Market Making with Random Linear Demand And Overnight Inventory Costs

José E. Figueroa-López

Washington University in St. Louis

figueroa-lopez@wustl.edu

USC Math Finance Colloquium
Sept. 21, 2020
Joint work with Agostino Capponi and Chuyi Yu

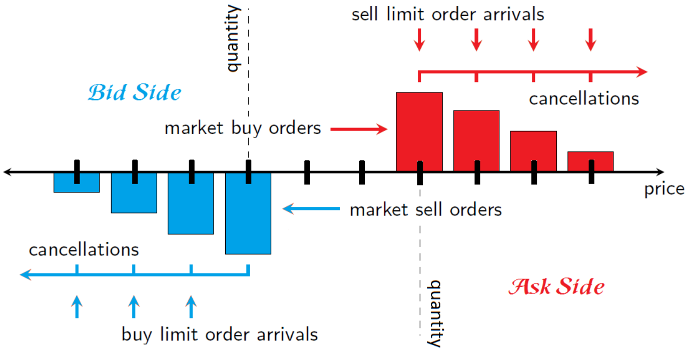
Outline

Limit Order Book (LOB)

A bid (ask) **limit order** (LO) is a request to buy (sell) a prescribed amount of an asset at a desired price.

A **market order** (MO) seeks to buy or sell a certain quantity of the asset immediately at the best available prices of the outstanding limit orders.

A **limit order book** lists all ask and bid limit orders, waiting to be matched with MOs, with their specific price and the corresponding volume.



What is Market Making?

- A Market Maker (MM) provides liquidity to the market by **continually submiting simultaneous bid and ask limit order**;
- If both (bid and ask) orders get executed or lifted, the MM gains the quoted spread, which will be at least as large as the market spread;
- There is a natural tradeoff when placing orders: deeper orders entail higher profits, if executed, but are less likely to be executed;
- In addition, there is also considerable **inventory risk**: the MM may accumulate large inventory that must be cleared aggressively at the end of day at suboptimal prices or even worst at market prices.

Selective Literature on Market Making I

Avellaneda and Stoikov (2008):

- Maximization of the exponential utility from terminal trading cashflow W_T and residual inventory I_T liquidation:

$$E[-e^{-\gamma(W_T+I_T S_T)}];$$

- Optimize bid/ask LO placements $S_t \pm L_t^{\pm}$ of **one unit share** under a Brownian midprice dynamics $S_t = \sigma B_t$ and Poisson MOs arrival times with lifting rates $\pi_S(L^-)$ and $\pi_B(L^+)$.

Selective Literature on Market Making II

Cartea and Jaimungal (2015):

- Optimize terminal trading cashflow W_T and residual inventory liquidation $I_T S_T$, while penalizing terminal and running inventory:

$$\sup_{(L_s^{\pm})_{t\leq s\leq T}} E\left[W_T + I_T S_T - \lambda I_T^2 - \phi \int_t^T I_s^2 ds \middle| W_t = w, I_t = i, S_t = s\right]$$

- Still optimize bid/ask LO placements $S_t \pm L_t^{\pm}$ of one unit, which is lifted with probability $e^{-\kappa_{\pm} L_t^{\pm}}$ when a Poissonian MO arrives;
- Midprice follows a Brownian dynamics between MOs arrival times but with i.i.d. jumps at those MO times (due to price impact of MOs);

Selective Literature on Market Making III

Adrian, T., Capponi, A., Vogt, E., & Zhang, H. (2020):

- Optimize

$$\sup_{(L_s^{\pm})_{t \le s \le T}} E\left[W_T + I_T S_T - \lambda I_T^2 \middle| W_t = w, S_t = s, I_t = i \right],$$

  but allows placement of orders of arbitrary volume;

- Use of exogenously specified deterministic linear demand and supply functions to determine number of lifted shares at the arrival of each MO;
- Brownian midprice dynamics and indep. Poissonian arrival times of MOs;

Market Making Model I

- Finite time horizon $[0, T]$ with discrete action times $0 = t_0 < t_1 < \cdots < t_N = T$;
- At time t_k, the market maker places a bid and ask LO with price levels

$$b_k = S_{t_k} - L_k^-, \qquad a_k = S_{t_k} + L_k^+, \qquad k = 0, \dots, N-1,$$

  relative to the stock's fundamental price[1] S_{t_k} at time t_k;.
- Arrivals of market orders occur according to the probabilities

$$\mathbb{P}(\mathbf{1}_k^+ = j^+, \mathbf{1}_k^- = j^- \big| \mathcal{F}_{t_{k-1}}) = \pi_k(j^+, j^-), \quad j^+, j^- \in \{0, 1\},$$

  where

$$\mathbf{1}_k^+ = \begin{cases} 1, & \text{if one or more buy MOs arrives during } [t_{k-1}, t_k); \\ 0, & \text{otherwise;} \end{cases} \qquad \mathbf{1}_k^- = \begin{cases} 1, & \text{if at least one sell MOs arrives during } [t_{k-1}, t_k); \\ 0, & \text{otherwise;} \end{cases}$$

[1] S_{t_k} is the market maker's opinion of the true price of the stock; often the midprice.

Market Making Model II

Linear Demand/Supply Functions

- The number of filled shares from the MM in the bid side during $[t_k, t_{k+1})$ is

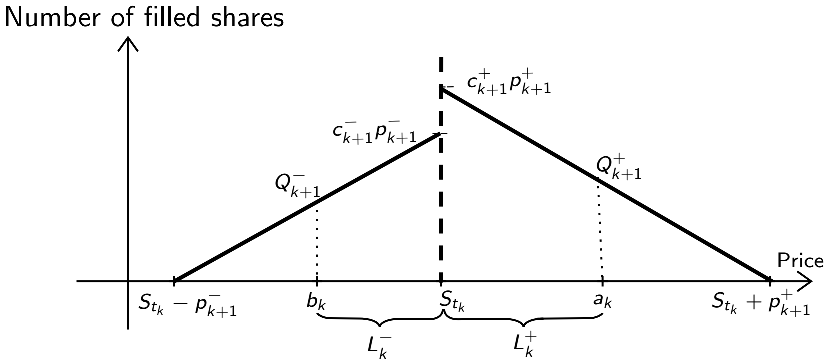
$$\begin{aligned} Q^-_{k+1} &= \mathbf{1}^-_{k+1} c^-_{k+1} \left[b_k - (S_{t_k} - p^-_{k+1}) \right] \\ &= \mathbf{1}^-_{k+1} c^-_{k+1} \left[p^-_{k+1} - L^-_k \right]; \end{aligned}$$

- The number of filled shares from the MM in the ask side during $[t_k, t_{k+1})$ is

$$\begin{aligned} Q^+_{k+1} &= \mathbf{1}^+_{k+1} c^+_{k+1} \left[(S_{t_k} + p^+_{k+1}) - a_k \right] \\ &= \mathbf{1}^+_{k+1} c^+_{k+1} \left[p^+_{k+1} - L^+_k \right]; \end{aligned}$$

- Here, $p^\pm_{k+1}, c^\pm_{k+1} \in \mathcal{F}_{t_{k+1}}$ are positive random variables, unknown to the market maker at time t_k.

Demand/Supply Functions



- $S_{t_k} - p^{-}_{t_{k+1}} =$ lowest price that sell MOs can reach during $[t_k, t_{k+1})$.
- $S_{t_k} + p^{+}_{t_{k+1}} =$ highest price that buy MOs can reach during $[t_k, t_{k+1})$.
- The number of filled shares $Q^{\pm}_{k+1}$ increases as the market maker places her limit orders closer to the mid-price S_{t_k}.

Relationship With Literature

- [Hendershott and Menkveld, 2014] assumes the demand/supply is normally distributed with a linear mean on the spreads;
- [Adrian et. al., 2018] uses deterministic linear demand functions (i.e., $c_k \equiv c$ and $p_k \equiv p$, for some constants c and p);
- Relationship with exponential lifting probabilities (e.g., Cartea and Jaimungal (2015)):
  - Market orders arrive at rate λ;
  - An order placed at $S \pm L^{\pm}$ is lifted with probability $e^{-\kappa_{\pm} L^{\pm}}$
  - The average number of orders filled during $[t_k, t_{k+1})$ is

$$\lambda \times (t_{k+1} - t_k) \times e^{-\kappa_{\pm} L^{\pm}} \approx (1 - \kappa_{\pm} L^{\pm}) \times \lambda \times (t_{k+1} - t_k),$$

which is linear in $L^{\pm}$.

Cash Holding and Inventory Dynamics

▷ Cashflow Holding at time t_{k+1}:

$$\begin{aligned} W_{t_{k+1}} &= W_{t_k} + a_k Q^{+}_{k+1} - b_k Q^{-}_{k+1} \\ &= W_{t_k} + (S_{t_k} + L^{+}_k)\mathbf{1}^{+}_{k+1} c^{+}_{k+1}\left[p^{+}_{k+1} - L^{+}_k\right] \\ &\qquad - (S_{t_k} - L^{-}_k)\mathbf{1}^{-}_{k+1} c^{-}_{k+1}\left[p^{-}_{k+1} - L^{-}_k\right] \end{aligned}$$

▷ Inventory at time t_{k+1}:

$$\begin{aligned} I_{t_{k+1}} &= I_{t_k} - Q^{+}_{k+1} + Q^{-}_{k+1} \\ &= I_{t_k} - \mathbf{1}^{+}_{k+1} c^{+}_{k+1}\left[p^{+}_{k+1} - L^{+}_k\right] \\ &\qquad + \mathbf{1}^{-}_{k+1} c^{-}_{k+1}\left[p^{-}_{k+1} - L^{-}_k\right] \end{aligned}$$

The Optimal Market Making Problem

We study the problem:

$$V_{t_k} = \sup_{(L_\cdot^+, L_\cdot^-) \in \mathcal{A}} \mathbb{E}\left[W_T + S_T I_T - \lambda I_T^2 \middle| \mathcal{F}_{t_k} \right],$$

where $\mathcal{A}$ consists of all adapted processes $L_\cdot^\pm$ and $\lambda > 0$.

Interpretation:

- $S_T I_T - \lambda I_T^2 = I_T(S_T - \lambda I_T)$ is interpreted as the cost for liquidating the inventory I_T at time T;
- That is, $S_T - \lambda I_T$ is the average price per share for liquidating an inventory of size $|I_T|$;
- The offset $\lambda |I_T|$ is due to the price impact of submitting a market order to liquidate $|I_T|$ shares.

Optimal Placement Policy

Assumptions:

- The distribution of $\{(c_k^+, p_k^+, c_k^-, p_k^-)\}_{k=1,\dots,N}$ conditional on $\mathcal{F}_{t_k}$ and $(\mathbf{1}_{k+1}^+, \mathbf{1}_{k+1}^-)$ does not depend on k;
- $\{\pi_k(j^+, j^-)\}_{k=1,\dots,N}$ are deterministic;
- The stock's midprice process $\{S_t\}_{t\geq 0}$ is a martingale:

$$\mathbb{E}(S_{t_{k+1}}|\mathcal{F}_{t_k}) = S_{t_k}, \quad k = 0, \dots, N-1.$$

Notation:

- $\pi_k^\pm := \mathbb{P}(\mathbf{1}_k^\pm = 1|\mathcal{F}_{t_{k-1}})$, probability of a MO arrival during $[t_{k-1}, t_k)$;
- $\mu_c^\pm := \mathbb{E}(c_{k+1}^\pm \,|\, \mathcal{F}_{t_k}, \mathbf{1}_{k+1}^\pm = 1)$
- $\mu_{c^2}^\pm := \mathbb{E}((c_{k+1}^\pm)^2 \,|\, \mathcal{F}_{t_k}, \mathbf{1}_{k+1}^\pm = 1)$
- $\mu_{c^\ell p^m}^\pm := \mathbb{E}\left((c_{k+1}^\pm)^\ell (p_{k+1}^\pm)^m \,\middle|\, \mathcal{F}_{t_k}, \mathbf{1}_{k+1}^\pm = 1\right)$

Optimal Placement Policy (case $\pi(1,1)=0$)

When $\pi(1,1)=0$ (no simultaneous arrival of sell and buy market orders),

$$a_k^* = S_{t_k} + \overbrace{\frac{\mu_c^+ \alpha_{k+1}}{\mu_c^+ - \alpha_{k+1}\mu_{c^2}^+} I_{t_k} + \frac{\mu_{cp}^+ - 2\alpha_{k+1}\mu_{c^2p}^+}{2[\mu_c^+ - \alpha_{k+1}\mu_{c^2}^+]} + \frac{\mu_c^+ h_{k+1}}{2[\mu_c^+ - \alpha_{k+1}\mu_{c^2}^+]}}^{L_k^{*+}}$$

$$b_k^* = S_{t_k} + \underbrace{\frac{\mu_c^- \alpha_{k+1}}{\mu_c^- - \alpha_{k+1}\mu_{c^2}^-} I_{t_k} - \frac{\mu_{cp}^- - 2\alpha_{k+1}\mu_{c^2p}^-}{2[\mu_c^- - \alpha_{k+1}\mu_{c^2}^-]} + \frac{\mu_c^- h_{k+1}}{2[\mu_c^- - \alpha_{k+1}\mu_{c^2}^-]}}_{-L_k^{*-}}$$

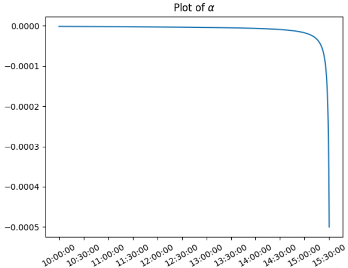
Optimal Placement Policy (case $\pi(1,1)=0$)

Here, $\alpha_N = -\lambda$ and $h_N = 0$, and, for $k = 1, \ldots, N-1$,

$$\alpha_k = \alpha_{k+1} + \sum_{\delta=\pm} \pi^{\delta}_{k+1} \frac{(\alpha_{k+1}\mu^{\delta}_{c})^2}{\mu^{\delta}_{c} - \alpha_{k+1}\mu^{\delta}_{c^2}}$$

$$h_k = h_{k+1} + \alpha_{k+1} \sum_{\delta=\pm} \pi^{\delta}_{k+1} c^{\delta}_{k}$$

$$c^{\pm}_{k} = \frac{(\mu^{\pm}_{c})^2 h_{k+1} \mp \mu^{\pm}_{c}\mu^{\delta}_{cp} \mp 2\alpha_{k+1}\mu^{\pm}_{c^2p}\mu^{\pm}_{c} \pm 2\alpha_{k+1}\mu^{\pm}_{cp}\mu^{\pm}_{c^2}}{\mu^{\pm}_{c} - \alpha_{k+1}\mu^{\pm}_{c^2}}$$



Analysis (case $\pi(1,1)=0$)

Set $p_k^{+*} = a_k^*$ and $p_k^{-*} = b_k^*$:

$$p_k^{\pm *} = S_{t_k} + \frac{\mu_c^{\pm}\alpha_{k+1}}{\mu_c^{\pm} - \alpha_{k+1}\mu_{c^2}^{\pm}} I_{t_k} \pm \frac{\mu_{cp}^{\pm} - 2\alpha_{k+1}\mu_{c^2p}^{\pm}}{2[\mu_c^{\pm} - \alpha_{k+1}\mu_{c^2}^{\pm}]} + \frac{\mu_c^{\pm} h_{k+1}}{2[\mu_c^{\pm} - \alpha_{k+1}\mu_{c^2}^{\pm}]}$$

- The 2nd term is the inventory adjustment, whose coefficient is negative:
  positive (negative) inventory $\Longrightarrow$ lower (higher) price levels
  $\Longrightarrow$ facilitate selling (dampen buying)
- α_k is close to 0 most of time and decreases rapidly to $-\lambda$ near T;
  h_k is close to 0 most of time;
- Most of the time $p_k^{\pm *} \approx S_{t_k} \pm \frac{1}{2}\frac{\mu_{cp}^{\pm}}{\mu_c^{\pm}} = S_{t_k} \pm \frac{1}{2}\mu_p^{\pm} \pm \frac{\mathrm{Cov}(c^{\pm},p^{\pm})}{2\mu_c^{\pm}}$
- The covariance between c and p, and variability c play crucial roles; e.g., the 2nd term can be written as

$$\frac{\alpha_{k+1}\mu_c^{\pm}}{\mu_c^{\pm} - \alpha_{k+1}\mu_{c^2}^{\pm}} I_{t_k} = \frac{\alpha_{k+1}}{1 - \alpha_{k+1}\mu_c^{\pm} - \alpha_{k+1}\mathrm{Var}(c^{\pm})/\mu_c^{\pm}} I_{t_k}$$

Optimal Placement Policy (General Case)

$$p_k^{\pm *} = S_{t_k} + \frac{\beta_k^{\pm}}{\gamma_k}\alpha_{k+1} I_{t_k} \pm \frac{\eta_k^{\pm}}{2\gamma_k} + \frac{\beta_k^{\pm}}{2\gamma_k} h_{k+1},$$

where

$$\begin{aligned}
\beta_k^{\pm} &:= -\pi_{k+1}^{+}\pi_{k+1}^{-}\mu_c^{\pm}(\mu_c^{\mp} - \alpha_{k+1}\mu_{c^2}^{\mp}) - \pi_{k+1}^{\mp}\mu_c^{\pm}(\mu_c^{\mp})^2\alpha_{k+1}\pi_{k+1}(1,1),\\
\gamma_k &:= -\pi_{k+1}^{+}\pi_{k+1}^{-}(\mu_c^{+} - \alpha_{k+1}\mu_{c^2}^{+})(\mu_c^{-} - \alpha_{k+1}\mu_{c^2}^{-})\\
&\quad + \left[\mu_c^{+}\mu_c^{-}\alpha_{k+1}\pi_{k+1}(1,1)\right]^2\\
\eta_k^{\pm} &:= -\pi_{k+1}^{+}\pi_{k+1}^{-}(\mu_c^{\mp} - \alpha_{k+1}\mu_{c^2}^{\mp})(\mu_{cp}^{\pm} - 2\alpha_{k+1}\mu_{c^2p}^{\pm})\\
&\quad - 2\pi_{k+1}^{\mp}\mu_c^{\pm}\mu_{cp}^{\mp}(\mu_c^{\mp} - \alpha_{k+1}\mu_{c^2}^{\mp})\alpha_{k+1}\pi_{k+1}(1,1)\\
&\quad + \mu_c^{+}\mu_c^{-}\pi_{k+1}^{\mp}(\mu_{cp}^{\mp} - 2\alpha_{k+1}\mu_{c^2p}^{\mp})\alpha_{k+1}\pi_{k+1}(1,1)\\
&\quad + 2\mu_c^{+}\mu_c^{-}\mu_{cp}^{\pm}\mu_c^{\mp}\alpha_{k+1}^2\pi_{t_{k+1}}(1,1)^2
\end{aligned}$$

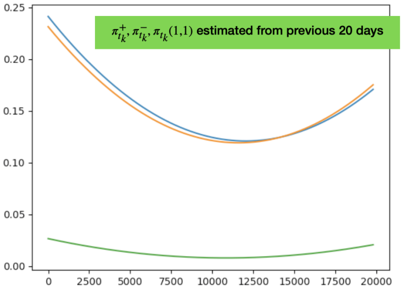
Remark: Under ultra high-frequency setting (i.e., $\max\{t_k - t_{k-1}\} \approx 1$ sec), $\pi(1,1) \approx 0$; the general case is need for $\max\{t_k - t_{k-1}\} \geq 3$ sec.

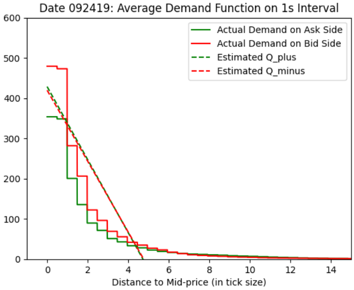
Data

- We implement the optimal strategy on historical LOB data of MSFT stock from April 2018 to May 2019 (252 days).
- The data records the type, volume and price of each actual LOB change that took place in NASDAQ.
- We reconstruct the top 20 levels of the LOB.
- We treat each day as an independent sample from the experiment.
- We take as trading frequency $t_k - t_{k-1} = 1$ second.
- Assume the investor trades from 10:00 a.m. to 3:30 p.m.

Training

- The parameters are trained based on the last 20 days of each test day.
- The $\pi_k^{\pm}$ is estimated based on the relative frequencies of MOs of the last 20 days. A quadratic pattern was uncovered and implemented.
- For the demand function $Q^{\pm}$, we fit a linear regression (see below) to the actual demand function for each day and then average the estimated $\mu_c^{\pm}$, $\mu_{c^2}^{\pm}$, and the relevant $\mu_{c^\ell p^m}^{\pm}$ from the last 20 days.



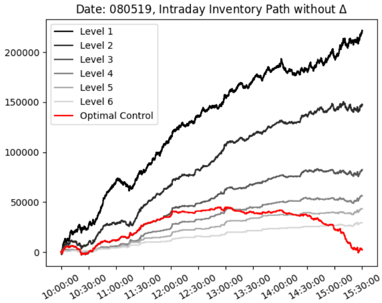


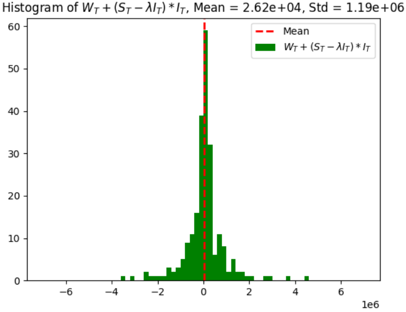
Other Parameters

- For each test day, we train the data using the previous 20 days and implement the general optimal placement strategy with $\pi(1,1) \neq 0$. We have 232 testing days.
- We take $\lambda = 0.0005$, which seems to match well the actual average liquidation cost well.
- We assume each submitted limit order has a volume of 500 shares, which roughly matches the average size of MOs.
- We compare the average and standard deviation of the 232 revenues to that obtained by following a constant placement policy (e.g., always place in level I, or level II, etc.)

Results

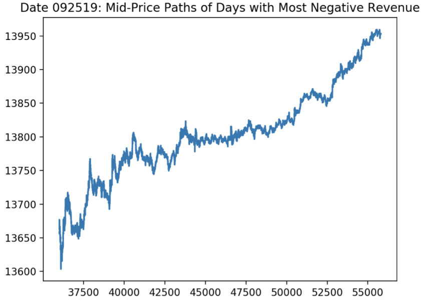
- ▷ The prototypical trajectory of the inventory in a day is shown below. The optimal strategy seems to control very well the inventory at the end of the trading period.
- ▷ The histogram of revenues is shown below. The average revenue for the 232 days is \$26, 200. However, there are large positive and negative revenues that produce a large standard deviation.

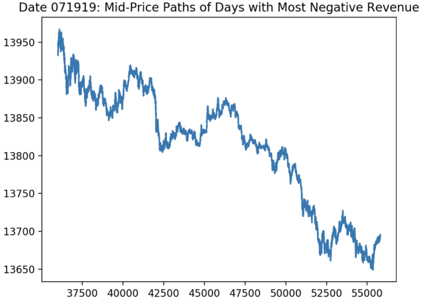




Analysis

- Below, we show the midprice trajectory for two prototypical days in which the final revenue is large negative.
- In both cases, the martingale assumption may be an issue. This suggests to relax this assumption.





Optimal Placement With Drift

- We consider the following model

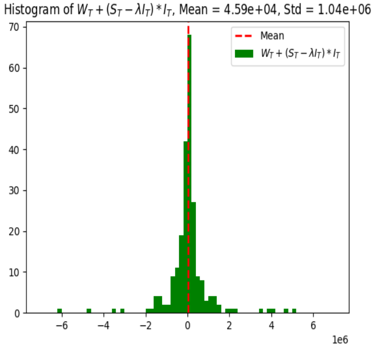
$$\mathbb{E}(S_{t_{k+1}} - S_{t_k} | \mathcal{F}_{t_k}) = \Delta_k, \quad \text{with} \quad \Delta_k \in \mathcal{F}_{t_k}$$
$$\mathbb{E}(S_{t_{k+\ell+1}} - S_{t_{k+\ell}} | \mathcal{F}_{t_k}) = 0, \quad \text{for} \quad \ell > 0.$$

- To estimate Δ_k, we take the average of the last 20 increments:

$$\widehat{\Delta}_k = \frac{1}{20} \sum_{i=1}^{20} (S_{t_{k-i+1}} - S_{t_{k-i}}).$$

Results With Drift

- ▷ The histogram of revenues is shown below. The average revenue for the 232 days almost doubled to \$45,900. However, there are still large positive and negative revenues.
- The revenues for the optimal strategies (with and without drift) and other fixed placement strategies is shown in the table below.



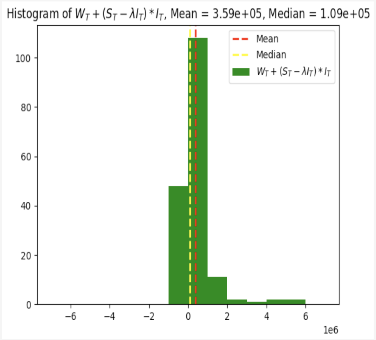
		Optimal Strategy	Level 1	Level 2	Level 3	Level 4
mean (std)	Without Delta	2.62E+04 (1.19e+06)	-7.79E+06 (1.52e+07)	-1E+06 (4.49e+06)	-1.14E+05 (2.01e+06)	-3.78E+04 (1.07e+06)
	With Delta	4.59E+04 (1.04e+06)				
ratio of days when optimal strategy >= benchmark	Without Delta		0.79	0.60	0.51	0.56
	With Delta		0.78	0.59	0.42	0.52

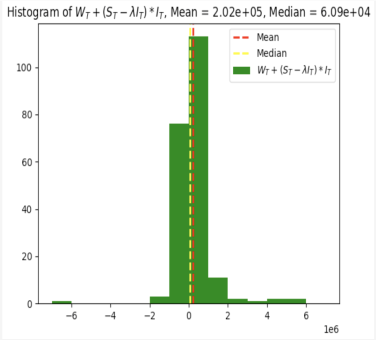
Model Risk

- Further analysis reveals that the days with large negative and positive values corresponds to days where the used trained parameter values (based on the last 20 days) are quite different from the actual values for the testing day.
- Therefore, as with any other financial model, model risk may be a key factor.
- A specific factor that is found to play a key role in the performance of the placement strategy is the accurate estimation of the demand function for placements near the midprice.

Result Excluding Days With High Model Risk

- Below, we exclude those days with high model risk (days with large error in the estimation of the demand function near the midprice).
- Left panel keeps days with errors within the 25 and 75 quartile (178 days kept out 232). Mean and SD are 3.59E+05 and 8.55E+05, resp.
- Right panel keeps days with errors within the 5 and 95 quartile (211 days kept out 232). Mean and SD are 2.02E+05 and 9.52E+05, resp.





Future Work

- Look to incorporate end-of-the day and intraday inventory penalty:

$$\max \mathbb{E}(W_T + I_T S_T - \lambda I_T^2 - \phi \int_0^T I_t^2 dt).$$

- Incorporate adaptive intensity models for the arrival of MO's:

$$\pi_k(j^+, j^-) = \mathbb{P}(\mathbf{1}_k^+ = j^+, \mathbf{1}_k^- = j^- \big| \mathcal{F}_{t_{k-1}}), \quad j^+, j^- \in \{0, 1\},$$

- Incorporate multiple LOB placements and adaptive online selection of the action times t_k's and submission volume per order (e.g., larger volume in high volatility and smaller in low volatility regimes).
- Implement model free reinforced learning methods.

References

Avellaneda, M., & Stoikov, S. (2008)
High-frequency trading in a limit order book
Quantitative Finance 8(3), 217-224.

Cartea, Á., & Jaimungal, S. (2015)
Risk metrics and fine tuning of high-frequency trading strategies.
Mathematical Finance 25(3), 576-611.

Adrian, T., Capponi, A., Vogt, E., & Zhang, H. (2018)
Intraday market making with overnight inventory costs.

Hendershott, T., & Menkveld, A. J. (2014)
Price pressures
Journal of Financial Economics 114(3), 405-423.