Signal Processing Models for Order Execution and the Microstructure of Financial Markets

Tutorial

Ilya Pollak
Pragma Securities LLC and Purdue University
ipollak@ecn.purdue.edu
www.ece.purdue.edu/~ipollak
www.pragmatrading.com

May 27, 2013
9:00-10:20am
Vancouver, Canada
Papers at ICASSP-2013 Related to Financial Signal Processing


Outline

I. Market microstructure
II. Order execution algorithms
III. Market impact
IV. Performance analysis
Why Trading Happens: An Example

- Employees contribute part of their paycheck to pre-tax retirement accounts and select mutual funds to invest into.
- Mutual funds (TIAA-CREF, Fidelity, ...) buy and sell securities to achieve their goals (e.g., following a market index).
- When a portfolio manager at a mutual fund needs to rebalance a portfolio, he will send the trades to his trader.
- The trader will select brokers to route the trades to, and will give them instructions (e.g., how many shares to be bought/sold, whether the trade is required to complete, by what time this needs to get done, etc).
- Brokers will split the order into small chunks, select order type and various other parameters, and will send the smaller orders to be executed at various trading venues.
Life Cycle of a Trade

Asset management company

Investor → PM → Trader → Broker → Trading Venue
Life Cycle of a Trade

Proprietary trading company

Investor → PM → Trader → Broker → Trading Venue
Life Cycle of a Trade

In this tutorial, we zoom in on the last part
Order Placement

• An order sent to a trading venue must contain
  – the instrument;
  – the size, i.e., the number of shares or lots or contracts to be traded;
  – the side, i.e., buy or sell;
  – other trading instructions, such as the price at which to trade.
Order Types

• Every trading venue supports many different trading instructions, i.e., has many different order types.
  
  – [https://www.directedge.com/Membership/FeeSchedule/EDGAFeeSchedule.aspx](https://www.directedge.com/Membership/FeeSchedule/EDGAFeeSchedule.aspx)
  – [https://www.directedge.com/Membership/FeeSchedule/EDGXFeeSchedule.aspx](https://www.directedge.com/Membership/FeeSchedule/EDGXFeeSchedule.aspx)

• In this tutorial, we consider only limit orders --- the most widely used order type.

• We only consider stocks.
The *order book* for a stock at a fixed time instant:

- **ASK**
  - $11.76  
  - 100, 1000, 200

- **BID**
  - $11.74  
  - 3000, 500
  - $11.73  
  - 5000

Someone offering to sell 100 shares for $11.77 per share

Someone bidding $11.73 per share to buy 5000 shares

Orders constantly arrive, get filled, get canceled
Example 1

A marketable limit order arrives: sell 1000 at $11.73 or higher

A tick gets printed: volume of 1000 was traded at the price of $11.74
Example 2

A marketable limit order arrives: buy 600 at $11.76 or lower

Two ticks get printed: 100, 400 were traded at $11.76
Example 3

A limit order to buy 600 at $11.74 arrives
Example 4

A limit order to buy 600 at $11.75 arrives
Example 5

A limit order to sell 600 at $11.75 arrives
Order Execution Problem

• Inputs: name, size, side, quantity, and possibly other parameters.

• Decisions to be made:
  – How fast to trade.
  – Which venues to use.
  – What order types to use.
Order Execution Problem: Two-Tiered Solution

• “Macro” step: decide on the overall execution strategy
  – Get instructions from the client, e.g.:
    • By what time will it have to be completed?
    • Is it required to execute 100% or is partial completion OK?
  – Develop a strategy to partition the order into smaller pieces to be executed over small time intervals.

• “Micro” step: a specific algorithm for sending small orders.

• “Macro” and “micro” decisions might be interdependent and form one integrated algorithm.
Example

• Buy 100,000 shares of IBM on May 28, 2013 from 9:30am until 4pm.
Possible “Macro” Solutions

- How to select a good solution?
- How to analyze performance after the fact?
One Possible Solution

**ASK**

- **$206.17** 100
- **$206.16** 100, 600, 200

Place a marketable buy order for 100,000 shares at a very high price

**BID**

- **$206.18** 99000
- **$206.15** 3000, 500
- **$206.14** 5000

**BID**

- **$206.18** 99000
- **$206.15** 3000, 500
- **$206.14** 5000

• Results of placing a huge marketable buy order:
  – Wipe out the “ask” side of the order book.
  – Invite everyone to sell to you at inflated prices and create a feeding frenzy.
  – It is likely that the order will be executed very quickly, but the average execution price will be far greater than the $206.16 best ask at the time of placing the order.
What to do?

- Partition the order into small pieces and send the smaller pieces one by one.
- However, doing this too slowly is risky: the market price may change if we wait too long.
Typical Macro Strategies

• Time-Weighted Average Price (TWAP)
• Volume-Weighted Average Price (VWAP)
• More complex algorithms which arise from casting order scheduling as an optimization problem:
  – a body of interesting academic literature
  – limited practical applications because of the difficulty in estimating certain parameters and in analyzing performance
Time-Weighted Average Price

• Let \([t_0, t_N]\) be a time interval.
• Suppose that during this time interval, the price of a stock changed at times \(t_1, t_2, \ldots, t_{N-1}\).
• Suppose the price during \([t_{n-1}, t_n]\) is \(p_n\).
• TWAP on \([t_0, t_N]\) is defined as

\[
\frac{\sum_{n=1}^{N} p_n (t_n - t_{n-1})}{t_N - t_0}
\]
TWAP Execution

\[ \sum_{n=1}^{N} p_n(t_n - t_{n-1}) \]

- TWAP on \([t_0, t_N]\) is \(\frac{\sum_{n=1}^{N} p_n(t_n - t_{n-1})}{t_N - t_0}\)
- How to execute an order so that the average execution price over \([t_0, t_N]\) is equal to TWAP?
- Execute at a uniform rate of \(\frac{V}{(t_N - t_0)}\) per unit time, where \(V\) is the total order volume.
- Total execution price over the interval \([t_{n-1}, t_n]\): \(p_n(t_n - t_{n-1})\frac{V}{(t_N - t_0)}\).
- Average execution price overall: TWAP (sum over all the intervals, divided by the total volume).
TWAP Trajectory

Cumulative quantity bought

0

100,000

9:30pm

4:00pm

Time

Ilya Pollak
TWAP Execution: Shortcomings

• TWAP may not be a good metric of the quality of execution.
  – Through executing our order, we may significantly influence TWAP.

• TWAP execution is not a good strategy if the volume is predictably time-varying
  – E.g., volume during 9:30-11am is usually much larger than during 1:30-3pm.
Volume-Weighted Average Price

• Let $[t_0, t_N]$ be a time interval.
• For each trade that occurred during this time interval, let $p_k$ be the price and $v_k$ be the volume traded.
• Let the total number of trades be $K$.
• VWAP on $[t_0, t_N]$ is defined as

$$\frac{\sum_{k=1}^{K} p_k v_k}{\sum_{k=1}^{K} v_k}$$
Example: TWAP vs VWAP

• Suppose there are two trades during 10-11am:
  – 100 shares for $10 at 10am
  – 10000 shares for $11 at 10:30am

• The TWAP is $10.50 because the last traded price was at $10 for 30 minutes and at $11 for 30 minutes

• The VWAP is $10.99 --- a more realistic reflection of the price one would likely get
VWAP Execution

\[
\text{VWAP} = \frac{\sum_{k=1}^{K} p_k v_k}{\sum_{k=1}^{K} v_k}
\]

- VWAP on \([t_0, t_N]\) is 

- Suppose we know ahead of time the fraction \(\sum_{k=1}^{K} v_k\) of the volume that will be traded over each small time interval \([t, t+\Delta t]\).

- During \([t, t+\Delta t]\), we execute the following fraction of our total order \(V\): 

\[
\frac{\Delta v_t}{\sum_{k=1}^{K} v_k} V
\]

- Average execution price: VWAP.
Typical VWAP Trajectory

Cumulative quantity bought

Time

9:30pm 4:00pm

0 100,000

Ilya Pollak
VWAP Execution: Shortcomings

• VWAP may not be a good metric of the quality of execution.
  – Through executing our order, we may significantly influence VWAP.
  – In fact, if our order is 100% of the trades, we are guaranteed to execute at VWAP!
• Need reasonable volume forecasts.
More Complex Approaches

- Many other papers.
Problem Formulation

• $s_t = \Phi_t(\tilde{p}_{t-1}) = \# \text{ shares bought at price } p_t, \ t=1,...,T$
• $\tilde{p}_{t-1} = \text{ the market price at time } t-1$
• Total buy order: $S = \Sigma s_t$
• Total execution price: $P = \Sigma p_t s_t$
• Minimize the expected execution price as well as some measure of the risk:
  
  $\min_{\phi_1,...,\phi_T} \left( E_0(P) + \lambda V_0(P) \right)$

• E.g., $V_0(P) = \text{ variance of } P$
• $E_0$ and $V_0$ denote conditional statistics given all price observations until time $t=0$. 
Model for Market Price

\[ \tilde{p}_t = \text{mid-quote price at time } t \]
\[ = (\text{best bid} + \text{best ask}) / 2 \]

\[ \tilde{p}_t = \tilde{p}_{t-1} + g(s_t) + \varepsilon_t \]

- \( \varepsilon_t = \text{white zero-mean noise, independent of past prices and execution quantities} \)
- \( g = \text{permanent market impact} \)
Model for Our Execution Price

\[ p_t = \tilde{p}_{t-1} + h(s_t) \]

- \( s_t \) = # shares executed at price \( p_t \)
- \( h \) = temporary market impact
- \( \tilde{p}_{t-1} \) = market mid-price at time \( t-1 \)
Stochastic Control Problem

\[
\min_{\phi_1, \ldots, \phi_T} \left( E_0 \left( \sum_{t=1}^{T} p_t \phi_t \left( \tilde{p}_{t-1} \right) \right) + \lambda V_0 \left( \sum_{t=1}^{T} p_t \phi_t \left( \tilde{p}_{t-1} \right) \right) \right)
\]

subject to \( \phi_t \left( \tilde{p}_{t-1} \right) = s_t \) and \( \sum_{t=1}^{T} s_t = S \)

and \( p_t = \tilde{p}_{t-1} + h(s_t) \)

and \( \tilde{p}_t = \tilde{p}_{t-1} + g(s_t) + \epsilon_t \)
What is $s_t$?

- Quantity traded at time $t$?
  - This is not directly controllable, unless we always use aggressive orders.
- Quantity *placed* at time $t$, traded at a possibly later time at price $p_t$ which is related to the market price at time $t-1$ through

$$
p_t = \tilde{p}_{t-1} + h(s_t)
$$
Complexity

• Optimal policy is computed during trading.
• Fast solutions are required, especially when many instruments are being traded at the same time.
• Possible strategies for lightening the complexity:
  – Use simple impact functions (e.g., linear).
  – Search only over open-loop control policies.
  – Simplify the cost function.
Example 1: Bertsimas-Lo ‘98

No risk term in the cost: $\min_{\phi_1, \ldots, \phi_T} E_0 \left( \sum_{t=1}^{T} p_t \phi_t (\tilde{p}_{t-1}) \right)$

No temporary impact: $p_t = \tilde{p}_t$

Linear permanent impact: $\tilde{p}_t = \tilde{p}_{t-1} + a\phi_t (\tilde{p}_{t-1}) + \varepsilon_t$

- TWAP = optimal closed-loop policy = optimal open-loop policy.
- Proof via dynamic programming --- DP possible since the cost is additive.
Example 2: Almgren-Chriss ‘01

Open-loop control: \( \min_{s_1, \ldots, s_T} \left[ E_0 \left( \sum_{t=1}^{T} p_t s_t \right) + \lambda V_0 \left( \sum_{t=1}^{T} p_t s_t \right) \right] \)

Linear temporary impact: \( p_t = \tilde{p}_{t-1} + bs_t \)
Linear permanent impact: \( \tilde{p}_t = \tilde{p}_{t-1} + as_t + \epsilon_t \)

• A fast procedure to pre-compute the trading trajectory off-line.
• Closed-form solution in the continuous limit.
Learning the Model from Internal Data

\[ p_t = \tilde{p}_{t-1} + h(s_t) \]

\[ \tilde{p}_t = \tilde{p}_{t-1} + g(s_t) + \varepsilon_t \]

• If you trade large volumes and store the trade data, you can take \( s_t \) to be your actual historical traded quantities.
• Look at the scatter plots of traded quantities vs the trade price and vs the market price, and design \( h \) and \( g \) (e.g., linear).
• Estimate parameters of \( h \) and \( g \) via regression using historical data.
• Weakness: most of market data is unused.
Learning the Model from Market Data

\[ p_t = \tilde{p}_{t-1} + h(s_t) \]
\[ \tilde{p}_t = \tilde{p}_{t-1} + g(s_t) + \varepsilon_t \]

• For each trade, infer from the data whether it was buyer-initiated or seller-initiated.

• Look at the scatter plots of traded quantities vs the trade price and vs the market price, and construct a form for \( h \) and \( g \) (e.g., linear).

• Estimate parameters of \( h \) and \( g \) via regression using historical data.

• Weakness: the model may be poorly matched to your flow.
Performance Analysis

• Popular benchmarks:
  – arrival price shortfall
  – VWAP shortfall

• But one needs to be careful.
Performance Analysis: Example 1

• Suppose you are a trader and you are evaluating two execution algorithms provided by two different brokers:
  – GreatAlgo’s past performance is claimed to be worse than VWAP, on average, by 2 basis points (i.e., by 0.02%).
  – SuperAlgo’s past performance is claimed to be better than VWAP, on average, by 10 basis points (i.e., by 0.1%).

• Which one would you choose?
Example 1

• [If possible, do life testing on your own flow!]
• If it is possible to beat VWAP by 10 bps, buy and sell at the same time and pocket the average profit of 20bps. Why bother with agency trading?
• Key question: are the trades always required to complete?
What if no completion required?

Only buy if the price is below the running average

Price,
Running VWAP, assuming uniform volume

• Beat VWAP
• But may not trade at all!
Performance Analysis and Completion

• This example shows that, unless all trades are required to complete (or a required completion percentage is specified), performance analysis is meaningless.

• Why does this example not necessarily imply the existence of arbitrage?
Performance Analysis: Example 2

• Suppose you are evaluating two execution algorithms on your own flow, and observe that:
  – ExcellentAlgo is worse than the arrival price, on average, by 2 basis points (i.e., by 0.02%).
  – FantasticAlgo is better than the arrival price, on average, by 10 basis points (i.e., by 0.1%).

• What do you do?
Example 2: Illustration

Arrival price = $100

Average buy price = $75

Trade more slowly! (E.g., start trading later.)
Performance Analysis: More Criteria

• Since both the arrival price shortfall and the VWAP shortfall have limitations as performance metrics, it is important to also look at other things
  – e.g., use data from before, during, and after trading to estimate market impact of trading

• Need a lot of data for performance analysis results to be meaningful.
Other Problems Arising in Order Execution

• Optimal allocation among venues (must consider fees/rebates, fill probabilities)
• Optimal order aggressiveness (e.g., cross the spread vs place a passive limit order)
• Optimal order type (e.g., dark vs lit)
• Order book modeling
References


