# DSMP Ideas from Readings:

<https://people.maths.ox.ac.uk/porterm/papers/gould-qf-final.pdf> Gould et al. Limit Order Books:

* Look at b/a prices relative to the b(t) and a(t).
* Also, look at b/a prices relative to the spot price (weighted mid-price).
* Look at price changes as a function of the aggregate order flow on the bid and ask sides – both orders and cancellations frequency have an impact on the price in near-terms.
  + Using data for 50 stocks traded on the NYSE, they performed (separately for each stock) an ordinary least-squares regression of the mean change in mid price over a time window of length 10 s onto the order flow imbalance over the same time window. For 43 of the stocks studied, the slope of the regression line was significantly different from 0 at the 95% level and was larger on average for those stocks with smaller mean values of nb(b(t), t) and na(a(t), t).
  + Regressions that used the trade imbalance size did significantly worse.
* <https://www.youtube.com/watch?v=l08LICz8Ink> – useful to know about volatility clustering and Hurst exponents to look at memory of markets. Then read 4.7.3.

<https://www.mdpi.com/2227-7390/10/8/1234>:

TRADES, QUOTES AND PRICES Textbook:

* Chapter 6.2: The Statistics of Price Changes: An Informal Primer
  + Need to plot a signature plot for the data to see how volatility changes over time. Increasing is trending, decreasing is mean reverting. Flat means statistically efficient.
* Need to look at the autocorrelation of volatility over a week with 30-min periods and over 6 months with 1-day lags.
* For both, see graphs in the chapter. This will help to see how statistically efficient the market for this stock is.

<https://arxiv.org/pdf/1907.06230.pdf>

Multi-Level Order-Flow Imbalance in a Limit Order Book: Ke Xu∗1, Martin D. Gould1, and Sam D. Howison1

* Text, letter

  Description automatically generated
* Text

  Description automatically generated
* Text

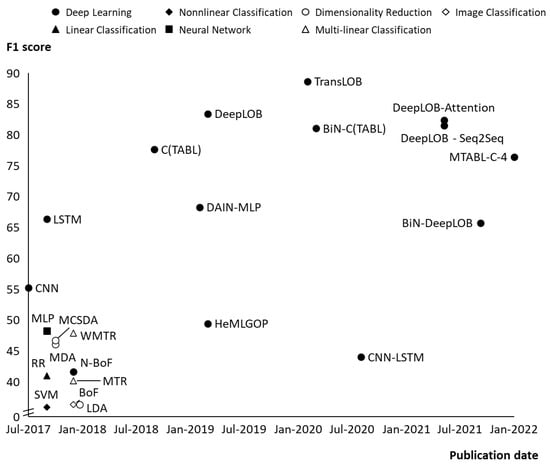
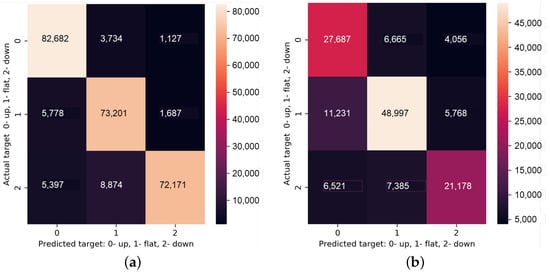
  Description automatically generated
* Text, letter

  Description automatically generated
* Pg 6 all about OFI

Ideas post-reading: Predicting Stock Price Changes Based on the Limit Order Book: A Survey (2022)

<https://www.mdpi.com/2227-7390/10/8/1234>

* Need to look at the minimum length of time between opportunities for a return to be made. Aka the spread to be jumped. Aka when the interval after which the bid price rises above the ask price. Or the ask price drops below the bid price. These are the opportunities for a return to be made.
* Then from there, can determine the optimal prediction window and look at varying the time intervals for predictors before that. Also, use multiple days to train?
* Need to have a look at the tape data to see whether transactions are actually taking place at the extremes of liquidity on the ask side. If so, can assume that limit orders would be fulfilled and the mid-price can actually be a good predictor. If not, maybe unsupervised learning could be a better option - could then use RNN or deep learning.
* Use below image in report as survey info on best models by performance in the past:
  + Table

    Description automatically generated
  + 
* Other ideas:
  + Filter out timestamps where the spread is above a certain limit and include in trading strategy this same logic maybe.
  + Extend order flow imbalance period to greater than 10s and play with that.

RNN:

<https://cs230.stanford.edu/projects_spring_2018/reports/8289864.pdf>

Have overlapping windows with four timestamp windows that roll over the dataset to avoid artificial grouping problems.