Ch 3: Optimal Organizations for optimal trading

3.1 organizing a trading structure to answer a fragmented landscape

large asset managers concentrated on mounting/unwinding positions to dealing desks

Dealing desk responsibilities

- relationship with intermediaries
- measuring Transaction cost analysis (TCA)
- choosing benchmark (VWAP, TWAP, ...)

Trading architecture

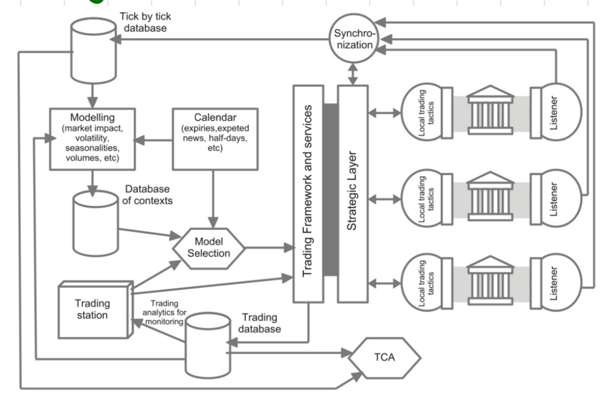


Figure 3.1. Stylized blueprint of a trading architecture.

Main inputs of trading tools

Market data

contains

- market trades (price, quantity, time stamp)
- limit orderbooks (bid/ask, quant, price level part., modifications)

connection to venues

trading algo connected to exchanges
- not natively synched (latency mismatch)
- two way
  - → one to send messages to trading venue
  - → recieve messages (can be used as analytic for dark pools (only have post transparency))

Historical data

Optimal trading relies on intraday risk control and rythms of market features
→ captured by models that use historical data tested params

Historical data should include:
- market data
- analytics
- recordings of messages
- context info

Models

- made of formula, estimation, accuracy comp.

3.1.2 Components of trading algos

components

- execution param

  define conditions that apply to the algo: side + quant, benchmark, aggressiveness, access to dark pools

- risk control layer

  trading core: uses model and expected market contexts to produce min/max trading envelope to balance market impact & market risk

- liquidity seeking tactics (SOR, ect.)

communication between risk layer and tactics: autonomous tactic as close as possible to venue, risk control layer located equal distance to robots.

3.1.3 Main output of automated trading system

Pre trade analytics

- allow for consistency of portfolio & generate automated directives to speed / slow

Pre trade analytics:

- bid / ask spread, traded volume, quantities on the book
- expected trading profiles
- expected context (news, vol, ect.)
- break down by sector, currency, country

- execution params reduce footprint in market & avoid exposure to specific events
- synchronization measures can be taken

Monitoring Indicators

- traders build on the fly diagnosis of algos
- algos need to make concrete decisions to improve quality of execution (on the fly modification params)

Performance indicators

- subset of monitoring indicators are direct measures of performance it trading would stop

- estimates of performance 'seen from now' if market conditions do not change

Average price

$$\boxed{\text{Avg price} = \text{Immediate price} + \text{Market moves} + \text{Market impact}}$$

benchmark by benchmark Avg. price

Avg Price = Immediate price
+ planned price   expected risk control on avg.
+ unexpected market moves
+ trading eff   success of liquidity seeking tactics
+ market impact

where to produce real time indicators

- algos should be indicator publishers, stateless agents with introspection capabilities

Post trade analysis

- pre trade analysis → expected avg. value
- monitoring gives: curr value, updated exp. avg. value, likelihood curr value

- gives assessment of value of bid/ ask spread

visualization break downs

- sector
- currency
- buy / sell

Transactional Cost Analysis

- used to compare different algo providers to compare efficiency

TCA : understand what happened bc of market context
Post-trade Analysis: understand performance from design of algo

<u>objective:</u> associate given algo with given market context / trading style

- user able to now send flow to proper algo provider
- liquidity must be taken into account

liquidity quantified by

- bid / ask spread
- tick size basis points
- quantity on the books
- daily turnover
- avg trade size
- free float
- trading rate

3.2 Market impact Measurements : Understanding PFP from one investor

- trade timing describes how an order is being traded compared to what happens in the market

3.2.1 Market Impact over trading period

Intraday analysis allows
- quantity Δ in market impact as order being executed
- dilution of impact at end of execution

Difference between price & arrival time

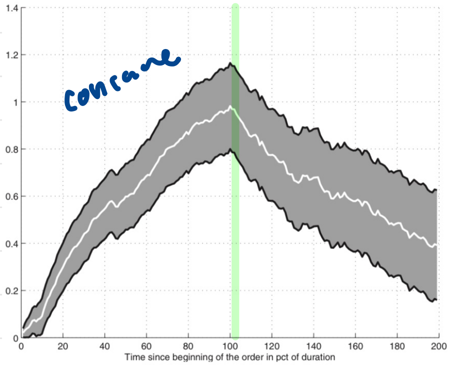


Figure 3.3. Intraday market impact in spreads.

terminates at 100

trend in return of a stock according to the arrival price in units of spread over time

- buy order
- white curve shows avg of stock returns according to arrival price (in spread)
- grey 95% confidence

(0, 100): concave over time
slice executed at beginning of order has more impact than one executed at end of order

(100, 200): price reverts to lower level than that of execution (100)

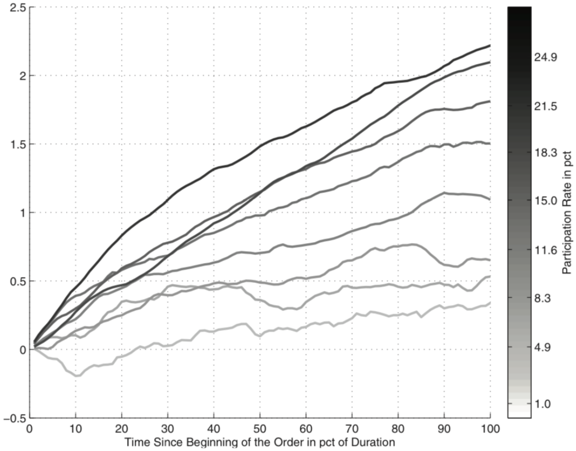
Market impact parameters

Stock specific:
- liquidity, vol, spread

order specific:
- Aggressiveness of execution
- quantity to execute
- expected duration

Period Liquidity Ratio

- dependency between market impact and the duration of an order
- observe: market impact increases when participation rate and duration increase even independently



execution at 100

- market impact increases as participation increases

- longer the duration the bigger the impact

$$\boxed{\text{Market impact} \propto \text{Duration}^{\alpha} \times \text{Participation}^{\gamma}}$$

3.2.2 Market impact on a longer horizon: Price anticipation & permanent market impact

permanent market impact: remaining price shift after decay has taken place

price anticipation or permanent market impact?

1. permanent impact as consequence of mechanical process

- stock prices move bc of trading from all market participants
- selling pressure ↑, price ↓

2. permanent impact as trace of new info in price

- prices move bc of new info made available to market participants

permanent market impact on a long term horizon

- focus on studies that debias post-execution variations on the daily

→ seperate into effect with permanent market impact

Brokmann: intensity of trading signal which triggered each meta order to remove info content of studied proprietary meta order database

- divergence between effective & expected price gives measure of permanent impact in absence of info effect

↳ observed no permanent impact on residual price moves

Waelbroeck and Gomes' differentiates meta orders

- informed trades: meta orders coming from portfolio rebalance
- uninformed / cash trades: triggered by heterogeneous & relatively exogenous to market info

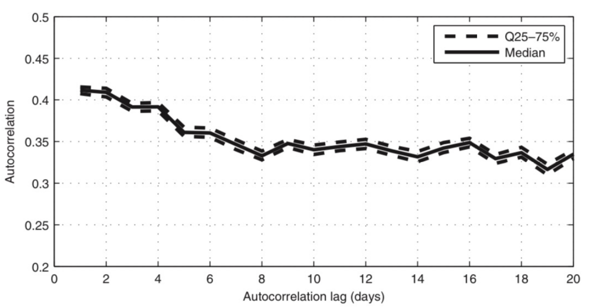
found only informed trades have permanent market impact.

Bacry:

- $\beta \rightarrow 1$, total portfolio made up of all meta orders not far from CAPM market portfolio
- after removing alpha from price moves no permanent market impact is left

Debiasing temp market impact of autocorrelated execution during post execution period

- remove autocorrelations of meta order flows to deinfluence post execution price variations



withdraw impact of metaorders
- simple market impact model
- deconvolute decay kernal
- fit sq root model of daily participation on temp market impact

CAPM decomposition of price variations

Bacry developed a CAPM framework to disentangle systematic component of market impact from idiosyncratic comp.

decomp. centered on execution day over 41 day period:

$$\forall d \in \{D-20, \dots D+20\},$$

$$\log(P_d) - \log(P_{d-1}) = \beta(\log(I_d) - \log(I_{d-1})) + \Delta W_d$$

D: metaorder execution date
P_d: stock close price on date d

β: beta of stock $[D-20, D+20]$

I_d: ref index on d

β estimated with LS Reg.

$$\boxed{\underbrace{w_d - w_{D-1}}_{} = \sum_{d=D}^{d} \Delta w_k}$$

idiosyncratic comp.
starting one day before execution

systematic component : $\boxed{\beta(\log(I_d) - \log(I_{D-1}))}$

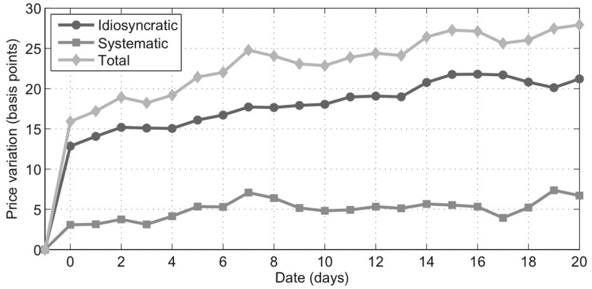


Figure 3.8. Post-execution profile relative to close price on the day before execution. Idiosyncratic component + systematic component = total component.

- not yet debiased post execution profiles

day 1 market impact jump visible for both comp.

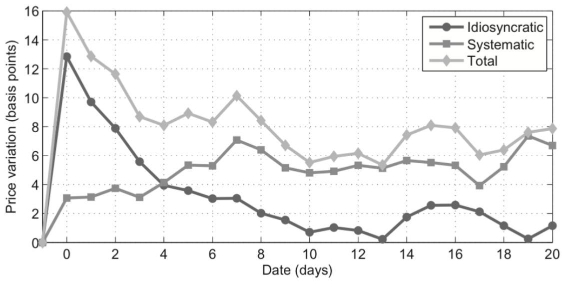


Figure 3.9. Post-execution profile without the impact of other metaorders. Price moves are considered relatively to close price the day before execution. Idiosyncratic component + systematic component = total component.

- debiased post execution profile.
- market impact day 1 removed.
- permanent market impact that remains day 20 entirely explained by systematic component

conclusion:
After removing temp market impact, corr, systematic component,

no remaining permanent market impact due to meta order itself

3.3 PFP and orderbook dynamics

· most trading takes place on centralized limit order book (CLOB)
[opposite of quote systems RFQ]

Attractiveness of limit order books

- anony and provide multilateral trading
- fast and synchronized, support fragmentation and competition between trading platforms

How can practitioners extract info from orderbook dynamics?

3.3.1 info reaching orderbooks

1. At beginning of the day orderbook is empty.
2. Traders & MM send orders if not matched with resting order
3. Matched if compatible price, generates transaction outputs 3 messages

   → public trade
   → 2 private trades

localization of datacenter

Members of trading facility connected to server by:

- private connection : send orders to matching engine
- public channel: see transaction & state of orderbook

Messages

· trader can insert, modify, cancel an order, add tags

the messages can
- remove liquidity via cancelation/transaction
- add liquidity via insertion

Transactions can only occur at first limits (top lowest buy
highest sell

Studying orderbook dynamics is about understanding
- insert
- cancel
- trade

are influenced by current state or by recent past of the order book

3.3.2 Understanding via conditioning

- conditioning for orderbooks focus on probability limit orders touching the first limit only and when size is large

3 trends of conditioning

- zero intelligence: few conditioning, focus on identifying dist. of variables of interest
- game theoretic: explain behavior of participants with accuracy and try to deduce laws for probabilities of each type of order
- empirical: estimate empirically transitions

Poisson modeling

- point process: probability of occurance of an event

λ: intensity of event

$\mathbb{E}(\Delta N)$: expectation see inverse of counting of event N during Δt

$$\boxed{\lambda = \lim_{\Delta t \to 0} \frac{\mathbb{E}(\Delta N)}{\Delta t}}$$

Avg. number of events observed during next ΔT : $\boxed{\lambda \times \Delta T}$

Analyze model orderbook dynamics

- isolate 3 intensities:
  - λ^+ limit order
  - λ^c cancellation
  - λ^x transaction

Plain homogenous Poisson modeling

homogenous Poisson framework:

- liquidity consuming or providing independent of state orderbook

Estimating λ: (Naive)

- Starting at t observe nature of next event,
- record timestamp in τ^+, τ^c, τ^x
- observe $\tau - t$ observations, average out for $1/\lambda$

simulate randomly dynamics of an orderbook and look at distributions of queue sizes

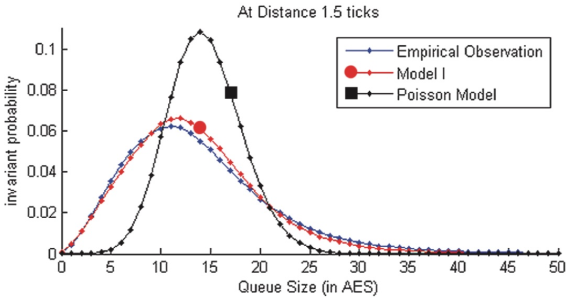


Figure 3.11. Asymptotic distribution of the size of the second queue in three cases: Empirical and "Model I" of [Huang *et al.*, 2015] are very close, the Poisson model (upper dark line), is very different. *x*-axis is observed queue size in AES, *y*-axis is the probability of occurrence of this size.

- without taking into account state of orderbook, modeling of atomic events,

simulation does not succeed in recovering observed config. of queue sizes on the same data

look at huang 2015

Improving the model with conditioning : Model I

Huang 2015 improve modeling by 3 intensities based on size of considered queue

keep record of one intensity for each queue size:

1. start at t
2. record queue size Q_t at t (avg. size order involved in t, c, x)
3. observe event at time τ^c (say cancel c)
4. keep track of $\tau^c - t$ in register associated with size of Q_t^c at t

at end have collection of interevent times $\tau - t$ for each event and possible size

But first and third limits not good with model I.

Models II & III

- improve stuff by conditioning intensities by current size of other queues

- condition on size of queue and neighbors

→ intensity of events on first queue function of current size for first, second, opposite first queue

- first bid: 4
- Second bid: $p := \lambda^c(4,2,1) \times 10$
- best ask: 2

predictive power of imbalance ?

- when 1st queue < opposite queue, it will be consumed
- when 1st queue large, 2nd queue will see more cancellations

predictive power of imbalance → liquidity imbalance at first limit gives good idea of future price direction

Asymptotic behaviors of first 2 queues

A/R Ratio: Arrival or departure ratio of each queue

$\boxed{\{\text{intensity of insertions}\} / \{\text{sum of intensities cancellations \& trades}\}}$

ratio > 1: queue will explode (size never stop increasing)

ratio < 1: queue will disappear

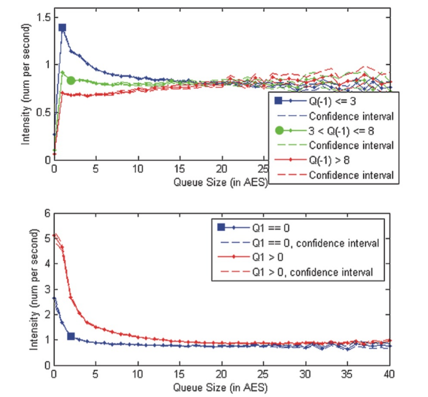


Figure 3.12. Ratio of incoming over outgoing liquidity on first limit (top) and second limit (bottom) as a function of its size (x-axes).

first limit

- first limit always depletes to 0

second limit

- second limit explodes

order book stableness

1. first queue goes to 0 while second queue explodes

2. once first queue disappears, second queue promoted to first queue

3. Starts to deplete, preventing explosion

liquidity dynamics vs. price dynamics

queue reactive model needs one more effect to model price trajectories.

promoted first queue leads to price reversion: imbalance drives price where it was before depletion

use hawke processes, or look at:

- Time to time: when first queue fully depletes
- all market participants accept newly set price as a new 'fair price' (model has to reset order book to randomly chosen state).

3.3.3 conclusion on orderbook dynamics

Queue reactive Model:

- predictive power of imbalance of provided liquidity (limit 1 & 2)

- heterogeneous Poisson process to model occurances of orderbook events (insertion, cancellation, transaction)

- condition intensity of arrival rates based by queue size

- exogenerous reset of orderbook state to prevent intense price mean reversion if tick large: liquidity more meaningful

[stats computed on large tick]

other models exist

- Hawkes process
- use event time (instead of seconds)
- propagator models

3.4 Optimal Trading Methods

3.4.1 Algo trading: Adapting trading style to investor needs

Trading feature has its own benchmark

- stealth trading point of fows (US) → liquidity seekers
- Asia: efficiency in local market important → VWAP, PoV
- Europe: avoid HFT and efficiency interaction with orderbooks → IS

Types of features of trading algos

PoV: percentage of volume

- Medium/large market depth
- long duration position
- follows current market flow, reactive, aggressive, price opportunity driven if max-min large

VWAP/TWAP:

- any market depth
- Hedge order, long duration position, delta hedging of fast evolving inventory

- follows usual market flow, passive, good for unexpected volumes

IS: Implementation shortfall (expected \$ - actual \$)

- medium liquidity depth
- alpha extraction, hedge non-linear position (gamma hedging)
- finish fast, cut losses

liquidity seeker:

- poor fragmented market depth
- alpha extraction, position mounting, split/scheduled order
- price oriented, capture liquidity, stealth (min leakage)

Customization multi feature trading styles

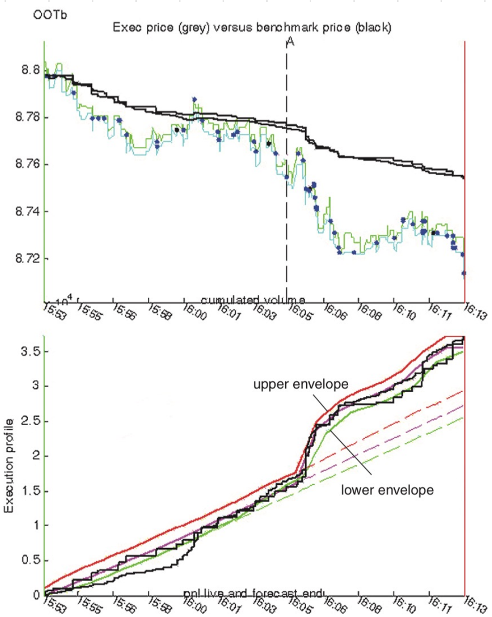


Figure 3.13. Intraday behavior of a VWAP with a min pct (activated at the end of the trading period). Top chart is for the market price (grey and dark lines are for the avg. price and the VWAP, dots for algo trades), bottom chart: The trading envelopes (upper and lower as indicated) and the trading curve (in grey) vs. the market curve (dark).

- combo of VWAP & min participation rate
- hard combo: bang bang control

zero / one policy applied when threshold crossed

- can be gamed/adversely selected

Trading envelopes

- building benchmark based policy is noisy: HFT activity in orderbook & uncertainty on params

optimal trading curve trade rate that has to be respected by algorithm.

- If obtain trades too fast/slow based on curve it must adjust.

inside envelope trading trajectory can't be a priori more efficient than another from risk control perspective

Inside trading envelope combo of liquidity capturing tactics can be plugged in without harming optimality

customization of trading algo process

1. understand desired risk portfolio and injecting needed constraints into envelope building process
2. understand mix of liquidity adapted to investment style, plug into envelopes wrt defined market conditions

3.4.2 liquidity seeking algos are no longer nice to have

SOR to liquidity seeking

- US: trading on official consolidated orderbook. Clearing & settlement have to be same for all trading venues

Impact of trading costs / fees

Pδ : tick size

f' : fees on V'

f : fees on V

P : price p

$f' = f - \phi$

Q : quantity

$$\boxed{p(1+\delta)(1+f-\phi)\,Q < p(1+f)Q}$$

equivilent of having tick size small enough:

$$\boxed{\delta < \left(1 - \phi/_{1+f}\right)^{-1} - 1 \simeq \phi}$$

tick size should be as small as fee difference

- tick size up to which it is worthwhile to use " worse price but better fees"

$\rho = \phi / f$

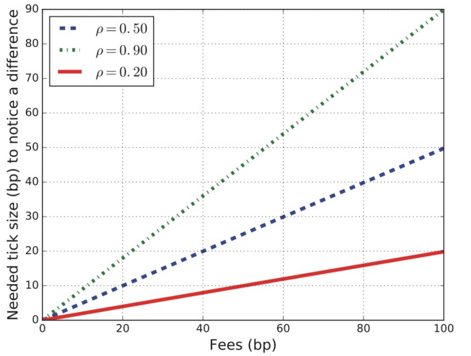


Figure 3.16. Relation between the fees on two venues and the corresponding critical tick size value. For a given value of fees on one venue (x-axis), the y-value is the maximum tick size giving advantage to another venue improving the fees by ρ percent (cf legend).

Naive remarks:

for liquidity providers → most rewarding to post first on a cheap / rewarding venue rather than on an expensive one

⇓

expect to see limit orders going first on cheapest venue

- smaller market orders on cheaper venue (if too large use up first limit)

Seeking optimal liquidity capturing scheme

- liquidity seeking take place on HF to compensate fragmentation

Dealing with marketable order (buy order price > best ask price)

- probability hidden liquidity in orderbook
- probability HFT took advantage into of market order to cancel liquidity providing orders

(limit orders have a lot of phenomena to take into account)

Passive split

Table 3.4. A stylized example of European orderbook on the bid side.

	Price	Euronext	Chi-X	Turquoise
	101.00	50	100	75
(Ask)	100.00	75		25
	...	...	...	...
(Bid)	97.00	50		25
	96.00	75	20	75
	95.00	325	80	100

If i want to insert order at 95:

450 ahead   120 ahead   200 ahead

Allocate less quantity on venues with more shares with better position in queue : Chi-X then Turquoise then Euronext

Ask side: market participants more aggressive on

Euronext → Turquoise → Chi X

immediate selling flow should also be taken into account

Resting time

$$\boxed{T_n(D) = \frac{D + Q_n}{r_n}}$$

D: size D
r_n: sell rate
Q_n: number shares better priority

Split 500 shares:

$500 \times f_1$
$500 \times f_2$
$500 \times f_3$

$\sum_{n=1}^{3} f_n = 1$

$T_n(D f_n)$ same on all venues

minimum waiting time

$$\boxed{T^* = \frac{D - \sum_n Q_n}{\sum_n r_n}}$$

$$\boxed{f_n = \frac{T^* r_n - Q_n}{D}}$$

Table 3.5. Some results for the passive posting example.

Variable	Euronext	Chi-X	Turquoise
$\tau(500)$	95 s	120 s	140 s
τ^*	62.5 s	62.5 s	62.5 s
f_n	35.0%	42.5%	22.5%
$f_n \cdot D$	175	212	113

Building a liquidity seeker

liquidity seekers take care of:

- removing liquidity optimally using estimates of hidden liquidity
- post limit orders to end order as fast as possible

- liquidity in pools are clustered