A Multiclass Queueing Model of Limit Order Book Dynamics

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: `ciamac@gsb.columbia.edu`

Joint work with Costis Maglaras.

Algorithmic Trading

Algorithmic trading of a large order is typically divided into three steps:

Algorithmic Trading

Algorithmic trading of a large order is typically divided into three steps:

- **Trade scheduling:** splits parent order into ~ 5 min “slices”
  relevant time-scale: minutes-hours
  schedule follows user selected “strategy” (VWAP, POV, IS, …)
  reflects urgency, “alpha,” risk/return tradeoff
  schedule updated during execution to reflect price/liquidity/…

Algorithmic Trading

Algorithmic trading of a large order is typically divided into three steps:

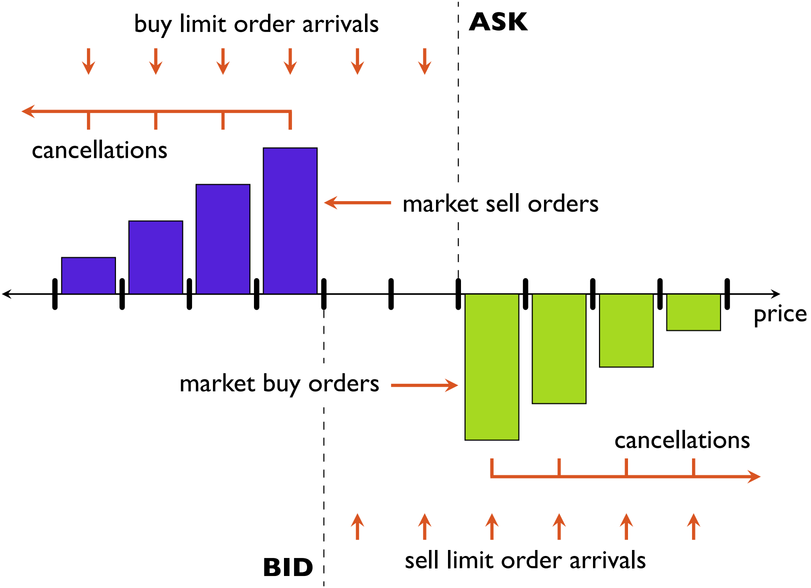
- **Trade scheduling:** splits parent order into ~ 5 min "slices"
  relevant time-scale: minutes-hours
  schedule follows user selected "strategy" (VWAP, POV, IS, …)
  reflects urgency, "alpha," risk/return tradeoff
  schedule updated during execution to reflect price/liquidity/…
- **Optimal execution of a slice:** further divides slice into child orders
  relevant time-scale: seconds–minutes
  strategy optimizes pricing and placing of orders in the limit order book
  execution adjusts to speed of LOB dynamics, price momentum, ...

Algorithmic Trading

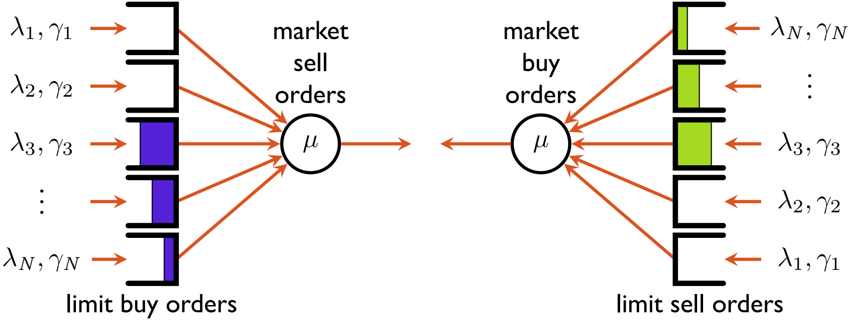
Algorithmic trading of a large order is typically divided into three steps:

- **Trade scheduling:** splits parent order into ~ 5 min "slices"
  relevant time-scale: minutes-hours
  schedule follows user selected "strategy" (VWAP, POV, IS, $\ldots$)
  reflects urgency, "alpha," risk/return tradeoff
  schedule updated during execution to reflect price/liquidity/$\ldots$
- **Optimal execution of a slice:** further divides slice into child orders
  relevant time-scale: seconds–minutes
  strategy optimizes pricing and placing of orders in the limit order book
  execution adjusts to speed of LOB dynamics, price momentum, ...
- **Order routing:** decides where to send each child order
  relevant time-scale: ~ 1–50 ms
  optimizes fee/rebate tradeoff, liquidity/price, latency, etc.

The Limit Order Book



LOB as Coupled Multi-Class Priority Queues



Arrival rates into two sides of book are coupled

- Arrival rates for buy orders depend on lowest non-empty queue with sell orders (ASK)
- Arrival rates for sell orders depend on highest non-empty queue with buy orders (BID)
- More complicated dependencies could be considered (e.g., distance from best bid/ask, queue length, time-of-day, etc.)

Questions

- **Stochastic analysis of multi-class priority queues**
  steady state characterization of coupled queues
  characterization of dynamics of price process
  fluid / diffusion approximations of queueing dynamics

- **Optimal execution**
  how to optimize order placement in LOB
  how to determine rate of trading
  how to estimate queueing delays
  design of market-making strategies that exploit LOB dynamics & transient behavior

- **Market impact modeling**
  microstructure-based model of market impact

Questions

- **Stochastic analysis of multi-class priority queues**
  steady state characterization of coupled queues
  characterization of dynamics of price process
  fluid / diffusion approximations of queueing dynamics

- **Optimal execution**
  how to optimize order placement in LOB
  how to determine rate of trading
  how to estimate queueing delays
  design of market-making strategies that exploit LOB dynamics & transient behavior

- **Market impact modeling**
  microstructure-based model of market impact

Related Literature

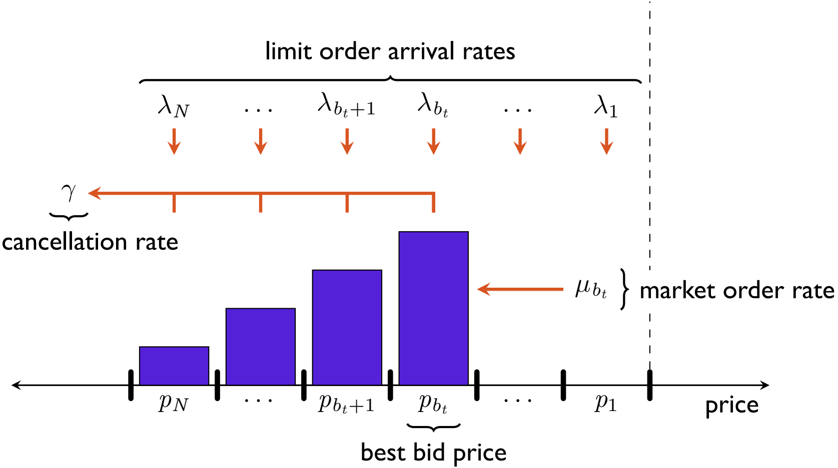
- Market microstructure: Kyle; Glosten-Milgrom; Glosten; …
- Empirical analysis of limit order books: Bouchaud et al.; Hollifield et al.; Smith et al.; …
- Econo-physics: Farmer (+ several coauthors); …
- Transaction cost modeling and estimation: Madhavan; Dufour & Engle; Holthausen et al.; Huberman & Stanzl; Almgren et al.; Gatheral; …
- Optimal execution / trade scheduling: Bertsimas & Lo; Almgren & Chriss; …
- Dynamic models and optimal execution in LOB: Obizhaeva & Wang; Cont et al.; Rosu; Alfonsi et al.; Foucault et al.; Parlour; …
- Stochastic models of multi-class queueing networks

One-sided LOB Fluid Model

- **Fluid model:** continuous & deterministic arrivals and cancellations

One-sided LOB Fluid Model

- **Fluid model:** continuous & deterministic arrivals and cancellations



One-sided LOB Fluid Model: Summary

- $Q_i(t)$ is the quantity of limit orders at price level p_i

$$p_N \leq \cdots \leq p_2 \leq p_1$$

- Limit orders arrive into queue i with rate $\lambda_i > 0$

- Market sell orders arrive to the system with rate μ_{b_t}, dependent on the *best bid queue* b_t

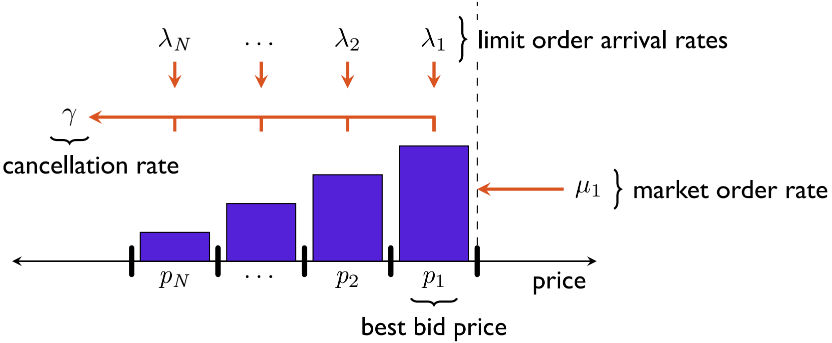
$$\mu_N \leq \cdots \leq \mu_2 \leq \mu_1$$

- Market orders hit resting limit orders according to price/time priority

queue $1 >$ queue $2 > \cdots >$ queue N; FIFO within each queue

- Orders individually cancelled at rate γ

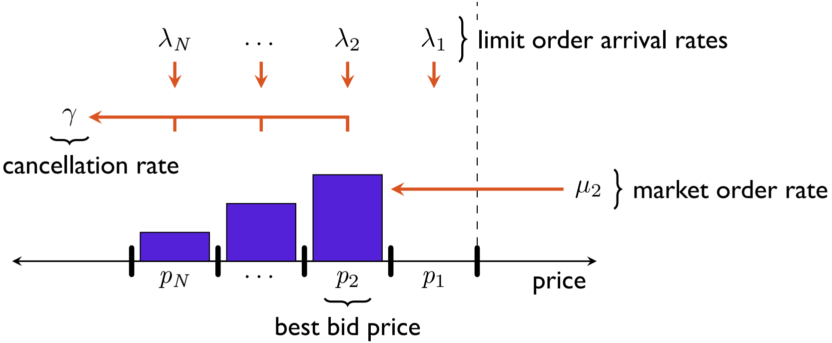
LOB Fluid Model Dynamics



For $0 \le t \le T_1$:

- $\dot{Q}_1(t) = \lambda_1 - \mu_1 - \gamma Q_1(t), \quad \dot{Q}_i(t) = \lambda_i - \gamma Q_i(t),\ i \ge 2$
- Until time $T_1 = \frac{1}{\gamma} \log \left(1 + \frac{\gamma}{\mu_1 - \lambda_1} Q_1(0)\right)$, when $Q_1(T_1) = 0$

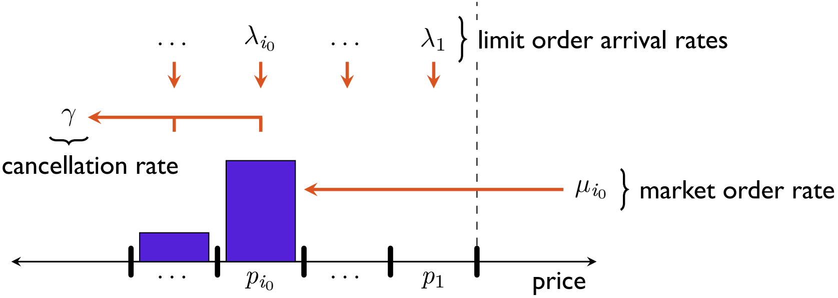
LOB Fluid Model Dynamics



For $T_1 \leq t \leq T_2$:

- $Q_1(t) = 0, \quad \dot{Q}_2(t) = \lambda_2 - (\mu_2 - \lambda_1) - \gamma Q_2(t), \quad \dot{Q}_i(t) = \lambda_i - \gamma Q_i(t),\ i \geq 3$
- Until time T_2 when $Q_2(T_2) = 0$
- ...

LOB Fluid Model Steady State



$$i_0 = \text{ min } \left\{ i \ : \ \Lambda_i \triangleq \textstyle\sum_{j \leq i} \lambda_j \geq \mu_i \right\}$$

$$Q(t) \to q^* \text{ as } t \to \infty, \text{ where } q_i^* = \begin{cases} 0 & \text{for } 1 \leq i < i_0 \\ \frac{\Lambda_{i_0} - \mu_{i_0}}{\gamma} & \text{for } i = i_0 \\ \frac{\lambda_i}{\gamma} & \text{for } i_0 < i \leq N \end{cases}$$

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$
\begin{aligned}
p^*(C, T) \triangleq \text{minimize} \quad & \text{(average price)} \\
\text{subject to} \quad & \text{(total time)} \leq T \\
& \text{(total quantity)} = C
\end{aligned}
$$

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$\begin{aligned} p^*(C,T) \triangleq \text{minimize} \quad & \text{(average price)} \\ \text{subject to} \quad & \text{(total time)} \leq T \\ & \text{(total quantity)} = C \end{aligned}$$

- purchase C total shares given a target average price $\bar{p}$ in minimum time

$$\begin{aligned} T^*(C,\bar{p}) \triangleq \text{minimize} \quad & \text{(total time)} \\ \text{subject to} \quad & \text{(average price)} \leq \bar{p} \\ & \text{(total quantity)} = C \end{aligned}$$

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$
\begin{aligned}
p^*(C,T) \triangleq \text{minimize} \quad & \text{(average price)} \\
\text{subject to} \quad & \text{(total time)} \leq T \\
& \text{(total quantity)} = C
\end{aligned}
$$

- purchase C total shares given a target average price $\bar{p}$ in minimum time

$$
\begin{aligned}
T^*(C,\bar{p}) \triangleq \text{minimize} \quad & \text{(total time)} \\
\text{subject to} \quad & \text{(average price)} \leq \bar{p} \\
& \text{(total quantity)} = C
\end{aligned}
$$

Efficient frontier:

$$
\mathcal{E} \triangleq \big\{(C,T,p) \;:\; p = p^*(C,T)\big\} \triangleq \big\{(C,T,p) \;:\; T = T^*(C,p)\big\}
$$

Optimal Limit Order Placement

Problem:

- purchase C total shares given a target average price $\bar{p}$ in minimum time

$$
\begin{aligned}
T^*(C, \bar{p}) \triangleq \text{minimize} \quad & \text{(total time)} \\
\text{subject to} \quad & \text{(average price)} \leq \bar{p} \\
& \text{(total quantity)} = C
\end{aligned}
$$

Controls:

$$
\begin{aligned}
L_i(t) &= \text{cumulative \# of orders placed in queue } i \text{ in } [0, t] \\
&= \text{RCLL process}
\end{aligned}
$$

Proposition. The optimal control places all orders at time $t = 0$.

Intuition: place orders at $t = 0$ to gain priority over future arriving limit orders. Exploit time priority to place limit orders in lower priority queues (better prices).

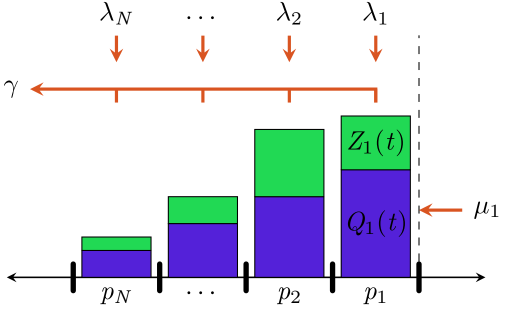
Optimal Limit Order Placement

Notation:

$$z_i = \text{initial \# of investor orders in queue } i$$

$$Z_i(t) = \text{remaining \# of unexecuted investor orders in queue } i$$

$$Q_i(t) = \text{\# of orders from other investors}$$



Optimal Limit Order Placement

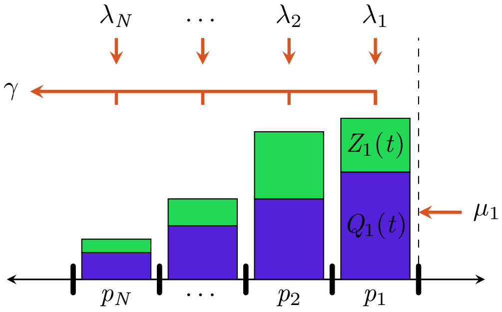
Notation:

$$z_i = \text{initial \# of investor orders in queue } i$$

$$Z_i(t) = \text{remaining \# of unexecuted investor orders in queue } i$$

$$Q_i(t) = \text{\# of orders from other investors}$$

$$\tau^{\ell}(z) \triangleq \inf \{t \geq 0 \ : \ Q_{i-1}(t) = 0, \ Z_i(t) = 0, \ i \leq \ell\}$$



Optimal Limit Order Placement

Notation:

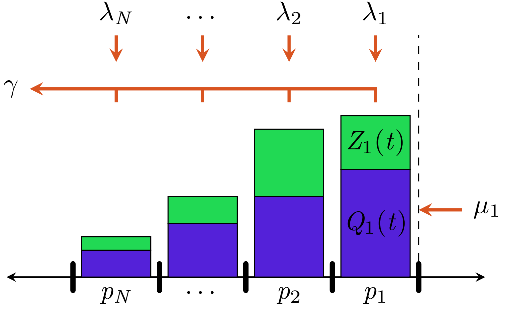
$$z_i = \text{initial \# of investor orders in queue } i$$

$$Z_i(t) = \text{remaining \# of unexecuted investor orders in queue } i$$

$$Q_i(t) = \text{\# of orders from other investors}$$

$$\tau^\ell(z) \triangleq \inf\{t \geq 0 \;:\; Q_{i-1}(t) = 0,\; Z_i(t) = 0,\; i \leq \ell\}$$

$$\tau^*(z) \triangleq \tau^{\ell(z)}(z), \quad \ell(z) \triangleq \max\{i \;:\; z_i > 0\}$$



Optimal Limit Order Placement

Notation:

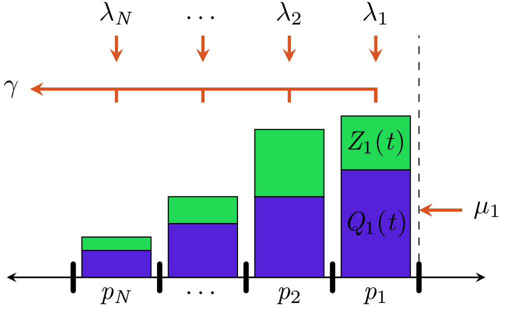
$$z_i = \text{initial \# of investor orders in queue } i$$

$$Z_i(t) = \text{remaining \# of unexecuted investor orders in queue } i$$

$$Q_i(t) = \text{\# of orders from other investors}$$

$$\tau^\ell(z) \triangleq \text{inf}\,\{t \geq 0 \;:\; Q_{i-1}(t) = 0,\; Z_i(t) = 0,\; i \leq \ell\}$$

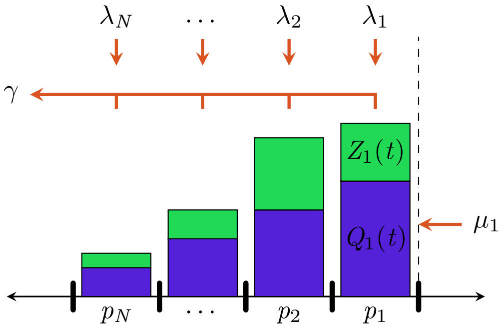
$$\tau^*(z) \triangleq \tau^{\ell(z)}(z), \qquad \ell(z) \triangleq \max\,\{i \;:\; z_i > 0\}$$



Problem:

$$\begin{array}{ll} \text{minimize} & \tau^*(z) \\ \text{subject to} & \mathbf{1}^\top z = C \\ & \mathbf{1}^\top p \leq \bar{p}C \\ & z \geq \mathbf{0} \end{array}$$

Optimal Limit Order Placement

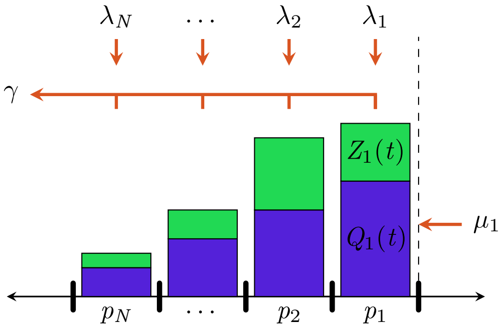


Problem:

$$
\begin{array}{ll}
\text{minimize} & \tau^\ell(z) \\
\text{subject to} & \mathbf{1}^\top z = C \\
& \mathbf{1}^\top p \leq \bar{p} C \\
& z \geq \mathbf{0}
\end{array}
$$

[fixed ℓ]

Optimal Limit Order Placement



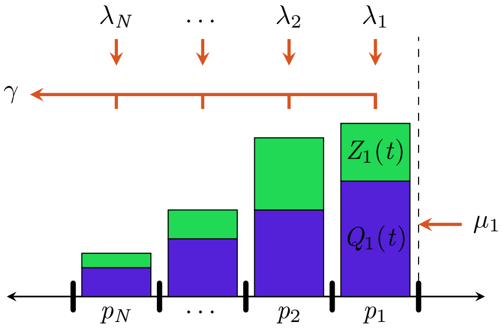
Problem:

$$\begin{array}{ll} \text{minimize} & \tau^\ell(z) \\ \text{subject to} & \mathbf{1}^\top z = C \\ & \mathbf{1}^\top p \le \bar{p} C \\ & z \ge \mathbf{0} \end{array}$$

[fixed ℓ]

Lemma. If $Q_i(0) \le \lambda_i/\gamma$, then $\tau^\ell(z)$ is a concave in z.

Optimal Limit Order Placement



Problem:

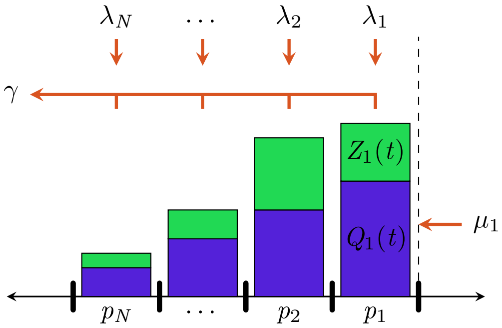
$$
\begin{array}{ll}
\text{minimize} & \tau^\ell(z) \\
\text{subject to} & \mathbf{1}^\top z = C \\
& \mathbf{1}^\top p \le \bar{p}C \\
& z \ge \mathbf{0}
\end{array}
$$

[fixed ℓ]

Lemma. If $Q_i(0) \le \lambda_i/\gamma$, then $\tau^\ell(z)$ is a concave in z.

Theorem. $\|z^{(\ell)}\|_0 \le 2$, i.e., optimal to place orders in at most two queues

Optimal Limit Order Placement

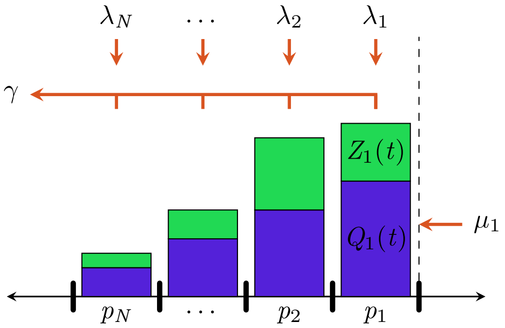


Problem:

$$\begin{array}{ll} \text{minimize} & \tau^*(z) \\ \text{subject to} & \mathbf{1}^\top z = C \\ & \mathbf{1}^\top p \leq \bar{p} C \\ & z \geq \mathbf{0} \end{array}$$

Note that z^* also minimizes $\tau^{\ell(z^*)}(\cdot)$

Optimal Limit Order Placement



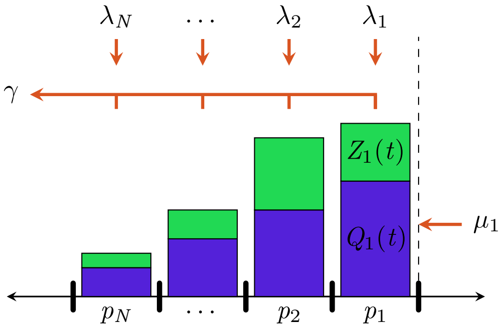
Problem:

$$
\begin{array}{ll}
\text{minimize} & \tau^*(z) \\
\text{subject to} & \mathbf{1}^\top z = C \\
& \mathbf{1}^\top p \leq \bar{p} C \\
& z \geq \mathbf{0}
\end{array}
$$

Note that z^* also minimizes $\tau^{\ell(z^*)}(\cdot)$

Corollary. $\|z^*\|_0 \leq 2$, i.e., optimal to place orders in at most two queues

Optimal Limit Order Placement



Problem:

$$\begin{array}{ll} \text{minimize} & \tau^*(z) \\ \text{subject to} & \mathbf{1}^\top z = C \\ & \mathbf{1}^\top p \leq \bar{p} C \\ & z \geq \mathbf{0} \end{array}$$

Note that z^* also minimizes $\tau^{\ell(z^*)}(\cdot)$

Corollary. $\|z^*\|_0 \leq 2$, i.e., optimal to place orders in at most two queues

Corollary. Optimal to place orders in two adjacent queues

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$p^*(C, T) \triangleq \begin{array}{ll} \text{minimize} & \text{(average price)} \\ \text{subject to} & \text{(total time)} \leq T \\ & \text{(total quantity)} = C \end{array}$$

Optimal Policy: $[Q(0) = \mathbf{0}]$

Recall $i_0 \triangleq \min \{i : \Lambda_i \triangleq \sum_{j \leq i} \lambda_j \geq \mu_i\}$. Define $\bar{C}_i \triangleq (\mu_i - \Lambda_{i-1})^+ T$.

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$p^*(C,T) \triangleq \begin{array}{ll} \text{minimize} & \text{(average price)} \\ \text{subject to} & \text{(total time)} \leq T \\ & \text{(total quantity)} = C \end{array}$$

Optimal Policy: $[Q(0) = \mathbf{0}]$

Recall $i_0 \triangleq \min \{ i \;:\; \Lambda_i \triangleq \sum_{j \leq i} \lambda_j \geq \mu_i \}$. Define $\bar{C}_i \triangleq (\mu_i - \Lambda_{i-1})^+ T$.

Then:

- $0 \leq C \leq \bar{C}_{i_0} \quad \Longrightarrow \quad z_j^* = 0$ unless $j = i_0$

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$
\begin{aligned}
p^*(C, T) \triangleq \text{minimize} \quad & \text{(average price)} \\
\text{subject to} \quad & \text{(total time)} \leq T \\
& \text{(total quantity)} = C
\end{aligned}
$$

Optimal Policy: $[Q(0) = \mathbf{0}]$

Recall $i_0 \triangleq \min\ \{i\ :\ \Lambda_i \triangleq \sum_{j \leq i} \lambda_j \geq \mu_i\}$. Define $\bar{C}_i \triangleq (\mu_i - \Lambda_{i-1})^+ T$.

Then:

- $0 \leq C \leq \bar{C}_{i_0} \quad \Longrightarrow \quad z_j^* = 0$ unless $j = i_0$
- $\bar{C}_i \leq C \leq \bar{C}_{i-1} \quad \Longrightarrow \quad z_j^* = 0$ unless $j \in \{i-1, i\}$

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$p^*(C, T) \triangleq \text{minimize} \quad \text{(average price)}$$
$$\text{subject to} \quad \text{(total time)} \leq T$$
$$\text{(total quantity)} = C$$

Optimal Policy: $[Q(0) = \mathbf{0}]$

Recall $i_0 \triangleq \min \{i : \Lambda_i \triangleq \sum_{j \leq i} \lambda_j \geq \mu_i\}$. Define $\bar{C}_i \triangleq (\mu_i - \Lambda_{i-1})^+ T$.

Then:

- $0 \leq C \leq \bar{C}_{i_0} \implies z_j^* = 0$ unless $j = i_0$
- $\bar{C}_i \leq C \leq \bar{C}_{i-1} \implies z_j^* = 0$ unless $j \in \{i-1, i\}$
- $C > \bar{C}_1 \triangleq \mu_1 T \implies$ infeasible! (must use market orders)

Optimal Limit Order Placement

Problem:

- purchase C total shares in time T for lowest possible average price

$$p^*(C,T) \triangleq \begin{array}{ll} \text{minimize} & \text{(average price)} \\ \text{subject to} & \text{(total time)} \leq T \\ & \text{(total quantity)} = C \end{array}$$

Optimal Policy: $[Q(0) = \mathbf{0}]$

Recall $i_0 \triangleq$ min $\{i \;:\; \Lambda_i \triangleq \sum_{j \leq i} \lambda_j \geq \mu_i\}$. Define $\bar{C}_i \triangleq (\mu_i - \Lambda_{i-1})^+ T$.

Then:

- $0 \leq C \leq \bar{C}_{i_0} \quad \Longrightarrow \quad z_j^* = 0$ unless $j = i_0$
- $\bar{C}_i \leq C \leq \bar{C}_{i-1} \quad \Longrightarrow \quad z_j^* = 0$ unless $j \in \{i-1, i\}$
- $C > \bar{C}_1 \triangleq \mu_1 T \quad \Longrightarrow$ infeasible! (must use market orders)

Easy to compute recursively the $\bar{C}_i$ thresholds starting from arbitrary $Q(0)$

Efficient Frontier

Price grid is uniformly spaced $p_i \triangleq p_1 - (i-1)\delta$ ($\delta =$ tick size)

$$
\begin{aligned}
\bar{C}_1 &= \mu_1 T && \text{shares at } p_1 \\
\bar{C}_2 &= (\mu_2 - \Lambda_1)\, T && \text{shares at } p_2 \\
&\vdots \\
\bar{C}_{i_0} &= (\mu_{i_0} - \Lambda_{i_0 - 1})^+ T && \text{shares at } p_{i_0}
\end{aligned}
$$

$$\lambda_i > 0 \quad \Longrightarrow \quad \bar{C}_1 > \bar{C}_2 > \cdots > \bar{C}_{i_0}$$

Efficient Frontier

Price grid is uniformly spaced $p_i \triangleq p_1 - (i-1)\delta$ ($\delta =$ tick size)

$$\begin{aligned} \bar{C}_1 &= \mu_1 T && \text{shares at } p_1 \\ \bar{C}_2 &= (\mu_2 - \Lambda_1) T && \text{shares at } p_2 \\ &\vdots \\ \bar{C}_{i_0} &= (\mu_{i_0} - \Lambda_{i_0-1})^+ T && \text{shares at } p_{i_0} \end{aligned}$$

$$\lambda_i > 0 \quad \Longrightarrow \quad \bar{C}_1 > \bar{C}_2 > \cdots > \bar{C}_{i_0}$$

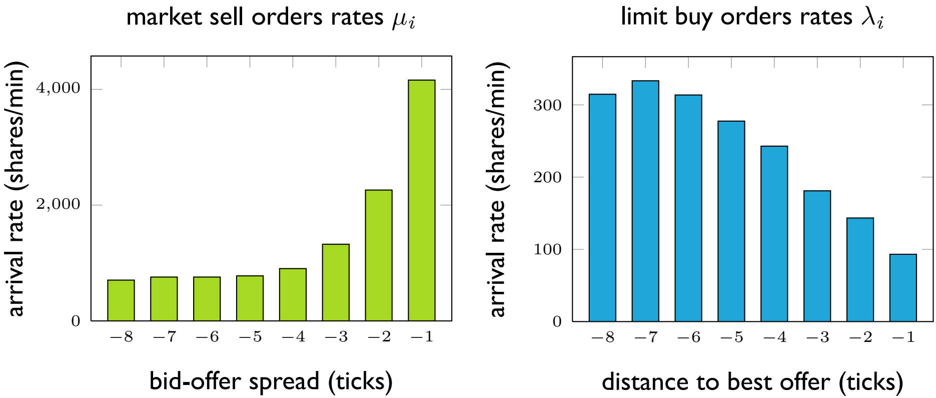
Shape of Efficient Frontier:

- Determines price impact as a function of trading rate C/T
- $\lambda_1 = \lambda_2 = \cdots = \lambda_{i_0-1} \quad \Longrightarrow$ "linear" cost-per-share
- $\lambda_1 > \lambda_2 > \cdots > \lambda_{i_0-1} \quad \Longrightarrow$ "*sub-linear*" cost-per-share

Numerical Example

First Solar, Inc. (ticker: FSLR.Q) $3/22/2010$

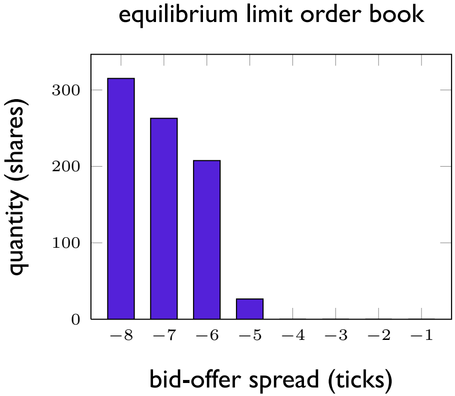
close price $= \$109.36$ average volume $= 2{,}699{,}000$ shares

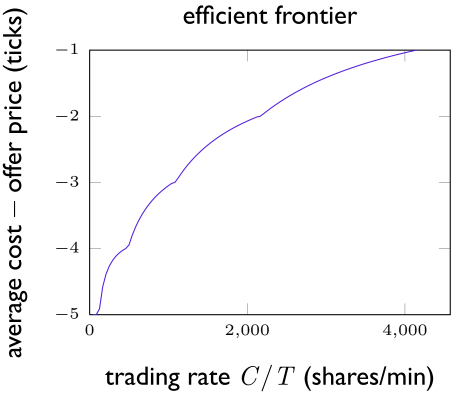


cancellation rate $\gamma = 6.03$ (min^{-1})

Numerical Example

First Solar, Inc. (ticker: FSLR.Q)





Implementation Issues

How to implement the fluid policy in the original, stochastic system?

- discrete review / model predictive control
- approximate dynamic programming

Implementation Issues

How to implement the fluid policy in the original, stochastic system?

- discrete review / model predictive control
- approximate dynamic programming

In our model, all orders are places at time $t = 0$, this may pose problems:

- information leakage
- execution price not averaged over time horizon (adverse selection)

Implementation Issues

How to implement the fluid policy in the original, stochastic system?

- discrete review / model predictive control
- approximate dynamic programming

In our model, all orders are places at time $t = 0$, this may pose problems:

- information leakage
- execution price not averaged over time horizon (adverse selection)

Optimal order placement suggests:

- Break up slice into n child orders to be executed "uniformly" over T (roughly every T/n time units)
- The price of each child order depends on:
  (a) state of LOB
  (b) speed of LOB
  (c) size of the child order $= C/n$
  (d) duration $= T/n$

Precise formulations possible ...

Model Shortcomings

- Non-stationarity

Model Shortcomings

- Non-stationarity
- Order placement influences arrival rates

Model Shortcomings

- Non-stationarity
- Order placement influences arrival rates
- Incorporate market orders and other side of LOB

Model Shortcomings

- Non-stationarity
- Order placement influences arrival rates
- Incorporate market orders and other side of LOB
- Not relevant for illiquid names

Model Shortcomings

- Non-stationarity
- Order placement influences arrival rates
- Incorporate market orders and other side of LOB
- Not relevant for illiquid names
- Incorporate multiple LOB's / order routing
  (≈ 5 LOB's matter in practice)

Conclusion

Fluid models are a useful characterization of LOB dynamics

Optimal Execution:

- Tractable order placement problem
- Allow estimation of queueing delays
- Determine optimal rate of trading

Market Impact Modeling:

- A microstructure-based model of market impact
- Exponent directly estimated from rate of events in LOB
- Suggests different exponents for limit orders versus market orders?