. Calibration of the parameter c remains challenging.

- **Realised Kernel Estimator:** Studies such as Barndorff-Nielsen et al. (2008) indicate that the realised kernel estimator delivers consistent, robust, and positive semi-definite estimates even in noisy environments. Empirical tests on US equities have shown it to be substantially more precise than methods based on 5 or 10-minute returns.
- **Lee-Ready Algorithm:** Although the Lee-Ready algorithm performs well in theory, its application to high-frequency data is hindered by systematic misclassifications, which could undermine subsequent volatility measures.

5 Applicability to Two-Level LOB Data

5.1 Adjusted-RV Method

Since Level 2 LOB data provides direct bid and ask prices, the Adjusted-RV method can be effectively applied. The accurate computation of the bid-ask spread supports the adjustment formula; however, the critical challenge remains in accurately calibrating the constant c.

5.2 Realised Kernel Estimator

The mid-point series derived from Level 2 LOB data, if sampled at high frequency, is a suitable input for the realised kernel estimator. This method's flexibility with sampling frequency and its robustness to noise make it an attractive option for such data.

5.3 Enhancing Data Quality

While not a direct noise reduction technique, employing the Lee-Ready algorithm for accurate trade classification can further improve the quality of the dataset prior to volatility estimation. This step is crucial in minimizing misclassification errors that may exacerbate the noise problem.

6 Discussion and Recommendations

Based on the theoretical and empirical review, the following recommendations are proposed:

1. **Adopt the Realised Kernel Estimator:** Its robust performance and empirical validation in reducing microstructure noise make it the most promising approach.

2. **Utilize Adjusted-RV as a Complement:** Despite its challenges in calibration, Adjusted-RV can serve as a useful complementary measure, particularly when detailed bid-ask data is available.

3. **Improve Trade Classification:** Implement an enhanced version of the Lee-Ready algorithm to improve the reliability of trade data, thereby supporting more accurate volatility estimates.

4. **Further Empirical Analysis:** Conduct additional studies using Level 2 LOB data to fine-tune these methods and assess their performance under various market conditions.

7 Conclusion

Accurately measuring volatility in high-frequency financial data requires effective noise reduction techniques. The realised kernel estimator emerges as the leading method due to its robustness and empirical support. The adjusted realised volatility method, while conceptually sound, needs precise calibration for practical use. Augmenting these methods with improved trade classification via the Lee-Ready algorithm can further enhance data quality. Future work should focus on empirical calibration and validation using actual Level 2 LOB data to optimize these approaches for real-world application.