Optimal High-Frequency Market Making

Takahiro Fushimi, Christian González Rojas,
and Molly Herman
{tfushimi, cgrojas, mrherman}@stanford.edu

June 11, 2018

Abstract

The paper implements and analyzes the high frequency market making pricing model by Avellaneda and Stoikov (2008). This pricing model is integrated with a proprietary inventory control model that dynamically adjusts the order size to mitigate inventory risk, the risk that we bear due to our inventory. Then, we develop a trading simulator to assess the P&L and inventory of our optimal pricing strategy in comparison to a baseline pricing model for five representative stocks. With the inventory model, the optimal pricing model outperforms the baseline in inventory management while ensuring profitability.

Contents

1 Introduction

Market makers are critical providers of liquidity in markets as they constantly place bid and ask orders in the limit order book such that any market order will always be capable of being filled. The goal of the market maker is to strategically place these bids and asks to capture the spread, the difference between the bid and ask price, while also earning a rebate for providing liquidity. On the other hand, market making has become one of the prevailing strategies for high-frequency traders who profit by turning over positions in an extremely short period. These high-frequency traders play integral roles in providing liquidity to markets, accounting for more than 50% of total volume in the US-listed equities (SEC, 2014).

Various pricing models for market making have been proposed in the academic literature. Ho and Stoll (1981) is one of the early studies that analyze the market making problem under a stochastic control framework. Avellaneda and Stoikov (2008) extends the model proposed by Ho and Stoll (1981), derives the optimal bid and ask quotes using asymptotic expansion and applies it to high-frequency market making. Furthermore, Guéant, Lehalle, and Fernandez-Tapia (2013) develops the model further by deriving the closed form solution of the optimal bid and ask spread with boundary conditions on inventory size.

In this paper, we implement the high frequency market making pricing model proposed by Avellaneda and Stoikov (2008). We choose this model for the ease of implementation and analysis, and unlike recent models such as Guéant et al. (2013), it does not restrict the permissible inventory size. Although this unconstrained inventory assumption lets the market maker to keep trading regardless of their position, it is a shortcoming since it increases the likelihood of accumulating a one-sided position and getting exposed to inventory risk, the risk that we bear due to inventory. To complement the pricing model, we develop an inventory control model that dynamically adjusts the order size based on the current position. This integrated approach allows us to effectively control inventory risk while ensuring profitability. A trading simulator is devised to assess the P&L and inventory of our optimal pricing strategy compared to a baseline pricing model for five representative stocks.

The rest of the paper is organized as follows: Section 2 describes the pricing model and the inventory model, section 3 explains a trading simulator on which the strategy is tested, section 4 discusses the experiment and results, while section 5 concludes.

2 Market Making Model

As a high-frequency market maker, we integrate the pricing framework proposed by Avellaneda and Stoikov (2008) and a proprietary order size dynamic model. The combination of an optimal quote and a dynamic order size strategy allows us to effectively control inventory risk and ensure profitability.

2.1 Pricing

We use the optimal market making model developed by Avellaneda and Stoikov (2008) as our bid and ask quote-setting strategy. The framework is based on a utility-maximizing market maker trading in a limit order book. This section presents a brief summary of the model.

We are interested in maximizing our expected exponential utility given our profit and loss at terminal time T. Assuming the risk-free rate is zero and the mid-price of a stock follows a standard brownian motion $dS_t = \sigma dW_t$ with initial value $S_0 = s$ and standard deviation σ, Avellaneda and Stoikov (2008) formulates the market maker problem as:

$$
\begin{aligned}
u(s,x,q,t) &= \max_{\delta^a,\delta^b} \ \mathbb{E}_t\left[-e^{-\gamma(X_T+q_T S_T)}\right] \\
\text{where} \quad & \delta^a, \delta^b \text{ are the bid and ask spreads} \\
& \gamma \text{ is a risk aversion parameter} \\
& X_T \text{ is the cash at time } T \\
& q_T \text{ is the inventory at time } T \\
& S_T \text{ is the stock price at time } T
\end{aligned}
$$

A few assumptions must be made before solving the stochastic optimal control problem. First, it is important to model inventory as a stochastic process, given that order fills are random variables. Therefore, we can model:

$$
\begin{aligned}
q_t &= N_t^a - N_t^b \\
\text{where} \quad & N_t^a \text{ is the amount of stock sold} \\
& N_t^b \text{ is the amount of stock bought}
\end{aligned}
$$

Based on this definition, we can model cash as a stochastic differential equation in the form:

$$
\begin{aligned}
dX_t &= p^a dN_t^a + p^b dN_t^b \\
\text{where} \quad & p^a, p^b \text{ are the bid and ask quotes}
\end{aligned}
$$

Avellaneda and Stoikov (2008) also provides a structure to the number of bid and ask executions by modeling them as a Poisson process. According to their framework, this Poisson process should also depend on the market depth of our quote. This is achieved through the following expression:

$$
\begin{aligned}
\lambda(\delta) &= Ae^{-\kappa\delta} \\
\text{where} \quad & \delta \text{ is the market depth}
\end{aligned}
$$

This framework to model execution intensity will also prove useful in the design of our trading simulator. Avellaneda and Stoikov (2008) then continue to solve the stochastic control problem using the following Hamilton–Jacobi–Bellman equation:

$$
\begin{aligned}
0 = u_t + \frac{1}{2}\sigma^2 u_{ss} &+ \max_{\delta^b} \lambda^b(\delta^b)[u(s, x - s + \delta^b, q + 1, t) - u(s,x,q,t)] \\
&+ \max_{\delta^a} \lambda^a(\delta^a)[u(s, x + s + \delta^a, q + 1, t) - u(s,x,q,t)]
\end{aligned}
$$

Then, this nonlinear partial differential equation is solved using an asymptotic expansion for a small inventory q. This results in the pricing equations that are relevant to our algorithm:

$$\begin{cases} r(s,t) &= s - q\gamma\sigma^2(T-t) \quad \rightarrow \quad \text{Indifference price} \\ \delta^a + \delta^b &= \gamma\sigma^2(T-t) + \ln\left(1 + \frac{\gamma}{\kappa}\right) \quad \rightarrow \quad \text{Spread around } r(s,t) \end{cases}$$

It is important to notice that, since Avellaneda and Stoikov (2008) defines T as the terminal time in which the trader optimizes its expected utility, the spread equation can be seen as a linear function of $(T-t)$ given by:

$$\delta^a + \delta^b = \underbrace{\gamma\sigma^2}_{\mathbf{A}}(T-t) + \underbrace{\ln\left(1 + \frac{\gamma}{\kappa}\right)}_{\mathbf{B}}$$

where **A** is the slope of the spread equation

B is the closing spread when $t = T$

If $\gamma > 0$, the spread equation becomes a decreasing function of time. The rationale behind this optimal strategy is that tighter optimal spread enables the market marker to liquidate their position before the market closes so as to avoid overnight risk. To implement the framework developed by Avellaneda and Stoikov (2008) we must compute our indifference price and set an optimal spread around it given by these two equations. We exploit the linearity of the spread equation and our market data in order to adjust our spread to the best bid-ask spread dynamics. The calibration strategy is explained in detail in the Experiments section of the paper.

2.2 Inventory

To complement the pricing model, we develop a proprietary dynamic order size framework. Unlike the strategy followed by Guéant et al. (2013), who stops quoting if the inventory reaches the maximum permissible level, we are able to keep trading and earning rebates by adjusting the order size based on our current position. Our order size model is given by the following equation:

$$\phi_t^{bid} = \begin{cases} \phi_t^{max} & \text{if } q_t < 0 \\ \phi_t^{max} \cdot e^{-\eta q_t} & \text{if } q_t > 0 \end{cases} \qquad \phi_t^{ask} = \begin{cases} \phi_t^{max} & \text{if } q_t > 0 \\ \phi_t^{max} \cdot e^{-\eta q_t} & \text{if } q_t < 0 \end{cases}$$

where $\phi_t^{bid}, \phi_t^{ask}$ are the bid and ask order size at time t

ϕ_t^{max} is the maximum order size at time t

η is a shape parameter

We select the shape parameter $\eta = -0.005$ to obtain dynamic order size model shown in Figure 1. As we will later see, this function is very effective at controlling inventory risk. The main reasoning behind its mechanics is that the framework controls inventory risk by placing smaller order sizes in the direction of excess position accumulation.

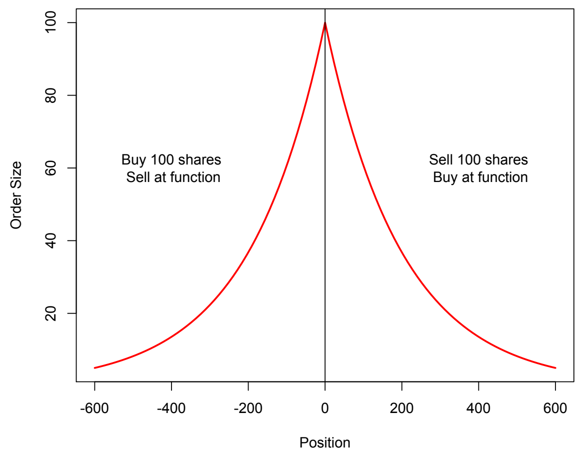


Figure 1: Dynamic order size function

2.3 Algorithm

As a market maker, we are interested in implementing an algorithm that places bid and ask quotes in the limit order book at all times. However, we are aware that in very brief periods, we must hold one-sided quotes for the sake of profitability. This situation occurs when both buy and sell orders are not filled at the same time interval.

Therefore, our strategy iterates as follows. During the trading day, we quote a bid and ask spread if we have no orders in the limit order book. If only one of these orders is filled, we wait for 5 seconds for the outstanding order to be executed. If this does not happen, then we cancel the order and place new bid and ask quotes. Finally, whenever we have two orders in the limit order book, we update our quotes every second. The summary of the trading algorithm is shown in Algorithm 1.

Algorithm 1 Market Making Algorithm

```
while current time < end time do
    if no orders in the book then
        Quote bid and ask prices
    else if 1 order in the book then
        if current time - execution time > waiting time then
            Cancel the outstanding order  Quote new bid and ask prices
        else
            Wait
        end
    else if 2 orders in the book then
        if current time - quote time > update time then
            Cancel both order  Quote new bid and ask prices
        else
            Wait
        end
    end
end
```

3 Trading Simulator

We build a trading simulator to assess our strategy. The principal constituent of the simulator is market order dynamics because as a market maker, we only use limit orders which are matched and filled by market orders. Given this model, we can simulate order executions and run our market making strategy. For simplicity, we make the following assumptions: 1) No latency, 2) No price impact, and 3) No competition with other market makers.

3.1 Market Order Dynamics

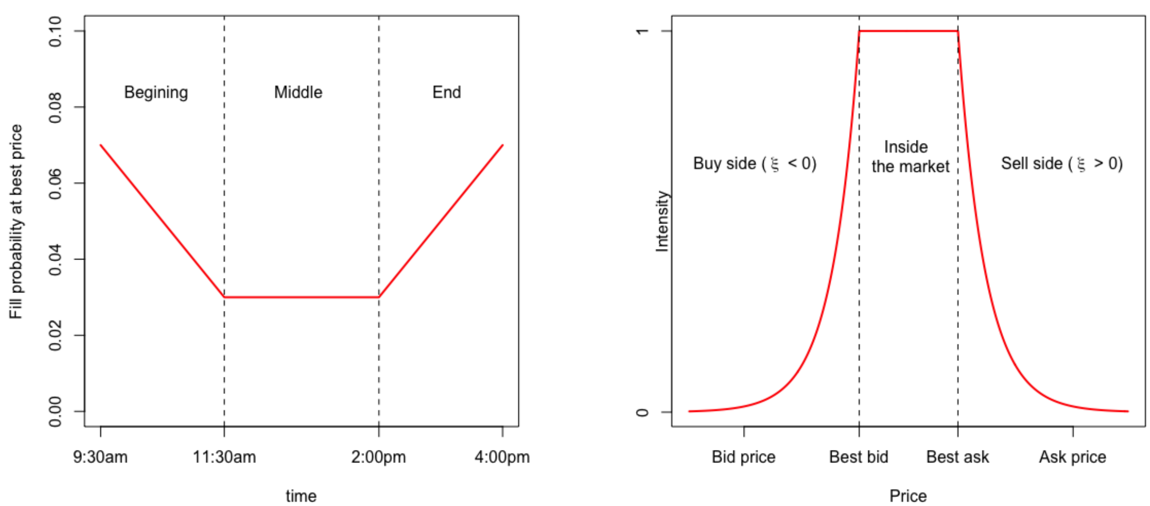
Let ξ denote the depth of our quote in the order book. We use a Time-Inhomogeneous Poisson Process to model the number of arrivals of market orders that are matched with the limit order at the depth of ξ:

$$N_t \sim Pois\left(\int_0^t \lambda(s,\xi)ds\right)$$

Analogous to the market making model, the intensity function is assumed to be a product of time and depth components in the following form:

$$\lambda(t,\xi) = \alpha_t e^{-\mu\xi}$$

A piece-wise linear bathtub shape is adopted for the time component α_t based on the empirical result of intraday volume pattern as discussed in Cartea, Jaimungal, and Penalva (2015). The depth component $e^{-\mu\xi}$ is a decreasing function of depth ξ since the deeper a quote is, the less likely it is for the order to be executed due to the lower priority. Figure 3 illustrates the shape of these two components respectively.



(a) Time component α_t **(b)** Depth component $e^{-\mu\xi}$

Figure 2: The shape of intensity function

3.2 Order Execution

The market order dynamics enables us to simulate an order execution as follow. At time t, we generate a Bernoulli random number $X \sim Ber(\lambda(t,\xi)\Delta)$ with the depth of a limit order ξ and time interval Δ. Then, $X = 1$ indicates the arrival of a market order and we assume that the order is executed. To make the experiment more realistic, we also allow an order to be partially filled by generating another random number from Gamma distribution $Y \sim Gamma(\kappa, \theta)$. Then, Y is multiplied by our order size to compute the executed order size. For instance, $Y \geq 1$ implies a full execution, while $Y < 1$ means a partial fill.

4 Experiments

We choose the week of June 12, 2017 to simulate our strategy, and trade from 9:30am - 4pm each day. As the market maker may deal with a wide variety of stocks, we choose the S&P500 as a baseline and then four stocks to represent combinations of high and low volume as well as high and low performance as shown in Table 1. With the technique described in section 2.1, the parameters are calibrated using the average of the opening and closing spreads from each stock in the previous week. The data is retrieved from a simulator provided by Thesys Technologies, and the time interval Δ is set to 1 second. The maximum order size ϕ_t^{max} is set to 100. The parameters of the simulator are set as $\mu = 100, \kappa = 2, \theta = 1/1.65$.

To assess the performance of the optimal strategy in section 2, we consider a baseline pricing strategy in which the market maker always quotes at the best bid and ask prices that are currently placed in the order book. Every other aspect of the baseline strategy remains identical to the optimal strategy. Then, a Markov Chain model is used to further examine probabilities that help measure performance of these strategies.

Table 1: Stocks to trade

	Volume	Performance	Open Spread	Close Spread
AAPL	high	high	0.05	0.01
AMZN	low	high	0.49	0.56
GE	high	low	0.04	0.01
IVV	low	high	0.03	0.01
M	low	low	0.09	0.01

4.1 Results

Our primary interest is the profitability and the inventory management of the strategies. Table 2 shows the average terminal P&L and position (inventory) of both strategies. The optimal strategy achieves higher profits in AAPL and AMZN and comparable profits in the rest of stocks, compared with the baseline strategy. Also, both strategies end each trading day with a small position on average, indicating the success of inventory management. Furthermore, the optimal strategy accomplishes the variance reduction of profits and position per day as shown in Table 3. Not only does the optimal strategy reduce the terminal position, but it also manages the inventory during a trading session while consistently making profits.

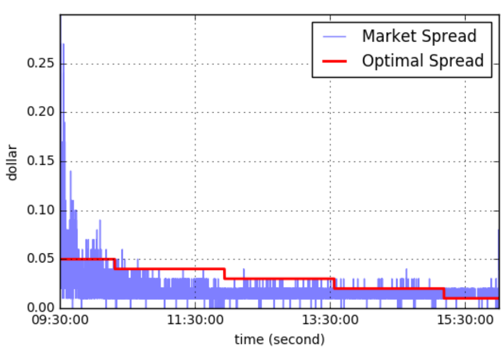
Table 2: Average terminal P&L and position

	Optimal		Baseline	
	P&L	Position	P&L	Position
AAPL	-1,378.01	-29.0	-1,625.25	-45.6
AMZN	58,331.04	8.4	13,522.2	-34.6
GE	703.12	-49.2	708.48	-38.4
IVV	217.58	-41.8	547.52	-36.8
M	534.04	-46.0	587.09	23.8

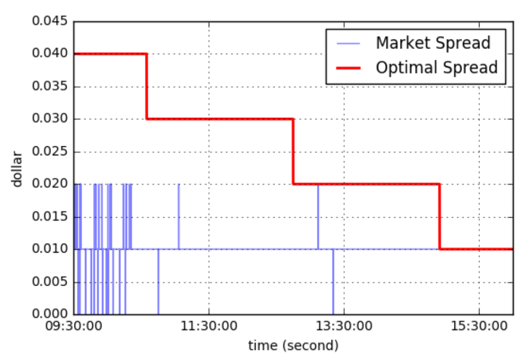
Table 3: The mean and standard deviation of profits and position per day

	Optimal Strategy				Baseline Strategy			
	Profits		Position		Profits		Position	
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
AAPL	-988.54	289.82	0.86	63.66	-1093.60	357.66	7.53	112.2
AMZN	32,426.72	16,157.0	48.52	438.33	4,889.20	4,202.4	2.96	126.94
GE	245.0	192.92	-2.41	60.92	248.96	191.43	11.97	109.19
IVV	23.14	129.9	-0.49	67.9	152.0	196.6	-1.38	109.59
M	144.26	146.78	-0.83	46.14	192.59	167.24	-3.86	105.93

Figure 3 demonstrates the market and optimal spreads for AAPL and GE on June 12, 2017. Note that the spreads are rounded to the nearest cent. As discussed in section 2, the spread of the optimal pricing model is a linear function of time. At the beginning, the optimal spread is wider than the market spread. Later in the trading day, around 2pm, there is a turning point when the optimal spread becomes narrower than the market spread, meaning that more orders are likely to be filled. This aggressiveness helps unwind the accumulated position before the market closes and boost our profits if spreads are large enough.

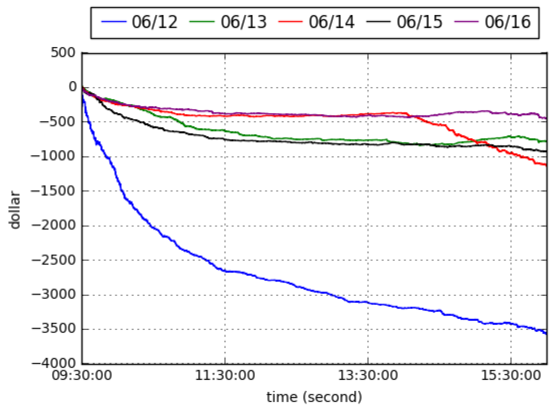


(a) AAPL

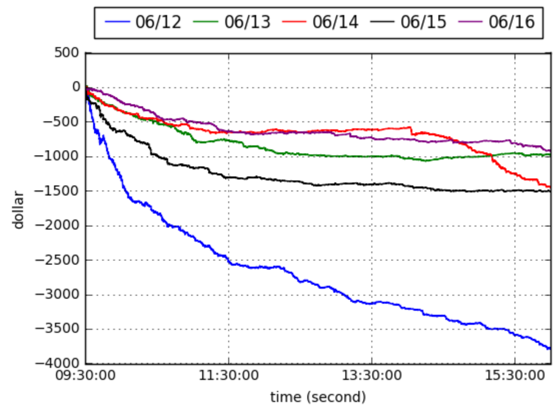


(b) GE

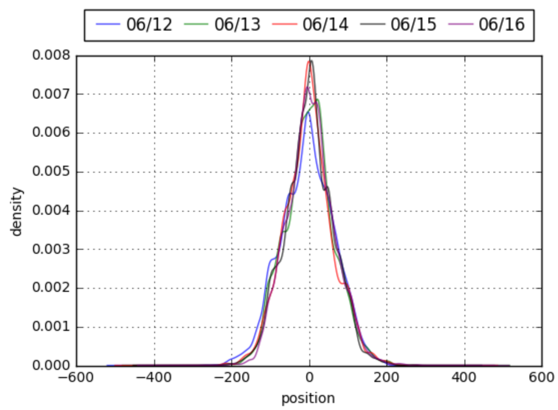
Figure 3: Market and optimal spreads



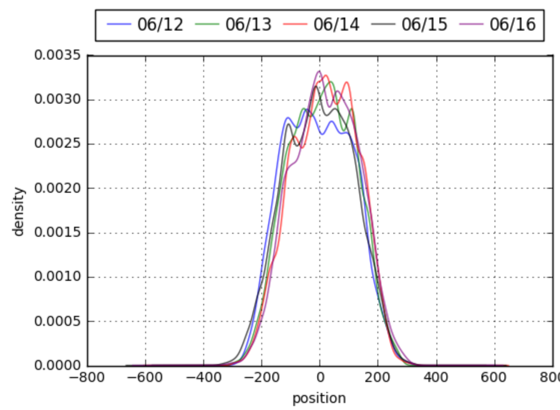
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy



(c) The density of inventory size in the optimal strategy

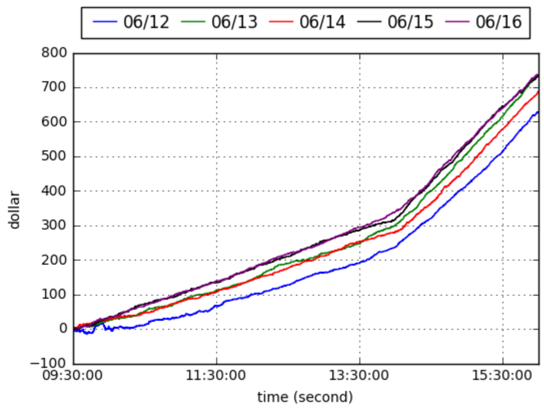


(d) The density of inventory size in the baseline strategy

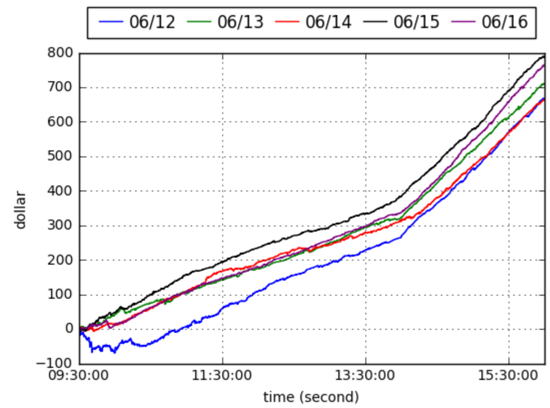
Figure 4: AAPL

Figure 4 displays P&L and inventory size for APPL on each day in both strategies. Though we observe the poorest performance overall on AAPL, we observe that the inventory size is centered around 0. For the optimal strategy, the position remains between ± 200 and the density has sharper peak than that of the baseline strategy, which suggests more control of inventory risk. In contrast to AAPL, for GE, P&L consistently increases over time. The same trend of profits is observed in the baseline strategy, but the variance of profits is lower in the optimal strategy. As is the case of AAPL, the optimal strategy outperforms the baseline strategy in terms of controlling the inventory size.

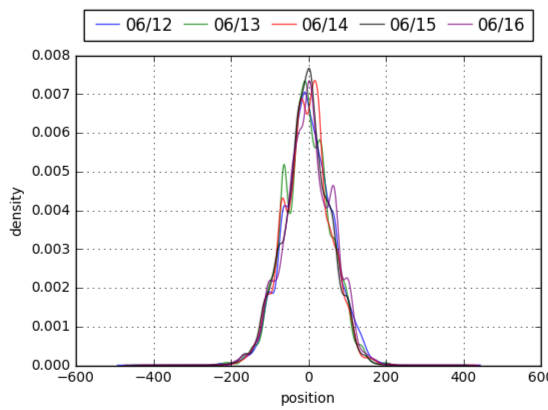
We also compute the average number of buy and sell orders executed, shares bought and sold, and quotes per day as shown in Table 4 and 5. For both strategies, the number of buy and sell orders, as well as the number of shares bought and sold are well balanced, indicating that the position stays around 0. In terms of quoting, the baseline strategy updates and posts prices more frequently than the optimal strategy. However, the optimal strategy quotes more efficiently because it has more buy (sell) orders executed per quote than the baseline strategy.



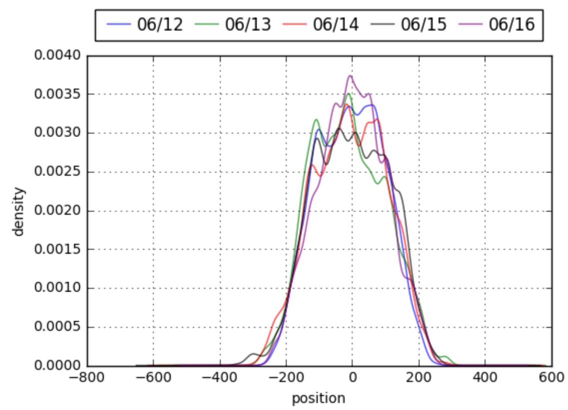
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy



(c) The density of inventory size in the optimal strategy



(d) The density of inventory size in the baseline strategy

Figure 5: GE

Table 4: The average number of orders, shares, and quotes per day in the optimal strategy

	buy orders	sell orders	shares bought	shares sold	quotes
AAPL	6,085	6,105	446,921	447,066	64,645
AMZN	14,530	14,394	696,907	696.865	42,130
GE	4,411	4,456	324,148	324,394	82,138
IVV	7,613	7,625	548,505	548,714	63,349
M	2,927	3,003	221,074	221,304	91,292

Table 5: The average number of orders, shares, and quotes per day in the baseline strategy

	buy orders	sell orders	shares bought	shares sold	quotes
AAPL	7,969	7,895	528,109	528,337	472,437
AMZN	7,833	7,850	500,537	500,710	528,028
GE	6,438	6,239	422,223	422,415	601,208
IVV	7,361	7,394	486,863	487,047	666,219
M	6,421	6,487	430,305	430,186	738,390

4.2 Markov Chain Analysis

We apply a Markov Chain model to compute the probabilities that measure the performance of strategies. Let us assume that the state space $\mathcal{S}$ consists of $\{0,1,2\} = \{Quoting, Waiting, Spread\}$ as depicted in Figure 6. *Quoting* means bid and ask prices are quoted, *Waiting* means one of the orders is filled and the market maker is waiting for the outstanding order to be filled, and *Spread* means both buy and sell orders are filled and the spread is captured. Note that the transition probability from *Spread* to *Quoting* is 1, and there is no arrow from *Spread* to *Waiting* because after making the spread, the market maker always quotes new bid and ask prices. Table 6 summarizes the interpretation of each transition probability.

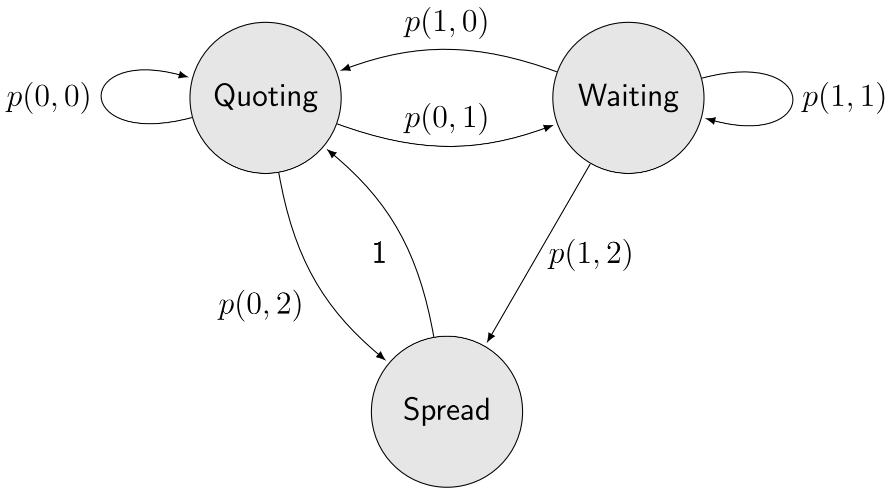


Figure 6: The Markov Chain model

Table 6: The interpretation of the transition probabilities

$p(0,0)$	Update bid/ask prices
$p(0,1)$	One of the orders is filled
$p(1,0)$	Cancel the outstanding order after waiting
$p(1,1)$	Wait for the outstanding order to be filled
$p(0,2)$	Both buy/sell orders are filled
$p(1,2)$	The outstanding order is filled after waiting

A quantity of interest for market makers is the probability of capturing the spread since it is the main source of profit in the business. There are two scenarios in which the market maker can capture the spread in each quote. The first case is that both buy and sell orders are filled within the next time interval. The other case is that one of the order is filled, and after 5 waiting periods, the outstanding order is filled. Therefore, the probability of making the spread is

$$p^\star = p(0,2) + \sum_{n=0}^{5} p(0,1)p(1,1)^n p(1,2).$$

Another quantity of interest would be the probability of one-sided fill, which is the case where only buy or sell is filled and the remaining order is cancelled after the waiting time. The higher likelihood of one-side fill increases the inventory risk since the fill is imbalanced. The probability of one-side fill after 5 waiting periods is defined as

$$q^{\star} = p(0,1) + p(0,1)p(1,1)^{5}p(1,0).$$

Table 7: The probability of making the spread

	AAPL	AMZN	GE	IVV	M
Optimal strategy	2.6%	19.3%	1.8%	4.9%	0.9%
Baseline strategy	5.1%	4.7%	3.7%	4.7%	3.8%

Table 8: The probability of one-side fill

	AAPL	AMZN	GE	IVV	M
Optimal strategy	0.8%	1.9%	0.4%	0.7%	0.3%
Baseline strategy	0.9%	1.0%	0.5%	0.7%	0.5%

The estimated probabilities are summarized in Table 7 and 8. The baseline strategy has higher probability of making the spread than the optimal strategy because the baseline strategy always quotes at the best bid and ask prices, which tend to be more aggressive than the optimal prices. Yet, it is important to note that the higher probability of making the spread may not always be indicative of the higher profit as it also depends on the size of spreads. The optimal strategy, on the other hand, achieves lower probability of one-side fill. This finding demonstrates that the optimal strategy can quote more efficiently and prevent the position from accumulating.

5 Conclusions

In this project, we implement the high frequency market making pricing strategy proposed by Avellaneda and Stoikov (2008). The inventory strategy that mitigates the inventory risk is proposed to complement the pricing model. Furthermore, we develop the trading simulator to experiment with our strategy on real high frequency data. With the inventory control model, the optimal pricing model achieves more controlled inventory size while ensuring profitability. A possible extension of our strategy can be to use a model to predict mid price and market order arrivals so that the market makers can profit regardless of the market movement and the size of spreads.

References

Avellaneda, M., & Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, *8*(3), 217-224. Retrieved from https://doi.org/10.1080/14697680701381228

Cartea, A., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and high-frequency trading*. Cambridge, United Kingdom: Cambridge University Press.

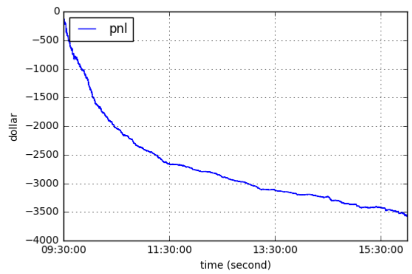
Guéant, O., Lehalle, C.-A., & Fernandez-Tapia, J. (2013). Dealing with the inventory risk: a solution to the market making problem. *Mathematics and Financial Economics*, *7*(4), 477–507. Retrieved from https://doi.org/10.1007/s11579-012-0087-0

Ho, T., & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, *9*, 47–73.

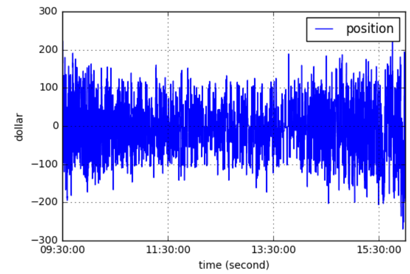
U.S. Securities and Exchange Commission. (2014). *Equity Market Structure Literature Review, Part II: High Frequency Trading.*

Appendix: Results for All Stocks

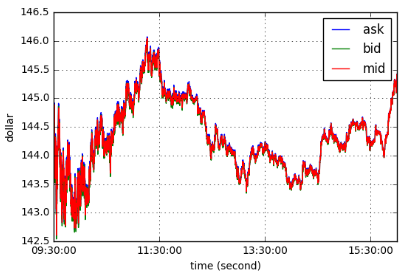
Optimal Strategy on June 12, 2017



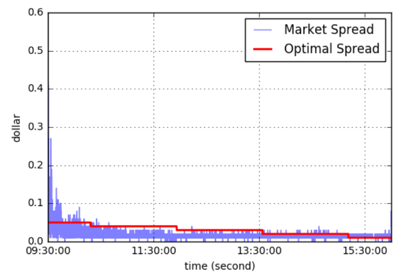
(a) P&L



(b) Position

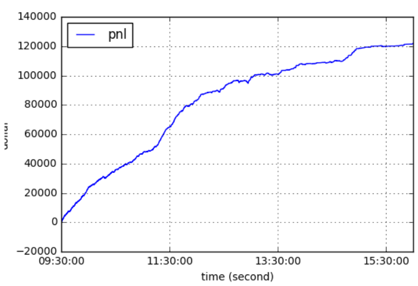


(c) Quoted Prices

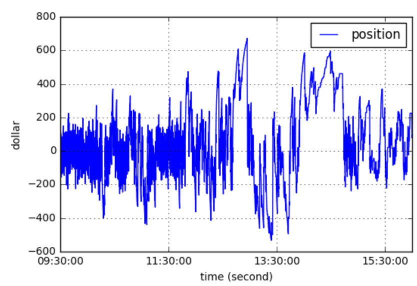


(d) Optimal bid-ask spread

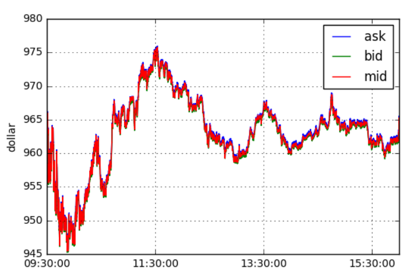
Figure 7: AAPL



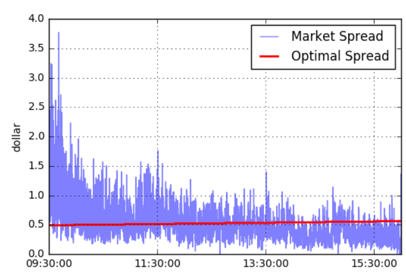
(a) P&L



(b) Position

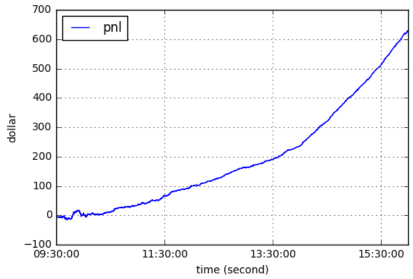


(c) Quoted Prices

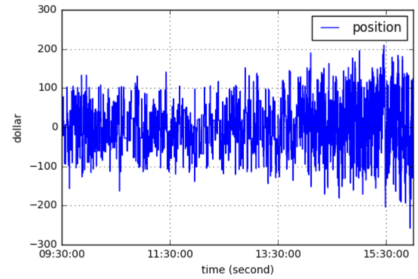


(d) Optimal bid-ask spread

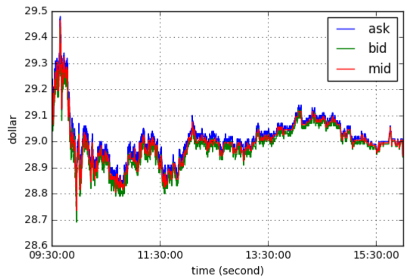
Figure 8: AMZN



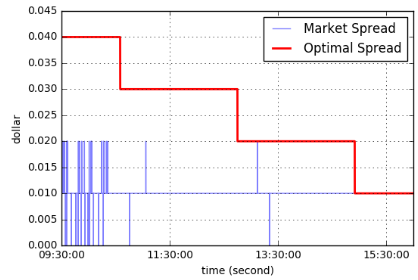
(a) P&L



(b) Position

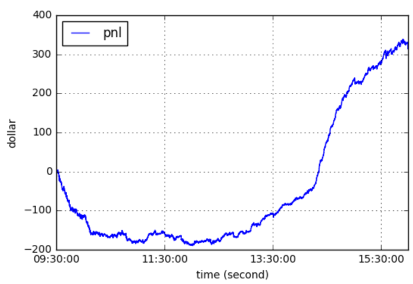


(c) Quoted Prices

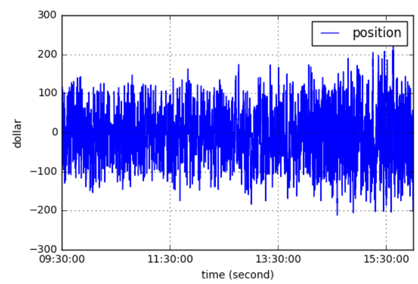


(d) Optimal bid-ask spread

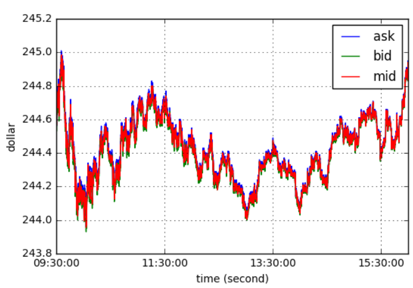
Figure 9: GE



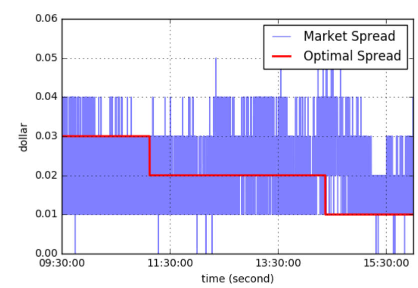
(a) P&L



(b) Position

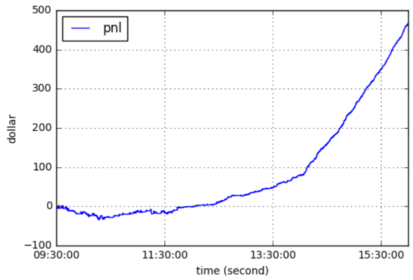


(c) Quoted Prices

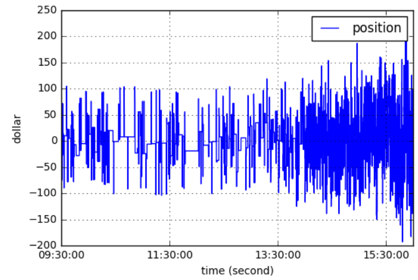


(d) Optimal bid-ask spread

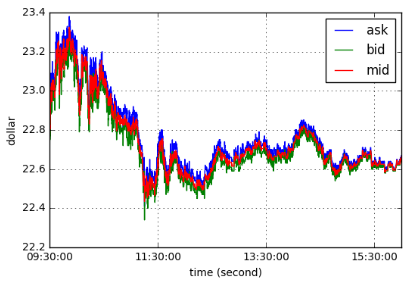
Figure 10: IVV



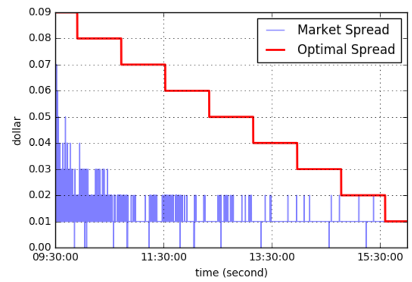
(a) P&L



(b) Position



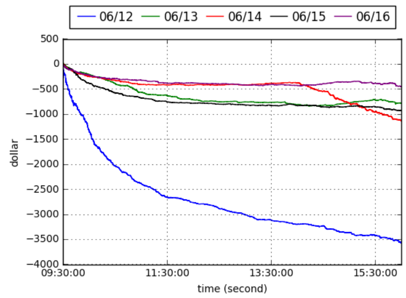
(c) Quoted Prices



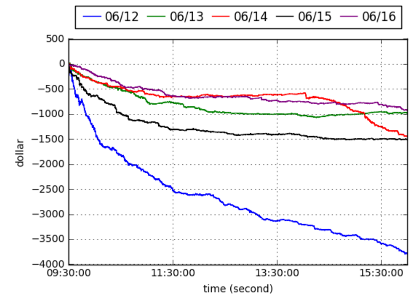
(d) Optimal bid-ask spread

Figure 11: M

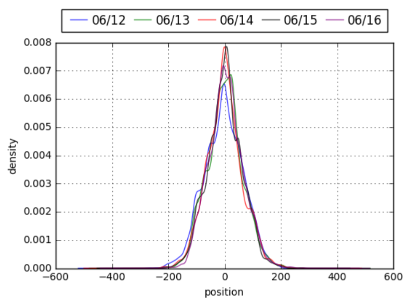
Results for the Complete Week: Optimal and Baseline



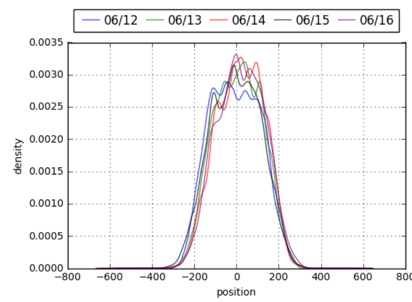
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy

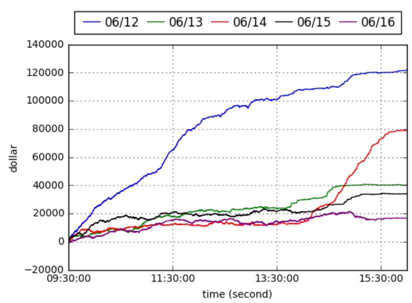


(c) The density of inventory size in the optimal strategy

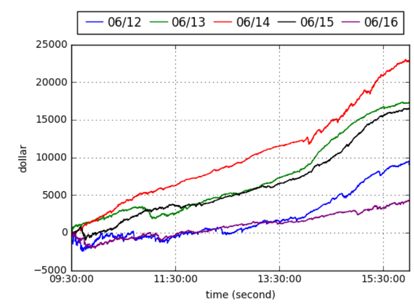


(d) The density of inventory size in the baseline strategy

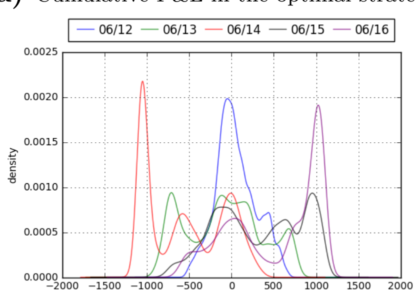
Figure 12: AAPL



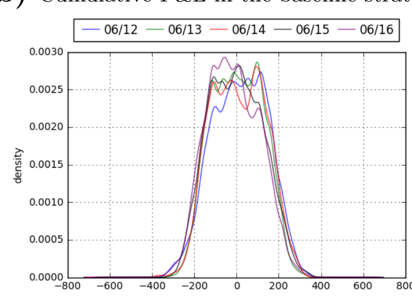
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy

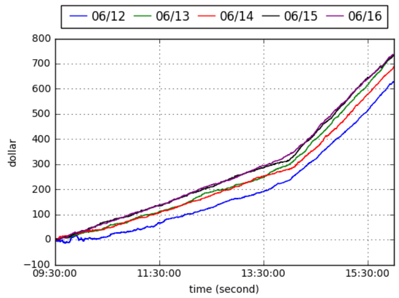


(c) The density of inventory size in the optimal strategy

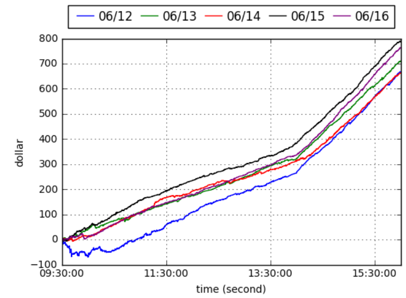


(d) The density of inventory size in the baseline strategy

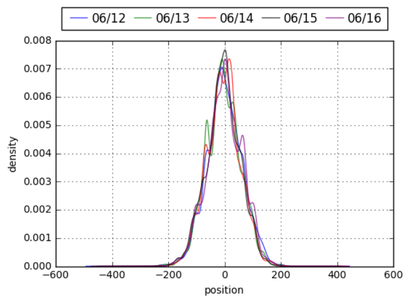
Figure 13: AMZN



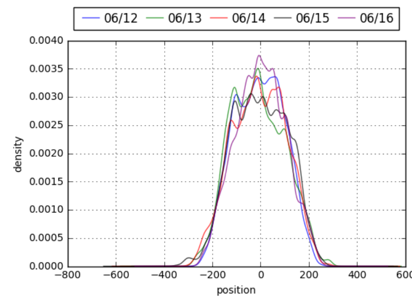
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy

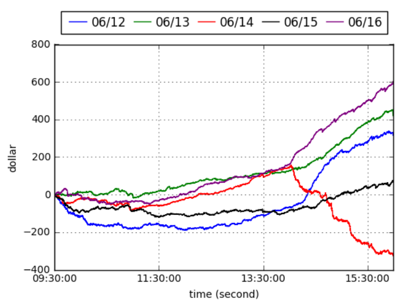


(c) The density of inventory size in the optimal strategy

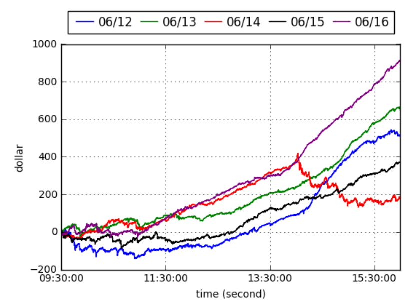


(d) The density of inventory size in the baseline strategy

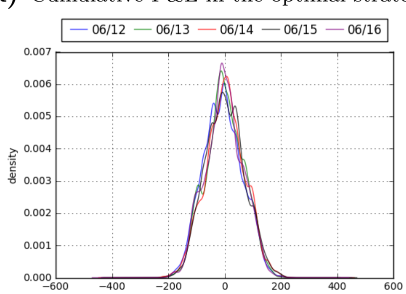
Figure 14: GE



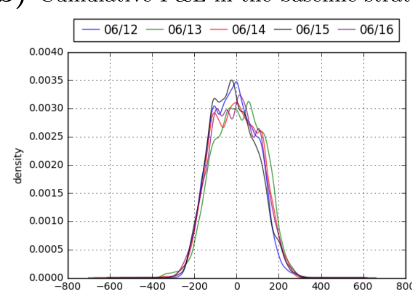
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy

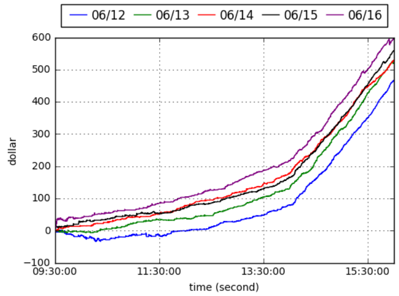


(c) The density of inventory size in the optimal strategy

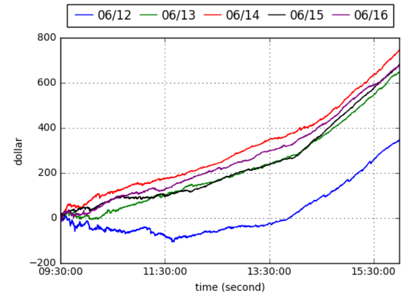


(d) The density of inventory size in the baseline strategy

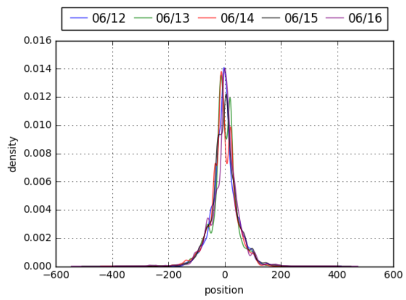
Figure 15: IVV



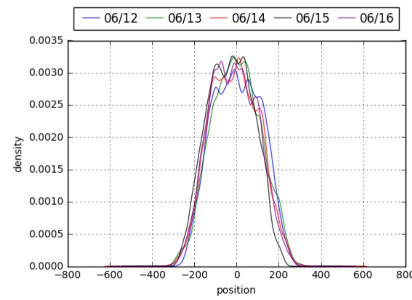
(a) Cumulative P&L in the optimal strategy



(b) Cumulative P&L in the baseline strategy



(c) The density of inventory size in the optimal strategy



(d) The density of inventory size in the baseline strategy

Figure 16: M