# INDENG 290: Applications of machine learning to electronic markets

Instructor: Dr. Svitlana Vyetrenko

# About the instructor

- 2011 - Caltech, PhD in Applied and Computational Mathematics

- 2011 - 2014 - Goldman Sachs, Equity Strategies, Associate

- 2014 - 2019 - JP Morgan, Macro Linear Quantitative Research, Vice President

- 2019 - Present, JP Morgan, AI Research, Executive Director

# Course Structure

- Outline
    - Market basics, stylized facts
    - Market impact
    - Reinforcement learning
        - Market making and algorithmic execution
    - Simulation
    - Synthetic data
        - Generating synthetic data (GANs)
        - Time series augmentation
- Syllabus on bcourses

# Research Project Examples

# What do you expect to learn from this class? Goals?

# Markets Basics

# Equity Securities

- **Common stocks** (AKA **equity securities** or **equities**) represent ownership share in corporation
    - Each share entitles owner to one vote at the annual meeting
    - Corporation controlled by board of directors elected by shareholder
    - Board selects managers who run the firm
- Common stock characteristics:
    - **Residual claim**
        - Stockholders are last in line to have claim on assets of the corporation in case of liquidation
        - If no liquidation, shareholders have claim to the operating income left after interest, taxes and other obligations (salaries, etc) are paid - managers can choose to pay residual income back to shareholders in form of dividends or reinvest residual income to increase the value of the shares
    - **Limit liability**
        - In the event of failure, the most shareholders can lose is the original investment

# Indices

- **DJIA** (Dow Jones Industrial Average)
    - Include 30 large “blue-chip” corporations
    - **Price-weighted average** of the 30 stocks
        - Index value = sum prices of 30 stocks and divide by 30
- **SPX** (Standard and Poor’s Composite 500)
    - **Market-value-weighted index**
        - Index return = percentage increase in the market value of 500 stocks from one day to another

# Example: Price-Weighted Average

| Stock | Initial Price | Final Price | Initial Market Value | Final Market Value |
|---|---|---|---|---|
| ABC | $25 | $30 | $500 | $600 |
| XYZ | $100 | $90 | $100 | $90 |
| **Total** | | | $600 | $690 |

**Price-Weighted Average:**

- **Portfolio:**
  - **Initial Value = $25 + $100 = $125**
  - **Final Value = $30 + $90 = $120**
  - **Percentage change = -5/125 = -4%**
- **Index:**
  - **Initial Value = (25+100)/2 = 62.5**
  - **Final Value = (30+90)/2 = 60**
  - **Percentage change = -2.5/62.5 = - 4%**

**Higher priced shares get more weight!**

# Example: Market-Value Weighted Average

| Stock | Initial Price | Final Price | Initial Market Value | Final Market Value |
|---|---|---|---|---|
| ABC | $25 | $30 | $500 | $600 |
| XYZ | $100 | $90 | $100 | $90 |
| **Total** | | | $600 | $690 |

**Market-Value Weighted average**

- **Index:**
    - **Initial Value = 100 (arbitrarily chosen)**
    - **Final Value = 100*(690/600) = 115**
    - **Percentage change = 15/115 = 15%**

**Higher market value shares get more weight!**

# Trading Mechanisms

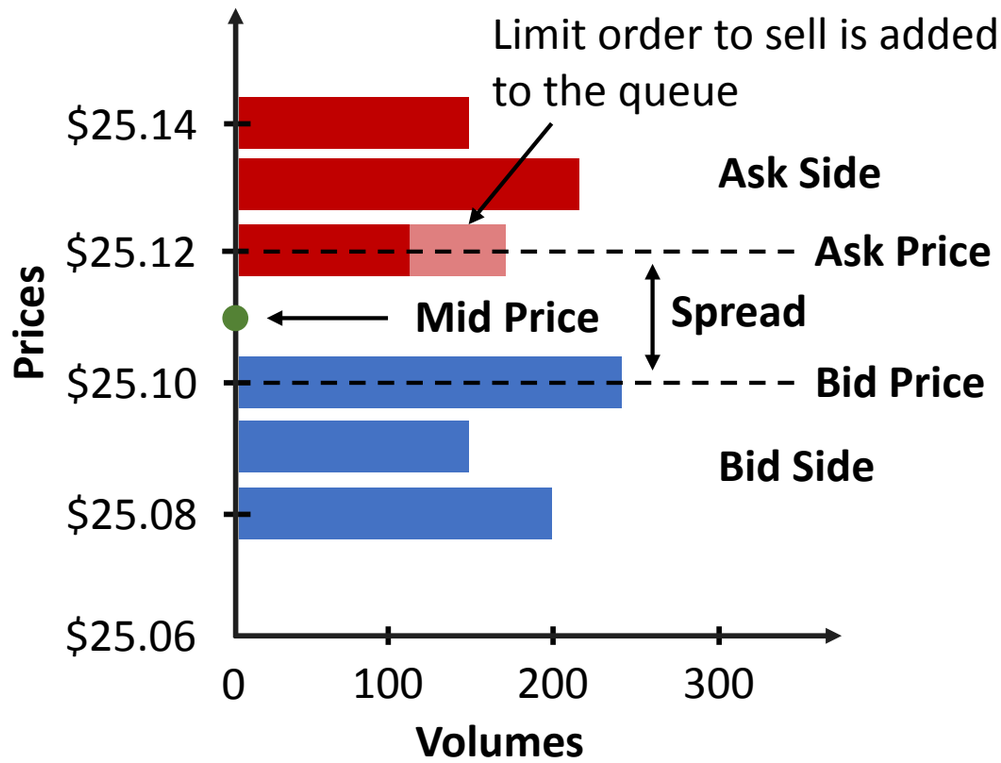
- An individual investor who want to buy or sell a share will place the order with a brokerage firm for a fee
- Broker has three avenues:
    - **Dealer market (Over-The-Counter or OTC)**
        - Dealers show bid and ask quotes (make money on the spread)
        - Brokers review dealer quotes and choose the best one
    - **ECNs (Electronic Communication Networks)**
        - Orders are put in the electronic order book and fulfilled automatically if there is a demand on the opposite side
    - **DMM (Designated Market Maker Markets)**
        - Obliged to provide “fair and orderly” markets in a particular security
        - Support “weak” side of the market
        - Especially useful in periods of high volatility, market stress, market open/close

# US Equities Markets

- **NASDAQ** (National Association of Security Dealers Automated Quotation)
    - Used to be OTC, now primarily electronic execution
    - Lists 3000 firms
- **NYSE** (New York Stock Exchange)
    - Typical daily volume - over 1 billion shares
    - Used to be human specialist, now primarily electronic market
    - NYSE Arca - fully electronic, exchange-traded funds trade there

# Limit Order Book: Definitions

- **<u>Limit Order Books</u>**: Queues of orders that a trading venue uses to record and match the interest of buyers (**bid side**) and sellers (**ask side**)
- Matching engines:
  - FIFO - first in, first out
  - Pro rata



# Limit Order Book: Definitions

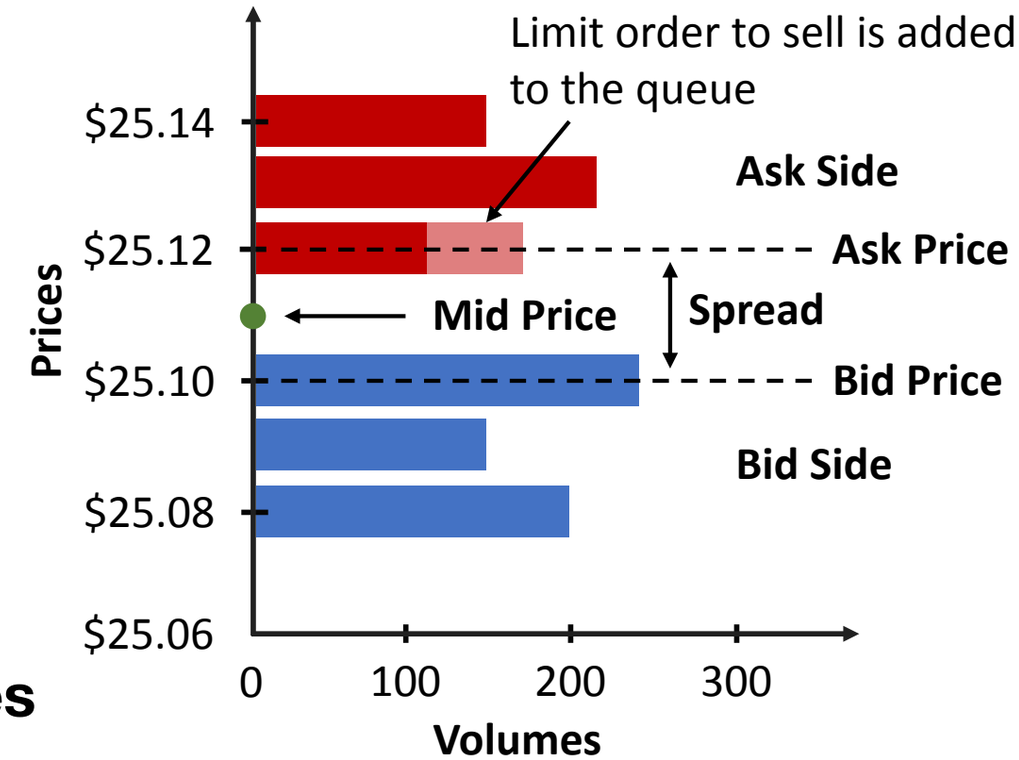
**spread = best ask price − best bid price**

$$\text{mid price} = \frac{\text{best bid price} + \text{best ask price}}{2}$$

**depth = number of levels of limit order book at which liquidity is offered**

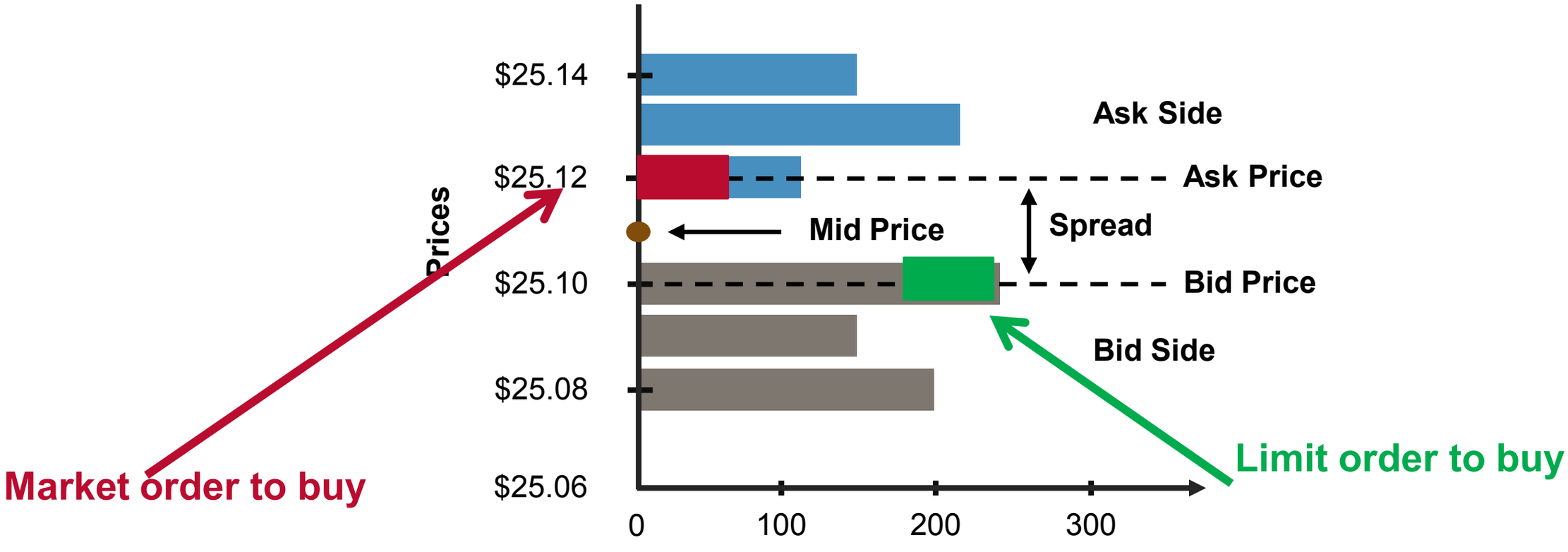
**tick size = minimum possible increment in prices**

**Examples?**



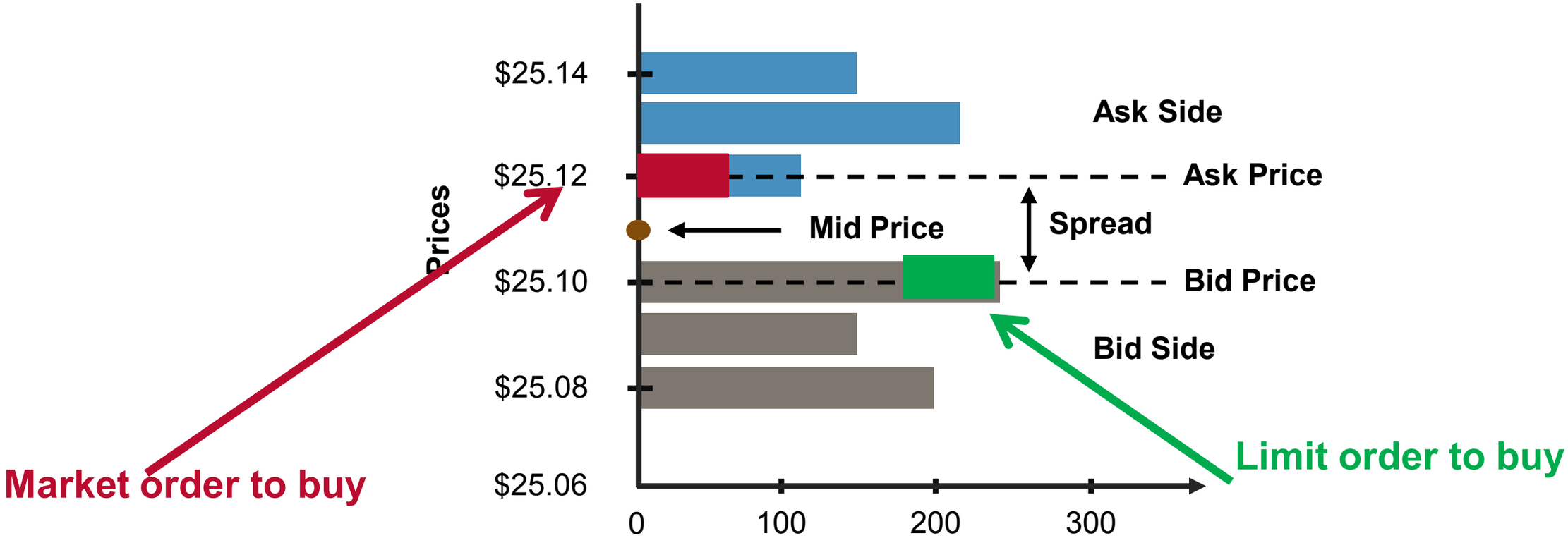
# Limit Orders

- Insertion of a new resting order in the LOB queue on the market participants's side of the market
  - Low cost (do not pay the speed)
  - High cost variance (market might move away)



# Market Orders

- Immediate consumption of available liquidity at the opposite side of the market
  - High cost (do not pay the speed)
  - Low cost variance (almost always order executes immediately)

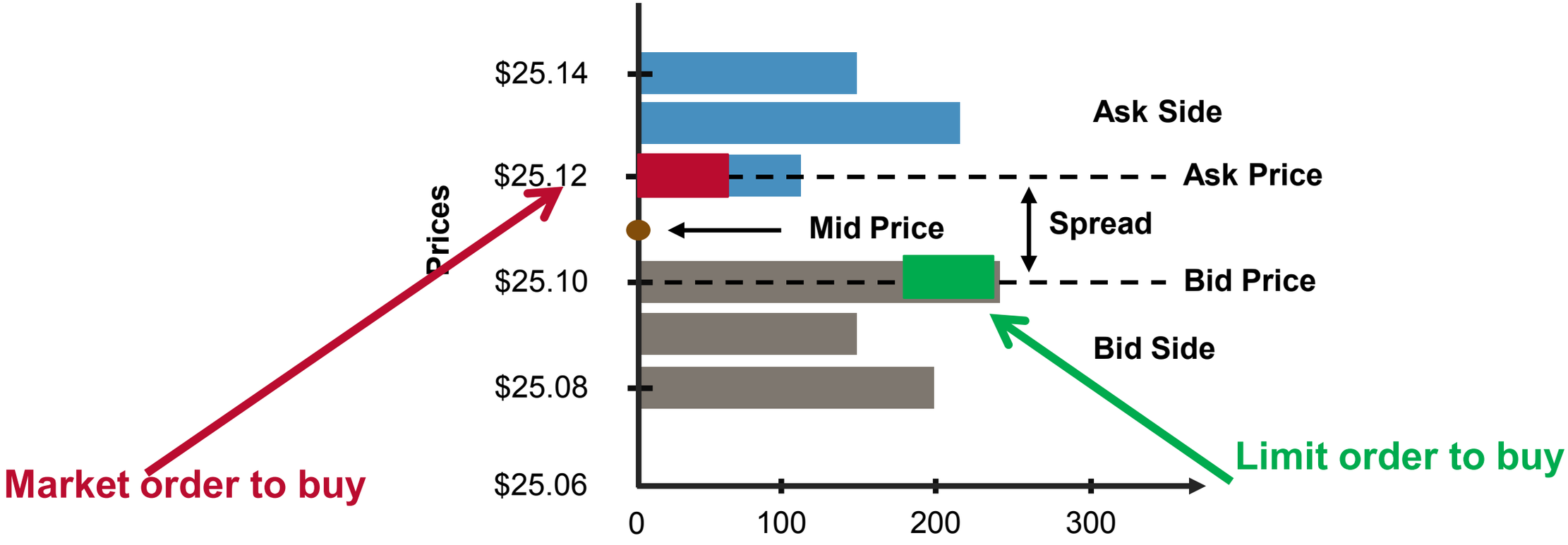


# Other Order Types

- Cancellations
- Modifications
- Replace (cancel and add in the same operation)

# Market Impact

- Market impact is the effect that the market participant has on the market due to its trading activity
  - Market orders have high market impact, “peel market away’
- Market impact grows with the size of the order
  - In practice it’s common to “slice” the order to “space out” execution



# Expected Returns

- *n* scenarios (recession, growth, severe recession, etc)
- *p(s)* - probability of each scenario
- Return in each scenario $r(s) = \dfrac{\textbf{ending price} - \textbf{beginning price}}{\textbf{beginning price}}$
- Expected return $E(r) = \sum_{s=1}^{n} p(s)r(s)$
- Variance $Var(r) = \sigma^2 = \sum_{s=1}^{n} p(s)\big(r(s) - E(r)\big)^2$
  - Variance is a measure of uncertainty, in finance it's a measure of volatility

# Historical Returns

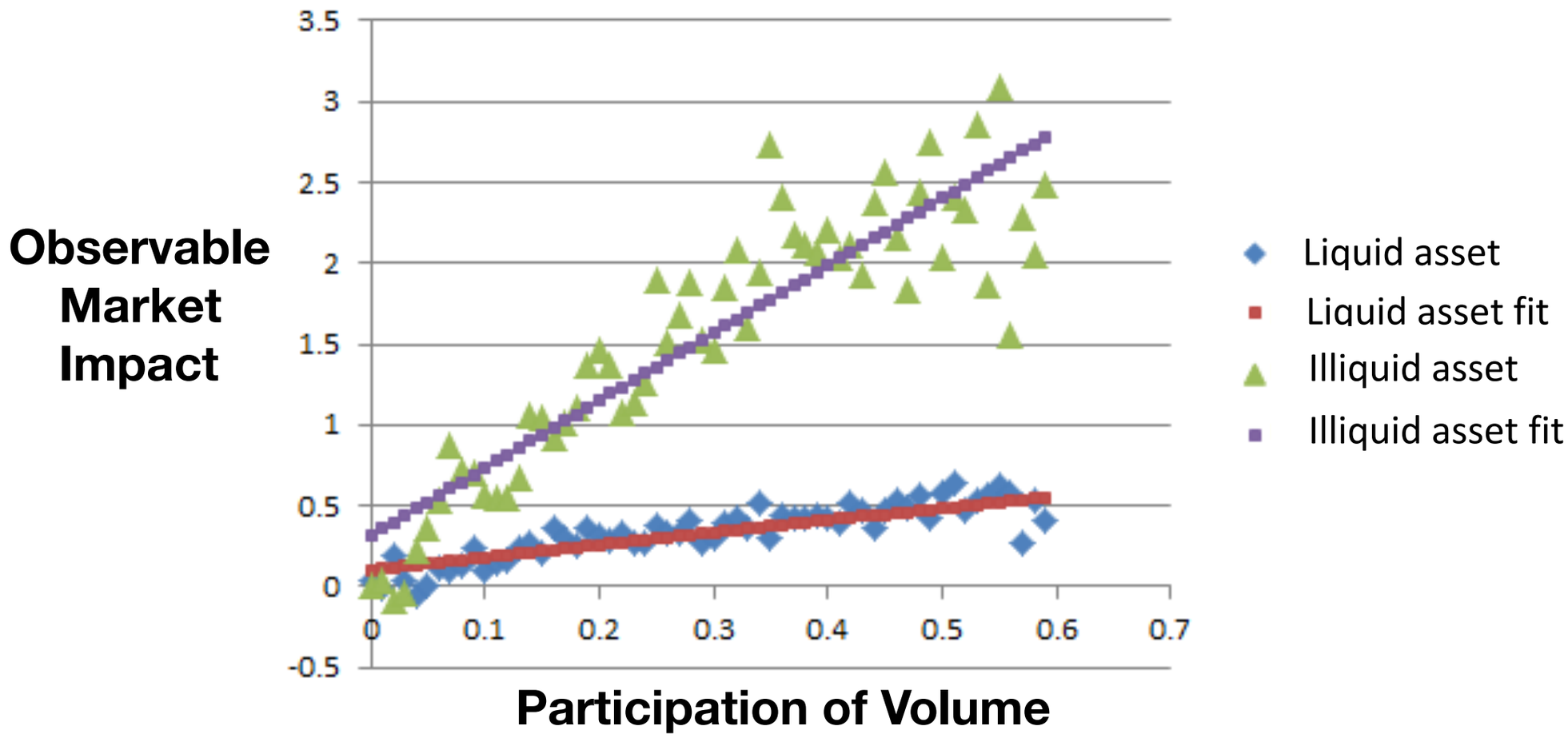
- When historical data comes in the form of time series, each observation is treated as an equally likely scenario

- Expected return can be estimated as

$$E(r) = \sum_{s=1}^{n} p(s)r(s) = \frac{1}{n}\sum_{s=1}^{n} r(s)$$

- Estimate variance as $\hat{\sigma}^2 = \frac{1}{n}\sum_{s=1}^{n}\left(r(s) - \bar{r}\right)^2$

# Empirical Market Impact Estimation



# Tick Data Example

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Time | Bid | BidVol | Ask | AskVol |
| 2 | 2018.03.05 13:49:00.674 | 1.23041 | 50 | 1.23042 | 25.1 |
| 3 | 2018.03.05 13:49:02.716 | 1.23041 | 75.1 | 1.23043 | 75.1 |
| 4 | 2018.03.05 13:49:02.900 | 1.23041 | 100.1 | 1.23043 | 75.1 |
| 5 | 2018.03.05 13:49:03.060 | 1.23042 | 25 | 1.23044 | 80 |
| 6 | 2018.03.05 13:49:03.541 | 1.23042 | 25 | 1.23043 | 25.1 |
| 7 | 2018.03.05 13:49:03.887 | 1.23042 | 25 | 1.23043 | 75.1 |
| 8 | 2018.03.05 13:49:04.318 | 1.23042 | 25 | 1.23043 | 74.9 |
| 9 | 2018.03.05 13:49:04.770 | 1.23042 | 24.9 | 1.23043 | 74.9 |
| 10 | 2018.03.05 13:49:05.300 | 1.23041 | 50 | 1.23043 | 75.1 |
| 11 | 2018.03.05 13:49:05.733 | 1.23042 | 25 | 1.23043 | 50 |
| 12 | 2018.03.05 13:49:06.284 | 1.23042 | 25 | 1.23043 | 75.1 |
| 13 | 2018.03.05 13:49:06.455 | 1.23041 | 50 | 1.23043 | 75.1 |
| 14 | 2018.03.05 13:49:06.973 | 1.23042 | 25 | 1.23043 | 50 |

# Empirical Market Impact Estimation Algorithm

## 0.1 Stylized facts about order market impact

Market impact of order placement is a expected to grow as a function of order volume. For each time interval $\tau$, define $V_{\text{buy},\tau}$ and $V_{\text{ask},\tau}$ to be buy and sell order volumes in $\tau$ respectively. Define participation of volume in $\tau$ as

$$P_\tau = \frac{|V_{\text{buy},\tau} - V_{\text{ask},\tau}|}{V_{\text{buy},\tau} + V_{\text{ask},\tau}}.$$

Note that $0 \leq P_\tau \leq 1$. Also define $\Delta m_\tau$ to be the observable mid-price move in $\tau$. Discretize the range for $P_\tau$ into bins $B_i, i = 1, \ldots, N$ such that $B_i = \{\tau : \frac{i-1}{N} \leq P_\tau \leq \frac{i}{N}\}$. For each $B_i$, define

$$M_i = \frac{1}{|B_i|} \sum_{\tau \in B_i} \Delta m_\tau \quad \text{and} \quad P_i = \frac{1}{|B_i|} \sum_{\tau \in B_i} \Delta P_\tau$$

to be the average price move and average participation of volume in bins with similar volume participation. One can then fit the relationship of the form $M_i \sim \alpha P_i^\beta$ through the data [1, 2, 3].

## References


[1] Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. Direct estimation of equity market impact. *RISK*, 18, 04 2005.

[2] J Farmer, Paolo Patelli, and Ilija Zovko. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102:2254–9, 03 2005.

[3] Jean-Philippe Bouchaud. Price impact. *Encyclopedia of quantitative finance*, 2010.