Milestone 1: Data Exploration

Dora Xu(dx33), Michael Cao(yc849), Kexin Deng(kd537), Yichen Gao(yg635)

Factor 1: Order Flow

Order flow refers to the difference between buyer-initiated and seller-initiated trading volumes within a given observation window. It serves as a measure of the imbalance between buying and selling pressure in the market and can help predict short-term price movements. Specifically, when informed market participants leverage their informational advantage to buy or sell at favorable moments, the order flow tends to skew toward that direction. Therefore, order flow can be seen as an indirect indicator of the market's consensus direction. Moreover, aggressive buy orders tend to lift the ask price, pushing the mid-price upward, while aggressive sell orders depress the bid price, causing the mid-price to move downward. As a result, order flow has a structural relationship with future price movements.

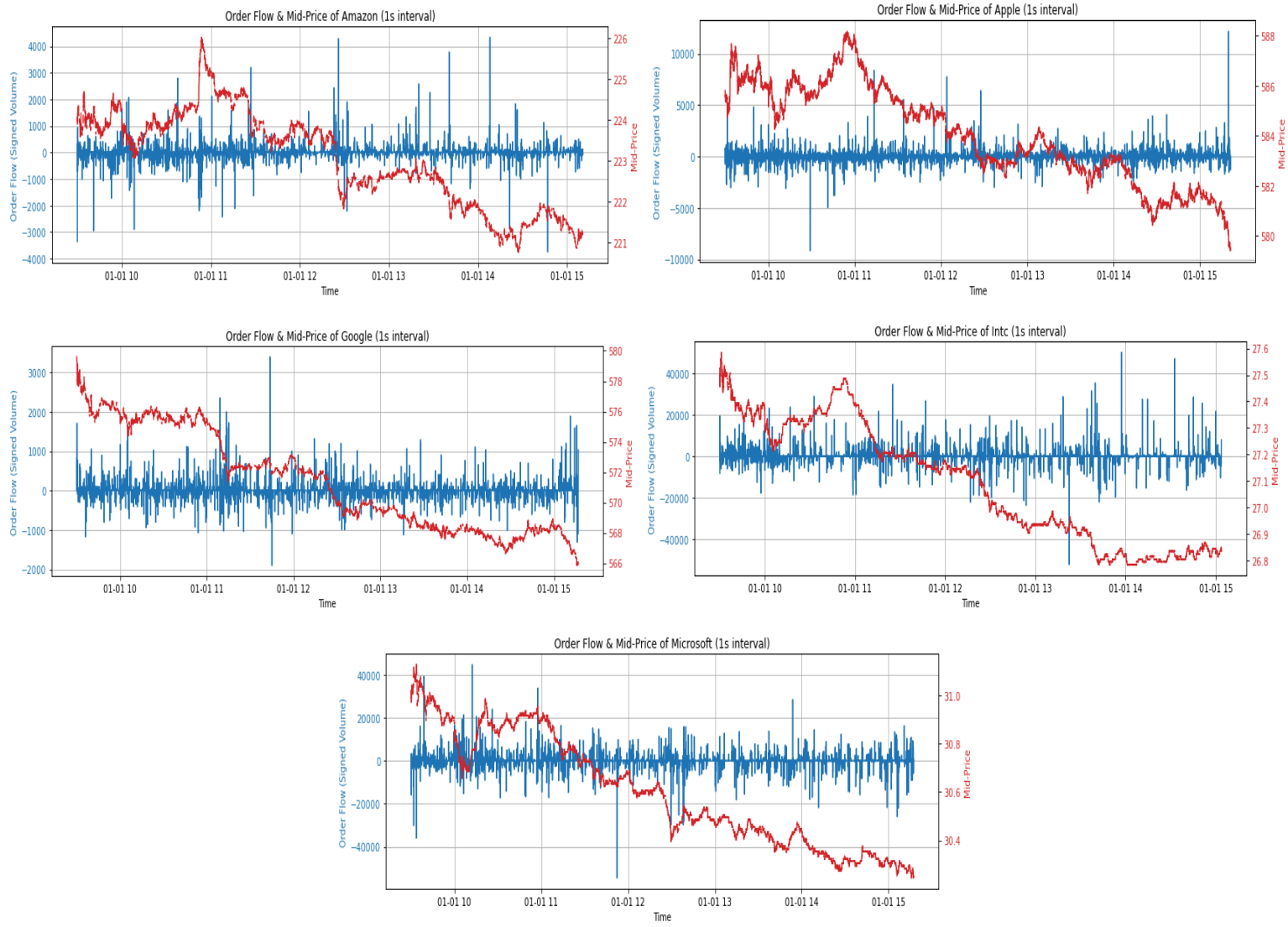

Fig 1 Order Flow

From the relationship between order flow and mid-price across the five stocks, it is observed that despite the overall downward trend in stock prices during the examined period, large positive order flows are often accompanied by short-term price rebounds or a slowdown in the downward trend. Kyle's Model (1985), which suggests that price changes are a linear function of order flow in markets with information asymmetry, supports this observation. Therefore, the order flow factor can inspire the following strategy insights:

- When the order flow exceeds a certain positive threshold, the stock price may rise in the short term, serving as a buy signal.
- When the order flow falls below a certain negative threshold, the stock price may drop significantly, indicating a sell signal.

Factor 2: Short term mid price momentum

The mid-price—defined as the average of the best bid and best ask—is a key reference point in limit order books. To anticipate short-term mid-price dynamics, we construct momentum features over 1-, 3-, and 5-second horizons (momentum_1s, momentum_3s, and momentum_5s), capturing recent directional shifts in mid-price. In high-frequency trading settings, momentum can serve as a predictive signal. These momentum features are computationally efficient and well-suited for integration into other time-series models. From the visualization: AAPL and GOOG exhibit stronger, more frequent momentum patterns. AMZN displays sharper, less consistent spikes. INTC and MSFT show limited momentum activity, indicating that longer horizon or composite features may be more informative than raw momentum alone.

Side Note: Additional momentum-based features could be developed using trend filtering techniques, which aim to extract the underlying signal from noisy mid-price movements. These filters—such as Kalman filters, exponential smoothing, or more advanced L1 trend filtering—can reveal smoother trend components that may enhance the robustness of momentum signals. While promising for trading signal generation, such methods are more algorithmic in nature and require careful tuning and real-time estimation, making them less suitable for inclusion as simple, static features at Milestone 1. Instead, they may be better positioned as part of a later-stage modeling or execution module in a full trading pipeline.

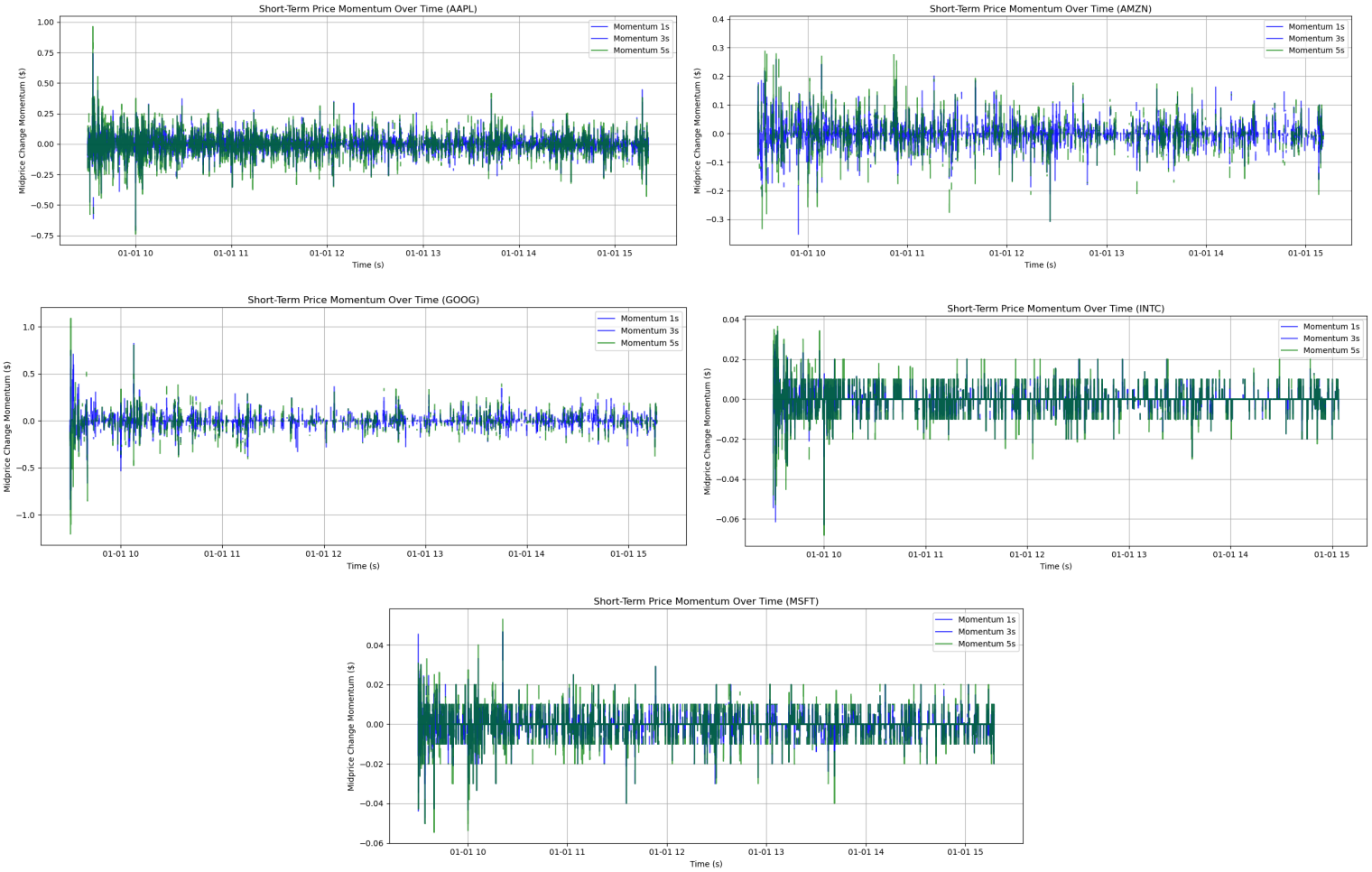

Fig 2 Mid Price Momentum

This analysis draws inspiration from Professor Sasha Stoikov's work on the Micro Price—a refined theoretical construct that approximates the fair market value by considering order book imbalance. Formally, the Micro Price is defined as the limit of the mid-price as observation frequency approaches infinity, effectively filtering out temporal noise and enhancing signal quality. Stoikov's framework demonstrates that Micro Price consistently outperforms traditional metrics such as the simple Mid Price and the Weighted Mid Price (WMid), particularly in environments characterized by bid-ask spread shifts and order flow asymmetry.

Motivated by these insights, we interpret persistent deviations in short-term mid-price momentum as actionable signals. We can have naive signals like below: define actionable buy/sell thresholds based on momentum magnitude, later adjusted by spread or volatility.

- Buy when momentum exceeds a positive threshold, indicating sustained upward imbalance;
- Sell when momentum falls below a negative threshold, reflecting pressure on the downside

Factor 3: Cancellation Rate

The cancellation rate refers to the ratio of canceled orders to total submitted orders within a specific time period. It provides insight into the behavior of traders who may retract their orders before they are executed. A high cancellation rate suggests increased market indecision, where traders are frequently adjusting their positions or pulling back from commitments. In our model, the cancellation rate could be a useful factor because it helps capture underlying market sentiment and liquidity conditions. Significant cancellations, particularly on the bid or ask side, can indicate potential price volatility or imbalances in supply and demand, which are important signals for predicting future market movements. By including this factor, we can enhance the model's ability to anticipate market shifts and improve its overall predictive accuracy.

For example, a high bid-side cancellation rate may signal weakening buy-side interest or growing bearish sentiment, potentially preceding a downward price move. Conversely, a high ask-side cancellation rate could reflect hesitation from sellers and foreshadow upward pressure on prices. By monitoring these patterns, traders can better anticipate short-term market shifts and adjust their strategies accordingly — such as avoiding aggressive buys when bid cancellations surge, or positioning long when ask-side liquidity is being pulled.

Another way we can use this information is by computing the z-score for cancellations occurred in the last time unit. This would assume a normal distribution of the cancellation activities in the past few minutes and serve as a metric for both bids and asks. For example, if a high z-score for cancelled asks and a low z-score for cancelled bids occur simultaneously, it could be a buy signal, and vice versa. If rolling z-scores doesn't work, we may also use rolling percentiles in the same way.

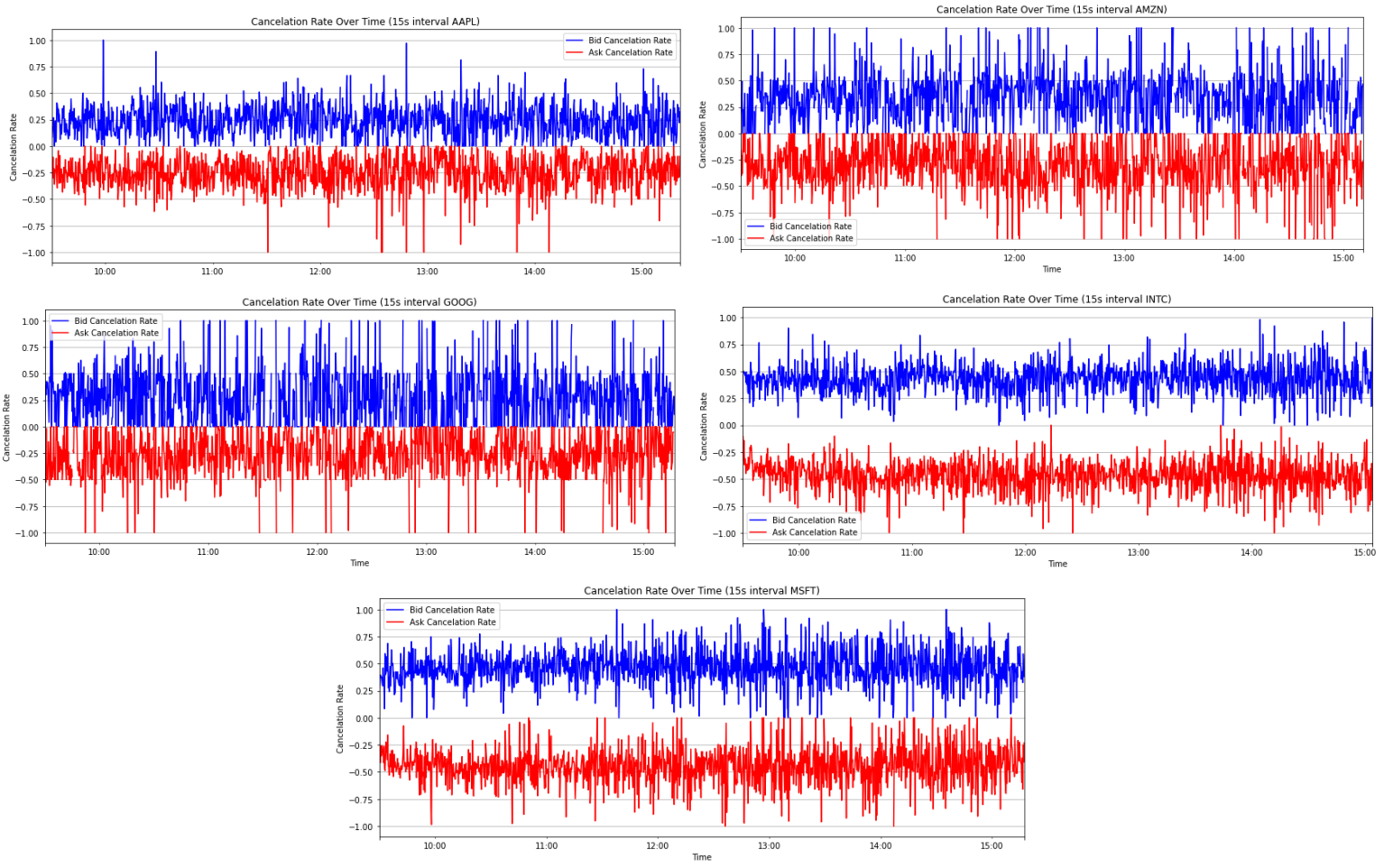

Fig 3 Cancellation Rate

Factor 4: Volatility-to-Spread Ratio

The volatility-to-spread ratio compares short-term price volatility to the bid-ask spread, which effectively measures how dynamic price movements are relative to the cost of executing trades. A high ratio suggests that the market is volatile even though the trade is low, implying a greater opportunity for execution since it doesn't cost much to enter or exit the market. A low ratio suggests that the market might be relatively stable or that the spread is wide, both of which indicate a less efficient and more illiquid market. Including this factor in our model would provide useful insights on when to be more or less aggressive in the market.

Based on the graphs, it can be seen that AAPL, AMZN, and GOOGL show frequent spikes that often go above 1.0 or even reach 2.0-3.0 at times. In these moments when volatility is significantly higher relative to the spread, it signals active trading windows with relatively low transaction costs. On the other hand, MSFT and INTC display much lower overall values and fewer spikes. With the low volatility or wider spreads, these markets are less efficient for quick trades. Thus, we can build the following trade strategy to maximize timing execution:

- If there is a high ratio (>1.0), it would be ideal for market order or small trade blocks. Potentially frontloading a TWAP strategy would also help because one can execute more at these points. If the ratio is paired with a signal from order book imbalance:
    - High ratio + high buy imbalance: we would want to enter long with a market order
    - High ratio + high sell imbalance: we would want to sell quickly and ride the momentum
- If there is a low ratio (<0.3), it would be best to avoid aggressive execution by waiting or submitting limit orders.

Thus, the volatility-to-spread ratio is a powerful signal for identifying windows of high market activity and low transaction cost. When integrated with other microstructure signals like order book imbalance, it can help optimize execution timing, reducing costs and improving price performance.

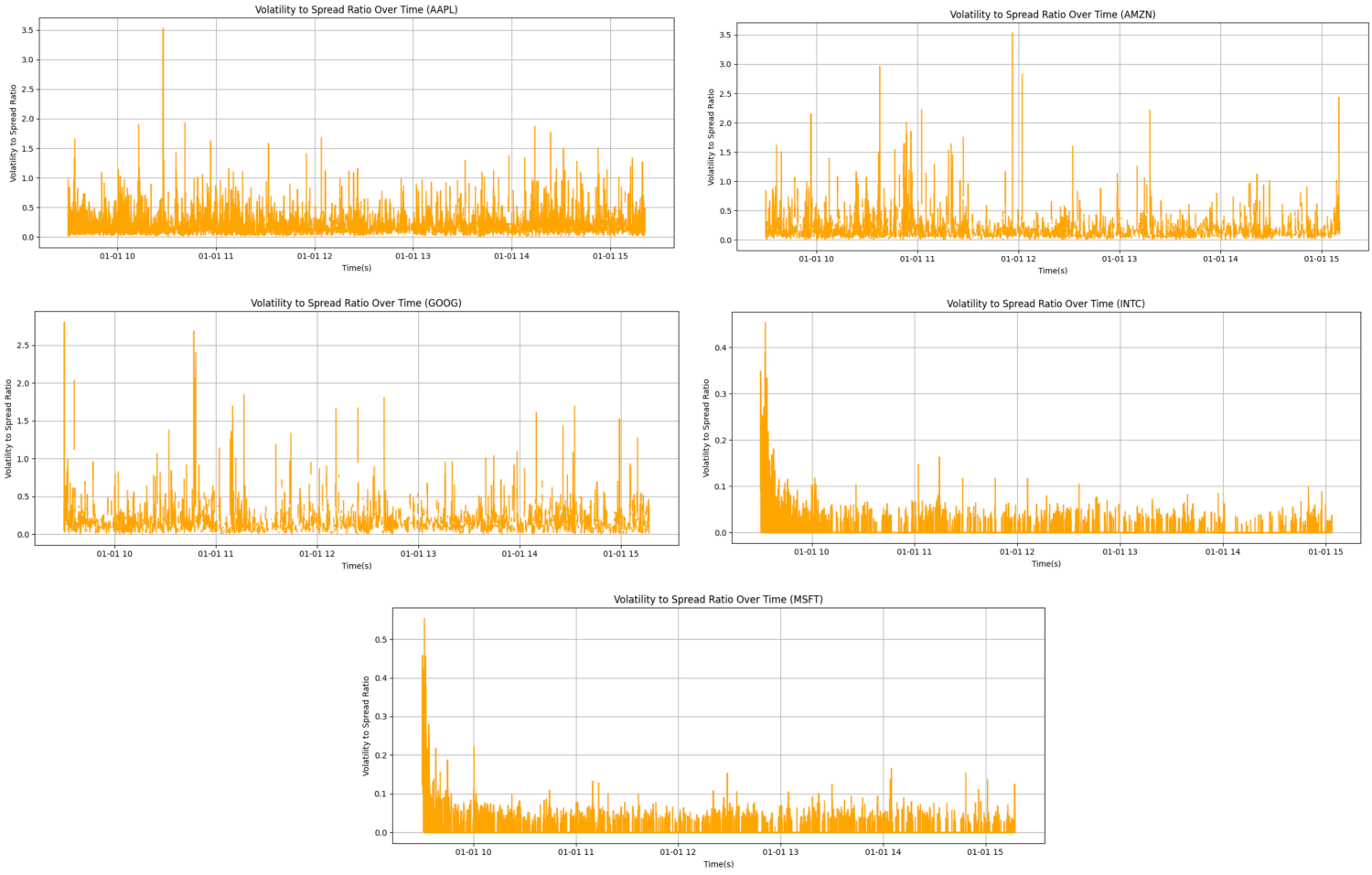

Fig 4 Volatility-to-Spread Ratio