MA668: Algorithmic and High Frequency Trading
Lecture 10

Prof. Siddhartha Pratim Chakrabarty
Department of Mathematics
Indian Institute of Technology Guwahati

Price Dynamics

1. For brevity: The interest rate is set to zero.
2. Let t denote the time index and let T be the time at which the cash value of the asset is determined.
3. Probabilities and Expectations are adjusted to incorporate the incorporation of public information from trade, as captured by the filtration $\mathcal{F}_t$.
4. As MM's observe different sequences of buy and sell orders they adjust (using Bayes' rule) the estimation of distribution of v. In particular, they set: $p_t = \mathbb{P}(v = V_H|\mathcal{F}_t)$ and $\mu_t = E[v|\mathcal{F}_t]$.
5. Then, bid and ask prices will adjust in response to the history of trading, so that:

$$a_t = \mu_t + \Delta_{a,t} = \mu_t + \frac{1}{1 + \frac{1-\alpha}{\alpha}\frac{1/2}{p_t}} \left(V_H - \mu_t\right),$$

and

$$b_t = \mu_t - \Delta_{b,t} = \mu_t - \frac{1}{1 + \frac{1-\alpha}{\alpha}\frac{1/2}{1-p_t}} \left(\mu_t - V_L\right).$$

Price Dynamics

1. The resulting bid-ask price display changes that reflect the public information embedded in the order flow.
2. Note also that at every execution, the execution price (a_t/b_t) is equal to the expectation of the underlying asset, conditional on the history of the order flow $\mathcal{F}_t$, and also the information in the execution.
3. It can be shown that the realized price process (the price process at the execution times) is a martingale.

Price Sensitive Liquidity Traders

1. An interesting extension of static model: Allow liquidity traders to avoid trading, if the half-spread Δ is too high.
2. A direct way to do this is to assume that liquidity traders get an additional (exogenous) value from executing their desired trade.
3. Trader i gets a cash equivalent utility gain of c_i if she/he manages to execute her/his desired trade.
4. Thus, if the transaction cost imposed by the half spread is too high (higher than c_i), then trader i will prefer not to execute the trade.
5. Assume that the distribution of parameter c_i in the population of liquidity traders is described by the CDF F, that is, $F(c)$ is the proportion of liquidity traders that have $c_i \leq c$.
6. c_i is referred to as the agent's urgency parameter.

Price Sensitive Liquidity Traders

1. Now, we can recompute the expected profit of the MM from setting an ask price of $a = \mu + \Delta_a$, which will now be given by:

$$\frac{(1 - F(\Delta_a))(1 - \alpha)/2}{\alpha p + (1 - F(\Delta_a))(1 - \alpha)/2} \Delta_a + \frac{\alpha p}{\alpha p + (1 - F(\Delta_a))(1 - \alpha)/2} (\Delta_a - (V_H - \mu)) .$$

2. In the expression, we have incorporated the fact that whenever a liquidity trader wants to buy $(1 - \alpha)/2$, only $(1 - F(\Delta_a))$ will have sufficient urgency to execute the trade with a buy-half-spread equal to Δ_a.

3. Introducing the parameter increases the half-spreads, which are now implicitly defined by the expressions:

$$\Delta_a = \frac{1}{1 + \frac{1-\alpha}{\alpha} \frac{(1-F(\Delta_a))/2}{p}} (V_H - \mu) ,$$

and

$$\Delta_b = \frac{1}{1 + \frac{1-\alpha}{\alpha} \frac{(1-F(\Delta_b))/2}{1-p}} (\mu - V_L) .$$

Price Sensitive Liquidity Traders

1. Key issue: As the MM increases the half-spread, she/he faces a smaller population of liquidity traders.
2. If the urgency parameters in the population are relatively small, the MM may find that the above expressions have only the extreme solutions $\Delta_a = V_H - \mu$ and $\Delta_b = \mu - V_L$.
3. These extreme solutions correspond to the solutions without liquidity traders and represent market collapse.
4. With those spreads no one gains anything from trade and any trade that may occur will come from the informed agents who are indifferent to either trading or not trading.
5. However, any trade will reveal the underlying value of the asset and the price will be strong efficient.

Empirical and Statistical Evidence: A Prelude

1. Focus: Empirical analysis of various aspects trading, including prices, returns, spreads, volume etc. primarily using millisecond stamped data.
2. Begin with: Daily data that will give us a general overview of the main issues.
   - (A) Prices and Returns.
   - (B) Volume and Market Quality Measures (Spreads, Volatility or Depth).

Quality

The Data

1. Data is used from several sources.
   - A. Daily and monthly data: Publicly available aggregated date from Yahoo! Finance "plus" Data from the Center for Research in Security Prices (CRSP).
   - B. Millisecond (one-thousandth of a second) data: Publicly available industry standard data, similar to the direct data feed.
2. The data was converted into tabular format for ease of processing and is in binary form for speed and storage reasons.
3. For illustrative purpose: Converted to more human readable form.

The Data (Contd ...)

The data are made up of the following fields (after dropping two fields which are irrelevant in the context of the current discussion):

1. Timestamp: Number of millisecond after midnight.
2. Order ID: Unique order ID.
   - B: Add buy order.
   - S: Add sell order.
   - E: Execute outstanding order in part.
   - C: Cancel outstanding order in part.
   - F: Execute outstanding order in full.
   - D: Delete outstanding order in full.
   - X: Bulk volume for the cross event.
   - T: Execute non-displayed order.
3. Shares: Order quantity (Zero for "F" and "D" messages).
4. Price: Zero for cancellations and executions.
5. Ticker: The ticker associated with the asset in question.
6. MPID: Market Participant ID associated with the transaction [a].
7. Exchange: ID of the current market (NASDAQ=1).

[a] This information is usually missing from the public feed

The Data (Contd ...)

1. These messages record events that affect the Limit Order Book (LOB).
2. Essentially they capture what happens to Limit Order (LOs).
3. LOs are posted (B,S) and later on they are cancelled (C,D) or executed (E,F).
4. So Market Orders (MOs) are not recorded, but must be deduced from observing how they are executed against standing LOs (or against non-displayed/hidden orders, T).

The Data (Contd ...)

Row	Time	Order ID	Type	Qty.	Price	Ticker	Ex.
1:	33219784	4889087	B	1900	345800	TZA [a]	1
2:	33219784	4887036	C	200	0	FMS [b]	1
3:	33219784	4879129	D	0	0	QQQQ [c]	1
4:	33219784	4889088	S	2000	454800	QQQQ	1
5:	33219784	4879130	D	0	0	QQQQ	1
6:	33219784	4889089	S	500	454800	QQQQ	1
7:	33219785	4882599	D	0	0	QQQQ	1
8:	33219785	4888889	F	0	0	STD[d]	1

Table: A LOB example (the row numbers have been added to facilitate the discussion and MPID have been dropped).

[a] Direxion Daily Small Cap Bear 3X Shares
[b] Fresenius Medical Care AG & Co.
[c] Invesco QQQ Trust Series 1
[d] Standard Chartered PLC

The Data (Contd ...)

1. Messages are sent to market between 33219784 and 33219785 ms from midnight of July 13, 2010, that is between 09:13:39.784 and 09:13:39.785.
2. Several messages for ETF QQQQ, one each for ETF TZA, and stocks FMS and STD (STD is SAN now).
3. Row 1: It is for TZA ETF and should be read as: Message recorded at 33219784 ms from midnight (09:13:39.784), with order number 4889087, the LO is posted to buy (B) for 1,900 shares at a price of \$34.58 (all prices are in \$ times 10,000), with the number 1 in the last column representing the market code of NASDAQ.
4. For QQQQ:
   - A. A LO is cancelled (Row 3).
   - B. Posting of a sell LO (Row 4).
   - C. Another LO cancellation (Row 5).
   - D. A second sell LO posted (Row 5).
   - E. A third LO cancelled (Row 6).

The Data (Contd ...)

1. The posted sell orders include the quantity and price for the orders (2,000 at \$45.48 and 500 also at \$45.48), while cancelled orders must be matched with their original posted orders (ID 4879130 and 4882599) in order to identify the corresponding prices and quantities.
2. Similarly for the full execution of the Order ID 4888889 for STD.
3. For FMS (Row 2) we see a partial cancellation of 200 units from Order ID 4887036 (the price needs to be read off the original posted order).
4. From this data, one can construct the complete order book at any point in time, and study how the market changes over time using different variables and methods.