STAT430: Machine Learning for Financial Data

LarryHua.com/teaching

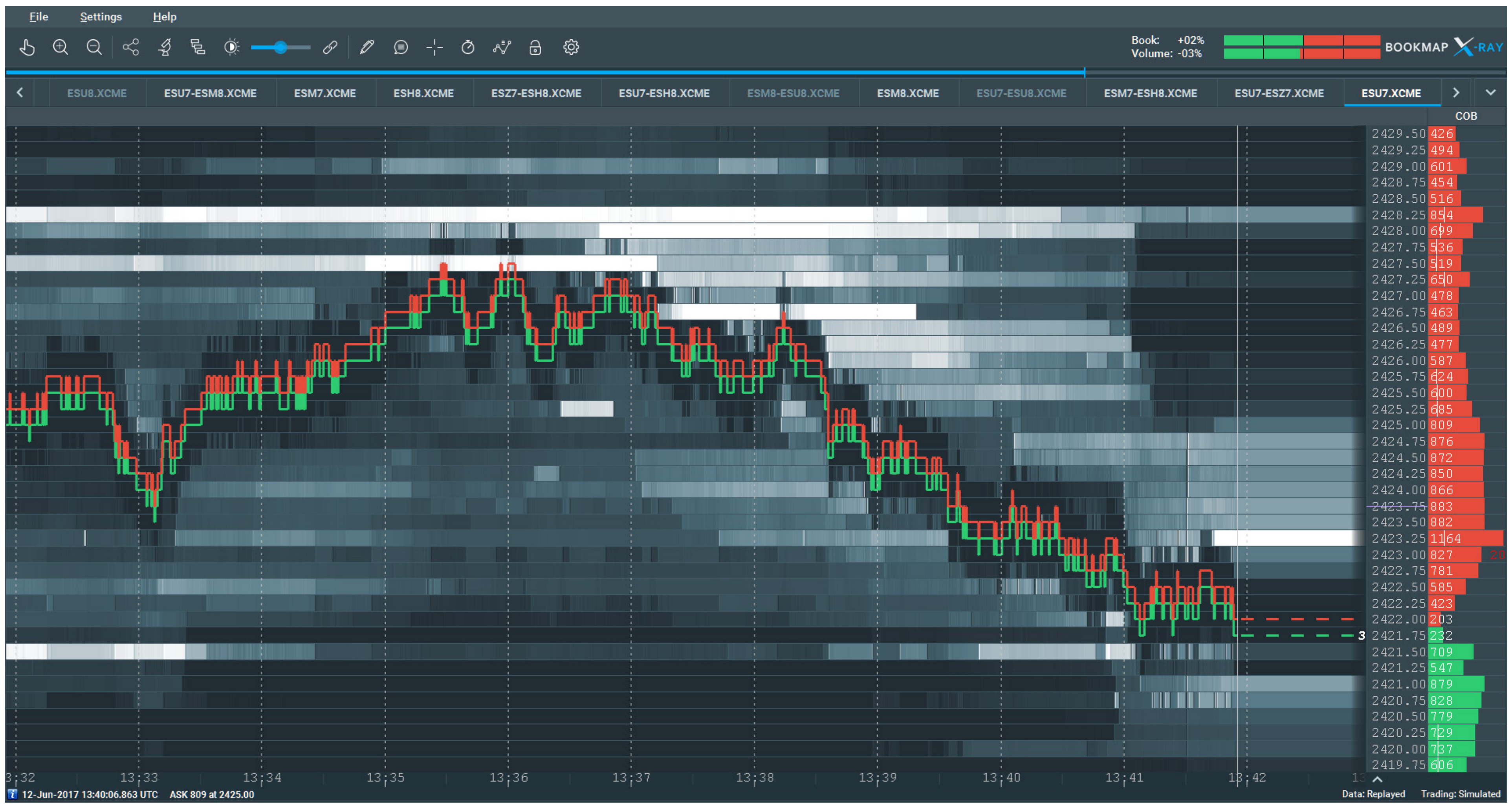
Spring 2019

Microstructural Features

Motivation

- Microstructural data contains primary information about auctioning process, such as limit order book, order cancellation
- It provides footprints for how market participants conceal and reveal their intentions
- Microstructural data is one of the most important ingredients for building predictive ML features

Motivation

- A video of order flows
  - 
  - (source: https://bookmap.com/bm-nanotick/)

1st Generation: price sequences

- Estimating the bid-ask spread and volatility of prices as proxies for illiquidity
  - Liquidity describes the degree to which an asset or security can be quickly exchanged without affecting the price
- The tick rule
  - $b_t \in \{1, -1, b_{t-1}\}$ depending on the price changes Δp_t
  - Informative features can be constructed based on those b_t's

Examples of features based on b_t

- Structural breaks based on Kalman filters on $E_t[b_{t+1}]$
- Entropy of b_t sequence
  - Lower entropy, more predictable
- t-values from Wald-Wolfowitz's tests of runs on b_t
  - a test for the randomness of b_t sequence
  - under null, the number of runs, given the numbers of 1 and -1, follows a normal distribution
- Fractional differentiation of c_t series, $c_t = \sum_{i=1}^{t} b_i$

2nd Generation: strategic trade models

- Focus on understanding and measuring **illiquidity**
  - Illiquidity is a risk that has an associated premium
  - Explain trading as the strategic interaction between informed and uninformed traders
  - Prefer features based on t-values over features based on mean values

Kyle's Lambda - illiquidity measure

- Kyle 1985, an Econometrica paper
- A risky asset with terminal value $v \sim N(p_0, \Sigma_0)$
- A noise trader trades a quantity $u \sim N(0, \sigma_u^2)$ with $u \perp v$
- An informed trader knowing v demands a quantity x, through a market order
- The informed trader believes that the market maker adjusts price based on $p = \lambda(x + u) + \mu$, where μ is the current price, and λ is an inverse measure of liquidity thus a measure of market impact
- The informed trader's profit is $(v - p)x$, which is maximized at $x = (v - \mu)/(2\lambda)$, with $\lambda > 0$ (solve a quadratic function)

Kyle's Lambda - illiquidity measure

- The market maker believes that the informed trader's demand is $x = \alpha + \beta v$, therefore the informed trader's profit is maximized when $\alpha = -\mu/(2\lambda)$ and $\beta = 1/(2\lambda)$
- Lower liquidity $\Rightarrow$ higher λ $\Rightarrow$ lower demand $|x|$
- In order to maximize profit and market efficiency: $\lambda = (1/2)\sqrt{\Sigma_0/\sigma_u^2}$
    - Illiquidity increases with uncertainty about v and decreases with the amount of noise
    - Estimate λ by a simple regression: $\Delta p_t = \lambda(b_t V_t) + \epsilon_t$, where $b_t V_t$ is the net order flow between $t-1$ and t

Kyle's Lambda - illiquidity measure

- Expected profit of the informed trader is $\frac{(v-p_0)^2}{2}\sqrt{\sigma_u^2/\Sigma_0}$
- Three sources of profit:
  - The security's mispricing: $(v-p_0)^2$
  - The variance of the noise trader's net order flow σ_u^2
  - The reciprocal of the terminal security’s variance Σ_0

Other versions of illiquidity measures

- Amihud's Lambda
  - Positive relationship between absolute returns and illiquidity
  - $|\Delta \log p_\tau| = \lambda \sum_{t \in B_\tau} (p_t V_t) + \epsilon_\tau$ $|\Delta \log p_\tau| = \lambda \sum_{t \in B_\tau} (p_t V_t) + \epsilon_\tau$
- Hasbrouck's Lambda
  - Similar idea for multiple securities

3rd Generation: sequential trade models

- Focusing on arrival rates of noise traders and informed traders
- Probability of Information-based Trading
  - Let $S_0 S_0$ be present price, $\alpha_t \alpha_t$ be the probability of new information, $S_B S_B$ be the price under bad news, $S_G S_G$ be the price under good news, and $\delta_t \delta_t$ be the probability of bad news given there is news
  - $E(S_t) = (1 - \alpha_t)S_0 + \alpha_t(\delta_t S_B + (1 - \delta_t)S_G)$
    $E(S_t) = (1 - \alpha_t)S_0 + \alpha_t(\delta_t S_B + (1 - \delta_t)S_G)$
  - Based on Poisson distribution, informed traders arrive at a rate $\mu \mu$, and uninformed traders arrive at a rate $\epsilon \epsilon$
  - Breakeven bid-ask spread: ($B_t B_t$ for bid, $A_t A_t$ for ask)

$$
\begin{aligned}
&E(A_t - B_t) \\
&= \frac{\mu\alpha_t(1-\delta_t)}{\epsilon + \mu\alpha_t(1-\delta_t)}(S_G - E[S_t]) + \frac{\mu\alpha_t\delta_t}{\epsilon + \mu\alpha_t\delta_t}(E[S_t] - S_B)
\end{aligned}
$$

$$
\begin{aligned}
&E(A_t - B_t) \\
&= \frac{\mu\alpha_t(1-\delta_t)}{\epsilon + \mu\alpha_t(1-\delta_t)}(S_G - E[S_t]) + \frac{\mu\alpha_t\delta_t}{\epsilon + \mu\alpha_t\delta_t}(E[S_t] - S_B)
\end{aligned}
$$

Additional microstructural features

Distibution of Order Sizes

- Frequency rates of trades per trade size decay in trade size
- Abnormal frequency at round trade sizes: 5, 10, 15, 20, …
- Proportions of round-sized trades differentiate human traders from "silicon traders"

Additional microstructural features

Cancellation Rates, Limit Orders, Market Orders

- Predatory algorithms utilize quote cancellations and various order types to adversely select market makers
  - Quote stuffers: quickly entering and then withdrawing large orders to slow down competing algorithms
  - Quote danglers: sends quotes that force a squeezed trader to chase a price against her interests
  - Liquidity squeezers: trade in the same direction of distressed traders to drain as much liquidity as possible
  - Pack hunters: a group of predators pretend to trade independently

Additional microstructural features

Time-Weighted Average Price Execution Algorithms

- A TWAP algorithm slices a large order into small ones, submitted at regular time intervals, to achieve a pre-defined time-weighted average price
- The largest concentrations of volume within a minute tend to occur during the first few seconds, for almost every hour of the day
- Especially at the open of Asian / UK / European / US markets, and at the close of US market
- A useful ML feature may be to evaluate the order imbalance at the beginning of every minute

Additional microstructural features

Some other features

- Options Markets
  - There are disagreements between bid-ask range implied by the put-call parity quotes and the actual bid-ask range of the stock
  - Option quotes do not contain as much economically significant information as stock quotes
  - Option quotes can remain irrational for prolonged periods
- Serial correlation of signed order flow
- More research on microstructural features
  - https://papers.ssrn.com
  - https://arxiv.org/archive/q-fin
- Back to Course Scheduler