Adaptive Market Making with Inventory Constraints via Online Learning

Shan Xue[1], **Ye Du**[2*], **Liang Xu**[3]

[1] School of Economics and Management, Leshan Normal University, Leshan, China
[2] Southwestern University of Finance and Economics, Chengdu, China
[3] School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China
xueshanads123@gmail.com, henry,duye@gmail.com, arecxuliang1112@gmail.com

Abstract

A market maker is a specialist who provides liquidity by continuously offering bid and ask quotes for a financial asset. The market maker's objective is to maximize profit while avoiding the accumulation of a large position in the asset to control inventory risk. To achieve model-free results, online learning has been applied to design market-making strategies that make no assumptions on the dynamics of the limit order book and asset price. However, existing work primarily focuses on profit rather than inventory risk. To address this limitation, this paper develops market-making strategies with inventory constraints within the online learning framework. To manage inventory risk, we propose two classes of market-making strategies with fixed bid-ask spreads that serve as reference strategies. Each reference strategy can ensure that the inventory remains under control, which enables the online learning algorithms designed for each class of reference strategies to satisfy inventory constraints. Different from the standard online learning model where the gain in each period is assumed to lie within a fixed bounded interval, the gain in our model depends on a state variable (i.e., the inventory size). Thus, a key challenge in analyzing the regret bounds is to bound the difference between the gains of any two reference strategies, which becomes significantly more complicated compared with scenarios without inventory constraints. By tackling these difficulties, we show that these algorithms achieve low regrets. Experimental results illustrate the superior performance of our algorithms in inventory risk control.

Extended version — https://papers.ssrn.com/abs=5110826

1 Introduction

Market making is a financial activity where an individual or firm, known as a market maker, provides liquidity by continuously offering bid and ask quotes for a financial asset. The market maker stands ready to buy the asset at the bid price and sell it at the ask price, with the difference between the two prices known as the bid-ask spread. By doing so, the market maker provides liquidity for other traders while profiting from the simple principle of buying low and selling high. However, this profit does not come without risk. In the presence of a strong price trend, such as rapidly rising prices, the market maker may accumulate a large number of short positions. If these positions are liquidated at the peak of the trend, significant losses may be incurred. This risk is referred to as *inventory risk* (Avellaneda and Stoikov 2008; Guilbaud and Pham 2013). Thus, the objective of a market maker is to maximize the profit and loss (PnL) of the trading while keeping her inventory low to minimize inventory risk.

To determine the optimal market-making strategy, there are three primary approaches in the existing literature. The first approach models the problem as a stochastic optimal control problem, which is then solved using the well-known Hamilton-Jacobi-Bellman equation (Avellaneda and Stoikov 2008; Guéant, Lehalle, and Fernandez-Tapia 2013; Obizhaeva and Wang 2013; Cartea, Jaimungal, and Penalva 2015). This method often involves making assumptions on the dynamics of the limit order book (LOB) and asset price, such as Brownian motion. The second approach frames it as a reinforcement learning (RL) problem (Patel 2018; Spooner et al. 2018; Zhang and Chen 2020), whose recent progress is comprehensively reviewed by (Hambly, Xu, and Yang 2021). Spooner and Savani (2020) develop adversarial RL based on the model of (Avellaneda and Stoikov 2008), and produce market-marking agents that are robust to adversarial and adaptively-chosen market conditions. Although their strategies are more robust to misspecification, they still rely on assumptions on the dynamics of the asset price and LOB. In contrast to the first two approaches, the third approach explores a model-free or robust version of market making, without imposing any assumptions on the dynamics of the LOB and asset price. The online learning or no-regret learning framework is well-suited to this scenario. Abernethy and Kale (2013) are the pioneer in this line of work. They develop a class of spread-based market-making strategies parametrized by a minimum quoted spread. Online learning algorithms are then designed for these strategies to achieve no-regret. However, rather than on inventory risk, their primary focus is on the PnL of these algorithms. Their algorithms will be exposed to significant inventory risk if the asset price rises or falls consecutively.

To address the above limitations, we focus on both PnL and inventory risk for the market maker within the framework of online learning. We extend the work of (Abernethy and Kale 2013) and provide model-free market-making strategies by incorporating inventory risk control. To man-

*Corresponding author.

age inventory risk, one naive approach is to impose a prespecified limit on the maximum absolute value of the inventory that the market maker is allowed to hold. We call it the *hard-constraint* approach. The market maker meets this constraint by limiting the price levels at which she places orders in the order book. Another approach, which is more subtle and relevant to finance literature, is to choose the bid and ask prices negatively proportional to the current inventory level. For instance, if the market marker currently holds a large short position, she may raise her bid prices to buy back some of the short position. This is referred to as the *soft-constraint* approach. Based on the two approaches, two classes of market-making strategies with fixed bid-ask spread: hard-constraint and soft-constraint strategies are proposed as reference strategies. We then design online learning algorithms for each class of reference strategies to develop adaptive market-making strategies.

Next, we present our main result informally in the following theorem. Different from the standard online learning model, where the gain in each period is assumed to lie within a fixed bounded interval, the gain in our model depends on a state variable (i.e., the inventory size). Since the inventory size is influenced by specific parameters of the reference strategies as well as the asset price path, the regret bound of adaptive market-making strategies is necessarily a function of these parameters.

Theorem 1. ***(Informal)*** *The adaptive market-making strategies have regrets bounded by $c\sqrt{T \ln N}$ after T periods to the best of N constraint strategies, where c is a constant depending on specific parameters of the strategies (see Theorem 10-13 for details).*

1.1 Contributions

Our contributions are summarized as follows:

i) To the best of our knowledge, we are the first to investigate inventory risk control of market making within the context of online learning. To accommodate the inventory risk, we introduce a new model. Its basic idea is to develop two types of reference strategies: hard-constraint and soft-constraint strategies. We demonstrate that, for any given path of the asset price, the inventory for both reference strategies can be bounded by a constant in each period. This feature naturally enables adaptive (dynamic) strategies to effectively control inventory risk.

ii) Unlike the standard online learning model, in which the gain (loss) in each period is assumed to lie within a fixed bounded interval, such as [0,1], the upper and lower bounds of the gain in our model depend on a state variable (i.e., the inventory size), which is path-dependent and time-varying. After accommodating inventory constraints, we make nontrivial efforts to bound the difference between the inventories (Lemma 4, 6 and Corollary 7) and further the difference between the gains (Lemma 5 and 8) of any two reference strategies. These results lead to the no-regret properties of our adaptive market-making strategies.

iii) We run comprehensive experiments on the tick-by-tick data of all component stocks of the China CSI500 index over a one-month period, which includes over 10,882 stock price paths. The experimental results show that our adaptive strategies could indeed achieve no-regret. Meanwhile, compared with market-making strategies without inventory constraints, our adaptive strategies significantly improve inventory risk control, especially in markets with strong upward or downward trends.

1.2 Related Literature

Perhaps the most famous model of market making in finance is the Glosten-Milgrom model (Glosten and Milgrom 1985), which investigates the market-making problem in a market with asymmetric information. Subsequent studies have sought to to explore the optimal behavior under various market dynamics settings. Among them, the work of (Avellaneda and Stoikov 2008) is well known and widely used in quantitative finance. Based on the assumption that the market maker with an exponential utility function has perfect knowledge of the market dynamics, they provided a closed-form solution for the optimal market-making strategy. The follow-up work derived the optimal solution for some other utility functions or price processes (Fodra and Labadie 2012; Guéant, Lehalle, and Fernandez-Tapia 2013; Cartea, Jaimungal, and Penalva 2015; Guéant 2017). Cartea, Donnelly, and Jaimungal (2017) considered the impact of model misspecification in the model of (Avellaneda and Stoikov 2008), and provided an analytical solution for the robust optimal strategies.

Within the AI community, a significant body of literature has studied the market-making problem with the RL approach. Chan and Shelton (2001) were the first to apply RL to market making, which developed explicit market-making strategies and tested them under a simulated environment. Later on, Spooner et al. (2018) generalized the results in (Chan and Shelton 2001) and designed a temporal-difference RL algorithm to improve the performance of market making. Some recent work has paid attention to improving the robustness of market-making strategies to model uncertainty (Spooner and Savani 2020; Gašperov and Kostanjčar 2021). Furthermore, RL algorithms can also be applied to high-dimensional, multi-asset market making (Guéant and Manziuk 2019; Baldacci et al. 2019), and multi-agent with different competitive scenarios (Patel 2018; Ganesh et al. 2019).

A substantial body of literature has explored online learning with constraints. Ding et al. (2013) investigated multi-armed bandit problems with random costs subject to a budget constraint and developed two algorithms for this setting. Mannor, Tsitsiklis, and Yu (2009) considered an online learning setting where a decision maker aims to maximize her average reward while ensuring that the average penalty adheres to a specified constraint. In the work of (Paternain et al. 2020), a constrained online optimization problem in networks was examined, where the constraints can vary arbitrarily over time. However, the constraint specifications in these models do not align with the framework employed in our study.

2 The Model

2.1 The Market Trading Framework

Following the work of (Chakraborty and Kearns 2011; Abernethy and Kale 2013), we consider a discrete-time trading model with T periods, where a period is indexed by $t \in \{0, 1, 2, \cdots, T\}$. There is a stock in the market. Denote by P_t the market price of the stock at the end of period t. Let δ be the minimum price variation or tick size, and M be some reasonable upper bound on the stock price. Thus, the set of all possible stock prices can be denoted by $\mathbb{P} = \{\delta, 2\delta, \cdots, M\}$. Assume that $|P_t - P_{t-1}| \leq \nabla$ for all t, where ∇ is a given sufficiently large constant.

A market maker exists in the market who interacts with a continuous double auction via an order book. The market maker focuses on both PnL and inventory risk. She trades the stock over T periods by submitting both limit and market orders to the order book. At the end of each period t, her trading strategy submits a limit order schedule $Q_t : \mathbb{P} \to \mathbb{R}$. For each $P \in \mathbb{P}$, the absolute value of $Q_t(P)$ represents the number of shares she intends to buy (if $Q_t(P) > 0$) or sell (if $Q_t(P) < 0$) at price P. In period $t + 1$, all limit buy orders offered at prices no less than P_{t+1} and all limit sell orders offered at prices no greater than P_{t+1} from period t are assumed to be executed. Furthermore, the strategy may also actively trade the stock to adjust its stock inventory level (i.e., the amount of the stock it holds) via a market order at the end of period t, which only specifies the amount of the stock it is willing to buy or sell. Unlike a limit order, a market order is assumed to be executed immediately at the current price P_t. Denote by H_t and C_t the amount of the stock inventory and cash the strategy holds at the end of period t, respectively. Initially, we set $H_0 = C_0 = 0$. If H_t is positive (resp., negative), the strategy holds a long (resp., short) position in the stock. Let $V_t = H_t P_t + C_t$, which is the total value of the strategy's holdings at the end of period t. The gain of the strategy in period t is naturally defined as $\Delta V_t = V_t - V_{t-1}$.

Assume that there are N reference strategies, the market maker constructs her trading strategy by viewing every reference strategy as an expert and running an online learning algorithm for learning with expert advice. This strategy is referred to as the adaptive strategy hereafter. Following the work of (Chakraborty and Kearns 2011; Abernethy and Kale 2013), we make some assumptions on the trading mechanism as follows:

- Neither transaction nor borrowing costs exist.
- The stock is perfectly divisible, and the market maker may purchase and sell fractional shares of the stock.
- The stock price is exogenously determined, which means that the market maker's trades do not affect the stock price.
- The limit orders submitted by the market maker at the end of each period t will be canceled at the end of the next period if unexecuted.

Although the assumptions outlined above are also present in the work of (Chakraborty and Kearns 2011; Abernethy and Kale 2013), some of them are relatively rigid and unrealistic, particularly the first and third assumptions. In real markets, transaction costs can affect both the market maker's profit and the bid-ask spread. To reduce the frequency of costly trades, the market maker may widen the spread. Furthermore, stock prices are not always determined exogenously, especially when the market maker provides a significant portion of market liquidity. We will relax these assumptions in future work.

2.2 Reference Strategies

To manage inventory risk, we propose two classes of market-making strategies with fixed bid-ask spreads that serve as reference strategies. The first class, known as hard-constraint strategies, ensures that the inventory level at any time remains within a pre-specified interval. While the second class, known as soft-constraint strategies, mitigates inventory risk by adjusting quoted bid and ask prices in a manner negatively correlated with the current inventory level.

Hard-Constraint Strategies To achieve inventory control, one approach is to strictly limit the absolute inventory in each period t, (i.e., $|H_t|$), to no more than some predetermined level R. Our hard-constraint strategies build upon the work of (Abernethy and Kale 2013). Different from their strategies, our hard-constraint strategies restrict the lowest quoted price for limit buy orders and the highest quoted price for limit sell orders in terms of their inventory position to ensure $|H_t| \leq R$ for all t. Specifically, consider a class of hard-constraint strategies parameterized by a window size $b \in \{\delta, 2\delta, \cdots, \nabla\}$. For a given hard-constraint strategy $S(b)$, let $H_t(b)$, $C_t(b)$, and $V_t(b)$ represent its inventory, cash, and total value, respectively. At the end of period t, the strategy $S(b)$ chooses a window of size b, denoted as $[a_t(b), a_t(b) + b]$, where $a_0(b) = P_0$ and $a_t(b)$ for $t \geq 1$ is determined by the following rules:

$$a_t(b) = \begin{cases} P_t - b & \text{if } P_t > a_{t-1}(b) + b \\ a_{t-1}(b) & \text{if } P_t \in [a_{t-1}(b), a_{t-1}(b) + b] \\ P_t & \text{if } P_t < a_{t-1}(b). \end{cases}$$

It then submits a limit buy order of one share at each price $P \in [\max\{a_t(b) - \delta(R - H_t(b)), \delta\}, a_t)$ and a limit sell order of one share at each price $P \in (a_t(b)+b, \min\{a_t(b)+b+\delta(H_t(b)+R), M\}]$. In this way, the strategy will buy no more than $R - H_t(b)$ shares or sell no more than $R + H_t(b)$ shares in period $t + 1$, which ensures that $|H_{t+1}| \leq R$ for all t.

Specially, if $R = +\infty$, our hard-constraint strategies degenerate into those presented in (Abernethy and Kale 2013), which we will refer to as **non-constraint strategies** hereafter.

Soft-Constraint Strategies In addition to imposing strict constraints on the inventory level, another approach for the market maker to controlling inventory risk is to dynamically adjust the quoted bid and ask prices in terms of the current inventory H_t. In the absence of consideration of the trade size and failure conditions, inventory risk should affect bid and ask prices, but not the size of the bid-ask spread

Algorithm 1: Hard-constraint strategy $S(b)$

Input: The window size b and the initial stock price P_0.

Initialize $a_0(b) := P_0$ and $H_0(b) := 0$. Submit limit order Q_0: $Q_0(P) = 1$ if $P \in [\max\{a_0(b) - \delta R, \delta\}, a_0(b))$, $Q_0(P) = -1$ if $P \in (a_0 + b, \min\{a_0 + b + \delta R, M\}]$, and $Q_0(P) = 0$ otherwise.
for $t = 1, 2, \cdots, T$ **do**
  Execute any limit orders from the previous period and observe the stock price P_t. The inventory position changes from $H_{t-1}(b)$ to $H_t(b)$.
  if $P_t > a_{t-1}(b) + b$ **then**
    $a_t(b) \leftarrow P_t - b$
  else if $P_t < a_{t-1}(b)$ **then**
    $a_t(b) \leftarrow P_t$
  else
    $a_t(b) \leftarrow a_{t-1}(b)$
  end if
  Submit limit order Q_t: $Q_t(P) = 1$ if $P \in [\max\{a_t(b) - \delta(R - H_t(b)), \delta\}, a_t(b))$, $Q_t(P) = -1$ if $P \in (a_t(b) + b, \min\{a_t(b) + b + \delta(H_t(b) + R), M\}]$, and $Q_t(P) = 0$ otherwise.
end for

(Amihud and Mendelson 1980; Stoll 1978; Grossman and Miller 1988). If the market maker has a long position in the stock, minimizing inventory risk is achieved by lowering both bid and ask prices. Contrarily, if she has a short position, inventory is controlled by raising both bid and ask prices. Therefore, building on the work of (Das 2005), we develop a class of soft-constraint strategies that adjust the bid and ask prices linearly based on the current inventory. Specifically, consider a class of soft-constraint strategies parameterized by a window size $b \in \{\delta, 2\delta, \cdots, \nabla\}$. At the end of period t, the strategy $S(b)$ selects a window of size b (i.e., $[a_t(b), a_t(b) + b]$), where $a_t(b) = P_t - b - \gamma H_t(b)$ and γ is a nonnegative parameter representing a risk-aversion coefficient. It then submits a limit buy order of one share at every price $P \in [\delta, a_t(b))$ and a limit sell order of one share at every price $P \in (a_t(b) + b, M]$. Note that when $\gamma H_t(b)$ is not a multiple of δ, it is rounded to the nearest multiple of δ toward $+\infty$ (resp., $-\infty$) if $H_t(b) > 0$ (resp., $H_t(b) < 0$) and then used to compute $a_t(b)$. For the case of $\gamma > \delta$, if the stock price in period t does not continue to move in the same direction as the previous period, all inventory will be liquidated at unfavorable prices. This inevitably results in significant losses in non-trend markets. Thus, we mainly focus on the case of $0 < \gamma \leq \delta$ hereafter.

2.3 The Regret of Adaptive Strategies

For each class of reference strategies, Denote by $\mathbb{B}$ the set of possible values of b and $V_T^{\mathcal{A}}$ the total value of the adaptive strategy's holdings using algorithm $\mathcal{A}$ at time T. The regret of an adaptive strategy using algorithm $\mathcal{A}$ at time T is defined as the total value of the best reference strategies in hindsight minus that of the adaptive strategy. Formally,

$$reg(\text{AS}) = \max_{b \in \mathbb{B}} V_T(b) - V_T^{\mathcal{A}}.$$

Algorithm 2: Soft-constraint strategy $S(b)$

Input: The window size b, risk-aversion coefficient γ, and the initial price P_0.

Initialize $a_0(b) := P_0 - b$. Submit limit order Q_0: $Q_0(P) = 1$ if $P \in [\delta, a_0(b))$, $Q_0(P) = -1$ if $P \in (a_0(b) + b, M]$, and $Q_0(P) = 0$ otherwise.
for $t = 1, 2, \cdots, T$ **do**
  Execute any limit orders from the previous period and observe the stock price P_t. The inventory position changes from $H_{t-1}(b)$ to $H_t(b)$.
  Update $a_t(b) \leftarrow P_t - b - \gamma H_t(b)$.
  Submit limit order Q_t: $Q_t(P) = 1$ if $P \in [\delta, a_t(b))$, $Q_t(P) = -1$ if $P \in (a_t(b) + b, M]$, and $Q_t(P) = 0$ otherwise.
end for

3 The Bounds of Gains

In the standard online learning model, the gain (or loss) in each period is typically assumed to lie within a fixed bounded interval. However, in our model, the upper and lower bounds of the gain depend on the inventory size, which is path-dependent and time-varying. In this section, we attempt to bound the difference between the inventories and further the difference between the gains of any two reference strategies within the same class. This constitutes our main technical contribution. Due to space limitations, all proofs are included in the appendix.

3.1 Hard-Constraint Strategies

Lemma 2. *For any hard-constraint strategy $S(b)$, if its position at the end of period t satisfies $|H_t(b)| < R$, then we have $H_t(b) - H_{t-1}(b) = \frac{1}{\delta}[a_{t-1}(b) - a_t(b)]$.*

For any two hard-constraint strategies $S(b^1)$ and $S(b^2)$, to avoid confusion, we will use the notations H_t^i, a_t^i, V_t^i, etc., to refer to $H_t(b^i)$, $a_t(b^i)$, $V_t(b^i)$, etc., with $i = 1$ or 2, respectively. We have the following result.

Lemma 3. *For any two hard-constraint strategies $S(b^1)$ and $S(b^2)$ with $b^1 < b^2$, denote by $[a_t^1, a_t^1 + b^1]$ and $[a_t^2, a_t^2 + b^2]$ the windows selected by $S(b^1)$ and $S(b^2)$ at the end of period t, respectively. Then we have $[a_t^1, a_t^1 + b^1] \subset [a_t^2, a_t^2 + b^2]$ for all t.*

Why is it more complicated to bound $H_t^1 - H_t^2$ for hard-constraint strategies? For non-constraint strategies presented in (Abernethy and Kale 2013) (i.e., hard-constraint strategies with $R = +\infty$), the condition in Lemma 2 is naturally satisfied. Thus, it is straightforward to prove that $H_t(b) = \sum_{j=1}^{t} H_j(b) - H_{j-1}(b) = \frac{1}{\delta}[a_0(b) - a_t(b)]$ by Lemma 2, and further that $H_t^1 - H_t^2 = \frac{1}{\delta}(a_t^2 - a_t^1) \in [\frac{b^1 - b^2}{\delta}, 0]$ by Lemma 3. However, the equation $H_t(b) = \frac{1}{\delta}[a_0(b) - a_t(b)]$ does not always hold for hard-constraint strategies with $R < +\infty$. The reason is that the inventory level may reach the limit of $\pm R$ before period t, thereby violating the condition in Lemma 2. This makes it more complicated to bound $H_t^1 - H_t^2$. To illustrate this, consider the following scenario with $P_0 = 1$, $b = \delta = 0.5$, and

t	P_t	$a_t(b)$	$\frac{1}{\delta}[a_0(b) - a_t(b)]$	$H_t(b)$	
				$R = +\infty$	$R = 2$
1	2	1.5	-1	-1	-1
2	3	2.5	-3	-3	-2
3	4	3.5	-5	-5	-2

Table 1: An example

$R = 2$. As shown in Table 1, if the stock price increases by one in each period, we have $a_0(b) = 1$, $a_3(b) = 3.5$, and $H_3(b) = -2 \neq \frac{1}{\delta}[a_0(b) - a_3(b)]$ for the hard-constraint strategy with $R = 2$. For general hard-constraint strategies, we have the following result.

Lemma 4. *For any two hard-constraint strategies $S(b^1)$ and $S(b^2)$ with $b^1 < b^2$, we have*

$$|H_t^1 - H_t^2| \leq \frac{1}{\delta}(b^2 - b^1)$$

for all t.

Proof. [Sketch] Define $BS_t^i = H_t^i - H_{t-1}^i - \frac{a_{t-1}^i - a_t^i}{\delta}$ with $i = 1$ or 2. For any two hard-constraint strategies $S(b^1)$ and $S(b^2)$ with $b^1 < b^2$, we first prove $0 \leq \sum_{j=1}^t BS_j^1 - BS_j^2 \leq \frac{1}{\delta}(b^2 - b^1)$ for all t. It is easy to verify $H_t^1 - H_t^2 = \sum_{j=1}^t (H_j^1 - H_{j-1}^1) - \sum_{j=1}^t (H_j^2 - H_{j-1}^2) = \frac{1}{\delta}(a_t^2 - a_t^1) + \sum_{j=1}^t BS_j^1 - BS_j^2$ by definition. Since $[a_t^1, a_t^1 + b^1] \subset [a_t^2, a_t^2 + b^2]$ for all t by Lemma 3, we have $b^1 - b^2 \leq a_t^2 - a_t^1 \leq 0$. It follows that $|H_t^1 - H_t^2| \leq \frac{1}{\delta}(b^2 - b^1)$ for all t.

□

We now consider N hard-constraint strategies. Let b^{min} and b^{max} be the minimum and maximum values of all $b \in \mathbb{B}$. Note that $b^{max} \leq \nabla$. With the assumption of $|P_t - P_{t-1}| \leq \nabla$, we can use Lemma 4 to bound the difference between the gains of any two hard-constraint strategies in each period. We have the following result.

Lemma 5. *Define $G^h = \frac{\nabla(b^{\max} - b^{\min})}{\delta} + \min(2R\nabla, \frac{\nabla^2}{\delta})$. For any two hard-constraint strategies $S(b^1)$ and $S(b^2)$ with $b^1, b^2 \in \mathbb{B}$, we have*

$$|(V_t^1 - V_{t-1}^1) - (V_t^2 - V_{t-1}^2)| \leq 2G^h$$

for all t.

3.2 Soft-Constraint Strategies

Lemma 6. *For any soft-constraint strategy $S(b)$, if $0 < \gamma \leq \delta$, we have*

$$-\frac{\nabla}{\gamma} \leq H_t(b) \leq \frac{\nabla - b}{\gamma}$$

for all t.

By Lemma 6, the following Corollary can be directly obtained.

Corollary 7. *For any two soft-constraint strategies $S(b^1)$ and $S(b^2)$ with $b^1 < b^2$, if $0 < \gamma \leq \delta$, we have*

$$|H_t^1 - H_t^2| \leq \frac{2\nabla - b^1}{\gamma}$$

for all t.

By Lemma 6 and Corollary 7, the difference between the gains of any two soft-constraint strategies in each period can be bounded. We thus have the following result.

Lemma 8. *Define $G^s = \frac{(2\nabla - b^{\min})(3\nabla - b^{\min})}{\gamma}$. If $0 < \gamma \leq \delta$, then for any two soft-constraint strategies $S(b^1)$ and $S(b^2)$ with $b^1, b^2 \in \mathbb{B}$, we have*

$$|(V_t^1 - V_{t-1}^1) - (V_t^2 - V_{t-1}^2)| \leq 2G^s$$

for all t.

4 The Regret Bounds of Adaptive Strategies

Thus far, we have introduced two classes of constraint strategies. Within each class, there are N strategies whose gains in each period are bounded by Lemma 5 or Lemma 8. We now attempt to design an *adaptive strategy* for each class via online learning such that it can earn almost as much as the best fixed spread strategy that considers the inventory constraints, respectively.

4.1 The Adaptive Strategies

Following the work of (Abernethy and Kale 2013), we view each constraint strategy $S(b)$ within either the first or second class as an expert and run an online learning algorithm for learning with expert advice to these strategies. Denote by $w_t(b)$ the weight assigned to the reference strategy $S(b)$ by the online learning algorithm at the end of each period $t-1$. Each reference strategy $S(b)$ is executed in proportion to its assigned weight $w_t(b)$ in period t. Furthermore, at the end of each period t, the adaptive strategy submits a market buy order of $\sum_{b \in \mathbb{B}} H_{t-1}(b)[w_t(b) - w_{t-1}(b)]$ shares such that its inventory equals $\sum_{b \in \mathbb{B}} H_t(b) w_t(b)$. The specific strategy is illustrated in Algorithm 3.

We consider two classic online learning algorithms to develop the adaptive strategies: the multiplicative weights (MW) (Littlestone and Warmuth 1994) and follow-the-perturbed-leader (FPL) (Kalai and Vempala 2005). The adaptive strategies based on MW and FPL are referred to as ASMW and ASFPL, respectively. The adaptive strategy starts with an initial weight $w_0(b) = \frac{1}{N}$ for each strategy $S(b)$. At the end of period t, ASMW with time-varying parameters η_t updates the weight for each $b \in \mathbb{B}$ as follows:

$$w_{t+1}(b) = \frac{w_t(b) \mathrm{e}^{\eta_t [V_t(b) - V_{t-1}(b)]}}{\sum_{b' \in \mathbb{B}} w_t(b') \mathrm{e}^{\eta_t [V_t(b') - V_{t-1}(b')]}}.$$

While ASFPL with parameters η updates the weight for each $b \in \mathbb{B}$ as follows:

$$w_{t+1}(b) = Pr[V_t(b) + f(b) \geq V_t(b') + f(b') \text{ for } \forall b' \in \mathbb{B}],$$

where $f(b)$ and $f(b')$ are samples from the exponential distribution with mean $1/\eta$.

Algorithm 3: The adaptive strategy

1: Run every constraint strategy S(b) in class one or two parallel such that $H_t(b)$, $C_t(b)$ and $V_t(b)$ for each strategy can be computed at the end of each period t.
2: Start an online learning algorithm $\mathcal{A}$ with one expert corresponding to each strategy $S(b)$. Denote by $w_t(b)$ the weight assigned to $S(b)$ at the end of $t-1$th period.
3: **for** $t = 1, 2, \cdots, T$ **do**
4: Execute the limit orders from the previous period: a $w_t(b)$ weighted combination of the limit orders of the reference strategies.
5: Submit and execute a market buy order of $\sum_{b\in\mathbb{B}} H_{t-1}(b)[w_t(b) - w_{t-1}(b)]$ shares at the price P_t.
6: Compute $\Delta V_t(b)$ for each strategy $S(b)$.
7: Update $w_{t+1}(b)$ for each strategy $S(b)$ from $\mathcal{A}$ according to $\Delta V_t(b)$.
8: Submit a $w_{t+1}(b)$ weighted combination of the limit orders of the strategies S(b).
9: **end for**

For a given online learning algorithm $\mathcal{A}$, denote by $C_t^{\mathcal{A}}$ and $H_t^{\mathcal{A}}$ the amount of the cash and inventory held by the adaptive strategy based on $\mathcal{A}$ at the end of period t. Thus, the total value of the adaptive strategy is given by $V_t^{\mathcal{A}} = H_t^{\mathcal{A}} P_t + C_t^{\mathcal{A}}$. Since $H_t^{\mathcal{A}} = \sum_{b\in\mathbb{B}} H_t(b) w_t(b)$, for all t we have

$$|H_t^{\mathcal{A}}| \leq R$$

for adaptive strategies on hard-constraint strategies, and

$$-\frac{\nabla}{\gamma} \leq H_t^{\mathcal{A}} \leq \frac{\nabla - b^{\min}}{\gamma}$$

for adaptive strategies on soft-constraint strategies. This demonstrates that our adaptive strategies, whether based on soft- or hard-constraint strategies, effectively control inventory risk.

4.2 The Regret Bounds

Define $reg(\mathcal{A}) = \max_{b\in\mathbb{B}} V_T(b) - \sum_{t=1}^{T}\sum_{b\in\mathbb{B}}[V_t(b) - V_{t-1}(b)]w_t(b)$, which is the regret of the underlying algorithm $\mathcal{A}$. We next bound the regret of the adaptive strategy in terms of $reg(\mathcal{A})$. Denote by $H^\star$ the upper bound of the difference between the inventory of any two reference strategies within the same class in each period. We have the following result.

Lemma 9. *The regret of the adaptive strategy satisfies:*

$$reg(AS) \leq reg(\mathcal{A}) + H^\star \nabla \sum_{t=2}^{T}\sum_{b\in\mathbb{B}} |w_t(b) - w_{t-1}(b)|.$$

Adaptive Market Making on Hard-Constraint Strategies. For the hard-constraint strategies, we have $H^\star = \frac{1}{\delta}(b^{\max} - b^{\min})$ by Lemma 4. Thus, the key of deriving the regret bound of the adaptive strategy is to bound $\sum_{b\in\mathbb{B}} |w_t(b) - w_{t-1}(b)|$. Define $c^h = \frac{9\nabla(b^{\max}-b^{\min})}{\delta} + 5\min(2R\nabla, \frac{\nabla^2}{\delta})$. We have the following results.

.

Theorem 10. *If $\eta_t = \frac{1}{G^h}\min\{\sqrt{\frac{\ln N}{t}}, \frac{1}{2}\}$, then the regret of ASMW on hard-constraint strategies is bounded from above by $2c^h\sqrt{T\ln N}$.*

Theorem 11. *If $\eta = \frac{1}{G^h}\sqrt{\frac{\ln N}{T}}$, then the regret of ASFPL on hard-constraint strategies is bounded from above by $c^h\sqrt{T\ln N}$.*

Adaptive Market Making on Soft-Constraint Strategies. For the soft-constraint strategies, we have $H^\star = \frac{2\nabla - b^{\min}}{\gamma}$ by Corollary 7. The proof of the regret bounds of the adaptive strategies on soft-constraint strategies is similar to that on hard-constraint strategies. Define $c^s = (2\nabla - b^{\min})(19\nabla - 5b^{\min})/\gamma$. We have the following results.

Theorem 12. *If $\eta_t = \frac{1}{G^s}\min\{\sqrt{\frac{\ln N}{t}}, \frac{1}{2}\}$, then the regret of ASMW on soft-constraint strategies is bounded from above by $2c^s\sqrt{T\ln N}$.*

Theorem 13. *If $\eta = \frac{1}{G^s}\sqrt{\frac{\ln N}{T}}$, then the regret of ASFPL on soft-constraint strategies is bounded from above by $c^s\sqrt{T\ln N}$.*

5 Experiments

The performance of our market making algorithms is examined via data from the Chinese stock exchange. We collect from http://www. cdqianlong.com the tick-by-tick data of the CSI 500 index component stocks on each trading day from December 1, 2022 to December 31, 2022. The data include high-frequency information, such as intraday transaction time, traded price, and volume for each stock from the opening time, 9:25 a.m., to the closing time, 3:00 p.m., on each trading day. Stocks with the daily high price minus low price of less than 6 cents are excluded, which leaves 10,882 observed stock price paths in the sample period.[1] The number of trades in each path (i.e., T) ranges from 1,005 to 406,363.

The experimental setup is as follows. We treat the implementation of an adaptive strategy on one stock of a single day as one experiment. Since the tick size δ is one cent, the window size b is specified in cents. The set of possible values for b is $\mathbb{B} = \{1, 2, 4, 6, 8, 10, 15, 20\}$, resulting in $N = 8$ reference strategies in our adaptive market making. We set $R \in \{10, 20, 30, 40, 50\}$ for the hard-constraint strategies and $\gamma \in \{0.1\delta, 0.2\delta, 0.3\delta, 0.4\delta, 0.5\delta\}$ for the soft-constraint strategies. The learning rate is set as $\eta_t = \min\{1, 4\sqrt{\ln N/t}\}/G_t$ for ASMW and $\eta = 4\sqrt{\ln N/T}/G_t$ for ASFPL, where $G_t = \max_{1\leq s\leq t, b^1, b^2\in\mathbb{B}} |V_s(b^1) - V_{s-1}(b^1) - V_s(b^2) + V_{s-1}(b^2)|$. The weight w_t in ASFPL is estimated by averaging 100 independently drawn perturbations.

[1] Our experiments are much more comprehensive than those in (Abernethy and Kale 2013) which are carried out on only three

Algorithm	Hard-constraint strategies					Soft-constraint strategies				
	R (absolute inventory limit)					γ (risk-aversion coefficient)				
	10	20	30	40	50	0.1δ	0.2δ	0.3δ	0.4δ	0.5δ
ASMW	3.32	3.54	3.11	3.48	3.05	0.71	0.73	0.68	0.65	0.74
ASFPL	1.74	1.78	1.72	1.75	1.69	0.35	0.32	0.17	0.26	0.29

Table 2: Our adaptive strategies with inventory constraints could indeed achieve no-regret. This table reports $\mathcal{G}^{\min}$ of adaptive strategies with inventory constraints, which is defined as the minimum of the realized average regret per period of an adaptive strategy minus its theoretical upper bound implied by theorems in Section 4 across experiments. The positive values of $\mathcal{G}^{\min}$ imply that, for any stock path and algorithm setup, the realized regret is always lower than the corresponding theoretical regret bound.

Algorithm	Index	Hard-constraint strategies					Soft-constraint strategies					Non-constraint strategies in (Abernethy and Kale 2013)
		R (absolute inventory limit)					γ (risk-aversion coefficient)					
		10	20	30	40	50	0.1δ	0.2δ	0.3δ	0.4δ	0.5δ	
ASMW	$\overline{E}$	0.12	0.14	0.15	0.17	0.18	0.14	0.12	0.1	0.08	0.06	0.19
	$\sigma(E)$	0.57	0.61	0.67	0.72	0.8	0.62	0.55	0.51	0.47	0.4	1.44
	Sharp ratio	**0.21**	**0.23**	**0.22**	**0.24**	**0.23**	**0.23**	**0.22**	**0.20**	**0.17**	**0.15**	0.13
	$\overline{H}$	5.52	8.52	11.72	14.08	18.25	12.15	10.22	8.53	6.95	5.71	38.42
ASFPL	$\overline{E}$	-0.74	-0.68	-0.57	-0.42	-0.34	-0.39	-0.42	-0.47	-0.51	-0.55	1.08
	$\sigma(E)$	4.35	5.11	6.48	7.06	7.83	6.85	5.98	5.01	4.12	3.17	12.84
	Sharp ratio	-0.17	-0.13	-0.09	-0.06	-0.04	-0.06	-0.07	-0.09	-0.12	-0.17	0.08
	$\overline{H}$	5.33	8.47	11.56	14.29	18.01	13.49	10.87	8.92	7.25	6.01	46.32

Table 3: The gain and risk of adaptive strategies with(out) inventory constraints. E denotes the average gain per period of an adaptive strategy in one experiment. $\overline{E}$ and $\sigma(E)$ are the mean and standard deviation of E across experiments. $\overline{H}$ denotes the mean of the final absolute inventory held by the adaptive strategy across experiments. Both $\overline{E}$ and $\sigma(E)$ are measured in units of cents. Bolded values indicate higher Sharp ratios than (Abernethy and Kale 2013). These results show that ASMW on both constraint strategies outperforms adaptive strategies on non-constraint strategies in terms of higher Sharp ratios.

5.1 Validation of No-Regret

We first compare the performance of our adaptive strategies with that of the best reference strategy in hindsight. Given a stock price path, let $reg^R = (\max_{b\in\mathbb{B}} V_T(b) - V_T^{\mathcal{A}})/T$, which is the *realized average regret* per period of the adaptive strategy based on an algorithm $\mathcal{A}$ in one experiment with T periods, and $\mathcal{G}$ be the corresponding theoretical upper bound of average regret per period implied by theorems in Section 4 minus reg^R. The distributions of reg^R are presented in the appendix. To valid the no-regret of our adaptive strategies, we define $\mathcal{G}^{\min} = \min_{\text{stock path}} \mathcal{G}$ whose values are reported in Table 2. The left-hand side of Table 2 concerns our adaptive strategies on the hard-constraint strategies. The positive values of $\mathcal{G}^{\min}$ imply that for any stock path and algorithm setup, the realized regret is always lower than the corresponding theoretical regret bound. This demonstrates that our adaptive strategies on hard-constraint strategies could indeed achieve no-regret. The comparison of adaptive strategies on the soft-constraint strategies reported in the right-hand side of Table 2 shows the same results.

stocks.

5.2 Comparison with (Abernethy and Kale 2013)

The gain and inventory risk control of our adaptive strategies are further examined by comparing them with those on non-constraint strategies presented in (Abernethy and Kale 2013). Let $E = V_T^{\mathcal{A}}/T$, which is the average gain per period of the adaptive strategy based on an algorithm $\mathcal{A}$ in one experiment with T periods, and H be its absolute inventory at the end of period T. Define Sharp ratio= $\overline{E}/\sigma(E)$ to measure the risk-adjusted gain, where $\overline{E}$ and $\sigma(E)$ are the mean and standard deviation of E across experiments, respectively. The $\overline{E}$, $\sigma(E)$, Sharp ratio, and the mean of H across experiments (i.e., $\overline{H}$) for adaptive strategies on constraint and non-constraint strategies in the full sample are reported in Table 3. These results show that our adaptive strategies on constraint strategies outperform those on non-constraint strategies in terms of risk control. First, adaptive strategies on both constraint strategies have lower $\overline{H}$ and $\sigma(E)$ than those on non-constraint strategies. It is worth noting that, consistent with Lemma 6, $\overline{H}$ for adaptive strategies on soft-constraint strategies decreases as γ increases from 0.1δ to 0.5δ. Furthermore, with respect to the gain, ASMW outperforms ASFPL and makes a profit for all R and γ cases. Note that inventory control inevitably leads to a decrease in

gain in no-trend paths, which account for the vast majority of our sample. Thus, adaptive strategies on non-constraint strategies unsurprisingly have the highest gain in the full sample. Finally, for any R and δ, ASMW on both constraint strategies outperforms adaptive strategies on non-constraint strategies in terms of higher Sharp ratios.

To further examine the performance of our strategies during the upward and downward markets, we classify all the price paths in our sample into three subsamples. Specifically, for a given stock price path on a trading day, denote by P_0, $P_{\min}$, $P_{\max}$, and P_T the daily opening, low, high, and closing prices, respectively. The price path is called under an uptrend if $\frac{P_T}{P_0} \geq 1.03$ and $\frac{P_T - P_0}{P_{\max} - P_{\min}} \geq 0.8$, under a downtrend if $\frac{P_T}{P_0} \leq 0.97$ and $\frac{P_T - P_0}{P_{\max} - P_{\min}} \leq -0.8$, and no trend otherwise. The comparisons in subsamples are reported in the appendix. In both uptrend and downtrend samples, adaptive strategies on both constraint strategies have larger $\overline{E}$, lower $\sigma(E)$, and larger Sharp ratio than those on non-constraint strategies for all R and γ. Specially, adaptive strategies on both constraint strategies almost always make a profit in both subsamples. In contrast, adaptive strategies on non-constraint strategies make a loss in both subsamples.

6 Conclusions

In this paper, we introduce soft- and hard-constraint strategies for inventory risk control. Two classic online learning algorithms, namely, MW and FPL, are used to develop adaptive strategies on both constraint strategies that achieve no-regret. We want to emphasize that our results are technically interesting. The reason is that in the standard online learning framework, the gain in each period is assumed to lie in a fixed constant interval and does not depend on any state variables. In contrast, the gain in our model depends on the past inventory, which is path-dependent and time-varying. Thus, we have to bound the difference between the gains of any two reference strategies, which is more challenging than the work of (Abernethy and Kale 2013). Furthermore, different from most of the existing work (Avellaneda and Stoikov 2008; Spooner and Savani 2020), our adaptive strategies are model-free in the sense that no assumptions on the dynamics of the LOB and stock price are required.

Our work leaves a few interesting open problems. In our model, only one parameter (i.e., the spread) is learned. In a more general model, the distance between the bid and middle price and that between the ask and middle price can be learned as two independent parameters of reference strategies. In addition, there are some other engineering problems in finance, such as the optimal execution with path-dependent constraints, which can also be investigated within the framework of online learning.

Acknowledgments

The authors are grateful for many valuable comments from the anonymous reviewers. This study was partially supported by the National Natural Science Foundation of China (Grant Numbers 11501464, 11761141007, 71971177, 72342012) and Leshan Normal University Scientific Research Start-up Project for Introducing High-level Talents (Grant Number RC2024031).

References

Abernethy, J.; and Kale, S. 2013. Adaptive market making via online learning. *Advances in Neural Information Processing Systems*, 26.

Amihud, Y.; and Mendelson, H. 1980. Dealership market: Market-making with inventory. *Journal of Financial Economics*, 8(1): 31–53.

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

Baldacci, B.; Manziuk, I.; Mastrolia, T.; and Rosenbaum, M. 2019. Market making and incentives design in the presence of a dark pool: a deep reinforcement learning approach. *arXiv preprint arXiv:1912.01129*.

Cartea, Á.; Donnelly, R.; and Jaimungal, S. 2017. Algorithmic trading with model uncertainty. *SIAM Journal on Financial Mathematics*, 8(1): 635–671.

Cartea, Á.; Jaimungal, S.; and Penalva, J. 2015. *Algorithmic and high-frequency trading*. Cambridge University Press.

Chakraborty, T.; and Kearns, M. 2011. Market making and mean reversion. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, 307–314.

Chan, N. T.; and Shelton, C. 2001. An electronic market-maker.

Das, S. 2005. A learning market-maker in the Glosten–Milgrom model. *Quantitative Finance*, 5(2): 169–180.

Ding, W.; Qin, T.; Zhang, X.-D.; and Liu, T.-Y. 2013. Multi-armed bandit with budget constraint and variable costs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, 232–238.

Fodra, P.; and Labadie, M. 2012. High-frequency market-making with inventory constraints and directional bets. *Quantitative Finance*.

Ganesh, S.; Vadori, N.; Xu, M.; Zheng, H.; Reddy, P.; and Veloso, M. 2019. Reinforcement learning for market making in a multi-agent dealer market. *arXiv preprint arXiv:1911.05892*.

Gašperov, B.; and Kostanjčar, Z. 2021. Market making with signals through deep reinforcement learning. *IEEE Access*, 9: 61611–61622.

Glosten, L. R.; and Milgrom, P. R. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1): 71–100.

Grossman, S. J.; and Miller, M. H. 1988. Liquidity and market structure. *The Journal of Finance*, 43(3): 617–633.

Guéant, O. 2017. Optimal market making. *Applied Mathematical Finance*, 24(2): 112–154.

Guéant, O.; Lehalle, C.-A.; and Fernandez-Tapia, J. 2013. Dealing with the inventory risk: a solution to the market making problem. *Mathematics and Financial Economics*, 7: 477–507.

Guéant, O.; and Manziuk, I. 2019. Deep reinforcement learning for market making in corporate bonds: beating the curse of dimensionality. *Applied Mathematical Finance*, 26(5): 387–452.

Guilbaud, F.; and Pham, H. 2013. Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, 13(1): 79–94.

Hambly, B.; Xu, R.; and Yang, H. 2021. Recent advances in reinforcement learning in finance. *arXiv preprint arXiv:2112.04553*.

Kalai, A.; and Vempala, S. 2005. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3): 291–307.

Littlestone, N.; and Warmuth, M. K. 1994. The weighted majority algorithm. *Information and Computation*, 108(2): 212–261.

Mannor, S.; Tsitsiklis, J. N.; and Yu, J. Y. 2009. Online Learning with Sample Path Constraints. *Journal of Machine Learning Research*, 10(3).

Obizhaeva, A. A.; and Wang, J. 2013. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1): 1–32.

Patel, Y. 2018. Optimizing market making using multi-agent reinforcement learning. *arXiv preprint arXiv:1812.10252*.

Paternain, S.; Lee, S.; Zavlanos, M. M.; and Ribeiro, A. 2020. Distributed constrained online learning. *IEEE Transactions on Signal Processing*, 68: 3486–3499.

Spooner, T.; Fearnley, J.; Savani, R.; and Koukorinis, A. 2018. Market Making via Reinforcement Learning. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 434–442.

Spooner, T.; and Savani, R. 2020. Robust market making via adversarial reinforcement learning. In *IJCAI International Joint Conference on Artificial Intelligence*, volume 2021, 4590–4596.

Stoll, H. R. 1978. The supply of dealer services in securities markets. *The Journal of Finance*, 33(4): 1133–1151.

Zhang, G.; and Chen, Y. 2020. Reinforcement learning for optimal market making with the presence of rebate. *Available at SSRN 3646753*.