FINS3666 Activity 1

Rohan Phadke

Link to Code and Data Created: <https://github.com/phadke555/Activity-1> (Code provided in the .ipynb extensions; Data as .csv files)

1. Data Processing

I preprocessed both the trades data and the limit order book data according to time only keeping in observations between the period of 10:10 AM – 16:00 PM (AEST). This was to avoid outlier observations from opening and closing auctions as well as pre and post market trading that occurs.

1. Market Microstructure Analysis

To analyze market microstructure, I looked at trade feed metrics, order book metrics and mid-price calculations. Some interesting observations from my analysis:

Trade Feed

* Volume traded was highest around market open time (approx. 10:30 AM) and second highest around market close time (approx. 3:30 PM).
* The traded price of PXA fluctuates about the Cumulative VWAP with 5 major swings throughout the trading day. (Here a swing is defined as a consecutive period of the traded price being above / below the VWAP)
* Out of all completed trades on this day, ~73% of happened on AX and ~27% on CHA.

Order Book

* Bid-Ask spread on average narrows over the course of the day. The highest spread is observed around market open going as far as $0.05 (5 cents). The lowest spread is $0.01 (1 cent). When the spread is 1 cent wide, hidden orders (between the bid and ask) seem to be executed at fractions of cents. This might be strategy of concealing large orders.
* Consecutive trades occurring on the bid result in the bids and asks lowering and the opposite occurs when there are consecutive trades occurring on the ask.
* Order depth imbalance is greatest at the start and end of the trading day. Imbalance lies closer to 0 during the middle of the day.

Mid-Price Calculations

* Three different methods of mid-price calculation were used for all levels of order book data (L1 – L10): Simple, Volume Weighted, Spread Crossing Volume Weighted. L1 or top of the order book seems to track the best with the traded price. This intuitively makes sense.

1. Statistical Analysis

I use the metrics Mean Absolute Error, Mean Squared Error, Tracking Error, Correlation Coefficient, and R-Squared (from linear regression) to examine the relationships between traded price and mid-price.

Traded Price and L1 Simple Mid Price have the following metrics:

* MAE = 0.004812, MSE = 0.000036, TE = 0.005949, r = 0.987

Traded Price and L1 VW Mid Price have the following metrics:

* MAE = 0.006855, MSE = 0.000068, TE = 0.00825, r = 0.975

Traded Price and L1 Spread Crossing VW Mid Price have the following metrics:

* MAE = 0.005119, MSE = 0.000043, TE = 0.006371, r = 0.9856

The simple mid-price seems to have the best statistical measures overall over the trading day. Lowest MAE, MSE, TE and highest r.

1. Critical Reflection

Strengths:

* The analysis is done at a granular level and if done over multiple days could provide strong evidence for the development of certain trading strategies. I already see potential for a mean reversion (proprietary) trading strategy using the Market VWAP as a measure of mean. When the simple mid-price of the bids and asks goes below the Market VWAP, the asset can be bought and when it goes above the asset can be sold.

Opportunity

* By taking 2 different exchanges into account there is potential for identifying cross-exchange arbitrage opportunities if there are market inefficiencies

Limitations

* In the US, dark pool execution is a pretty large market. This does not take into account trades occurring there.
* We only use data for a single stock here. This stock may be affected by other macro-economic factors or display co-movement with other stocks.