Introduction to Market Microstructure

Christopher Ting

http://cting.x10host.com/TMU/TMU.html

http://www.mysmu.edu/faculty/christophert/
✉: christophert@smu.edu.sg

June 9, 2018

Broad Lesson Plan

From Investment to Trading

- Investment: 3 broad **steps**—investment research, portfolio management, and trading.
- Portfolio managers should obtain best possible trades for their clients.
- Important: understanding of the types of traders and their venues; how to measure trading costs; available trading techniques; and market microstructure.
- **Market microstructure** refers to the particular characteristics of trading venues that affect transaction volumes and prices. The structure of a market may also cause **market friction** such that a security may not be correctly valued, presenting the astute manger with trading opportunities.

Market Structures

- Markets serve several purposes:
  - liquidity: minimal cost and timely trading
  - transparency: correct and up-to-date trade and market information
  - assurance of trade completion: trouble-free trade settlement

- There are 3 main categories of markets:
  - **quote-driven:** investors trade with dealers
  - **order-driven:** investors trade with each other directly
  - **brokered markets:** investors use brokers to locate the counterparty to a trade

- A hybrid market is a combination of the other three markets.

Three Properties of a Liquid Market

- Kyle (1985) identified three key properties of a liquid market.
- **Tightness**: the cost of turning around a position over a short period of time
- **Depth**: the size of an order-flow innovation required to change prices a given amount
- **Resiliency**: the speed with which prices recover from a random, uninformative shock

Liquid Market

- A liquid market has
  - ★ small bid-ask spreads
  - ★ sufficient market depth
  - ★ resilience
- What are necessary for a market to be liquid?
  - ★ An abundance of buyers and sellers: Traders know they can quickly reverse their trade if necessary.
  - ★ Diversity of investor characteristics: If every investor had the same information, valuations, and liquidity needs, there would be little trading.
  - ★ A convenient trading platform: Reliable and large-capacity trading engines increase investor activity and liquidity.
  - ★ Integrity. All investors receive fair treatment.
- In a liquid market, traders with information trade more frequently and security prices are more efficient, and liquidity risk premium for securities is reduced.

What is an Electronic Order?

- Orders are short messages to the exchange through the broker.
- An order is a set of instructions that the trader gives to the exchange. It must contain at least the following instructions:
  1. Trade instrument (or instruments): What to trade
  2. Trade direction: Buy or sell or cancel
  3. Trade price
  4. Trade size: How many shares or contracts to trade or cancel
  5. Lit or unlit: Display publicly or not if order is not executed immediately
  6. Validity: When the order may be filled

Three Examples of Limit Order Book

NKD Sep17			NIY Sep17			NK Sep17		
Bid Size	Price	Ask Size	Bid Size	Price	Ask Size	Bid Size	Price	Ask Size
	19970	24		19935	42		19930.00	104
	19965	27		19930	49		19925.00	81
	19960	32		19925	51		19920.00	82
	19955	29		19920	59		19915.00	91
	19950	29		19915	49		19910.00	122
	19945	31		19910	47		19905.00	95
	19940	31		19905	54		19900.00	95
	19935	37		19900	50		19895.00	98
	19930	37		19895	43		19890.00	92
	19925	14		**19890**	32		**19885.00**	191
3	**19920**		3	**19885**		71	**19880.00**	
31	19915		22	19880		381	19875.00	
42	19910		40	19875		48	19870.00	
31	19905		60	19870		258	19865.00	
23	19900		56	19865		55	19860.00	
30	19895		43	19860		264	19855.00	
24	19890		50	19855		120	19850.00	
20	19885		55	19850		55	19845.00	
18	19880		36	19845		53	19840.00	
13	19875		35	19840		50	19835.00	

Rule-Based Order-Matching system

- Order precedence rules (algorithm)
  1. Price: Which order's price is most competitive?
  2. Display: Lit orders ahead of unlit orders
  3. Time: Earliest order gets filled first
- The execution priority rule used by most exchanges is price, then display, and then time.
- Highest bids and lowest offers always execute before lower bids and higher offers.
- Among equally priced orders, displayed orders always get executed before non-displayed orders.
- Among displayed and non-displayed orders at the same price level, time of arrival determines an order's priority.

Limit Order and Market Order

Definition

A **limit order** $x = (p_x, \omega_x, t_x)$ *submitted at time* t_x *with price* p_x *and signed size* $\omega_x < 0$ *(*$\omega_x > 0$*) is a commitment to buy (sell) up to* $|\omega_x|$ *units of the traded asset at a price no more (less) than* p_x*.*

- A market buy order is $x = (\infty, \omega_x, t_x)$ when $\omega_x < 0$.
- A market sell order is $x = (0, \omega_x, t_x)$ when $\omega_x > 0$.
- **Market orders** are executed immediately because their prices are super-competitive.

Trading Matching Algorithm

- When a buy (sell) order x is submitted, an LOB's **trade-matching algorithm** checks whether it is possible to match x to some other previously submitted sell (buy) order. If so, the matching occurs immediately and a trade occurs.

- Otherwise, x becomes **active**, and it remains active until either it becomes matched to another incoming sell (respectively, buy) order or it is cancelled.

Limit-Order Book (LOB)

Definition

*An **LOB** $\mathcal{L}(t)$ is the set of all active orders in a market at time t.*

- For a limit order $x = (p_x, \omega_x, t_x)$ that becomes active upon arrival, it holds that $x \in \mathcal{L}(t_x), x \notin \lim_{t' \uparrow t_x} \mathcal{L}(t')$.
- The active orders in an LOB $\mathcal{L}(t)$ can be partitioned into the set of active buy orders $\mathcal{B}(t)$, for which $\omega_x < 0$, and the set of active sell orders $\mathcal{A}(t)$, for which $\omega_x > 0$.
- An LOB can be considered as a set of queues, each of which consists of active buy or sell orders at a specified price.

Lot Size and Tick Size

Definition

The **lot size** σ *of an LOB is the smallest amount of the asset that can be traded within it. All orders must arrive with a size* $\omega_x \in \{\pm k\sigma | k = 1, 2, \ldots\}$.

Definition

The **tick size** π *of an LOB is the smallest permissible price interval between different orders within it. All orders must arrive with a price that is specified to the accuracy of* π.

Bid, Ask, and Bid-Ask Spread

Definition

The **best bid price** $b(t)$ *at time* t *is the highest stated price among active buy orders. The* **best ask price** $a(t)$ *at time* t *is the lowest stated price among active sell orders.*

$$b(t) := \max_{x \in \mathcal{B}(t)} p_x, \quad a(t) := \min_{x \in \mathcal{A}(t)} p_x. \tag{1}$$

Definition

The **bid-ask quoted spread** *at time* t *is* $s(t) := a(t) - b(t)$*.*

Definition

The **mid price** *at time* t *is* $m(t) := [a(t) + b(t)]/2$*.*

Effective Spread

Definition

The **effective spread** *is twice the absolute difference between the execution price and the true price of the asset transacted.*

- Typically the midpoint is used as the proxy for the true price.
- Effective spreads are often averaged over all transactions during a period in order to calculate an average effective spread.
- Lower average effective spreads indicate better liquidity for a security.

Depth

Definition

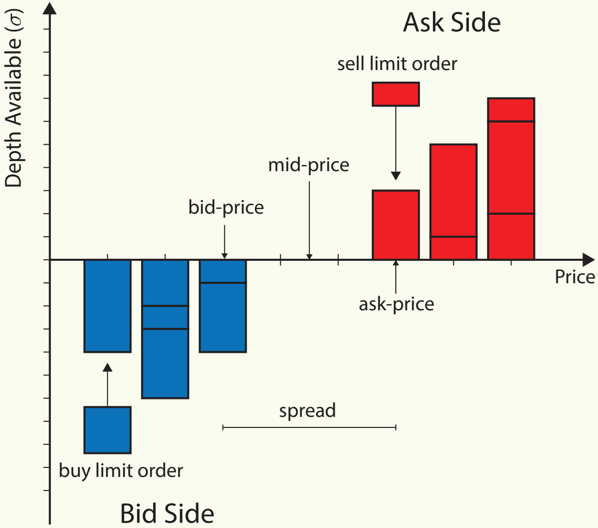
The **bid-side depth** *available at price p and at time t is*

$$n^b(p,t) := \sum_{\{x \in \mathcal{B}(t) | p_x = p\}} \omega_x. \qquad (2)$$

The **ask-side depth** *available at price p and at time t, denoted $n^a(p,t)$, is defined similarly using $\mathcal{A}(t)$.*

- The depth available is often stated in multiples of the lot size.
- Because $\omega_x < 0$ for buy orders and $\omega_x > 0$ for sell orders, it follows that $n^b(p,t) \leqq 0$ and $n^a(p,t) \geqq 0$ for all prices p.

Schematic of an LOB



Depth Profile

Definition

The **bid-side depth profile** *at time t is the set of all ordered pairs $\left(p, n^b(p,t)\right)$. The* **ask-side depth profile** *at time t is the set of all ordered pairs $\left(p, n^a(p,t)\right)$.*

Definition

The **mean bid-side depth** *available at price p between times t_1 and t_2 is*

$$\overline{n}^b(p, t_1, t_2) := \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} n^b(p,t)\mathrm{d}t. \quad (3)$$

The **mean ask-side depth** *available at price p between times t_1 and t_2, denoted $\overline{n}^a(p, t_1, t_2)$, is defined similarly using the ask-side depth available.*

Sampling

- Consider the time series $m(t_1), \ldots, m(t_n)$, for some times $t_1, \ldots, t_n$.
- When the t_i are spaced regularly in time, with τ seconds between successive samplings, such a time series is said to be sampled on a **τ-second timescale**.
- When the t_i are chosen to correspond to arrivals of orders, the t_i may be spaced irregularly in time. Such a time series is said to be sampled on an **event-by-event timescale**.
- When the t_i are chosen to correspond to trades (i.e., matchings in an LOB), the t_i may also be spaced irregularly in time. Such a time series is said to be sampled on a **trade-by-trade timescale**.

Arrival of an Order

- If $p_x \leqq b(t)$, then x is a limit order that becomes active upon arrival.
- If $b(t) < p_x < a(t)$, then x is a limit order that becomes active upon arrival. It causes $b(t)$ $\big(a(t)\big)$ to increase (decrease) to p_x at time t_x.
- If $p_x \geqq a(t)$, then x is a **marketable limit order** that immediately matches to one or more active sell orders upon arrival.

When Matching Occurs

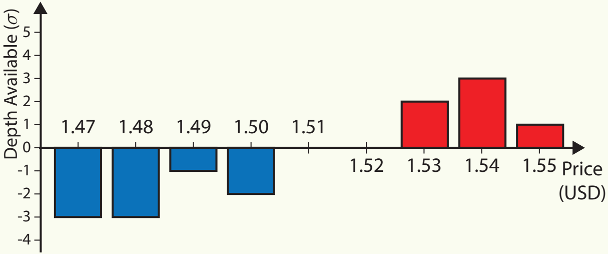
- Matching occurs at the price of the active order.
- Whether or not such a matching causes $a(t)$ or $b(t)$ to change at time t_x depends on $n^a(a(t), t)$ or $n^b(b(t), t)$, and ω_x.
- In particular, the new bid price $b(t_x)$ immediately after the arrival of a sell market order x is

$$\max(p_x, q), \text{ where } q = \arg\max_{k'} \sum_{k=k'}^{b(t)} \left| n^b(k, t) \right| > \omega_x. \quad (4)$$

- Similarly, the new ask price $a(t_x)$ immediately after the arrival of a buy market order x is

$$\min(p_x, q), \text{ where } q = \arg\min_{k'} \sum_{k=a(t)}^{k'} n^a(k, t) > |\omega_x|. \quad (5)$$

LOB Price Formation



Arriving order x	Values after arrival (USD)			
	$b(t_x)$	$a(t_x)$	$m(t_x)$	$s(t_x)$
Initial Values	1.50	1.53	1.515	0.03
$(\$1.48, -3\sigma, t_x)$	1.50	1.53	1.515	0.03
$(\$1.51, -3\sigma, t_x)$	1.51	1.53	1.52	0.02
$(\$1.55, -3\sigma, t_x)$	1.50	1.54	1.52	0.04
$(\$1.55, -5\sigma, t_x)$	1.50	1.55	1.525	0.05
$(\$1.54, 4\sigma, t_x)$	1.50	1.53	1.515	0.03
$(\$1.52, 4\sigma, t_x)$	1.50	1.52	1.51	0.02
$(\$1.47, 4\sigma, t_x)$	1.48	1.53	1.505	0.05
$(\$1.50, 4\sigma, t_x)$	1.49	1.50	1.495	0.01

Returns

Definition

The bid-price return between times t_1 and t_2 is $R^b(t_1, t_2) := \dfrac{b(t_2) - b(t_1)}{b(t_1)}$. The **ask-price return** *between times t_1 and t_2, denoted $R^a(t_1, t_2)$, and the* **mid-price return** *between times t_1 and t_2, denoted $R^m(t_1, t_2)$, are defined similarly.*

Definition

The **bid-price logarithmic return** *between times t_1 and t_2 is $r^b(t_1, t_2) := \log\left(b(t_2)/b(t_1)\right)$. The* **ask-price logarithmic return** *between times t_1 and t_2, denoted $r^a(t_1, t_2)$, and the* **mid-price logarithmic return** *between times t_1 and t_2, denoted $r^m(t_1, t_2)$, are defined similarly.*

Execution Immediacy Versus Uncertainty

- Traders use market orders when they want a trade *immediately* to fill their orders.
- Limit orders allow traders to have their orders filled at the specified price or better.

	Market Orders	Limit Orders
Order Execution	Guaranteed	Uncertain
Time to Execution	Short	Uncertain
Order Resubmission	No	Yes
Price Improvement	Not Applicable	Yes
Price Impact	Yes	No
Market View	Momentum	Contrarian
Liquidity	Taker	Provider

Major Approaches to Modeling LOB

Perfect Rationality

1. Cut-off strategies
2. Fundamental values and informed traders
3. Minimizing market impact

Zero Intelligence

1. Random-walk diffusion models
2. Discrete-time models
3. Continuous-time models

Agent-Based Approach

- Large number of possibly heterogeneous agents interact in a specified way.

Perfect Rationality Versus Zero Intelligence

- Orders are submitted because *perfectly rational* traders attempt to maximize their "utility" by making trades in markets driven by "information".
- In the zero-intelligence approach, aggregated order flows are assumed to be governed by specified stochastic processes whose rate parameters are conditional on other variables such as $\mathcal{L}(t)$.
- Many models rely exclusively on Monte Carlo simulation to produce output.
- Although such simulations still motivate interesting observations, it is often difficult to trace exactly how specific model outputs are affected by input parameters.

State-Space Complexity

- It is a well-established empirical fact that current order flows depend on both $\mathcal{L}(t)$ and on recent order flows.
- A problem with studying conditional behaviour is that the state space of an LOB is huge.
- If there are P different choices for price in a given LOB, then the state space of the current depth profile alone, expressed in units of the lot size σ, is $\mathbb{Z}^P$.
- This makes it very difficult to investigate conditional dependences, as the number of variables is so large. Therefore, a key modeling task is to find a way to simplify the evolving, **high-dimensional state space**, while retaining an LOB's important features.

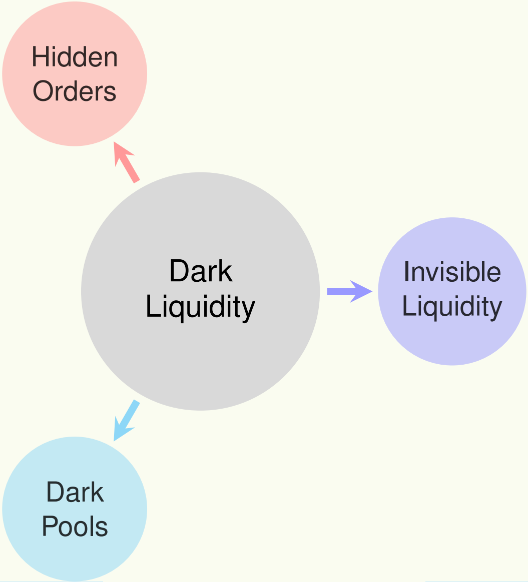
Feedback and Coupling

- In addition to traders' actions depending on $\mathcal{L}(t)$, the state of $\mathcal{L}(t)$ also clearly depends on traders' actions.
- These mutual dependences induce feedback between $\mathcal{L}(t)$ and trader behaviour.
- Also, $b(t)$ determines the boundary condition for sell limit order placement because any sell order placed at or below $b(t)$ at least partially matches immediately.
- A similar role is played by $a(t)$ for buy orders.
- Therefore, **order flow creates a strong coupling** between $b(t)$ and $a(t)$.

Trade Intention

- An LOB $\mathcal{L}(t)$ reflects only the **subset of trading intentions** that traders have announced up to time t.
- However, the fact that no trader has submitted a limit order at a given price does not imply that none of them want to trade at this price, because they could be keeping their intentions private by submitting orders only when absolutely necessary.
- A typical snapshot of $\mathcal{L}(t)$ at a given time is often very sparse, containing few active orders. However, this is not an indication that few people wish to trade; it is merely an indication that they have not yet announced any intention to do so.
- Indeed, some traders choose to submit only market orders and do not submit limit orders at all.

Dark Liquidity



Realized Volatility

Definition

Given the bid-price series $b(t_1), b(t_2), \ldots, b(t_n)$ *sampled at regularly spaced times, the* **bid-price realized volatility** *is* $v^b(t_1, t_2, \ldots, t_n) := \text{st.dev.}\left(\left\{r^b(t_i, t_{i+1}) \mid i = 1, 2, \ldots, n-1\right\}\right)$. *The* **ask-price realized volatility**, *denoted* $v^a(t_1, t_2, \ldots, t_n)$, *and the* **mid-price realized volatility**, *denoted* $v^m(t_1, t_2, \ldots, t_n)$, *are defined similarly.*

- Realized volatility depends on the frequency at which price series are sampled.
- It is a useful measure for comparing the variability of return series sampled with the same frequency, but it is not appropriate to compare the realized volatility of a once-daily price series for one stock to a once-hourly price series for another.

Bid-Price Realized volatility Per Trade

- To compare how prices vary on a trade-by-trade basis, realized volatility per trade is more appropriate.

Definition

Given the bid-price series $b(t_1), b(t_2), \ldots, b(t_n)$ sampled at the times at which n consecutive sell market orders arrive, the **bid-price realized volatility per trade** *is*

$$V^b(t_1, t_2, \ldots, t_n) := \text{st.dev.}\left(\left\{r^b(t_i, t_{i+1}) \mid i = 1, 2, \ldots, n-1\right\}\right).$$

The **ask-price realized volatility per trade***, denoted $V^a(t_1, t_2, \ldots, t_n)$, is defined similarly using n consecutive buy market order arrival times.*
The **mid-price realized volatility per trade***, denoted $V^m(t_1, t_2, \ldots, t_n)$, is defined similarly using n consecutive market order arrival times.*

Intra-Day Volatility

Definition

For a given trading day D, the **bid-price intra-day volatility** *is*

$$\rho^b(D) := \log\left(\max_{t\in D} b(t) / \min_{t\in D} b(t)\right).$$

The **ask-price intra-day volatility***, denoted $\rho^a(D)$, and the* **mid-price intra-day volatility***, denoted $\rho^m(D)$, are defined similarly.*

- Intra-day volatility is useful for estimating the probability of very large price changes in a given day.
- It is particularly important for *day traders*, who unwind their trading positions before the end of each trading day.

Power Law

Definition

A random variable Z is said to have a **power-law tail with exponent α** *if there exists some $\alpha \in \mathbb{R}$ such that $f_Z(z) \sim O\left(z^{-\alpha}\right)$ as $z \to \infty$.*

- If there exists a $z_{\min} > 0$ such that $f_Z(z)$ is proportional to $z^{-\alpha}$ for all $z \geqq z_{\min}$, then clearly Z has a power-law tail.
- When attempting to fit power-law tails to empirical observations, it is often assumed that such a $z_{\min}$ exists.
- Under this assumption, you can use an algorithm for consistent estimation of α and $z_{\min}$ via maximum likelihood techniques, and for testing the hypothesis that the data really does follow a power law for $z \geqq z_{\min}$.
- Different estimators often produce vastly different estimates of α, making such inference difficult.

Auto-Correlation Function and Long Memory

- We denote by $\boldsymbol{X}$ a real-valued, second-order stationary time series of length k, $\boldsymbol{X} = X(t_1), X(t_2), \ldots, X(t_k)$

Definition

The **autocorrelation function** A *of a time series* $\boldsymbol{X}$ *is given by*

$$A_{\boldsymbol{X}}(l) := \frac{1}{k-l} \sum_{i=1}^{k-l} \left(X(t_i) - \langle \boldsymbol{X} \rangle\right)\left(X(t_{i+l}) - \langle \boldsymbol{X} \rangle\right), \tag{6}$$

where $\langle \boldsymbol{X} \rangle$ *is the mean of the series.*

Definition

A time series $\boldsymbol{X}$ *is said to exhibit* **long memory** *if there exists some* $\alpha \in (0,1)$ *such that* $A_{\boldsymbol{X}}$ *decays like a power law,*

$$A_{\boldsymbol{X}}(l) \sim O\left(l^{-\alpha}\right), \text{ as } l \to \infty. \tag{7}$$

Hurst Exponent

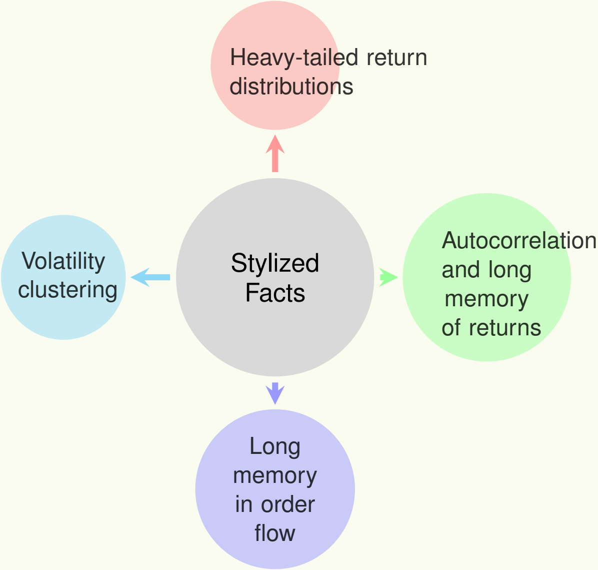
- An alternative way to characterize long memory is via the diffusion properties of the integrated series $\boldsymbol{Y}$,

$$Y(l) = \sum_{i=1}^{l} X(t_i). \tag{8}$$

- If $\boldsymbol{X}$ is a long-memory process, then the standard deviation of $\boldsymbol{Y}$ scales as $O\left(l^H\right)$, with $\frac{1}{2} < H \leq 1$.
- If $\boldsymbol{X}$ does not have long memory, then the standard deviation of $\boldsymbol{Y}$ scales as $O\left(l^{1/2}\right)$.
- The exponent H is called the **Hurst exponent**, and is related to α by

$$H = 1 - \frac{\alpha}{2}. \tag{9}$$

Stylzed Facts about LOB



What is Price Impact?

- Price impact refers to the correlation between an incoming order (to buy or to sell) and the subsequent price change.
- A buy trade pushes the price up and a dour reality for traders is that their second buy trade is on average more expensive than the first because of their own impact on the market.
- Isn't a transaction a fair deal between a buyer and a seller? So why is there price impact?

Three possibilities

- ★ Traders successfully forecast short term price movements and trade accordingly.
- ★ The impact of trades reveal some private information.
- ★ Impact is a statistical effect due to order flow fluctuations.

Instantaneous Impact

Definition

The **instantaneous bid-price impact** *of a market event at time t' is the difference:*

$$b(t') - \lim_{t \uparrow t'} b(t). \tag{10}$$

Definition

The **instantaneous bid-price logarithmic return impact** *of a market event at time t' is*

$$\log b(t') - \lim_{t \uparrow t'} \left[\log b(t) \right]. \tag{11}$$

Impact Functions

Definition

The **instantaneous bid-price impact function** $\phi^b(\omega_x)$ *outputs the mean instantaneous bid-price impact for a buy market order of size* $|\omega_x|$.

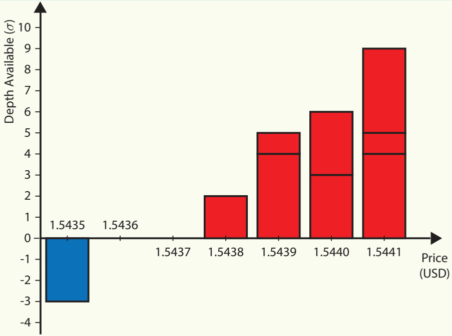
Definition

The **instantaneous bid-price logarithmic return impact function** $\Phi^b(\omega_x)$ *outputs the mean instantaneous bid-price logarithmic return impact for a buy market order of size* $|\omega_x|$.

$\phi^a(\omega_x)$ and $\Phi^b(\omega_x)$ for a sell market order are defined analogously.

Market Reality: What event is this?

Market Impact



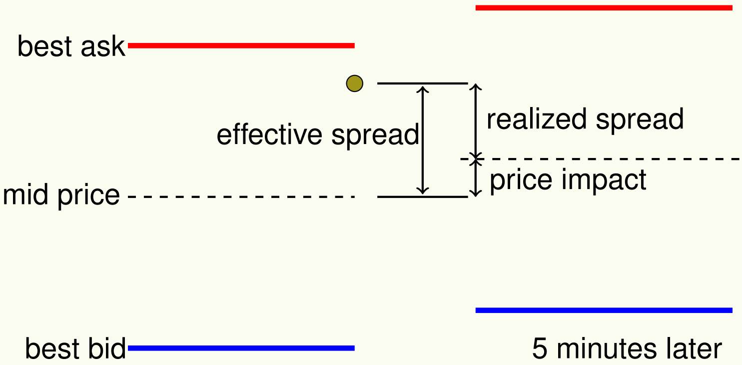
Definition

The **instantaneous market impact of a market event** *at time t' is*

$$\mathcal{L}(t') \setminus \lim_{t \uparrow t'} \mathcal{L}(t), \qquad (12)$$

where $\setminus$ denotes the difference of the two sets.

Effective Spread versus Realized Spread



Trade Imbalance or Order Flow

Definition

The **trade imbalance count** *during time interval $T = [t_1, t_2]$, denoted $\Omega^c(T)$, is the difference between the total number of incoming buy market orders and the total number of incoming sell market orders that arrive during time interval T.*

Definition

The **trade imbalance size** *during time interval $T = [t_1, t_2]$, denoted $\Omega^\omega(T)$, is the difference between the total absolute size of all incoming buy market orders and the total size of all incoming sell market orders that arrive during time interval T.*

Flash Crash on May 6, 2010 Source: www.nanex.net

At about 14:43:31 HFT systems detected a sudden price drop and automatically went short to ride the downward momentum, causing a feed-back loop to develop and panic ensued.

Takeaways

- A limit order book is essentially a file on a computer that contains all orders sent to the market, along with their characteristics such as the sign of the order, price, quantity, and a time stamp.
- Matching priority is price, display, and time for order-driven market
- Market order takes liquidity provided by limit order.
- There are substantial challenges associated with studying historical LOB data.
- Modeling the dynamics of LOB is very difficult.

Reference

1. **Limit Order Book** by Gould et al (2013)