Shira Rozenthal and Alex Olteanu

Professor Maus

CMPS4010

Due September 14th, 2023

**Milestone 1**

For our capstone project, we will approach a former [Kaggle competition](https://www.kaggle.com/competitions/optiver-realized-volatility-prediction/overview) posted by [Optiver](https://optiver.com/), a leading electronic market maker.

Volatility, or the magnitude of a stock price’s fluctuations, is directly related to the price at which options on that stock trades. To further sophisticate its pricing algorithms, Optiver published hundreds of millions of rows of high-frequency market data and challenged the public to build models that more effectively predict short-term volatility.

**Problem Description**

The dataset is broken down into four files:

1. **Order Book**: Effectively representative of interest in the market, this file contains the ticker, ask/bid price, ask/bid size, and time of a received market order.
2. **Trades**: Derived from the file above, this file contains the ticker, price, size, and time of those orders that actually executed.
3. **Training Data**: Organized by the corresponding ticker and time IDs from above, this file contains the target realized volatility computed in the 10-minute window that follows the feature data.
4. **Test Data**: Formatted like the training data, a small portion is available for download to map the above files to unseen data. The rest is released upon submission of our model.

In this capstone, we will explore the relationship between market interest, order execution, and short-term changes in volatility in the provided high-frequency data. The ultimate goal is to predict where a stock’s volatility will realize in the 10 minutes following a market order, and do so as accurately as possible market-wide. We will measure our effectiveness using Root Mean Square Percent Error (RMSPE) in coordination with the competition’s evaluation standards.

**Motivation**

The magnitude of a stock’s price movements is a primary determinant of the likelihood an option on that stock will land in the money. Accordingly, the stock’s volatility is a driving input when calculating the price of its option contracts. Effectively modeling short-term moves in volatility is the most strategic approach to options trading in today’s markets.

As two Computer Science and Finance double majors, we are eager to overlay our technical skillset with our understanding of markets from the business school – particularly the graduate course in Algo and Quant Trading we are both auditing this semester. Further, Shira is particularly excited to model volatility, as she has previously interned on and will continue to work as a trader for an options market maker.

Finally, public high-frequency data is extremely hard to come across, and we are particularly excited by the opportunity to explore modern approaches market research as undergraduates. Regardless of the outcome, the two of us are guaranteed to learn and benefit professionally from the nature of this project and exposure to the dataset itself.

**Previous Work**

*Modeling Volatility Dynamics*

<https://www.newyorkfed.org/research/staff_reports/research_papers/9522.html>

*Multivariate High-Frequency-Based Volatility (HEAVY) Models*

<https://scholar.harvard.edu/files/multiHeavy.pdf>

*What is a Good Volatility Model?*

<https://web-static.stern.nyu.edu/rengle/EnglePattonQF.pdf>

*Bayesian Analysis of Intraday Stochastic Volatility Models of High-Frequency Stock Returns* <https://www.mdpi.com/1911-8074/14/4/145>

*Forecasting Volatility and Tail Risk in Electricity Markets* [https://www.mdpi.](https://www.mdpi.com/1911-8074/14/4/145)[com/1911-8074/14/7/294](https://www.mdpi.com/1911-8074/14/7/294)

*Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts* <https://www.jstor.org/stable/2527343>

*Forecasting Volatility in FInancial Markets: A Review*

[https://](https://www.newyorkfed.org/research/staff_reports/research_papers/9522.html)[/www.aeaweb.org/articles?id=10.1257/002205103765762743](https://www.aeaweb.org/articles?id=10.1257/002205103765762743)

**Challenges**

We expect to encounter the following challenges:

1. **Inexperience with High-Frequency Data**: Though we both have a strong background working on various data science projects, we’ve never worked with this *much* data. We’re expecting to struggle managing millions of data points across various files, especially given that time is documented in a complicated way (each tuple has a timeID, then separately # of seconds in bucket value — essentially time is non-chronological).
2. **Randomness in Markets**: Markets follow a random walk, so we will be training our model on (price) data that by nature is relatively noisy.
3. **Underlying Math of Volatility Calculations**: Many of the leaders in the high-frequency derivatives space have an academic background in physics – which speaks to the systemic complexity of these assets in a dynamic environment. While we would not need to perform notebooks of calculations for this project, our limited math backgrounds may set back our understanding high-level publications in this space.

.

**Feasibility**

Given that this is a Kaggle competition with thousands of submissions many of which were successful, we can safely assume that this is a *feasible* (incredibly difficult, but feasible) project.

We will work with Professor Hamm, and plan to regularly check-in with him on [TBD] and formally on the week of October 2nd with a running naive model, access to the competition test data, and a write up on the top submissions’ strategies.

**Roles**

While we plan to work very closely throughout the entirety of the project, we plan to play to our strengths: Alex interned for an AI startup this summer and Shira for an options market maker. We will rely on one another’s understanding to collectively explore the data and brainstorm approaches to modeling it. Because we’re still in the early stages, we will have a more formal delegation of tasks once we have a clearer picture of our model and the steps we’ll have to take to improve it.