

Algorithmic Trading: A Machine Learning Approach

D. N. Sotiropoulos

Department of Management Science & Technology
Athens University of Economics and Business

dsotirop@gmail.com

March 29, 2016

Summary

- 1 Fundamental Concepts of Market Microstructure
- 2 Trading Positions
- 3 Trading Problems

Limit Order Book I

- **Limit Order Books** (LOBs) are the fundamental financial tool used by exchanges where all participants are able to trade assets through the use of buy and sell orders.
- Also, they all have access to **top** few **buy** and **bottom** few **sell** orders from other traders on which they can base their trading strategy.
- These orders can take the following forms:
 - **Marker Order**: an order to buy or sell a certain amount of an asset for the **best available price**, resulting in an **immediate** matching.
 - Market orders are considered to **consume liquidity**.
 - **Limit Order**: an order to buy or sell a certain amount of an asset for a **specified price** and will remain active until it is matched to a new market order, or is canceled by the user.
 - Limit orders are considered to **provide liquidity**.

Definition (Sell Order)

An order x with price p_x and size $\omega_x > 0$ is a **sell** order corresponding to the commitment of its owner to sell up to $|\omega_x|$ units of the asset at a price **no less** than p_x . That is, $p_{sell} \geq p_x$.

Definition (Buy Order)

An order x with price p_x and size $\omega_x < 0$ is a **buy** order corresponding to the commitment of its owner to buy up to $|\omega_x|$ units of the asset at a price **no greater** than p_x . That is, $p_{buy} \leq p_x$.

Limit Order Book III

- Whenever a trader submits a **buy** order x , an LOB's trade-matching algorithm checks whether it is possible to match to an **active sell** order y such that $p_y \leq p_x$.
- Since x is a buy order, it seeks to ideally be executed at a price $p_{buy} \leq p_x$.
- Accordingly, y is a sell order which seeks to ideally be executed at a price $p_{sell} \geq p_y$.

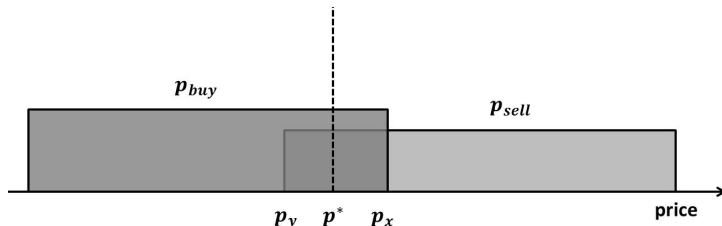


Figure : Buy Order Matching

Limit Order Book IV

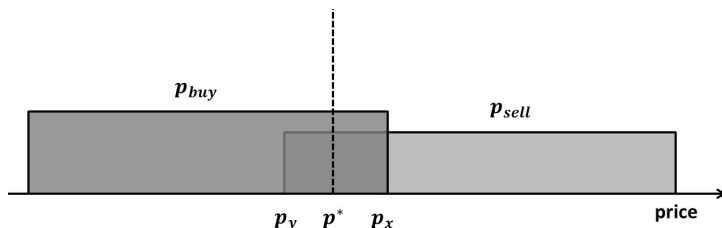


Figure : Buy Order Matching

- If the owners of the relevant orders agree ($p_{buy} = p_{sell} = p^*$) a trade is executed for the specified amount at the specified price. (i.e. $p^* \leq p_x$ and $p^* \geq p_y$, which yields $p_y \leq p^* \leq p_x \Rightarrow p_y \leq p_x$)
- If not, then x becomes active at the price p_x and it remains active until it matches an incoming sell order or is canceled by its owner.

Limit Order Book V

- Whenever, a trader submits a **sell** order x , an LOB's trade-matching algorithm checks whether it is possible to match to an **active buy** order y such that $p_y \geq p_x$.
- Since x is a sell order, it seeks to ideally be executed at a price $p_{sell} \geq p_x$.
- Accordingly, y is a buy order which seeks to ideally be executed at a price $p_{buy} \leq p_y$.

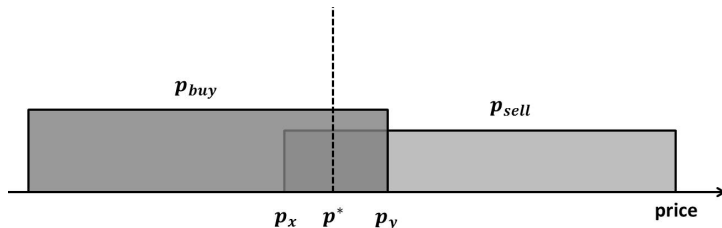


Figure : Sell Order Matching

Limit Order Book VI

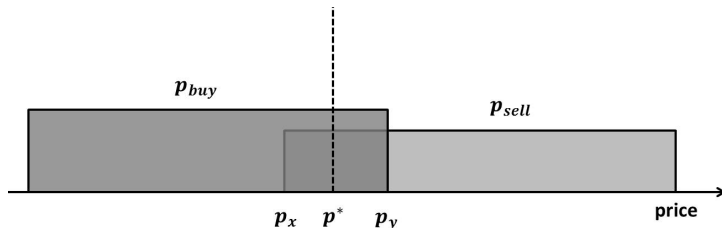


Figure : Sell Order Matching

- If the owners of the relevant orders agree ($p_{buy} = p_{sell} = p^*$) a trade is executed for the specified amount at the specified price. (i.e. $p^* \geq p_x$ and $p^* \leq p_y$, which yields $p_x \leq p^* \leq p_y \Rightarrow p_x \leq p_y$)
- If not, then x becomes active at the price p_x and it remains active until it matches against an incoming buy order or is canceled by its owner.

Limit Order Book VII

- Orders that match upon arrival are called **marker orders**.
- Orders that do not match upon arrival are called **limit orders**, and become *active* in the LOB.

Definition (Limit Order Book)

An LOB $\mathcal{L}(t)$ is the collection of all active orders for a given asset on a given platform at a given time t .

- Limit buy (or **bid**) orders are listed in decreasing order. Therefore, the best bid is the highest bid.
- Limit sell (or **ask**) orders are listed in increasing order. Therefore, the best ask is the lowest ask.

Limit Order Book VIII

■ Given $\mathcal{L}(t)$:

- the **bid-price** b_t is the highest among active buy orders at time t .
- the **ask-price** a_t is the lowest among active sell orders at time t .

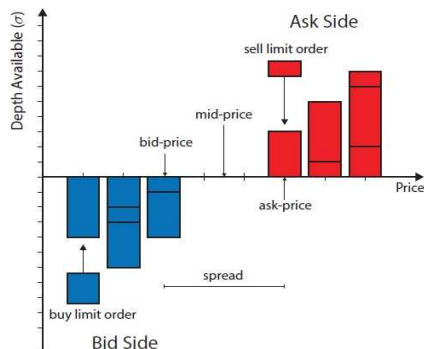


Figure : Schematic of an LOB

Definition (Best Quotes)

The bid and ask prices are collectively known as the **best quotes**.

Definition (Bid-Ask Spread)

The difference $s_t = a_t - b_t$ is called the **bid-ask spread**.

Definition

The mean value $m_t = \frac{a_t + b_t}{2}$ is called the **mid-price**.

- For a given price q and time t we say that q is:
 - on the **bid-side** if $q \leq b_t$.
 - on the **ask-side** if $q \geq a_t$.
 - **inside the spread** if $b_t < q < a_t$.

Definition (Tick Size)

An LOB's tick size $\pi > 0$ is the smallest permissible price interval between different orders. All orders must arrive with a price that is an integer multiple of the tick size.


- Because LOBs implement a tick size $\pi > 0$, it is common for several different active orders to reside at the same price at a given time.
- To help traders evaluate the state of the market, electronic trading platforms typically summarize the information in $\mathcal{L}(t)$ by disseminating a feed that lists the aggregate quantities offered for purchase or sale at a set of price levels.

Limit Order Book XI

Definition (Queueing Priority)

Most exchanges implement a **price-time** priority rule. That is, for **active buy** (respectively, **sell**) orders, priority is given to the orders with the **highest** (respectively, **lowest**) price, and ties are broken by selecting the active order with the earliest submission time.

Order book



Ask ¹		Bid ¹		Ask ²		Bid ²		Ask ³		Bid ³		...
Price	Vol.	Price	Vol.	Price	Vol.	Price	Vol.	Price	Vol.	Price	Vol.	...
585.69	16	585.44	167	585.71	118	585.40	50	585.72	2	585.38	22	...
585.71	118	585.44	167	585.72	2	585.40	50	585.74	18	585.38	22	...
...
585.71	118	585.70	66	585.72	2	585.44	167	585.75	4	585.40	50	...
585.71	118	585.70	66	585.72	2	585.44	167	585.80	100	585.40	50	...
...
585.71	100	585.44	167	585.80	100	585.40	50	585.81	100	585.38	22	...
585.71	100	585.45	18	585.80	100	585.44	167	585.81	100	585.40	50	...
585.68	18	585.45	18	585.71	100	585.44	167	585.80	100	585.40	50	...

Figure : Time Evolution of an LOB

Limit Order Book XII

■ Order Matching and Price Evolution:

- Consider a buy (respectively, sell) order x that arrives immediately after time t .
- If $p_x \leq b_t$ (respectively, $p_x \geq a_t$), then x is a **limit order** that becomes **active** upon arrival and does not cause b_t or a_t to change.
- If $b_t < p_x < a_t$, then x is a **limit order** that becomes **active** upon arrival and causes b_t to increase (respectively, a_t to decrease) to p_x .
- If $p_x \geq a_t$ (respectively, $p_x \leq b_t$), then x is a **market order** that matches to one or more **active sell** (respectively, **buy**) orders upon arrival.

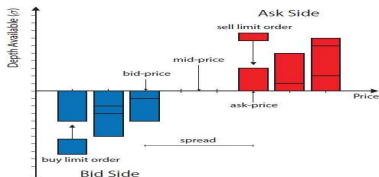


Figure : Schematic of an LOB

Limit Order Book XIII

■ Order Matching and Price Evolution (contd):

- When such a matching occurs, it does so at the price of the active order which is not necessarily equal to p_x .
- Whether or not such a matching causes a_t (respectively, b_t) to change depends upon whether or not $|\omega_x|$ exceeds the total size available at sale at a_t (respectively, for purchase at b_t).
- Price changes also occur if the total size available for sale at a_t (respectively, for purchase at b_t) is canceled.

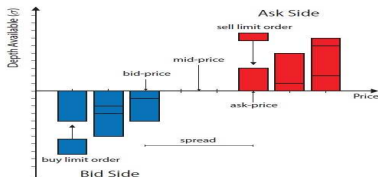


Figure : Schematic of an LOB

Limit Order Book XIV

- Orders constantly **arrive**, get **filled**, get **canceled**.

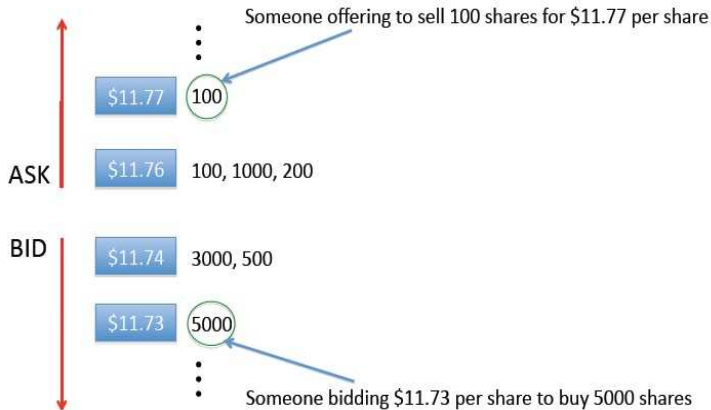


Figure : LOB Alternative Representation

Limit Order Book XIV

- A tick gets printed: volume of 1000 was traded at the price of \$11.74

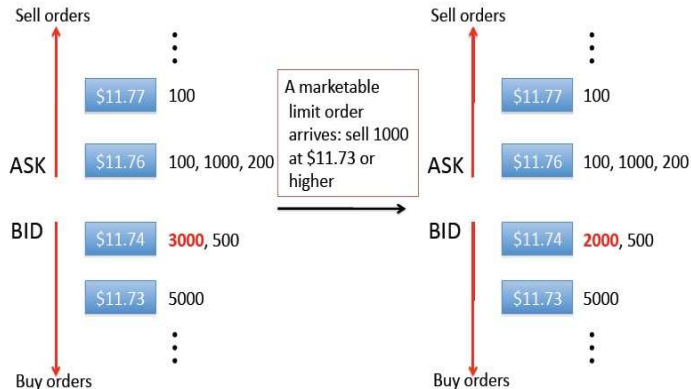


Figure : Market Ask-Order Arrival

Limit Order Book XV

- Two ticks get printed: volumes of 100, 400 were traded at the price of 11.76.

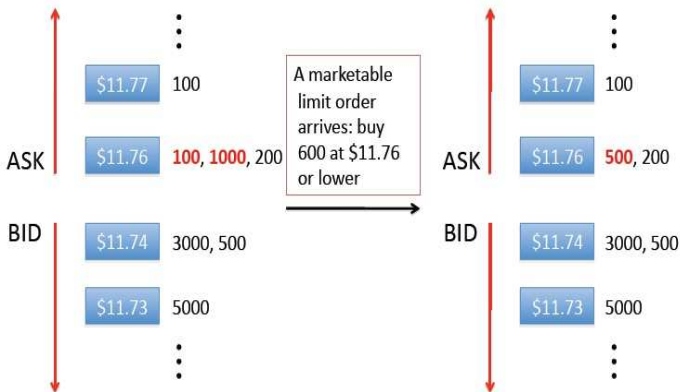


Figure : Market Buy-Order Arrival

Limit Order Book XVI

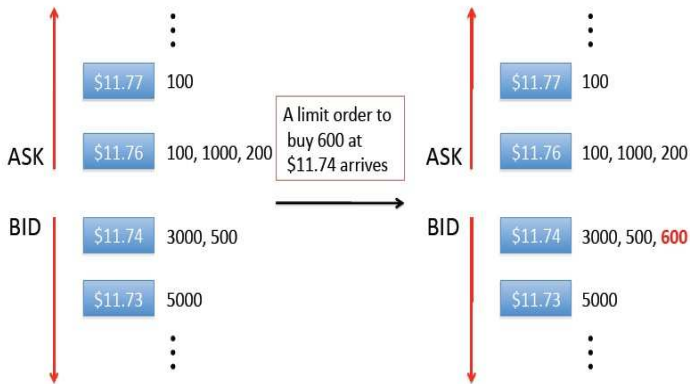


Figure : Limit Buy-Order Arrival

Limit Order Book XVII

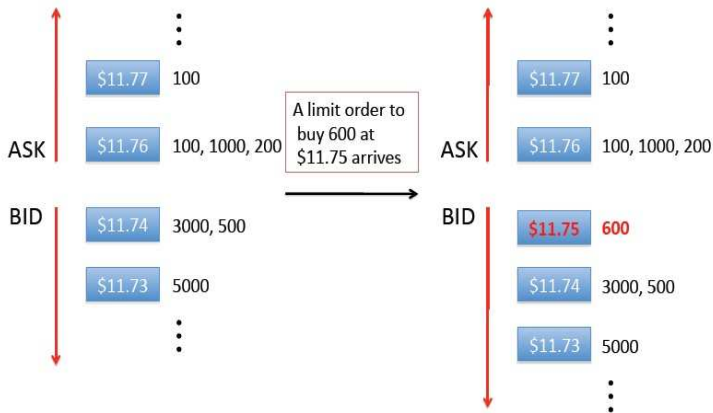


Figure : Limit Buy-Order Arrival

Limit Order Book XVIII

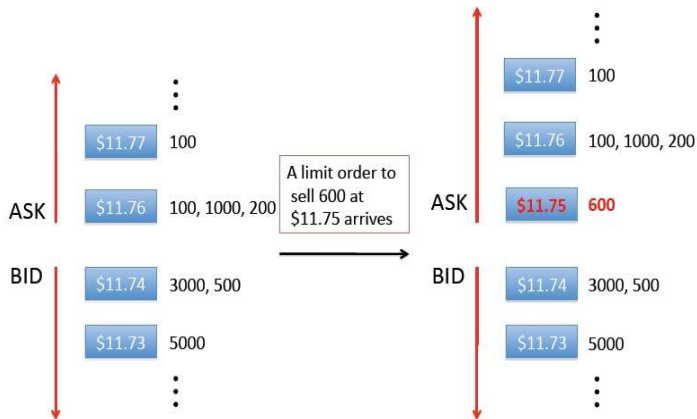


Figure : Limit Sell-Order Arrival

Limit Order Book XIX

- Many exchanges allow traders to submit so-called **hidden orders**. An example of this are **iceberg orders**.

Definition (Iceberg Orders)

Iceberg orders are limit orders which contain a **visible (disclosed)** part and a **hidden** part. The disclosed part has the same priority as a regular limit order. The hidden part does not become visible unless the disclosed part is executed.

Definition (Dark Pools)

Dark pools are LOBs where all active orders have hidden parts.

- Dark pools:
 - can be advantageous to for instance large investment banks, as they can make massive transactions without influencing the price too much.
 - can be very risky, as it becomes harder and harder to regulate the market.

■ Trading Day Description:

- **Pre-opening phase:** An opening auction at the beginning of the day (from 7 : 15 to 9 : 00 for the Euronext), where orders are placed in the order book, but no transactions take place. At the end of this phase, all orders that can be executed against an opposing order, and the resulting order book determines the opening price.
- **Main trading session:** Continuous trading takes place where all incoming orders are checked to be matched with standing limit orders.
- **Closing phase:** Similar to the pre-opening phase, at the end of the day (from 17 : 35 to 17 : 35 for the Euronext) as still orders can be submitted but not matched. After this a closing price is determined which will serve as a guideline on the pre-opening phase of the next trading day.

Trading Positions I

Definition (Long Position: Buy Low, Sell High)

Buying assets on an **Long Position** is the action of purchasing the assets anticipating that their price will **rise** over time.

- **Long Position Example:** Gary wants to purchase 100 shares of Nike, Incorporated at present time with the anticipation that the price per share will appreciate in the near future.
- Therefore, Gary buys **100** shares at today's closing price of **\$82.00** a share.

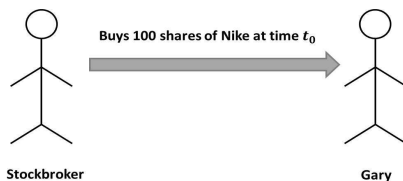


Figure : Long Position: Buy Phase

Trading Positions II

- His initial investment (not including the broker's fee) is $100 \times \$82.00 = \$8,200.00$

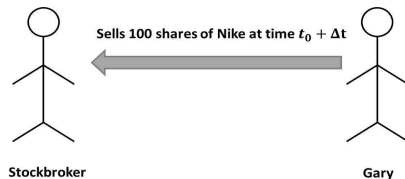


Figure : Long Position: Sell Phase

- One year later the price of the Nike stock is **\$87.00** a share, an increase of **\$5.00** per share from Gary's initial investment.
- The value of Gary's investment would now be as follows:
 $100 \times \$87.00 - \$8,200.00$ (a gain of **\$500.00**, not including the broker's fee if he decides to sell).

Trading Positions IV



Figure : Long Position Example

Trading Positions V

- **Long Position Case I:** Buy a quantity of x shares at t_0 for P_{t_0} a share and sell a quantity of x shares at $t_0 + \Delta t$ for $P_{t_0 + \Delta t}$ a share.
 - Return at time t_0 : $R_{t_0} = -x \cdot P_{t_0}$
 - Return at time $t_0 + \Delta t$: $R_{t_0 + \Delta t} = +x \cdot P_{t_0 + \Delta t}$
 - Overall return: $R_{\text{overall}} = R_{t_0} + R_{t_0 + \Delta t} = x \cdot (P_{t_0 + \Delta t} - P_{t_0})$.
 - Since $P_{t_0 + \Delta t} < P_{t_0}$ the overall return will be negative ($R_{\text{overall}} < 0$) and the trader will suffer an amount of loss which will be proportional to the price difference $\Delta P = P_{t_0 + \Delta t} - P_{t_0}$.
- **Long Position Case II:** Buy a quantity of x shares at t_0 for P_{t_0} a share and sell a quantity of x shares at $t_0 + \Delta t'$ for $P_{t_0 + \Delta t'}$ a share.
 - Return at time t_0 : $R_{t_0} = -x \cdot P_{t_0}$
 - Return at time $t_0 + \Delta t'$: $R_{t_0 + \Delta t'} = +x \cdot P_{t_0 + \Delta t'}$
 - Overall return: $R_{\text{overall}} = R_{t_0} + R_{t_0 + \Delta t'} = x \cdot (P_{t_0 + \Delta t'} - P_{t_0})$.
 - Since $P_{t_0 + \Delta t'} > P_{t_0}$ the overall return will be positive ($R_{\text{overall}} > 0$) and the trader will gain an amount of profit which will be proportional to the price difference $\Delta P = P_{t_0 + \Delta t'} - P_{t_0}$.

Trading Positions VI

Definition (Short Position: Sell High, Buy Low)

The **Short Position** is a technique used when an investor anticipates that the value of a stock will decrease in the short term. In a **short sell** transaction the investor borrows the shares of stock from the investment firm to sell to another investor. The investor must eventually return the stock they borrow. The intent is to **borrow** the stock for sale at a high price, then **buy** them back later at a lower price and return them to the stockbroker.

- **Short Position Example:** Jill decides to short sell 100 shares of Ford Motor Company because she has heard rumors of massive recall of their minivans. Jill anticipates that the value of Ford's stock will decrease in the near future because of the high costs of the recall and the negative publicity.

Trading Positions VII

- Therefore, Jill borrows **100** shares of Ford stock from her broker and sells it to another investor for today's closing price of **\$34.00**.

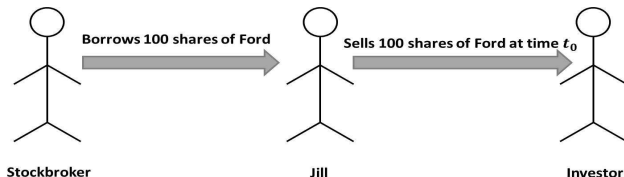


Figure : Short Position: Sell Phase

- Jill sold **100** shares at **\$34.00**: $100 \times \$34.00 = \$3,400.00$ (**Short Selling**)

Trading Positions VIII

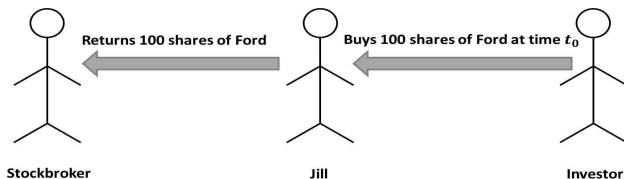


Figure : Short Position: Buy Phase

- Two weeks later after notices of the recall have been publicized and other investors have reacted by negatively by selling their Ford stock, the price has fallen to **\$28.00** a share.
- Jill decides to purchase **100** shares of Ford stock now to replace what she has borrowed from her broker: $100 \times \$28.00 = \$2,800.00$ (**Short Cover**).
- Overall, Jill made a profit of $\$3,400.00 - 2,800.00 = \600.00 (not including the broker's fee).

Trading Positions IX



Figure : Short Position Example

Trading Positions X

- **Short Position Case I:** Sell a quantity of x shares at t_0 for P_{t_0} a share and buy them back at $t_0 + \Delta t$ for $P_{t_0 + \Delta t}$ a share.
 - Return at time t_0 : $R_{t_0} = +x \cdot P_{t_0}$
 - Return at time $t_0 + \Delta t$: $R_{t_0 + \Delta t} = -x \cdot P_{t_0 + \Delta t}$
 - Overall return: $R_{\text{overall}} = R_{t_0} + R_{t_0 + \Delta t} = x \cdot (P_{t_0} - P_{t_0 + \Delta t})$.
 - Since $P_{t_0} < P_{t_0 + \Delta t}$ the overall return will be negative ($R_{\text{overall}} < 0$) and the trader will suffer an amount of loss which will be proportional to the absolute price difference $|\Delta P| = |P_{t_0 + \Delta t} - P_{t_0}|$.
- **Short Position Case II:** Sell a quantity of x shares at t_0 for P_{t_0} a share and buy them back at $t_0 + \Delta t'$ for $P_{t_0 + \Delta t'}$ a share.
 - Return at time t_0 : $R_{t_0} = +x \cdot P_{t_0}$
 - Return at time $t_0 + \Delta t'$: $R_{t_0 + \Delta t'} = -x \cdot P_{t_0 + \Delta t'}$
 - Overall return: $R_{\text{overall}} = R_{t_0} + R_{t_0 + \Delta t'} = x \cdot (P_{t_0} - P_{t_0 + \Delta t'})$.
 - Since $P_{t_0} > P_{t_0 + \Delta t'}$ the overall return will be positive ($R_{\text{overall}} > 0$) and the trader will gain an amount of profit which will be proportional to the absolute price difference $|\Delta P| = |P_{t_0 + \Delta t'} - P_{t_0}|$.

Trading Problems I

- Characteristics of order-driven trading systems change the dynamics of markets and demand new trading strategies that can capture the short-term behavior of underlying assets.
- A great amount of research focuses on providing a Machine Learning framework for capturing the dynamics of high frequency limit order books (financial equity markers).
- Automated real-time prediction of metrics such as:
 - **Mid-Price Movement**: a statistical indicator of potential trading profits. However, it is not considered to be a guarantee.
 - **Price Spread-Crossing**: is a **less frequent** occurrence that does assure a profit if correctly identified in advance.

Trading Problems II

■ Quantities Measured:

- P_t^{ask} : the best (minimum) ask (sell) price at time t .
- P_t^{bid} : the best (maximum) bid (buy) price at time t .
- $P_t^{\text{mid}} = \frac{P_t^{\text{ask}} + P_t^{\text{bid}}}{2}$: the **mid-price** at time t .
- $\Delta P_t^{\text{mid}} \in \{< 0, = 0, > 0\}$: the **mid-price movement** at time time t .
- **bid-ask spread-crossing** at time $t + \Delta t$:
 - **upward price spread-crossing**: $P_{t+\Delta t}^{\text{bid}} > P_t^{\text{ask}}$
 - **downward price spread-crossing**: $P_{t+\Delta t}^{\text{ask}} < P_t^{\text{bid}}$
 - **no price spread-crossing**: $P_{t+\Delta t}^{\text{ask}} \geq P_t^{\text{bid}}$ and $P_{t+\Delta t}^{\text{bid}} \leq P_t^{\text{ask}}$

Trading Problems III

- **Case I:** Given the correct prediction of an upward price spread crossing at time $t + \Delta t$ such that $b_{t+\Delta t} > a_t$:

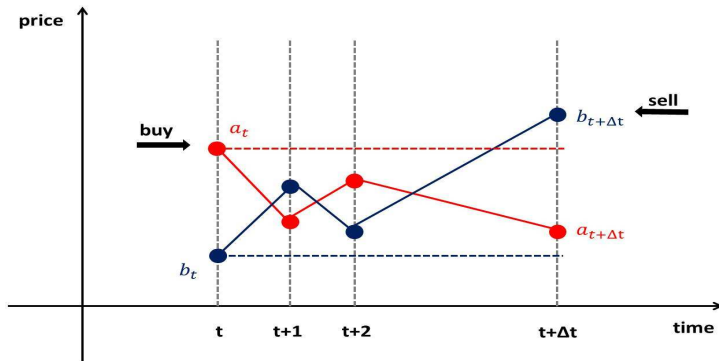


Figure : Upward Price Spread Crossing: Long Position

Trading Problems IV

- A profitable trading strategy would be the following (**long position**):
 - **buy at best ask price** at time t an amount of x shares inducing a return: $R_t = -x \cdot a_t$.
 - **sell at best bid price** at time $t + \Delta t$ an amount of x shares inducing a return: $R_{t+\Delta t} = +x \cdot b_{t+\Delta t}$.
 - Thus, the overall return would be given by:
$$R_{\text{overall}} = R_t + R_{t+\Delta t} = x \cdot (b_{t+\Delta t} - a_t) > 0$$

Trading Problems V

- **Case II:** Given the correct prediction of a downward price spread crossing at time $t + \Delta t$ such that $a_{t+\Delta t} < b_t$:

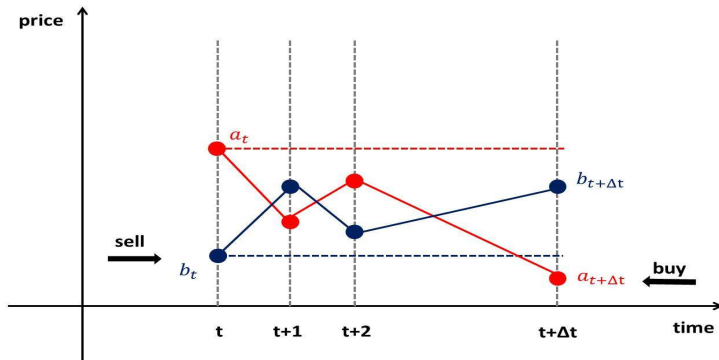


Figure : Downward Price Spread Crossing: Short Position

- A profitable trading strategy would be the following (**short position**):
 - **borrow and sell at best bid price** at time t an amount of x shares inducing a return: $R_t = +x \cdot b_t$.
 - **buy and return at best ask price** at time $t + \Delta t$ an amount of x shares inducing a return: $R_{t+\Delta t} = -x \cdot a_{t+\Delta t}$.
 - Thus, the overall return would be given by:
$$R_{\text{overall}} = R_t + R_{t+\Delta t} = x \cdot (b_t - a_{t+\Delta t}) > 0$$

Trading Problems VII

- Price fluctuations are captured by:
 - *occurrence* and *direction* of mid-price movement.
 - *occurrence* and *direction* of bid-ask spread crossing.
- Machine Learning approaches aim at predicting the direction that the price will move over a time horizon Δt .
- Any move to be predicted has to be significant enough to cross the spread.
 - $P_{t+\Delta t}^{bid} > P_t^{ask} \implies$ Buy the currency (long position)
 - $P_{t+\Delta t}^{ask} > P_t^{bid} \implies$ Sell the currency (short position)
 - $P_{t+\Delta t}^{bit} \leq P_t^{ast}$ and $P_{t+\Delta t}^{ask} \leq P_t^{bid} \implies$ Take no position

Trading Problems VIII

- **Trade Sign Prediction:** Corresponds to the problem of predicting whether a future market order in a time horizon Δt will be buyer or seller initiated.
- Let $I_{t+\Delta t}^{ts}$ be the trade-sign indicator function such that:

$$I_{t+\Delta t}^{ts} = \begin{cases} +1, & \text{market order at time } t + \Delta t \text{ is a **market-buy** order;} \\ -1, & \text{market order at time } t + \Delta t \text{ is a **market-sell** order.} \end{cases}$$

and \mathbf{S}_t a vector containing limit-order book related information until time t

- Machine learning approaches aim at constructing models that capture the following probabilities:

$$\mathbb{P}(I_{t+\Delta t}^{ts} = +1 | \mathbf{S}_t)$$

and

$$\mathbb{P}(I_{t+\Delta t}^{ts} = -1 | \mathbf{S}_t)$$

■ Price Jump Prediction:

- The problem of price jump prediction relates to the following two situations:
 - A sell market order arrival which is executed at a price smaller than the best bid price at the moment.
 - A buy market order arrival which is executed at a price larger than the best ask price at the moment.
- Price Jumps:
 - **inter-trade** price jumps.
 - **trade-through** price jumps.

- **Inter-trade Price Jump** (on the bid side): is defined as an event where a **market sell order** is executed at a price which is **smaller** than the **best limit price on the bid** just after the precedent market order arrival:

$$P_{t_{i+1}}^{mo} < P_{t_i}^{bid}$$

- An inter-trade price jump on the bid side permits a limit order submitted at the best bid just after a market order arrival to be surely by the next market order.

- **Inter-trade Price Jump** (on the ask side): is defined as an event where a **market buy order** is executed at a price which is **larger** than the **best limit price on the ask** just after the precedent market order arrival:

$$p_{t_{i+1}}^{mo} > p_{t_i}^{ask}$$

- An inter-trade price jump on the ask side permits a limit order submitted at the best ask just after a market order arrival to be surely by the next market order.

- **Trade-through Price Jump:** corresponds to the arrival of a new market order, the size of which is larger than the quantity available at the best limit on the Bid (for a sell order) or Ask (for a buy order) side of the order book.
- By nature such an order will imply an automatic and instantaneous price change, the value of which will be exactly the difference in monetary units before and after transaction on the relevant side of the order book.

Trading Problems XIII

- Trade-through price jump may be interpreted as the instantaneous price change triggered by a market order.
- Inter-trade price jump, is on the contrary, a post-trade market impact.
- Relevant research focuses on investigating whether the order book shape is informative for the inter-trade price-jump and whether the identification of trade-through events contributes to this problem.

Trading Problems XIV

- **Optimized Trade Execution Problem:** The goal in the trade-execution optimization problem is to sell (respectively buy) V shares of a given stock within a fixed time period (or horizon) H , in a manner that minimizes the revenue received (respectively, minimizes the capital spent).
- An optimal trader would buy an asset before the price rises, and sell the asset before its value declines.
- Let \mathbf{F}_t denote the position taken at time t and $\mathbf{r}_t = \mathbf{p}_t - \mathbf{p}_{t-1}$ the difference in value of the asset between the current period t and the previous period $t - 1$.

Trading Problems XVI

- F_t may be defined in the following way:

$$F_t \begin{cases} > 0, & \text{Long Position;} \\ = 0, & \text{Neutral Position;} \\ < 0, & \text{Short Position.} \end{cases}$$

- Trader takes the long position when he anticipates that the price appreciates by period $t + 1$ which is when the estimation for the price difference is $\hat{r}_{t+1} = \hat{p}_{t+1} - p_t > 0$.
- When taking the neutral position the outcome at period $t + 1$ has no effect on the trader's profits.
- Trader takes the short position when he anticipates that the price at period $t + 1$ will be lower which is when the estimation for the price difference is $\hat{r}_{t+1} = \hat{p}_{t+1} - p_t < 0$.

Trading Problems XVII

- F_t represents holdings at period t . That is, $n_t = \mu \cdot F_t$ shares are bought (long position) or sold (short position), where μ is the maximum possible amount of shares per transaction.
- Return at time t , considering the decision F_{t-1} is:

$$R_t = \mu \cdot (F_{t-1} \cdot r_t - \delta \cdot |F_t - F_{t-1}|)$$

where δ is the cost for a transaction at period t .

- If $F_t = F_{t-1}$, (no change in investment this period) then there will be no transaction penalty. Otherwise, the penalty is proportional to the difference in shares held.
- The trade execution problem relates to the optimization of the following criterion (**Sharpe Ratio**)

$$S_T = \frac{E[R_T]}{\sqrt{E[R_T] - E^2[R_T]}} = \frac{\text{Average}(R_T)}{\text{Volatility}(R_T)}$$