

Optimal Strategies of High Frequency Traders *

Jiangmin Xu[†]

First Draft: November 15, 2013

This Draft: November 14, 2015

ABSTRACT

This paper develops a model of the optimal strategies of high-frequency traders (HFTs) to rationalize their pinging activities. Pinging is defined as limit orders submitted inside the bid-ask spread that are cancelled shortly. The HFT in my model uses pinging to control his inventory or chase the short-term price momentum without learning or manipulative motives. I show that the model can match over 70% of the pinging activities observed in the data, and demonstrate how the HFT's pinging behaviors vary across stocks. The result confirms that pinging is rationalizable as part of the dynamic trading strategies of HFTs.

*I am deeply indebted to Yacine Aït-Sahalia for his advice, encouragement, and insights. I am also very grateful to Harrison Hong for his many suggestions. I also thank Valentin Haddad, Joel Hasbrouck, Gideon Saar, Mehmet Saglam, Christopher Sims, David Sraer, Wei Xiong, as well as seminar participants at Princeton University, the 2014 Princeton-QUT-SMU conference, the University of Warwick, Shanghai Advanced Institute of Finance, Peking University, Hong Kong University of Science and Technology, and the 2015 American Finance Association Meeting in Boston for helpful comments. In addition, I thank NASDAQ for providing the access to their historical TotalView-ITCH real-time limit order book message feed database.

[†]Guanghua School of Management, Peking University (email: jiangminxu@gsm.pku.edu.cn).

1 Introduction

A recent and ongoing heated debate concerns high-frequency traders and high-frequency trading activities (HFT stands for high-frequency trader/trading thereafter). The interest in this subject has grown significantly after the Flash Crash, because HFTs appear as a black-box mystery to the general public as well as to the academic world.¹ One type of HFT activity that has attracted a great deal of attention due to HFTs' speed advantage is the so-called pinging activity. Pinging, or the most aggressive fleeting order activity, is defined as the submission of limit orders inside the bid-ask spread that are cancelled very shortly.² These activities occur in the scale of seconds or milliseconds with extremely low latency, which is “the hallmark of proprietary trading by HFTs” (Hasbrouck and Saar (2013)).³

Regulators have expressed concerns over such pinging activities, the main one being that pinging done by HFTs might be manipulative. As pointed out in a concept release of the Securities and Exchange Commission (SEC), HFTs could use pinging orders to detect and learn about hidden orders inside the spread.⁴ Hidden orders are limit orders completely non-observable to other market participants. They have become increasingly popular in the past five years or so and are allowed by many stock exchanges around the world nowadays.⁵ This learning on hidden orders through pinging would enable HFTs to ascertain the existence of potential large trading interest in the market. Consequently, they would be able to trade ahead and capture a price movement in the direction of the large trading interest.⁶

¹For a comprehensive review of the Flash Crash, see the studies by Kirilenko, Kyle, Samadi, and Tuzun (2014) and Easley, López de Prado, and O'Hara (2011). For an illustration of the general public's interest in high-frequency trading, see Lewis (2014)'s New York Times Best Seller book: *Flash Boys: A Wall Street Revolt*.

²Existing studies (e.g. Hasbrouck and Saar (2009)) have used the term “fleeting orders” to denote submissions and subsequent quick cancellations of limit orders in general. Therefore, to be consistent with the literature, I use the term “pinging” throughout this paper to denote the most aggressive fleeting order activities, i.e. fleeting orders that take place inside the bid-ask spread.

³To give an illustration of the speed of such pinging activities, Hautsch and Huang (2012a) find that the median cancellation time is below one second for limit orders submitted inside the spread on NASDAQ.

⁴SEC concept release on equity market structure, January 21, 2010.

⁵According to Hautsch and Huang (2012b), on average over 20% of trades on NASDAQ are executed against hidden orders in October 2010.

⁶This type of front-running behavior and the worries associated with it are also described in Lewis (2014)'s book of *Flash Boys*.

However, despite the large amount of empirical evidence on pinging activities, few studies have shown *how* they can arise as a result of the optimal trading behaviors of HFTs. Given the public speculations and concerns surrounding such pinging activities, it is instructive to have an understanding on the nature of these activities, and my paper attempts to address this issue from a theoretical perspective.

In this paper, I aim to rationalize the pinging activity levels observable in the data through a theoretical setup without manipulative elements. It develops a continuous-time model of the optimal trading strategies of HFTs absent any learning or strategic feedback effects. The model exploits two well-known forces in the existing literature: inventory control (e.g., Ho and Stoll (1981)) and momentum (trend) chasing (e.g., Hirschey (2013)).

To achieve my purpose, I first incorporate the existence of hidden orders inside the bid-ask spread into a continuous-time model along the lines of Ho and Stoll (1981). As a result, the HFT in my model would ping for hidden orders inside the spread as a cheaper way to control his inventory compared to using market orders. This would produce pinging orders that execute against hidden orders. However, with inventory control as the sole motive for pinging, it is not possible to obtain a large number of *cancelled* pinging orders at the same time. The reason is that when the HFT uses pinging to control his inventory, he intends to fill his pinging orders and is not incentivized to cancel many of them.

To tackle this problem, I then introduce short-term price momentum into the model as a channel for cancelled pinging orders to occur. The momentum is modelled through the predictive power of depth imbalances for the directions of price movements. When the HFT sees a large depth imbalance and anticipates a likely directional price move, he could use pinging orders as directional bets to chase the price momentum. Moreover, if there is a subsequent large change in the depth imbalance, the HFT would cancel his pinging orders and adjust his strategy according to the variation in the price momentum. Hence the model will now give rise to both submissions and cancellations of pinging orders because of the HFT's momentum-chasing behaviors.

The model is solved numerically because of the number of state variables, and I use historical order book message feed data from NASDAQ to reconstruct limit order books and estimate model parameters.⁷ The optimal HFT strategies are characterized based on the viscosity solution to my model and the model parameter estimates.⁸ There are two main findings. First, for stocks whose order books have high depths with relatively stable movements and whose spreads tend to be narrow, approximately 20% of the HFT's optimal strategies are attributable to ping-pong. On the other hand, for stocks that tend to have low order-book depths, volatile order-book movements and wide spreads, ping-pong accounts for nearly 50% of the HFT's optimal strategies. These ping-pong percentages from the model are proven to match over 70% of the observable ping-pong activities in the data. Thus it implies that the ping-pong occurrences observed in reality can be mostly rationalized by my model without any learning or manipulative components.

Second, I show that the mechanism of momentum chasing plays a much more important role than that of inventory control in rationalizing ping-pong activities for low-depth and wide-spread stocks. In contrast, the two mechanisms carry around similar weights in rationalizing ping-pong activities for high-depth and narrow-spread stocks. Moreover, both mechanisms are proven to be necessary for the model to rationalize the ping-pong activities found in the data.

The result here in this paper that HFTs' ping-pong activities can be rationalized in a dynamic model without manipulation is reminiscent of the finding by Baruch and Glosten (2013). They demonstrate in a repeated-game setting that flickering quotes (rapid submissions and cancellations of limit orders) are an equilibrium outcome without manipulation. Their idea is that limit-order traders use flickering quotes to mitigate the undercutting risk of their orders when providing liquidity to the market, hence flickering quotes are not the result of these traders trying to manipulate the market.

However, my model differs from Baruch and Glosten (2013)'s in twofold. First, the

⁷I am grateful to NASDAQ for providing me with the access to their historical TotalView-ITCH real-time limit order book message feed database.

⁸The concept of viscosity solution is explained in the next Section.

HFT's motives of pinging are different. In my model, the HFT uses pinging not because he is concerned about his limit orders being undercut, but rather to control his inventory or chase the price momentum. Second, and more importantly, my model offers an explanation on how the HFT's pinging behaviors would *change* across different types of stocks because of the variations in the HFT's optimal trading strategies. This result is not seen in Baruch and Glosten (2013). Therefore, my model not only gives a rationale for the empirical findings in Hasbrouck and Saar (2009, 2013), but also shows how the inventory-control and momentum-chasing incentives of HFTs can lead to different pinging activities among stocks. This is my main contribution to the literature.

Beyond pinging rationalization, my model also generates a couple of additional auxiliary predictions. They concern about the directions of pinging activities and the frequencies of cancelled pinging activities with regard to the depth imbalances of order books. These predictions are also found to be largely consistent with the data, which further suggests that pinging is rationalizable as part of the dynamic trading strategies of HFTs.

The paper is organized as follows. I describe the related literature in Section 2. I show the model and my numerical solution in Section 3 and 4. Parameter estimations are depicted in Section 5. I present and discuss the main findings in Section 6. Finally, Section 7 concludes.

2 Related Literature

There is an important line of empirical studies (see, e.g., Brogaard, Hagströmer, Nordén, and Riordan (2014), Brogaard, Hendershott, and Riordan (2014), Hagströmer and Nordén (2013), Hasbrouck and Saar (2013), Hendershott and Riordan (2013), Hendershott, Jones, and Menkveld (2011) and Menkveld (2013)) that try to understand the effects of high-frequency and algorithmic trading activities on market quality.

On one hand, the majority of the empirical studies show that, on balance, HFTs are beneficial for market quality. For instance, Brogaard, Hendershott, and Riordan (2014) find

that HFTs enhance price discovery and market efficiency on NASDAQ. Hasbrouck and Saar (2013) also show that increased HFTs' low-latency activities are associated with lower posted and effective spreads, lower short-term volatility and increased market depth. Additionally, Hendershott, Jones, and Menkveld (2011) find that algorithmic tradings improve liquidity and make quotes more informative for NYSE stocks.

On the other hand, there are empirical analyses demonstrating that some HFT strategies are speculative and anticipatory in nature. For example, Hirschey (2013) finds that HFTs on NASDAQ tend to anticipate future order flows and trade ahead of them by taking liquidity from the market. Moreover, Kirilenko, Kyle, Samadi, and Tuzun (2014) show that HFTs exacerbated market volatility during the Flash Crash by trading in the direction of the downward price spiral. In addition, Baron, Brogaard, and Kirilenko (2014) find that aggressive, liquidity-taking HFTs earn short term profits at the expense of other market participants.

This paper focuses on the optimal trading strategies of HFTs and their pinging activities in particular, and attempts to show that such activities are rational and non-manipulative. It builds on the work of Ho and Stoll (1981), Avellaneda and Stoikov (2008), and Guilbaud and Pham (2013). In their models, an HFT performs the role of a pure market maker by optimally posting limit orders in the market. When allowed, the HFT would trade via market orders solely for the purpose of inventory control.

Compared to these previous configurations, my model makes two contributions. Firstly, hidden orders are introduced into the limit order book. Hence the HFT can ping to take the hidden liquidity when his limit orders improve the best bid or ask prices and are submitted inside the spread. This captures the idea of pinging strategies as identified in Hasbrouck and Saar (2009)'s empirical study. Secondly, I model the order book's depth imbalance at the best quotes as a stochastic process that can forecast the movement of the mid-price and the existence of hidden orders. Thus the HFT can use the depth imbalance as an imperfect signal to anticipate the likely price movement. Therefore, the HFT will utilize pinging or market orders to conduct directional tradings and chase the short-term price momentum.

This captures the idea of anticipatory strategies as identified in Hirschey (2013).

Since market orders are impulse controls in continuous time, the value function of the HFT in my model is not necessarily differentiable or even continuous everywhere. Thus I will apply the viscosity solution technique to my model. The viscosity solution concept, originated in Crandall and Lions (1983), is a generalization of the classical solution concept to Hamilton-Jacobi-Bellman equations (or partial differential equations in general) where value functions might not be everywhere differentiable or continuous. I will discuss the setup of my model and its solution in detail in the next two Sections.

3 The Model

The economy is defined in a continuous-time, finite horizon $\mathcal{T} = [0, T]$. There is a high-frequency trader (HFT) who trades a single risky stock in a limit order book (LOB) using either limit or market orders. The LOB has a tick size δ , so that prices of all orders are in multiples of δ .⁹ Characteristics of the LOB, such as the spread and the mid-price, evolve stochastically and are exogenous to the HFT.¹⁰

3.1 Processes of Limit Order Book Characteristics

I will fix a filtered probability space $(\Omega, \mathcal{O}, \mathbb{O}, \mathbb{P})$, $\mathbb{O} = (\mathcal{O}_t)_{t \geq 0}$ satisfying the usual conditions, so that all stochastic processes and random variables are defined on $(\Omega, \mathcal{O}, \mathbb{O}, \mathbb{P})$.

⁹Notice it is only the prices of all limit and market orders, not all prices, that are in multiples of δ , since the mid-price can be in multiples of $\delta/2$. As will be seen in Section 3.1, this essentially means that my model does not allow for hidden orders sitting exactly at the mid-price inside the spread when the mid-price is a multiple of $\delta/2$ (or when the spread is either δ or 3δ).

¹⁰Given an asset, a limit order of size q at price p is an order to buy or sell q units of the asset at the specified price p ; its execution occurs only when it meets a counterpart market order. The depth at a particular price p is the total number of accumulated buy or sell limit orders at that price. The best bid (resp. ask) price is the highest (resp. lowest) price among all active buy (resp. sell) limit orders in the order book. The spread is the difference between the best ask price and the best bid price, which is strictly positive. The mid-price is the mid-point between the best bid and the best ask price. A market order of size q is an order to buy q units of the asset at the best ask price or sell q units of the asset at the best bid price; its execution is immediate. For a more detailed explanation of limit order book variables, please refer to the non-technical survey of Gould, Porter, Williams, McDonald, Fenn, and Howison (2013).

To start with, the depth imbalance F at the best quotes is defined to be the log of the size of the depth at the best bid price minus that at the best ask price, so that $F > 0$ ($F < 0$) means bid (ask) side imbalance. It follows an Ornstein-Uhlenbeck process with mean zero:

$$dF_t = -\alpha_F F_t dt + \sigma_F dW_t, \quad (1)$$

where W is a standard Brownian motion, α_F measures the speed of mean-reversion and σ_F is a constant volatility parameter.

Next, let

$$S = \{S_t\}_{t \geq 0} \quad (2)$$

denote the bid-ask spread of the LOB. It follows a continuous-time Markov chain on the state space $\mathbb{S} = \{\delta, 2\delta, 3\delta\}$ with a constant jump intensity λ^S . The probability transition matrix of S is denoted by $\rho = (\rho_{ij})_{1 \leq i, j \leq 3}$, with $\rho_{ii} = 0$.

Furthermore, the mid-price P of the stock is assumed to follow a pure jump process:

$$dP_t = dJ_{1t} + dJ_{2t}, \quad P_t \in \mathcal{P} \text{ and } \mathcal{P} = \frac{\delta}{2} \cdot \mathbb{N}^+. \quad (3)$$

The first component, J_{1t} , has a constant jump intensity λ_1^J and jump sizes equal to $\delta/2$ and $-\delta/2$ with probabilities $\psi_1(F_t)$ and $1 - \psi_1(F_t)$ respectively. The second component, J_{2t} , has a constant jump intensity λ_2^J and jump sizes equal to δ and $-\delta$ with probabilities $\psi_2(F_t)$ and $1 - \psi_2(F_t)$ respectively.¹¹ The functions $\psi_i : \mathbb{R} \mapsto [0, 1]$ are assumed to have the form

$$\psi_i(u) = 1/(1 + \exp(-\beta_i u)), \quad \text{for } i = 1, 2, \quad (4)$$

where β_1, β_2 are positive constants. In addition, J_1 and J_2 are independent. This means the

¹¹The best bid and best ask prices are thus $P_t - S_t/2$ and $P_t + S_t/2$ respectively. The mid-price P can be in multiples of $\delta/2$, but when this is so, the spread S will certainly be either δ or 3δ . And when P is in multiples of δ , S will be 2δ for sure. Thus this consistency ensures that the best bid and ask prices are always in multiples of δ . In addition, the set \mathbb{N}^+ in (3) denotes the set of all positive integers.

depth imbalance F is a predictor of the directions of mid-price jumps. The reason is that the depth imbalance is informative about whether the depth at the best bid or ask is likely to become depleted first, leading to a mid-price change. Thus it is a signal for the HFT to forecast the direction of a future price movement.

The influence of the imbalance F on the directions of the mid-price jumps is interpreted as the short-term price momentum. If the parameter σ_F is large so that F is volatile, it leads to more frequent appearances of strong momentum (signals). Thus the HFT is more likely to engage in trend-chasing actions. This effect is amplified if the jump intensities λ_1^J and λ_2^J of the mid-price are higher, i.e. there are more realizations of the price momentum.

Besides these LOB variables, orders with limited pre-trade transparency have become increasingly popular on electronic trading platforms recently. Major US stock exchanges, such as NASDAQ and NYSE, permit submissions of hidden limit orders into their LOBs.¹² The price, size and location of these orders are completely concealed from other market participants, and they can be placed inside the observable bid-ask spread without affecting the visible best bid and ask quotes. This mechanism of hidden orders has led to a vast number of order activities as HFTs try to “ping” for hidden liquidity inside the spread by posting aggressive “fleeting orders” that are cancelled a few instants later if not executed.¹³

To account for this phenomenon, I will specify the existence of hidden orders inside the spread as follows. If the spread $S_t = 2\delta$, the probability of hidden bid orders sitting at P_t is $\varphi_1(F_t)$ and that of hidden ask orders at P_t is $\pi_1 - \varphi_1(F_t)$. Similarly, if $S_t = 3\delta$, the probabilities of hidden bid orders at $P_t - \delta/2$ and $P_t + \delta/2$ are $\varphi_1(F_t)$ and $\varphi_2(F_t)$ respectively, so that those of hidden ask orders at $P_t - \delta/2$ and $P_t + \delta/2$ are $\pi_1 - \varphi_1(F_t)$ and $\pi_2 - \varphi_2(F_t)$.¹⁴ Since the existing literature suggests that the probability of hidden orders inside the spread

¹²Hautsch and Huang (2012b) give a detailed empirical analysis of hidden orders on NASDAQ stocks. They find that, during the month of October 2010, on average 20.1% of all trades are executed against hidden orders. However, only a small proportion of the hidden depth gets executed, which implies that the share of undetected hidden depth is much greater.

¹³This pinging phenomenon was first documented in Hasbrouck and Saar (2009).

¹⁴This structure of the existence of hidden orders basically implies that hidden orders are not allowed to exist exactly at the mid-price when the spread S is equal to δ or 3δ .

is positively correlated with the own-side depth imbalance, I let the functions φ_i be

$$\varphi_i(u) = \pi_i / (1 + \exp(-\kappa u)), \quad \text{for } i = 1, 2, \quad (5)$$

where κ and π_i are positive constants and $\pi_i < 1$.¹⁵ When π_1 and π_2 are large, hidden orders are more likely to be found within the spread.

3.2 The HFT's Trading Strategies

At any time t , the HFT can submit limit buy and sell orders, which will be executed only when they are filled by incoming market orders. The quantity of the HFT's limit orders is fixed at one lot (100 shares).¹⁶ Alternatively, the HFT can send out market buy or sell orders trading at the best available price on the opposite side for an immediate execution.

3.2.1 Make Strategy

When using limit orders (a make strategy), the HFT can place his orders at the best available bid and ask prices, hence joining the existing order queues at these prices.¹⁷ This simply amounts to the traditional market-making behavior, which means that the HFT tries to passively capture the spread by posting limit orders at the best available quotes.

Furthermore, the HFT can also “ping” inside the spread by improving either the best bid or ask price by one tick whenever the spread S_t is greater than δ , i.e. submitting a buy (sell) limit order at $P_t - S_t/2 + \delta$ ($P_t + S_t/2 - \delta$). Such actions are done by HFTs in practice not only to capture the market order flow at the best quotes, but also to ping for hidden orders

¹⁵Buti and Rindi (2013) provide a theoretical model that gives rise to this positive correlation, and Hautsch and Huang (2012b) offer an empirical confirmation.

¹⁶The reason that I fix the quantity of limit orders to be one lot is motivated by a common finding that appears in a vast number of empirical studies, whether they examine HFT activities (e.g., Hasbrouck and Saar (2013)) or analyze limit order books (e.g., Hautsch and Huang (2012a)). The finding is that the average size of limit orders nowadays is just slightly bigger than 100 shares.

¹⁷A make strategy through limit orders can either be a traditional market-making strategy or a pinging strategy, and the HFT has to submit limit orders simultaneously at both the buy (bid) and sell (ask) sides of the LOB when using a make strategy. Quoting at one side of the LOB only is not allowed.

inside the spread (Hasbrouck and Saar (2009)). However, despite faster executions, these pinging limit orders receive worse execution prices than their queuing counterpart.

Because the submission, update or cancellation of limit orders entails no cost, it is natural to model the make strategy of the HFT as a continuous-time predictable control process:

$$\theta_t^{mk} = \{\theta_t^{mk,b}, \theta_t^{mk,a}\}, t \geq 0, \quad (6)$$

where $\theta_t^{mk,b} \in \{0, 1\}$ and $\theta_t^{mk,a} \in \{0, 1\}$. b and a stand for the bid and the ask side respectively. The predictable processes $\theta_t^{mk,b}$ and $\theta_t^{mk,a}$, with values equal to 0 or 1, represent the possible make strategy regimes: 0 indicates that the limit order is joining the queue at the best price, whereas 1 indicates the limit order is improving the best price by δ . Note that if the spread is at its minimum δ , $\theta_t^{mk,b}$ and $\theta_t^{mk,a}$ both can only take the value of 0, since improving the best bid or ask then will simply be considered as posting a market buy or sell order instead.

3.2.2 Take Strategy

Instead of limit orders, the HFT may also employ market orders (a take strategy) to obtain instant executions. However, unlike limit orders, market orders take liquidity from the LOB and are subject to a transaction fee of ϵ per share. The costly nature of market orders means that, if a take strategy is used continuously, the HFT will go bankrupt in finite time. Thus I will model the HFT's take strategy as an impulse control in continuous time:

$$\theta^{tk} = \{\tau_n, \zeta_n\}_{n \in \mathbb{N}}. \quad (7)$$

Here, $\{\tau_n\}$ is an increasing, finite sequence of stopping times indicating the times when the HFT uses market orders, and $\zeta_n \in [-\zeta_{max}, \zeta_{max}] \setminus \{0\}$ are \mathcal{F}_{τ_n} -measurable random variables denoting the number of shares (in lot size) bought ($\zeta_n > 0$) or sold ($\zeta_n < 0$) at these times. I restrict the market order sizes to be less than some small constant ζ_{max} , so that the HFT

remains small relative to the market and his market orders do not eat through the LOB.¹⁸

3.3 Order Execution Processes

For the HFT's limit orders submitted at the best bid and ask, they will be executed only if counterpart market orders arrive in the next instant and fill them completely. I model the arrivals of exogenous buy and sell market orders by two independent Poisson processes, M^b (for buy) and M^a (for sell), with intensities given by λ^b and λ^a respectively. In addition, when a buy (sell) market order arrives, the HFT's sell (buy) limit order in the queue will be filled with a probability given by the fill-rate function $h(F_t)$ ($h(-F_t)$), where

$$h(u) = 1/(1 + \exp(\varsigma_0 + \varsigma_1 u)), \quad \varsigma_0, \varsigma_1 \text{ are positive constants.} \quad (8)$$

This fill-rate function depicts the idea that the probability of execution for a queuing limit order becomes higher (lower) if the own-side queue is relatively shorter (longer), which illustrates the time priority structure of the LOB.

However, if a limit order jumps the queue instead (i.e. pings inside the spread), it will be executed instantly as long as there is a matching hidden order from the opposite side that resides inside the spread. Otherwise, this limit order will be fully matched if a counterpart market order arrives at the next moment, due to its time priority.

Turning to the HFT's market orders, they involve a fixed transaction fee of ϵ per share, but they will execute immediately by hitting either hidden orders inside the spread or limit orders at the best price on the opposite side.

3.4 Comments on the Model

There are two important aspects of the model that are new compared to the current literature on HFT strategies and are essential for rationalizing HFTs' pinging activities without

¹⁸Eating through the book means the size of an market order is larger than and thus exhausts the available depth at the best price(s), resulting in a price impact and a widening of the spread.

manipulation or learning. First, the use of limit orders to capture hidden liquidity has not been properly addressed in existing continuous-time models of HFTs, where the limit order strategy of the HFT is for market-making purposes only. In my model, limit orders can also be utilized by the HFT as a “pinging” strategy, due to the existence of hidden orders inside the spread. Compared to the traditional market-making behavior, pinging is more aggressive as it seeks a faster execution at a less favorable price. However, it is not as aggressive as using market orders since it does not cross the spread. Given the large number of pinging activities identified in empirical work (e.g., Hasbrouck and Saar (2009, 2013)), it is essential to include hidden orders and pinging strategies in a theoretical model of HFT behaviors.

Another aspect that has not been fully dealt with in the continuous-time literature concerns directional or momentum tradings by HFTs. In a standard setup, an HFT would use market orders only for inventory control purposes. However, in my model, the HFT would also implement market orders as a momentum strategy to aggressively take liquidity from the market and trade directionally when the price momentum is strong enough. There are empirical studies that have identified some most profitable HFTs as primarily liquidity takers (e.g., Baron, Brogaard, and Kirilenko (2014)) and have recognized the anticipatory and directional trading behaviors of HFTs (e.g., Hirschey (2013)). Hence it is crucial to incorporate momentum (directional) tradings into any theoretical models on HFT strategies.¹⁹

When both hidden orders and price momentum exist in the model, the HFT has two motives to carry out pinging strategies. One is to control the inventory. The HFT can ping inside the spread to hit hidden orders when he needs to unwind his inventory, which is cheaper than using market orders. Nonetheless, if hidden orders existed with high probabilities π_1 and π_2 and inventory control was the HFT’s only motive for pinging, the model would generate a large number of pinging orders filled by hidden orders. There will not be many cancelled pinging orders at the same time, contradicting with the numerous ones observed in reality.

The existence of the price momentum is then necessary to bring about cancelled pinging

¹⁹Since market orders are costly, if they are used solely for inventory-control purposes, it would be difficult to reconcile this with the findings that market order (take) strategies can be (very) profitable.

activities, because now the other motive for the HFT to ping is to chase price trends. The HFT can use pinging orders to establish directional positions when anticipating a likely price move. However, if the volatility parameter σ_F is large so that the depth imbalance (momentum) varies turbulently, it would induce the HFT to cancel his pinging orders frequently and change his strategies based on the momentum swings. This is escalated if the mid-price jump intensities λ_1^J and λ_2^J are also large. Large intensities prompt more trend-chasing pinging activities from the HFT as his directional bets have a higher chance to materialize, yet the HFT often needs to cancel his pinging orders due to volatile changes in the price momentum.

4 Equilibrium and Solution Method

4.1 Objective of the HFT

To derive the equilibrium of my model, I begin by stating the cash holding and the inventory processes of the HFT, and then define his optimization problem.

Let X and Y denote the cash holding and the inventory of the HFT. If a (limit order) make strategy θ_t^{mk} is used at t , the cash holding X and the inventory Y evolve according to:

$$dY_t = d\widetilde{M}_t^a - d\widetilde{M}_t^b \quad (9)$$

$$dX_t = - \left(P_t - \frac{S_t}{2} + \delta\theta_t^{mk,b} \right) d\widetilde{M}_t^a + \left(P_t + \frac{S_t}{2} - \delta\theta_t^{mk,a} \right) d\widetilde{M}_t^b, \quad (10)$$

where

$$d\widetilde{M}_t^a = \theta_t^{mk,b} \left(B_t^a dt + (1 - B_t^a) dM_t^a \right) + (1 - \theta_t^{mk,b}) h(F_t) dM_t^a \quad (11)$$

$$d\widetilde{M}_t^b = \theta_t^{mk,a} \left(B_t^b dt + (1 - B_t^b) dM_t^b \right) + (1 - \theta_t^{mk,a}) h(-F_t) dM_t^b. \quad (12)$$

Here, B_t^a (B_t^b) is an indicator for whether ask (bid) hidden orders exist inside the spread. B_t^a and B_t^b have distributions $\eta^a(F_t, S_t)$ and $\eta^b(F_t, S_t)$. η^a (η^b) matches the existence probability

of ask (bid) hidden orders at the best bid (ask) price plus δ , given F_t and S_t . Thus when the HFT's buy (sell) limit order joins the queue at the best bid (ask), its inventory increases (decreases) by one lot if sell (buy) market orders arrive at the next instant *and fill* the HFT's limit order. Alternatively, if the HFT's buy (sell) limit order is pinging inside the spread, its inventory increases (decreases) by one lot whenever the limit order hits an opposite-side hidden order. Otherwise, it rises (falls) by one lot if sell (buy) market orders arrive within the next instant. The cash holding thus increases (decreases) by an amount equal to the quoted price of the sell (buy) limit order multiplied by the order's uncertain execution state.

On the other hand, the dynamics of X and Y jump at t if the HFT exercises a (market order) take strategy θ_t^{tk} instead:

$$Y_t = Y_{t-} + \zeta_t, \quad (13)$$

$$X_t = X_{t-} - \left[\zeta_t P_t + |\zeta_t| \left(\frac{S_t}{2} - H_t + \epsilon \right) \right] \quad (14)$$

where ϵ is the fixed fee per share for market orders, H_t is an integer random variable with a probability mass function $G(\cdot | F_t, S_t)$ that takes the value of

$$H_t = \begin{cases} 0 & \text{if } S_t = \delta \\ \delta & \text{if } S_t > \delta \text{ and the market order hits a hidden order at } P_t + \text{sign}(\zeta_t) \left(\frac{S_t}{2} - \delta \right) \\ 2\delta & \text{if } S_t > 2\delta \text{ and the market order hits a hidden order at } P_t + \text{sign}(\zeta_t) \left(\frac{S_t}{2} - 2\delta \right) \end{cases} \quad (15)$$

and the distribution of H_t matches the existence probabilities of the inside-spread hidden orders given F_t and S_t . Hence when the HFT submits a market order, its inventory Y jumps up or down at t by the size of the order. Moreover, its cash holding X changes by the value of the order $\zeta_t P_t$ plus the cost associated with the market order. The cost consists of the uncertain part due to crossing the spread $(S_t/2 - H_t)$ and the fixed transaction fee ϵ .²⁰

²⁰In essence, the term $-H_t$ measures the reduction in the cost of market orders for the HFT given the possible existence of hidden orders inside the spread when $S_t > \delta$, i.e. it does not necessarily pay the full spread-crossing cost of $S_t/2$ (relative to the mid-price).

The objective of the HFT is to maximize over the finite horizon $[0, T]$ the cash earnings from his trades in the LOB and at the same time keep his inventory at bay. In addition, the HFT has to liquidate all his inventory at T . Hence the HFT's optimization problem is

$$\max_{\{\theta^{mk}, \theta^{tk}\}} \mathbb{E}_0 \left[X_T - \gamma \int_0^T Y_{t-}^2 d[P, P]_t \right], \quad s.t. \quad Y_T = 0, \quad (16)$$

where the maximization is taken over all admissible strategies Θ .²¹ The term $\gamma \int_0^T Y_{t-}^2 d[P, P]_t$ is a quadratic-variation penalization term for holding a nonzero inventory in the stock, where $\gamma > 0$ is a penalization parameter and $[P, P]_t$ is the quadratic variation of the mid-price P .

I can rewrite the above optimization problem with the constraint $Y_T = 0$ removed. To this end, I introduce the function

$$Q(x, y, p, f, s) = x + py - |y| \left(\frac{s}{2} - H + \epsilon \right), \quad (17)$$

which represents the total cash obtained after an immediate liquidation of the inventory y via a market order, given the cash holding x , the mid-price p , the depth imbalance f and the spread s .²² I can now reformulate the problem (16) equivalently as

$$\max_{\{\theta^{mk}, \theta^{tk}\}} \mathbb{E}_0 \left[X_T + P_T Y_T - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon \right) - \gamma \int_0^T Y_{t-}^2 d[P, P]_t \right]. \quad (18)$$

The proofs for all lemmas and propositions are shown in the Appendix.

Lemma 1. (16) and (18) are equivalent.

Having defined the objective, the value function of problem (18) for the HFT is then:

$$V(t, x, y, p, f, s) = \sup_{\{\theta^{mk}, \theta^{tk}\}} \mathbb{E}_t \left[X_T + P_T Y_T - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon \right) - \gamma \int_t^T Y_{u-}^2 d[P, P]_u \right], \quad (19)$$

for $\{\theta^{mk}, \theta^{tk}\} \in \Theta$, $t \in [0, T]$, and $(x, y, p, f, s) \in \mathbb{R}^2 \times \mathcal{P} \times \mathbb{R} \times \mathbb{S}$. Here, given $\{\theta^{mk}, \theta^{tk}\} \in \Theta$,

²¹The HFT's objective function is essentially very similar to a mean-variance utility function.

²² H is the same integer random variable defined before that measures the potential cost reduction for market orders due to hidden orders inside the spread.

\mathbb{E}_t denotes the expectation operator under which the solution (X, Y, P, F, S) to the processes (1)-(3) and (9)-(14), with initial state $(X_{t-}, Y_{t-}, P_{t-}, F_{t-}, S_{t-}) = (x, y, p, f, s)$, is taken.

Before defining the model's equilibrium, I state the following lemma, which demonstrates that the value function is finite and locally bounded.

Lemma 2. *There exists some constant C_0 and C_1 such that for all $(t, x, y, p, f, s) \in [0, T] \times \mathbb{R}^2 \times \mathcal{P} \times \mathbb{R} \times \mathbb{S}$,*

$$Q(x, y, p, f, s) \leq V(t, x, y, p, f, s) \leq x + py + C_0 + C_1. \quad (20)$$

Both bounds have an intuitive financial interpretation. The lower bound is the value from the particular strategy that eliminates all non-zero inventory through a market order, and then does nothing until the time reaches T . The upper bound is made of three terms. The first term, $x + py$, is the value of the HFT's cash holding plus his inventory evaluated at the mid-price. The second constant C_0 is the upper bound on the profit from the fictitious traditional market-making strategy that participates in every trade without any cost of controlling the inventory. The last constant C_1 denotes a bound on the profit from any directional costless market order strategy on a virtual asset always priced at the mid-price.

4.2 Equilibrium and Dynamic Programming Equations

The problem of (19) is a mixed regular-impulse stochastic control problem in a jump-diffusion model, and I will use dynamic programming to characterize the model's equilibrium. To this end, I introduce two mathematical operators. For any admissible (limit order) make strategy $\theta^{mk} = \{\theta^{mk,b}, \theta^{mk,a}\}$, I will define the second-order non-local operator \mathcal{L} :

$$\begin{aligned} \mathcal{L} \circ V(t, x, y, p, f, s) &= (\mathcal{L}^P + \mathcal{L}^F + \mathcal{L}^S) \circ V(t, x, y, p, f, s) \\ &\quad + g^a(f, s, \theta_t^{mk,b}) \cdot V(t, x - (p - s/2 + \delta\theta_t^{mk,b}), y + 1, p, f, s) \\ &\quad + g^b(f, s, \theta_t^{mk,a}) \cdot V(t, x + (p + s/2 - \delta\theta_t^{mk,a}), y - 1, p, f, s), \end{aligned} \quad (21)$$

where \mathcal{L}^P , \mathcal{L}^F , \mathcal{L}^S in the first term are the infinitesimal generators for the processes of the mid-price P , the depth imbalance F and the spread S respectively, and the next two terms are the non-local operators induced by the expected changes of the cash holding process X and the inventory process Y when the HFT applies a make strategy θ_t^{mk} at t . In addition,

$$g^a(f, s, \theta_t^{mk,b}) = \theta_t^{mk,b}(\eta^a(f, s) + (1 - \eta^a(f, s))\lambda^a) + (1 - \theta_t^{mk,b})\lambda^a h(f) \quad (22)$$

$$g^b(f, s, \theta_t^{mk,a}) = \theta_t^{mk,a}(\eta^b(f, s) + (1 - \eta^b(f, s))\lambda^b) + (1 - \theta_t^{mk,a})\lambda^b h(-f), \quad (23)$$

which are the expected execution rates for the HFT's bid and ask limit orders respectively.²³

The impulse control operator \mathcal{M} for an admissible (market order) take strategy θ^{tk} is

$$\begin{aligned} & \mathcal{M} \circ V(t, x, y, p, f, s) \\ &= \sup_{\zeta \in [-\zeta_{max}, \zeta_{max}]} \int_H V(t, x - \zeta p - |\zeta|(s/2 - H + \epsilon), y + \zeta, p, f, s) dG(H | s, f), \end{aligned} \quad (24)$$

which is generated by the jumps in X and Y due to θ_t^{tk} being used at t .

With \mathcal{L} and \mathcal{M} defined, the dynamic programming equation associated with the value function (19) is the Hamilton-Jacobi-Bellman quasi-variational inequality (HJB-QVI):

$$\max \left\{ \frac{\partial V}{\partial t} + \sup_{\{\theta^{mk}\}} \{\mathcal{L} \circ V\} - \gamma y^2 \frac{\mathbb{E}_t d[P, P]_t}{dt}, \mathcal{M} \circ V - V \right\} = 0, \text{ on } [0, T) \quad (25)$$

together with the terminal condition

$$V(T, x, y, p, f, s) = x + py - |y|(s/2 + \epsilon) + |y| \int_H H dG(H | s, f). \quad (26)$$

The next lemma shows that the HJB-QVI (25) has a well-defined expression by giving the full expressions for the infinitesimal generators (\mathcal{L}^P , \mathcal{L}^F , \mathcal{L}^S) and the value of $\mathbb{E}_t[P, P]_t$.

Lemma 3. *The HJB-QVI (25) admits a well-defined expression.*

²³The two terms, g^a and g^b , are basically the arrival rates of the “modified” market orders \widetilde{M}_t^a and \widetilde{M}_t^b that have been explained under equation (10).

I then present the equilibrium concept of my model in the following definition.

Definition 1. *In the above economy where the HFT trades a risky stock in an LOB governed by the stochastic processes laid out in Section 3.1-3.3, the partial equilibrium is defined by a value function $v(t; \bullet)$ and policy functions (strategies) $\{\theta^{mk}, \theta^{tk}\}_t, t \in [0, T]$, such that*

- (a) *The policy functions solve the HFT's maximization problem (18);*
- (b) *Given the policy functions, the value function v solves the HJB-QVI (25) for $t \in [0, T)$ with the terminal condition (26) at T .*

The next proposition shows that a solution to the partial equilibrium exists and is unique.

Proposition 1. *There exists a unique solution to the partial equilibrium of the model. In particular, the value function in (19) is the unique **viscosity solution** to (25) and (26).²⁴*

Before providing a numerical solution to the equilibrium of my model, I present a lemma that simplifies the value function (19) and reduces its dimensionality.

Lemma 4. *The value function (19) can be decomposed as*

$$V(t, x, y, p, f, s) = x + py + \nu(t, y, f, s). \quad (27)$$

Moreover, the reduced-form value function ν satisfies the quasi-variational inequality and the terminal condition shown below, which are simplified from (25)-(26) after decomposing V :

$$\begin{aligned} \max \left[\frac{\partial \nu}{\partial t} + y \frac{\mathbb{E}_t dP_t}{dt} + \mathcal{L}^{\mathcal{F}} \circ \nu + \mathcal{L}^{\mathcal{S}} \circ \nu - \gamma y^2 \frac{\mathbb{E}_t d[P, P]_t}{dt} + \right. \\ \left. \sup_{\theta^{mk}} \left\{ g^a(f, s, \theta_t^{mk, b}) \cdot (\nu(t, y + 1, f, s) - \nu(t, y, f, s) + s/2 - \delta \theta_t^{mk, b}) + \right. \right. \\ \left. \left. g^b(f, s, \theta_t^{mk, a}) \cdot (\nu(t, y - 1, f, s) - \nu(t, y, f, s) + s/2 - \delta \theta_t^{mk, a}) \right\}, \right. \\ \left. \sup_{\zeta} \left\{ \nu(t, y + \zeta, f, s) - |\zeta|(s/2 + \epsilon) + |y| \int_H H dG(H | s, f) \right\} - \nu \right] = 0, \text{ on } t \in [0, T) \quad (28) \end{aligned}$$

²⁴The viscosity solution concept is a generalization of the classical solution concept to a partial differential equation. For a classic reference, see Crandall, Ishii, and Lions (1992) or Fleming and Soner (2005).

with the terminal condition:

$$\nu(T, y, f, s) = -|y|(s/2 + \epsilon) + |y| \int_H H dG(H | s, f). \quad (29)$$

4.3 Numerical Solution

For the numerical solution to the value function V of (19), since $V(t, x, y, p, f, s) = x + py + \nu(t, y, f, s)$, I will provide a backward, finite-difference scheme that solves the quasi-variational inequality (28) and the terminal condition (29). My method is based on the schemes developed by Chen and Forsyth (2008) and Guilbaud and Pham (2013).

First, I use a time discretization on $[0, T]$ with a step $\Delta_T = T/N_T$ and a regular grid

$$\mathbb{T}_{N_T} = \{t_k = k\Delta_T, k = 0, \dots, N_T\}. \quad (30)$$

Second, I discretize and localize the spaces for Y and F on two regular grids, with bounds M_Y, M_F , and steps $\Delta_Y = M_Y/N_Y \equiv 1$, $\Delta_F = M_F/N_F$, where $N_Y, N_F \in \mathbb{N}$, so that

$$\mathbb{Y}_{N_Y} = \{y_i = i\Delta_Y, i = -N_Y, \dots, N_Y\}, \quad \mathbb{F}_{N_F} = \{f_j = j\Delta_F, j = -N_F, \dots, N_F\}. \quad (31)$$

Next, I define two finite-difference matrices, D_1 and D_2 , for calculating the first and the second order derivatives against F on the grid \mathbb{F}_{N_F} , where D_2 uses central differences and D_1 uses forward differences when $f_j < 0$ and backward differences when $f_j \geq 0$:²⁵

$$D_2\nu(t, y, f_j, s) = \frac{\nu(t, y, f_{j+1}, s) - 2\nu(t, y, f_j, s) + \nu(t, y, f_{j-1}, s)}{(\Delta_F)^2} \quad (32)$$

$$D_1\nu(t, y, f_j, s) = \begin{cases} \frac{\nu(t, y, f_{j+1}, s) - \nu(t, y, f_j, s)}{\Delta_F} & \text{if } f_j < 0 \\ \frac{\nu(t, y, f_j, s) - \nu(t, y, f_{j-1}, s)}{\Delta_F} & \text{if } f_j \geq 0 \end{cases} \quad (33)$$

²⁵The two matrices D_1 and D_2 are defined in a similar fashion to the finite difference space derivatives in Section 5.1 of Cont and Voltchkova (2005). The reason that I switch from forward to backward differences in D_1 when f_j becomes greater than 0 is detailed over there.

I now state the main part of the numerical scheme. To this end, I introduce the explicit-implicit operator for the time-space discretization of the quasi-variational inequality (28). I define, for any $(t, p, f, s) \in [0, T] \times \mathcal{P} \times \mathbb{R} \times \mathbb{S}$ and any real-valued function $\phi : \mapsto \phi(t, y, f, s)$,

$$\mathcal{A}(t, y, f, s, \phi) = \max \left\{ \tilde{\mathcal{L}}(t, y, f, s, \phi), \tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t, y, f, s, \phi) \right\}, \quad (34)$$

where²⁶

$$\begin{aligned} \tilde{\mathcal{L}}(t, y, \cdot, s, \phi) &= \left(I_{N_F \times N_F} - \Delta_T \sigma_F^2 D_2 - \Delta_T \alpha_F (\mathbb{F}_{N_F} \mathbf{1}'_{N_F}) \cdot D_1 \right)^{-1} \times \\ &\quad \left(\phi(t, y, \cdot, s) + \Delta_T y \frac{\mathbb{E}_t dP_t}{dt} + \Delta_T \mathcal{L}^S(\phi(t, y, \cdot, s)) - \Delta_T \gamma y^2 \frac{\mathbb{E}_t d[P, P]_t}{dt} + \right. \\ &\quad \left. \Delta_T \sup_{\theta^{mk}} \left\{ g^a(\cdot, s, \theta_t^{mk, b}) \odot (\phi(t, y+1, \cdot, s) - \phi(t, y, \cdot, s) + \frac{s}{2} - \delta\theta^{mk, b}) + \right. \right. \\ &\quad \left. \left. g^b(\cdot, s, \theta_t^{mk, a}) \odot (\phi(t, y-1, \cdot, s) - \phi(t, y, \cdot, s) + \frac{s}{2} - \delta\theta^{mk, a}) \right\} \right), \quad (35) \end{aligned}$$

and

$$\tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t, y, f, s, \phi) = \sup_{|\zeta| \leq \zeta_{max}} \left\{ \tilde{\mathcal{L}}(t, y+\zeta, f, s, \phi) - |\zeta| \left(\frac{s}{2} + \epsilon \right) + |y| \int_H H dG(H | s, f) \right\}. \quad (36)$$

When the inventory y is on the boundary of \mathbb{Y}_{N_Y} , i.e. $y = -M_Y$ or $y = M_Y$, make and take strategies are confined to the buy side or the sell side only, so that y does not go off from its grid. Then, I approximate the solution ν to (28)-(29) by the numerical solution ϖ on $\mathbb{T}_{N_T} \times \mathbb{Y}_{N_Y} \times \mathbb{F}_{N_F} \times S$ to the backward explicit-implicit finite difference scheme:

$$\varpi(T, y, f, s) = -|y| \left(\frac{s}{2} + \epsilon \right) + |y| \int_H H dG(H | s, f) \quad (37)$$

$$\varpi(t_k, y, f, s) = \mathcal{A}(t_{k+1}, y, f, s, \varpi), \quad k = N_T - 1, N_T - 2, \dots, 0, \quad (38)$$

²⁶In the operator $\tilde{\mathcal{L}}$, the colon $:$ denotes the column vector of \mathbb{F}_{N_F} , $I_{N_F \times N_F}$ is an N_F by N_F identity matrix, $\frac{\mathbb{E}_t dP_t}{dt}$ and $\frac{\mathbb{E}_t d[P, P]_t}{dt}$ are vectors evaluated on \mathbb{F}_{N_F} , $\mathbf{1}'_{N_F}$ is an $N_F \times 1$ vector of 1's, and the sign \odot denotes element-by-element product for vectors and matrices. $\tilde{\mathcal{L}}$ is expressed as a vector on the grid \mathbb{F}_{N_F} because of the implicit time step that I used when approximating the generator \mathcal{L}^F .

where (37) and (38) approximate (29) and (28) respectively.

The complete solution algorithm is summarized in the backward induction steps below:

1. At the terminal date $t_{N_T} = T$: for each combination of (y, f, s) , make $\varpi(T, y, f, s) = -|y|(\frac{\xi}{2} + \epsilon) + |y| \int_H H dG(H | s, f)$.
2. (*Backward Induction*) From time step t_{k+1} to t_k where k runs from $N_T - 1$ back to 0, for each combination of (y, f, s) :
 - (a) Calculate $\tilde{\mathcal{L}}(t_{k+1}, y, f, s, \varpi)$ from (35) and obtain $\theta^{mk,*}$.
 - (b) Calculate $\tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t_{k+1}, y, f, s, \varpi)$ from (36) and obtain $\theta^{tk,*}$.
 - (c) If $\tilde{\mathcal{L}}(t_{k+1}, y, f, s, \varpi) \geq \tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t_{k+1}, y, f, s, \varpi)$, set $\varpi(t_k, y, f, s) = \tilde{\mathcal{L}}(t_{k+1}, y, f, s, \varpi)$ and the policy (optimal strategy) at t_k is thus $\theta^{mk,*}$, given (y, f, s) .
 - (d) Otherwise, let $\varpi(t_k, y, f, s) = \tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t_{k+1}, y, f, s, \varpi)$, and $\theta^{tk,*}$ is taken to be the policy (optimal strategy) at t_k , given (y, f, s) .

Finally, I state the theorem for the convergence of my numerical solution ϖ to the reduced-form value function ν , and hence the convergence of the discretized policy functions to the equilibrium optimal strategies.

Proposition 2. *The solution ϖ to the numerical scheme (37)-(38) and the corresponding discretized policy functions converge locally uniformly to the reduced-form value function ν and the equilibrium optimal strategies on $[0, T] \times \mathbb{R} \times \mathbb{R}$ as $(\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F) \rightarrow (0, 0, \infty, 0, \infty)$, $\forall s \in \mathcal{S}$.*

5 Estimation

In order to quantify the optimal trading strategies of the HFT based on the numerical solution, I need to obtain values for the parameters that govern the stochastic processes of

the LOB variables. This Section concerns the estimation of these parameters. Panel (A) of Table 1 summarizes the parameters to be estimated.

5.1 Data

I use NASDAQ TotalView-ITCH 4.0 limit order book message feed data on three types of stocks listed on NASDAQ during the month of June 2012 (21 trading days). The three types consist of stocks with narrow spreads and high order-book depths, stocks with medium spreads and depth levels, and stocks with wide spreads and low order-book depths. I focus on three representative stocks, with one from each of the three types: INTC (Intel, narrow spread and high depth), QCOM (Qualcomm, medium spread and medium depth), and AMZN (Amazon, wide spread and low depth).

The TotalView data include all real-time messages of limit order submissions, cancellations, executions and hidden order executions for all NASDAQ-listed stocks on every trading day from 7:00am EST when NASDAQ's electronic limit order book system starts accepting incoming limit orders. The messages are stamped to millisecond precision. The system is initialized by an empty order book where all overnight limit orders are resubmitted automatically at the beginning of each day. The data record a unique identification for each limit order and I can identify the attribute of a limit order (cancelled, executed or neither) by tracking it through its order ID. Moreover, trades are identified via the records of limit order and hidden order executions. Since the trading directions of limit orders and hidden orders are recorded, I can exactly identify whether a trade is buyer-initiated or seller-initiated. I consolidate all trade transactions into one single trade from a market order if they are logged at the same time-stamp in the data and have the same initiation type. Therefore, the LOB for any NASDAQ stock at any time can be reconstructed (up to millisecond precision) through continuously updating the book system according to all reported messages.

Since the LOB variables of my model concern only the first level of the order book, I track the evolutions at the top of the book for estimation purposes. As in common practices,

I use data between 9:45am and 15:45pm to avoid certain erratic market movements that are likely to occur at market openings or closures.

5.2 Estimation

I will employ standard nonparametric estimators for all the intensity parameters and the transition matrix ρ .²⁷ The parameters of the depth imbalance F are estimated using the maximum likelihood method, since an Ornstein-Uhlenbeck process has an explicit transition density.²⁸ In addition, the parameters governing various probabilities are estimated by logistic regressions, as these probability distribution functions all have logistic forms. I conduct the estimations for all trading days in June 2012, and then calculate the averages of these daily estimates as my final estimated parameter values to be fed into my numerical solutions. The mean values of my daily parameter estimates are reported in Table 2, with Newey-West heteroskedasticity and autocorrelation consistent (HAC) standard errors in parentheses.

The estimation results have three aspects that are worth noticing. Firstly, compared to Amazon, Intel and Qualcomm have a less volatile imbalance process F as well as a higher tendency to stay at lower-spread positions. Secondly, due to its lower book depth, Amazon's mid-price jumps occur more often, as measured by the larger values of both λ_1^J and λ_2^J for Amazon. Thirdly, the estimates of hidden order parameters, market order arrival rates and limit order fill rates are more alike for those three stocks. We will see next that these dissimilarities in the parameter estimates will lead to a much different quantification of the optimal HFT strategies for Amazon as opposed to Intel or Qualcomm.

Besides the estimated parameters, the fixed parameters in Panel (B) of Table 1 are also used in my numerical quantification and simulation study of the optimal HFT strategies.

²⁷For a standard reference, see, for example, Karr (1991).

²⁸For a reference, see Aït-Sahalia (1999) and Aït-Sahalia and Mykland (2003).

6 Computation and Simulation Results

In this Section, I provide numerical results on the optimal HFT strategies computed via the implementation of my numerical scheme (37)–(38), for the three representative stocks – Intel (INTC), Qualcomm (QCOM) and Amazon (AMZN).²⁹ I use as inputs the estimated parameter values shown in Table 2 and the fixed parameters listed in Panel (B) of Table 1.

6.1 Percentage Attributions of Optimal Strategies

In this part, I conduct Monte Carlo simulations to quantify the properties of the HFT’s optimal strategies for the three stocks: Intel, Qualcomm and Amazon. The number of simulations is set to be $N_{MC} = 10000$ for each stock, and I use a standard Euler scheme to simulate the paths of the state variables (P, S, F, X, Y) and the exogenous market order arrivals (M^b, M^a) .

Table 3 summarizes the percentage attributions (averaged over simulations) of the optimal HFT strategies to three types of activity: traditional market making (submitting a limit order at the best available bid and ask prices) and inventory control (using market orders), pinging, and momentum (directional) trading via market orders. These percentages are calculated as follows. For each one of the Monte Carlo simulations, I compute the number of strategy choices attributable to each type of the activity, which are optimally made given the numerical solution and the simulated state variables at each point on the simulated path. Then I divide these numbers by the total number of strategy-choosing points in a simulation, which equals $T/\Delta_T = 7200$, to arrive at the corresponding percentage attribution numbers for one simulation. Finally, I average those percentage numbers across all 10000 simulations to obtain the percentage attributions shown in Table 3.

There are two main implications that we can learn from Table 3. **Firstly**, for stocks with

²⁹For each of the three stock types stated in Section 5.1, I checked 10 stocks, including the representative ones (Intel, Qualcomm, Amazon) being focused here. The percentage attributions are similar, as well as the optimal strategy profiles that will be presented later, i.e. the HFT is more of a market maker (resp. taker) and the pinging percentage is lower (resp. higher) if the stock under consideration has higher (resp. lower) order book depths and narrower (resp. wider) spreads.

narrow spreads on average and abundant order book depths such as Intel, the HFT behaves like a market maker in the traditional sense. He is providing liquidity to the market most of the time. As a result, pinging activities account for only 20% of the optimal strategies and approximately 70% of the strategies are in the realm of liquidity provision (market making) and inventory management by market orders.

Secondly, for stocks like Amazon with wide spreads on average and an order book that has low depths and volatile movements, the HFT looks less like a passive market maker. Instead, he acts more like an aggressive chaser for short-term price trends, since pinging and momentum trading together account for about 70% of the strategies for Amazon. In particular, pinging constitutes nearly 50% of the optimal strategies. Pinging is considered as a strategy that demands liquidity from the market when its objective is to build a directional inventory position. Hence for stocks with wide spreads and low book depths, the HFT is mainly a market taker and frequently taking liquidity to bet on the directional price moves.

Next, I compared the pinging percentages obtained from the model (as shown in Table 3) to the pinging percentages observable from the data to see how much pinging in the data can be rationalized by the model. To calculate pinging percentages in the data, I first compute the number of cancelled pinging activities for different stocks in a similar fashion to Hasbrouck and Saar (2009). This is defined as the number of limit orders submitted inside the spread and then cancelled in less than 2 seconds. Second, I need to compute the number of pinging orders that execute against hidden orders inside the spread. This is somewhat problematic since such pinging orders are identical to market orders that hit hidden orders. In order to deal with this issue, I treat non-consecutive orders executing on hidden orders as pinging orders. Therefore, the number of total pinging activities are calculated as the sum of the number of cancelled pinging activities and the number of pinging orders hitting hidden orders. I then divide the number of total pinging activities by the total number of

order book activities to arrive at the pinging percentages found in the data:

$$\text{Pinging \% in the data} = \frac{\text{Number of total pinging activities}}{\text{Total number of order book activities}}. \quad (39)$$

Finally, the model's pinging percentages are gauged against those from the data (expressed as a fraction of data's pinging percentage numbers) in Table 4.

As clearly seen from Panel (A) in Table 4, the pinging percentages produced by the model match quite closely to the percentages from the data. At least around 70% of the pinging activities observable from the data can be captured by the model, and the number is higher for stocks with high depths and low spreads (like Intel). Hence the result indicates that most of the pinging activities found in the data can be rationalized by the model with the two mechanisms of inventory control and momentum chasing.

Panel (B) in Table 4 further breaks down the roles that the mechanisms of inventory control and momentum chasing play in rationalizing pinging activities in the model. There are two noticeable features from Panel (B) in Table 4. Firstly, for stocks with high depths and narrow spreads such as Intel, inventory control and momentum chasing contribute comparably to the rationalization of pinging. This can be seen from the similar resulting pinging percentages produced by the model if either of the mechanism is shut down. Secondly, for stocks with low depths and wide spreads such as Amazon, momentum chasing carries more weight than inventory control in the rationalization of pinging. This is because the pinging percentage produced by the model decreases by a larger amount when the channel of price momentum is shut down than when the channel of hidden order is.

Finally, Table 4 shows that both the inventory control and the momentum chasing mechanisms are necessary and indispensable for rationalizing pinging activities, as the model can only capture less than 50% of the pinging in the data if either mechanism is turned off.

6.2 Optimal Strategy Profiles

To further grasp the intuition behind the percentage attributions of the optimal HFT strategies for different stocks, I now show the detailed characterizations of the strategies in Figures 1 to 6 and present the economic rationales. The optimal HFT strategies depend on t (time), Y (inventory), F (depth imbalance) and S (spread). Therefore, I will characterize the strategies as a function of Y and F , for $S = \delta$ and $S = 2\delta$, near $t = 0$ and $t = T$. The optimal strategies are mostly time invariant if the value of t is not extremely close to the terminal date T . In addition, the strategy profiles for Qualcomm are not exhibited since they are in between the ones for Intel and the ones for Amazon.³⁰

Figure 1 illustrates the optimal HFT strategies for Intel at $t = 10$, with the inventory and the depth imbalance levels shown on the horizontal and the vertical axes respectively. Consider Figure 1(A) on the left first, where the spread $S = \delta$. The central region of “market making” denotes submitting a limit order at both the best bid and ask prices. The two blue regions stand for inventory management: buy (sell) indicates using buy (sell) market orders to increase (decrease) the inventory towards zero. In particular, “partial inventory control” means unwinding the inventory partially, whereas “inventory control” means changing the inventory precisely back to zero. For instance, if the inventory is -5 lots, the strategy will specify purchasing z lots with $z < 5$ in the region of partial inventory control (buy), yet it will specify buying exactly 5 lots in the region of inventory control (buy). The “momentum” area means utilizing market orders to change the inventory and set up a directional position: buy and sell denote establishing positive and negative positions respectively. For example, under momentum (buy), if the inventory is -5 lots, the strategy will dictate a purchase of z lots with $z > 5$, and if the inventory is 1 lot, the strategy will dictate a purchase of z lots with $z \geq 1$. Finally, there is no pinging since it is not possible to submit a limit order inside the spread when the spread is at its minimum level of δ .

³⁰The reason is that the key parameter estimates $(\lambda^S, \alpha_F, \sigma_F, \lambda_1^J, \lambda_2^J)$ of Qualcomm are between those of Intel and Amazon, as pointed out in Table 2.

The economic intuition behind Figure 1(A) is explained as follows. The situation where the depth imbalance $F \leq 0$ will be focused on, since a similar but symmetrically reversed exposition can be applied to the opposite case where $F \geq 0$. To begin with, when the depth imbalance is mildly positive, the probability of a positive mid-price jump is somewhat higher than that of a negative one. If the HFT has a negative inventory, he will face an inventory risk as the mid-price is likely to move against his holdings. Thus he will reduce this risk via a market order.³¹ However, if the HFT holds a positive inventory, he would gain from a positive mid-price jump and thus do market making. Since a take strategy is expensive and the signal effect from the depth imbalance on the directional mid-price movement is not strong enough, the HFT would not use market orders to increase his positive inventory.

Furthermore, when the depth imbalance becomes considerably more positive, the probability of a positive mid-price jump is much higher than that of a negative one. If the HFT has a negative inventory, he faces a substantial inventory risk, yet also a clear opportunity to chase the upward price momentum. Therefore, the HFT aggressively takes the liquidity from the market through buy market orders to establish a directional (positive) inventory position, which is the momentum (buy) region in the graph. However, the HFT stops chasing the price momentum if his inventory is rather positive. As the mid-price jump intensities of Intel are not very large, building a substantially positive inventory would result in a large inventory risk on the opposite side if the mid-price does not jump up soon enough.³²

Next, consider Figure 1(B) on the right, where the spread is 2δ . The green and yellow areas on each side of the market making region represent pinging strategies, which the majority of the graph consists of. Here, “pinging on the bid (ask) side” denotes submitting a buy (sell) limit order inside the spread at the best bid plus δ (best ask minus δ), while

³¹Note that a make strategy (market making) here would not achieve this risk-reduction purpose, for two reasons. First, a market-making strategy implies that the HFT would still post a sell limit order at the best ask. Second, the effect of a positive depth imbalance on the fill-rate function means that a buy limit order of the HFT has a smaller chance of being filled. Consequently, instead of decreasing, the two together exacerbate the inventory risk faced by the HFT.

³²The depth imbalance is mean-reverting towards zero. Consequently, the anticipated price jump would become less likely if it does not occur very soon.

letting the sell (buy) limit order joining the queue at the best ask (best bid). To understand the intuition behind Figure 1(B) and compare it to the case of Figure 1(A), I will again concentrate on the scenario in which the depth imbalance is less than zero.

When the depth imbalance is modestly positive, if the HFT's inventory level is quite negative, he will remove the inventory risk by market orders. However, the inventory control is only partial since complete inventory controls are too costly under $S = 2\delta$. Alternatively, if the HFT's inventory level is closer to zero, he will instead use a pinging strategy from the bid side to reduce his inventory risk, for three reasons. Firstly, despite execution uncertainty, pinging is free. Since the inventory risk is small, it decreases the HFT's desire to use market orders. Secondly, a buy pinging order might hit a sell hidden order inside the spread. Thirdly, a positive depth imbalance implies a lower fill rate for queuing buy limit orders at the best bid, so that it is optimal for the HFT to jump the queue. On the other hand, if the HFT holds a significant level of positive inventory, he faces a large inventory risk due to the possibility of no upward price jump. Hence the HFT will ping on the ask side, which has a high chance of hitting a bid-side hidden order inside the spread and is costless.

When the depth imbalance is significantly positive, the HFT will reduce his inventory risk more aggressively by market orders if his inventory is below zero. However, if the inventory is positive but small, the HFT will pursue the price momentum by pinging on the bid side, which has a higher execution probability than submitting a buy limit order at the best bid. Chasing the momentum via a market order is suboptimal when $S = 2\delta$ as the expected gain is less than the cost of using market orders. Thus the pinging strategies would serve two different purposes – unwinding the HFT's inventory or pursuing the price momentum.

Next, **Figure 2** shows the optimal HFT strategies for Intel at $T - 3$. Consider Figure 2(A) on the left first, with the spread $S = \delta$. Since time t is close to the terminal date, inventory management becomes a large concern for the HFT as he must liquidate all positions at T . Therefore, if the HFT is holding a negative (positive) inventory and the mid-price is more likely to jump up (down) because of the positive (negative) depth imbalance, he

will aggressively unload his inventory through market orders to reduce the inventory risk. Conversely, when the HFT's inventory is in line with the likely price movement, he has an expected gain from holding a directional position. Hence the HFT will use a market-making strategy if his inventory and the depth imbalance are both positive (negative). In this case, it is not optimal for the HFT to use an aggressive momentum strategy to raise his directional positions, as inventory control is the prime objective when time approaches T .

Consider Figure 2(B) next. Similar to Figure 2(A), inventory management accounts for a large part of the optimal strategy profile, but there is only partial inventory control as market orders are very costly under $S = 2\delta$. Unlike in Figure 1(B), pinging strategies now have one sole objective: reducing the inventory risk. The HFT would only ping on the ask (bid) side if his inventory and the depth imbalance are both positive (negative), as pinging orders have a larger probability of executing against opposite-side hidden orders inside the spread and are free compared to market orders.

Let us now look at **Figure 3**, which compares the optimal strategy for Amazon (right) to that for Intel (left) in the case of time $t = 10$ and spread $S = \delta$. The major difference for Amazon is that the HFT takes liquidity a lot more often to carry out momentum strategies. The HFT is more aggressive since Amazon has much higher mid-price jump intensities as shown in Table 2. Thus the HFT is anticipating an upward (downward) directional price movement with a much larger likelihood if the depth imbalance is positive (negative), which causes the HFT to build a matching directional position via market orders. The expected reward of a price jump simply outweighs the cost of using market orders aggressively.

Figure 4 contrasts the optimal strategy for Amazon (right) to that for Intel (left) in the case of time $t = 10$ and spread $S = 2\delta$. There are two main changes on pinging strategies employed for Amazon. First, there is no pinging on the bid (ask) side when the HFT's inventory is negative (positive) and the depth imbalance is also moderately negative (positive). Second, pinging on the bid (ask) side is dominant when both the HFT's inventory and the depth imbalance are above (below) zero. These imply that when the HFT holds

a directional position in line with a likely mid-price move, his pinging strategies are used exclusively for chasing the price momentum. This is because Amazon has much higher mid-price jump intensities so that the HFT can be more aggressive. Nevertheless, it is not optimal for the HFT to deploy momentum strategies via market orders as it is too expensive and the HFT already holds a favorable directional position. However, if the depth imbalance is extremely high, and the HFT holds an inventory position against the likely direction of the mid-price jump, the HFT would use momentum strategies in the case of Amazon (purple areas). The large cost of market orders is absorbed by the reduction in the inventory risk and the almost certain benefit from the anticipated price jump.

Finally, **Figure 5** and **Figure 6** demonstrate the optimal HFT strategies for Amazon (right) towards the terminal date T and compare them to those for Intel (left). The optimal strategies for the two cases are similar, despite the differences seen in **Figure 3** and **Figure 4** when $t \ll T$. Once more, this is because inventory management concerns are dominant as time approaches T , so that pursuing the price momentum is no longer the HFT's objective.

6.3 Auxiliary Predictions

Besides pinging rationalization, my model yields a couple of additional auxiliary predictions regarding pinging activities with respect to depth imbalances. They are shown below.

The first auxiliary prediction of the model is related to the directions of pinging activities. The model implies that if the HFT sees a positive (negative) depth imbalance and hence a positive (negative) price momentum more often, he is more likely to ping from the buy (sell) side and take positive (negative) directional bets due to momentum chasing motives. Consequently, it yields the following prediction:

Prediction 1. *There is more pinging from the buy (sell) side if the depth imbalance is more frequently positive (negative).*

Now I check this prediction against data. To do this, I divide each trading day into

30-second intervals.³³ Next, I calculate the number of buy and sell pinging activities in each of these intervals in a similar manner to the computation of the number of pinging activities shown in Section 6.1. In addition, I also compute the durations (measured in seconds) of positive and negative imbalances in each interval. Then I use the following regression to measure and test the effect of imbalance durations on pinging activity directions:

$$PO_t = \alpha + \beta DOI_t + x_t' \gamma + \epsilon_t. \quad (40)$$

PO_t denotes the number of buy or sell pinging activities and DOI_t the duration (measured in seconds) of positive or negative depth imbalances in the 30-second interval t . x_t stands for other control variables which include DOI_{t-1} , the average spread and the volatility of the depth imbalance in the interval t , and ϵ_t is an error term.

The regression is performed separately for buy pinging activities on positive imbalances and sell pinging activities on negative imbalances, and across all trading days in June 2012. The final parameter estimates are calculated in the same way as in Table 2, i.e. time series averages of daily parameter estimates. Overall, I find that the parameter β is statistically significant for all three types of stocks, with a value around 0.28 on average. This shows that the first auxiliary prediction is confirmed in the data.

The second auxiliary prediction of the model concerns the influence of depth imbalance volatilities on the frequencies of cancelled pinging activities. The model implies that If the momentum strengthens (imbalance widens) by a large amount, contemporaneously the HFT would cancel his pinging orders and use market orders instead to chase the momentum. And if the momentum weakens by a large amount or even reverses (imbalance reduces by a large amount or reverses), contemporaneously the HFT would also cancel his pinging orders as his pinging orders risk being adversely hit. This implication thus yields the following prediction:

Prediction 2. *If the depth imbalance is more volatile, there should be more cancelled pinging*

³³The results are similar if I use 15-second or 1-minute intervals instead of 30-second intervals.

activities occurring at the same time.

I employ a similar procedure as before to check this prediction, i.e. I divide each trading day into 30-second intervals and compute the number of cancelled pinging activities in each interval. Then I use the following regression to measure and test the effect of imbalance volatilities on cancelled pinging orders:

$$CPO_t = \alpha + \beta VOD_t + x_t' \gamma + \epsilon_t. \quad (41)$$

CPO_t denotes the number of cancelled pinging orders and VOD_t the log of the depth imbalance volatility in the 30-second interval t . x_t again stands for other control variables which include VOD_{t-1} , the average spread and the number of market order arrivals in the interval t , and ϵ_t is an error term.

The regression is performed across all trading days in June 2012. The final parameter estimates are also calculated as the time series averages of daily parameter estimates. Overall, I find that the parameter β is statistically significant for all three types of stocks as well. Nevertheless, the magnitude of β is much higher for stocks with high depths and narrow spreads (with a value around 3.3) compared to stocks with low depths and wide spreads (with a value less than 1). Therefore, the result suggests that the second auxiliary prediction is by and large confirmed in the data too.

7 Conclusion

In this paper, I build a continuous-time, partial equilibrium model on the optimal strategies of HFTs without any learning or manipulative ingredients to rationalize pinging activities observed in the data. I improve on the works of Ho and Stoll (1981) and Guilbaud and Pham (2013) by introducing hidden orders inside the bid-ask spread and short-term price momentum into my model. The HFT then uses pinging orders inside the spread to either control his inventory or chase price trends. I demonstrate that for stocks with high order book depths

and narrow spreads, pinging accounts for 20% of the optimal strategies in the model, whereas this number goes up to 50% for stocks with low order book depths and wide spreads. I then compare these pinging percentages from the model to their corresponding counterparts in the data, and find that over 70% of the pinging activities in the data are captured by the model. The result thus suggests that most of the observable pinging activities from the data can be rationalized by my model.

Furthermore, I show that for low-depth and wide-spread stocks, the majority of the pinging in the model occurs due to the momentum-chasing motive of the HFT. However, for high-depth and low-spread stocks, the inventory-control motive and the momentum-chasing motive of the HFT play similar roles in rationalizing pinging activities. In addition, I also develop a couple of other auxiliary predictions based on the model's implications. They are both assessed on and found to be consistent with the data in general. Therefore, my model gives the overall message that pinging activities do not necessarily have to be manipulative and can be mostly rationalized as part of the standard dynamic trading strategies of HFTs.

8 Appendix

I show here the proofs for all lemmas and propositions, which are omitted in the paper.

Proof of Lemma 1: On one hand, I can show that the maximal value of (16) is smaller than that of (18). This is because, for any admissible strategy such that $Y_T = 0$, we immediately obtain $Q(X_T, Y_T, P_T, F_T, S_T) = X_T$. On the other hand, given an arbitrary admissible strategy θ and its associated state variable processes (X, Y, P, F, S) , I can consider an alternative strategy $\tilde{\theta}$, coinciding with θ up to time T and employing an market order that liquidates all the inventory Y_T at the terminal date T . The associated processes of the state variables $(\tilde{X}, \tilde{Y}, P, F, S)$ under $\tilde{\theta}$ satisfy $(\tilde{X}_t, \tilde{Y}_t, P_t, F_t, S_t) = (X_t, Y_t, P_t, F_t, S_t)$ for all $t < T$, and $\tilde{X}_T = Q(X_T, Y_T, P_T, F_T, S_T)$, $\tilde{Y}_T = 0$. Therefore, this shows that the maximal value of (18) is smaller than that of (16), hence (16) is equivalent to (18). \square

Proof of Lemma 2: The lower bound is obvious, since $V(t, x, y, p, f, s) = Q(x, y, p, f, s)$ by adopting the simple strategy that sets $\theta_t^{tk} = y$ and terminates the problem immediately at time t . On the other hand, we have

$$\begin{aligned} V &= (x + py) + \sup_{\theta} \mathbb{E}_t \left[(X_T - x) + (P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon \right) - \gamma \int_t^T Y_{u-}^2 d[P, P]_u \right] \\ &\leq (x + py) + \sup_{\theta} \mathbb{E}_t \left[(X_T - x) + (P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon \right) \right]. \end{aligned} \quad (42)$$

Since all the jump intensities $(\lambda^S, \lambda_{1,2}^J, \lambda^a, \lambda^b)$ are finite constants, the term

$$\sup_{\theta} \mathbb{E}_t \left[(X_T - x) + (P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon \right) \right] \quad (43)$$

cannot be greater than the finite, maximum profit achievable through a combination of a make strategy that participates in every trade when market orders arrive (an upper bound on $X_T - x$, denoted by C_0) and a directional, costless take strategy that bets on the mid-price jumps (an upper bound on $(P_T Y_T - py) - |Y_T| \left(\frac{S_T}{2} - H_T + \epsilon \right)$, denoted by C_1). \square

Proof of Lemma 3: As stated under (26) in the main paper, only the infinitesimal generators $(\mathcal{L}^P, \mathcal{L}^F, \mathcal{L}^S)$ for the value function V and the value of $\mathbb{E}_t d[P, P]_t$ require explicit expressions. They are given below:

$$\mathbb{E}_t d[P, P]_t = \left(\frac{1}{4} \lambda_1^J + \lambda_2^J \right) dt \quad (44)$$

$$\begin{aligned} \mathcal{L}^P \circ V(t, x, y, p, f, s) &= \left(V\left(t, x, y, p + \frac{\delta}{2}, f, s\right) \psi_1(f) + V\left(t, x, y, p - \frac{\delta}{2}, f, s\right) (1 - \psi_1(f)) \right) \lambda_1^J dt \\ &\quad + \left(V\left(t, x, y, p + \delta, f, s\right) \psi_2(f) + V\left(t, x, y, p - \delta, f, s\right) (1 - \psi_2(f)) \right) \lambda_2^J dt \end{aligned} \quad (45)$$

$$\mathcal{L}^F \circ V(t, x, y, p, f, s) = V_f(\alpha_F f) dt + \frac{1}{2} V_{ff} \sigma_F^2 dt \quad (46)$$

$$\mathcal{L}^S \circ V(t, x, y, p, f, s) = \left(\sum_{j=1}^3 \rho_{ij} [V(t, x, y, p, f, j\delta) - V(t, x, y, p, f, i\delta)] \right) \lambda^S dt, \quad (47)$$

where V_f and V_{ff} are the first and second order partial derivatives of V against the state variable F . In addition, $\mathbb{E}_t dP_t = \left(\lambda_1^J \frac{\delta}{2} (2\psi_1(F_t) - 1) + \lambda_2^J \delta (2\psi_2(F_t) - 1) \right) dt$. \square

Proof of Proposition 1: To prove Proposition 1, I need to show that the value function V in (19) is the unique viscosity solution to (25) and (26). Since I have established the necessary growth (boundness) conditions on V that are shown in Lemma 2, the proposition – the existence and the uniqueness of the viscosity solution V (a.k.a. the value function) – is then a direct application of standard arguments and results from stochastic control theory, e.g., Øksendal and Sulem (2007), Chapters 9 and 10. \square

Proof of Lemma 4: It is clear that the quasi-variational inequality (28) and the terminal condition (29) are simplified versions of (25) and (26), if I can decompose the value function $V(t, x, y, p, f, s)$ as $x + py + \nu(t, y, f, s)$. The required decomposition can be established by first noting that the mid-price P has constant jump intensities and jump size distributions depending only on the state variable F , and then extending to my scenario the argument for a simpler case shown in Guilbaud and Pham (2013). \square

Proof of Proposition 2: I prove Proposition 2 by first establishing three properties for my numerical scheme and then applying the theorem from Barles and Souganidis (1991).

Lemma. (*Monotonicity*)

For any $\Delta_T > 0$ such that $\Delta_T \leq (\lambda^S + (\pi_1 + (1 - \pi_1)\lambda^a) + (\pi_1 + (1 - \pi_1)\lambda^b))^{-1}$, the operator \mathcal{A} defined in equation (34) of the numerical scheme is non-decreasing in ϕ , i.e.

$$\text{if } \phi < \tilde{\phi}, \text{ then } \mathcal{A}(t, y, f, s, \phi) \leq \mathcal{A}(t, y, f, s, \tilde{\phi}), \quad \forall t, y, s \text{ and } f.$$

Proof. From the expression in equation (35) of the numerical scheme, it is clear that $g^a(f, s, \theta_t^{mk,b}) < \pi_1 + (1 - \pi_1)\lambda^a$, $\forall f, s, \theta_t^{mk,b}$, and $g^b(f, s, \theta_t^{mk,a}) < \pi_1 + (1 - \pi_1)\lambda^b$, $\forall f, s, \theta_t^{mk,a}$. Thus $1 - \Delta_T \lambda^S - \Delta_T g^a(f, s, \theta_t^{mk,b}) - \Delta_T g^b(f, s, \theta_t^{mk,a}) > 0$ as long as I choose the time grid Δ_T to be $\Delta_T \leq (\lambda^S + (\pi_1 + (1 - \pi_1)\lambda^a) + (\pi_1 + (1 - \pi_1)\lambda^b))^{-1} < (\lambda^S + g^a(f, s, \theta_t^{mk,b}) + g^b(f, s, \theta_t^{mk,a}))^{-1}$, which implies that $\tilde{\mathcal{L}}(t, y, f, s, \phi)$ is monotone in ϕ (as the sum of coefficients in front of ϕ is positive) and so is $\mathcal{A}(t, y, f, s, \phi)$. \square

Lemma. (*Stability*)

For any $\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F > 0$, there exists a unique solution ϖ to (37)-(38), and the sequence $\{\varpi\}$ is uniformly bounded.

Proof. By the definition of the backward scheme (37)-(38), the solution ϖ exists and is unique. The uniform bound follows from the growth condition (lower and upper bounds) on the reduced-form value function ν (a modification on Lemma 2, i.e. the bounds on V). \square

Lemma. (*Consistency*)

The scheme (37)-(38) is consistent, i.e. for all $(t, y, f) \in [0, T] \times \mathbb{R} \times \mathbb{R}$ and any smooth test function ϕ , as $(\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F) \rightarrow (0, 0, \infty, 0, \infty)$, and $(t', y', f') \rightarrow (t, y, f)$,

$$\begin{aligned} & \lim_{\Delta_T} \frac{1}{\Delta_T} [\tilde{\mathcal{L}}(t' + \Delta_T, y', f', s, \phi) - \phi(t', y', f', s)] \\ &= \frac{\partial \phi}{\partial t} + y \frac{\mathbb{E}_t dP_t}{dt} + \mathcal{L}^{\mathcal{F}} \circ \phi + \mathcal{L}^{\mathcal{S}} \circ \phi - \gamma y^2 \frac{\mathbb{E}_t d[P, P]_t}{dt} + \\ & \quad \sup_{\theta^{mk}} \left\{ g^a(f, s, \theta_t^{mk, b}) \cdot (\phi(t, y + 1, f, s) - \phi(t, y, f, s) + s/2 - \delta \theta_t^{mk, b}) + \right. \\ & \quad \left. g^b(f, s, \theta_t^{mk, a}) \cdot (\phi(t, y - 1, f, s) - \phi(t, y, f, s) + s/2 - \delta \theta_t^{mk, a}) \right\} \end{aligned} \quad (48)$$

and

$$\lim_{\zeta} \tilde{\mathcal{M}} \circ \tilde{\mathcal{L}}(t', y', f', s, \phi) = \sup_{\zeta} \left\{ \phi(t, y + \zeta, f, s) - |\zeta|(s/2 + \epsilon) + |y| \int_H H dG(H | s, f) \right\} \quad (49)$$

Proof. This follows from the result in Section 6.1.2 of Chen and Forsyth (2008). \square

Proposition. (*Convergence*) The solution ϖ to the numerical scheme of (37)-(38) and the corresponding discretized policy functions converge locally uniformly to the reduced-form value function ν and the equilibrium optimal strategies respectively on $[0, T] \times \mathbb{R} \times \mathbb{R}$ as $(\Delta_T, \Delta_Y, M_Y, \Delta_F, M_F) \rightarrow (0, 0, \infty, 0, \infty)$, $\forall s \in \mathcal{S}$.

Proof. Given the properties of monotonicity, stability and consistency of the numerical scheme, this is a direct application of the result of Barles and Souganidis (1991). \square

References

- Aït-Sahalia, Y., 1999, “Transition densities for interest rate and other nonlinear diffusions,” *Journal of Finance*, 54, 1361–1395.
- Aït-Sahalia, Y., and P. A. Mykland, 2003, “The effects of random and discrete sampling when estimating continuous-time diffusions,” *Econometrica*, 71, 483–549.
- Avellaneda, M., and S. Stoikov, 2008, “High-frequency trading in a limit order book,” *Quantitative Finance*, 8, 217–224.
- Barles, G., and P. E. Souganidis, 1991, “Convergence of approximation schemes for fully nonlinear second order equations,” *Asymptotic Analysis*, 4, 271–283.
- Baron, M., J. Brogaard, and A. A. Kirilenko, 2014, “Risk and return in high frequency trading,” Working paper, available at SSRN: <http://papers.ssrn.com/abstract=2433118>.
- Baruch, S., and L. R. Glosten, 2013, “Flickering quotes,” Working paper, available at the following address: <https://www0.gsb.columbia.edu/faculty/lglosten/papers/Baruch-Glosten-Flickering-Quotes-final.pdf>.
- Brogaard, J., B. Hagströmer, L. L. Nordén, and R. Riordan, 2014, “Trading fast and slow: Colocation and market quality,” Working paper, available at SSRN: <http://papers.ssrn.com/abstract=2317797>.
- Brogaard, J., T. Hendershott, and R. Riordan, 2014, “High-frequency trading and price discovery,” *Review of Financial Studies*, 27, 2267–2306.
- Buti, S., and B. Rindi, 2013, “Undisclosed orders and optimal submission strategies in a limit order market,” *Journal of Financial Economics*, 109, 797–812.
- Chen, Z., and P. A. Forsyth, 2008, “A numerical scheme for the impulse control formulation for pricing variable annuities with a guaranteed minimum withdrawal benefit (GMWB),” *Numerische Mathematik*, 109(4), 535–569.

- Cont, R., and E. Voltchkova, 2005, “A finite difference scheme for option pricing in jump diffusion and exponential Lévy models,” *SIAM Journal on Numerical Analysis*, 43(4), 1596–1626.
- Crandall, M. G., H. Ishii, and P.-L. Lions, 1992, “User’s guide to viscosity solutions of second order partial differential equations,” *Bulletin of the American Mathematical Society*, 27, 1–67.
- Crandall, M. G., and P.-L. Lions, 1983, “Viscosity solutions of Hamilton-Jacobi equations,” *Transactions of the American Mathematical Society*, 277, 1–42.
- Easley, D., M. López de Prado, and M. O’Hara, 2011, “The microstructure of the "Flash Crash": Flow toxicity, liquidity crashes, and the probability of informed trading,” *Journal of Portfolio Management*, 37(2), 118–128.
- Fleming, W. H., and H. M. Soner, 2005, *Controlled markov processes and viscosity solutions*. Springer, New York, 2nd edn.
- Gould, M. D., M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison, 2013, “Limit order books,” *Quantitative Finance*, 13, 1709–1742.
- Guilbaud, F., and H. Pham, 2013, “Optimal high-frequency trading with limit and market orders,” *Quantitative Finance*, 13, 79–94.
- Hagströmer, B., and L. L. Nordén, 2013, “The diversity of high-frequency traders,” *Journal of Financial Markets*, 16, 741–770.
- Hasbrouck, J., and G. Saar, 2009, “Technology and liquidity provision: The blurring of traditional definitions,” *Journal of Financial Markets*, 12, 143–172.
- , 2013, “Low-latency trading,” *Journal of Financial Markets*, 16, 646–679.

- Hautsch, N., and R. Huang, 2012a, “Limit order flow, market impact, and optimal order sizes: Evidence from NASDAQ TotalView-ITCH data,” in *Market Microstructure: Confronting Many Viewpoints*. Wiley, United Kingdom, pp. 137–161.
- Hautsch, N., and R. Huang, 2012b, “On the dark side of the market: Identifying and analyzing hidden order placements,” Working paper, available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2004231.
- Hendershott, T., C. M. Jones, and A. J. Menkveld, 2011, “Does algorithmic trading improve liquidity?,” *Journal of Finance*, 66, 1–33.
- Hendershott, T., and R. Riordan, 2013, “Algorithmic trading and the market for liquidity,” *Journal of Financial and Quantitative Analysis*, 48, 1001–1024.
- Hirschey, N., 2013, “Do high-frequency traders anticipate buying and selling pressure?,” Working paper, available at SSRN: <http://ssrn.com/abstract=2238516>.
- Ho, T., and H. R. Stoll, 1981, “Optimal dealer pricing under transactions and return uncertainty,” *Journal of Financial Economics*, 9, 47–73.
- Karr, A., 1991, *Point processes and their statistical inference*. CRC / Marcel Dekker, Inc., New York, 2nd edn.
- Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun, 2014, “The Flash Crash: The impact of high frequency trading on an electronic market,” Working paper, available at SSRN: <http://papers.ssrn.com/abstract=1686004>.
- Lewis, M., 2014, *Flash Boys: A Wall Street Revolt*. W. W. Norton & Company, New York.
- Menkveld, A. J., 2013, “High frequency trading and the *new market makers*,” *Journal of Financial Markets*, 16, 712–740.
- Øksendal, B., and A. Sulem, 2007, *Applied stochastic control of jump diffusions*. Springer, New York, 2nd edn.

Table 1: Estimated Model Parameters and Fixed Parameters

Panel (A) of this table summarizes the model parameters to be estimated that govern the exogenous evolutions of the limit order book variables, as well as their respective estimation procedures. S stands for the bid-ask spread, F for the depth imbalance at the best quotes, and P for the mid-price. Prob. is an abbreviation for the word probability. The estimates of these parameters are reported in Table 2. Panel (B) of this table depicts the values of the fixed, non-estimated parameters that are used in my numerical analysis of the optimal HFT strategies. The \$0.003 per share fee for market orders corresponds to the same transaction fee stated by NASDAQ for its stocks. Furthermore, the tick size δ is set to be \$0.01 for Intel and Qualcomm, which is the tick size for all NASDAQ stocks that have prices above \$1. For Amazon, the tick size is set to be \$0.1, and the reason is given in Table 2. The word MC stands for Monte Carlo.

Panel (A): Model parameters to be estimated			
Parameter	Definition	Estimation Procedure	
λ^S, ρ	Jump intensity and transition matrix of S	Non-parametric estimators	
α_F, σ_F	Mean-reversion and volatility parameters of F	Maximum likelihood	
$\lambda_{1,2}^J$	Jump intensities of P	Non-parametric estimators	
$\beta_{1,2}$	Prob. distribution parameters of directions of P jumps	Logistic regressions	
$\kappa, \pi_{1,2}$	Prob. distribution parameters of hidden orders	Logistic regressions	
λ^a, λ^b	Exogenous sell (a) and buy (b) market order arrival intensities	Non-parametric estimators	
$\varsigma_{0,1}$	Parameters of limit order fill rates when hit by market orders	Logistic regressions	

Panel (B): Fixed parameters			
Parameter type	Parameter	Description	Value
Discretization and localization parameters	T	Time length in seconds	3600
	Δ_T	Size of time step in seconds	0.5
	M_Y	Inventory grid bound (in lot)	30
	Δ_Y	Inventory grid step size (in lot)	1
	M_F	Depth imbalance grid bound	10
	Δ_F	Depth imbalance grid step size	0.01
Model constants	δ	Tick size	0.01
	ϵ	Per share fee	0.003
	γ	Inventory penalization	2
	ζ_{max}	Max market order size (in lot)	10
Simulation parameters	N_{MC}	Number of MC simulation paths	10000
	X_0, Y_0	Initial cash and inventory of HFT	0
	P_0	Initial mid-price of stock	10

Table 2: Estimates of Limit Order Book Parameters

This table shows the average values of the daily estimates of the limit order book parameters in the month of June 2012 for the three representative stocks Intel (INTC), Qualcomm (QCOM) and Amazon (AMZN), using order book data reconstructed from NASDAQ TotalView-ITCH 4.0 message feeds. Newey-West HAC standard errors are in parentheses. Standard errors of the transition matrix ρ estimates are not shown as they are close to zero. Intensity parameters are all measured at per second frequency. In addition, I normalize the tick size δ to be \$0.1 for Amazon, since on average, the spread and limit order prices of Amazon tend to be in multiples of \$0.1 instead of the minimum tick size \$0.01 used for Intel and Qualcomm. For definitions of all the parameters, please refer to Table 1.

		INTC	QCOM	AMZN
Spread	λ^S	0.161/s (0.010)	0.312/s (0.017)	0.578/s (0.027)
	ρ	$\begin{pmatrix} 0 & 1 & 0 \\ 0.99 & 0 & 0.01 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.97 & 0.03 \\ 0.95 & 0 & 0.05 \\ 0.08 & 0.92 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.743 & 0.257 \\ 0.222 & 0 & 0.778 \\ 0.138 & 0.862 & 0 \end{pmatrix}$
Imbalance	α_F	0.308 (0.016)	0.547 (0.027)	0.734 (0.035)
	σ_F	0.777 (0.029)	1.429 (0.053)	2.336 (0.074)
Mid-price	λ_1^J	0.161/s (0.010)	0.307/s (0.018)	0.522/s (0.027)
	β_1	2.744 (0.077)	2.651 (0.076)	2.610 (0.075)
	λ_2^J	0.052/s (0.003)	0.075/s (0.005)	0.121/s (0.007)
	β_2	4.766 (0.244)	2.921 (0.257)	1.881 (0.129)
Hidden order	κ	1.196 (0.034)	1.036 (0.030)	0.992 (0.021)
	π_1	0.238 (0.010)	0.233 (0.010)	0.230 (0.009)
	π_2	0.125 (0.006)	0.117 (0.006)	0.114 (0.006)
Market order	λ^a	0.110/s (0.006)	0.130/s (0.006)	0.181/s (0.012)
	λ^b	0.110/s (0.004)	0.130/s (0.006)	0.179/s (0.010)
Fill rate	ς_0	1.320 (0.039)	1.454 (0.055)	1.648 (0.066)
	ς_1	0.399 (0.019)	0.462 (0.022)	0.673 (0.031)

Table 3: Percentage Breakdown of the Optimal HFT Strategies

The table reports, for the three representative stocks Intel (INTC), Qualcomm (QCOM) and Amazon (AMZN), the percentage attributions (averaged over simulations) of the optimal HFT strategies to three types of activity: traditional market making & inventory control, pinging, and momentum (directional) trading via market orders. For the explanation on how these percentage attributions are obtained, please refer to Section 6.1 in the main article.

	INTC	QCOM	AMZN
Traditional market making & Inventory control	70.5%	58.3%	31.8%
Pinging	19.7%	28.1%	46.7%
Momentum trading via market orders	9.8%	13.6%	21.5%

Table 4: Model’s Pinging v.s. Data’s Pinging

This table compares the pinging percentages (pinging as a percentage of all limit order book activities) obtained from the model to the pinging percentages observable from the data for the three representative stocks Intel (INTC), Qualcomm (QCOM) and Amazon (AMZN), in order to illustrate how much pinging in the data can be rationalized by the model (i.e. the % of data’s pinging captured by the model). Panel (A) reports the pinging percentages produced by the model that are the same as the corresponding numbers shown in Table 3. These are expressed as a fraction of the pinging percentage numbers observed in the data, which is then the percentage of data’s pinging captured by the model. Panel (B) breaks down the pinging percentages produced by the model when one of the limit order book features is eliminated from the model. No hidden orders means the existence probabilities of the hidden orders in the model are set to zero, i.e. the parameters of π_1 and π_2 are fixed at zero. No price momentum means the depth imbalance F of the limit order book does not have any predictive power for the directions of the mid-price jumps in the model, i.e. the parameters β_1 and β_2 are set to zero, which implies that the mid-price has equal probabilities of jumping up or down. For the detailed explanation on how pinging percentages from the data are calculated, please refer to Section 6.1 in the main article.

Panel (A): Pinging percentages from the model and the data			
	INTC	QCOM	AMZN
Pinging percentage produced by model	19.7%	28.1%	46.7%
Pinging percentage observed in data	23%	39%	70%
Percentage of data’s pinging captured by model	86%	72%	67%
Panel (B): Further breakdown on pinging percentages from the model			
	INTC	QCOM	AMZN
Model’s pinging percentage, no hidden orders	8%	15%	31%
Model’s pinging percentage, no price momentum	12%	13%	16%

Figure 1: Optimal Strategy, INTC, $t = 10$, $\text{spread} = \delta$ (left) / 2δ (right)

The figure shows the optimal HFT strategy profile for Intel (INTC) at time $t = 10$, with the spread S equal to δ (left panel) and 2δ (right panel). The HFT's inventory and the depth imbalance levels are shown on the horizontal and the vertical axes respectively. The optimal strategy is computed via the implementation of my numerical scheme (37)–(38), using the parameter estimates in Table 2 and the fixed parameters in Panel (B) of Table 1 as inputs. For given values of the inventory level and the depth imbalance, each colored region denotes a particular type of HFT strategy that is stated in the figure. Specifically, “market making” is the strategy of submitting a limit order at both the best bid and ask prices. “Partial inventory control” means using market orders to unwind the inventory partially, whereas “inventory control” means using market orders to change the inventory precisely back to zero, and buy (sell) indicates using buy (sell) market orders to increase (decrease) the inventory. “Momentum” is the strategy of utilizing market orders to change the inventory and set up a directional position: buy and sell denote establishing positive and negative directional inventory holding positions respectively. “Pinging on the bid (ask) side” is the strategy of submitting a buy (sell) limit order inside the spread at the best bid plus δ (best ask minus δ), while letting the sell (buy) limit order joining the queue at the best ask (best bid).

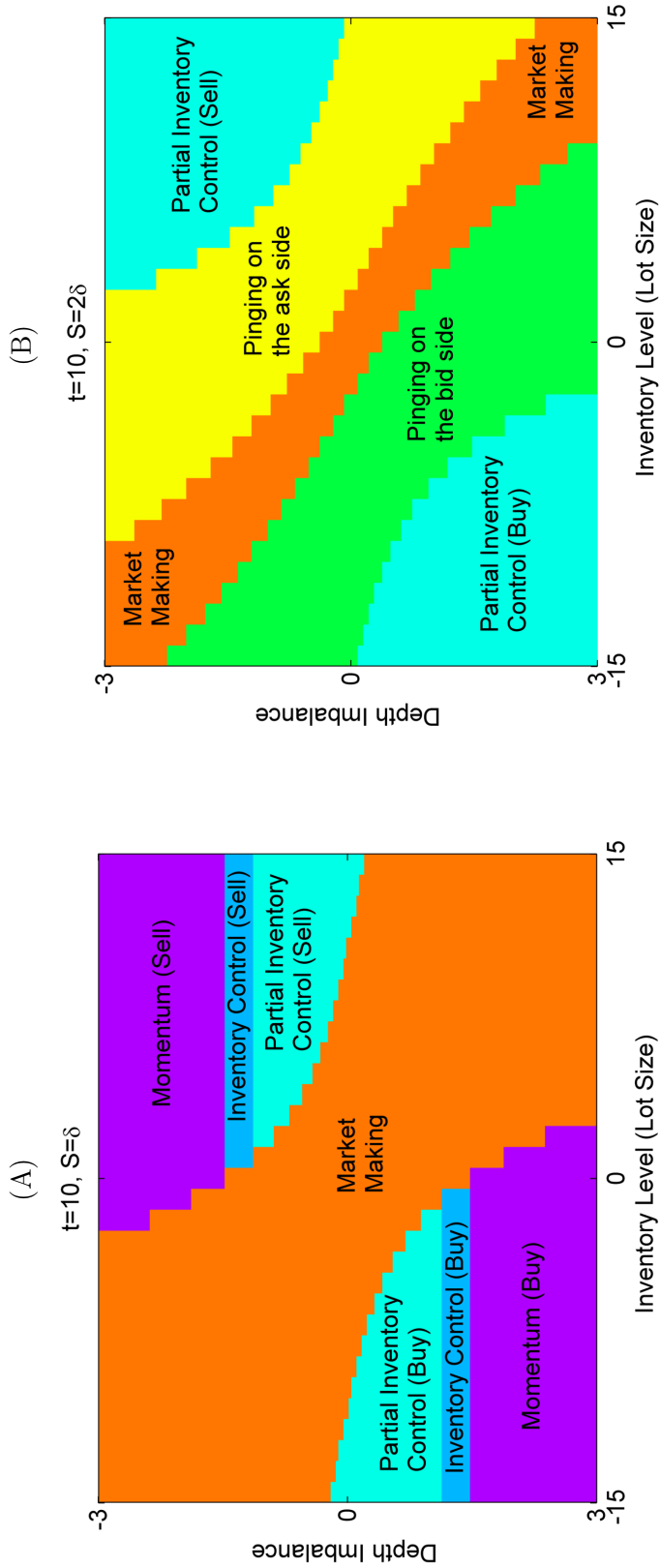


Figure 2: Optimal Strategy, INTC, $t = T - 3$, spread = δ (left) / 2δ (right)

This figure shows the optimal HFT strategy profile for Intel (INTC) at time $t = T - 3$, with the spread S equal to δ (left panel) and 2δ (right panel). For given values of the inventory level and the depth imbalance, each colored region represents a particular type of HFT strategy, which is stated in the figure. Specifically, “market making” is the strategy of submitting a limit order at both the best bid and ask prices. “Partial inventory control” means using market orders to unwind the inventory partially, whereas “inventory control” means using market orders to change the inventory precisely back to zero, and buy (sell) market orders to increase (decrease) the inventory. “Pinging on the bid (ask) side” is the strategy of submitting a buy (sell) limit order inside the spread at the best bid plus δ (best ask minus δ), while letting the sell (buy) limit order joining the queue at the best ask (best bid). “Mkt MK” in the right panel stands for market making.

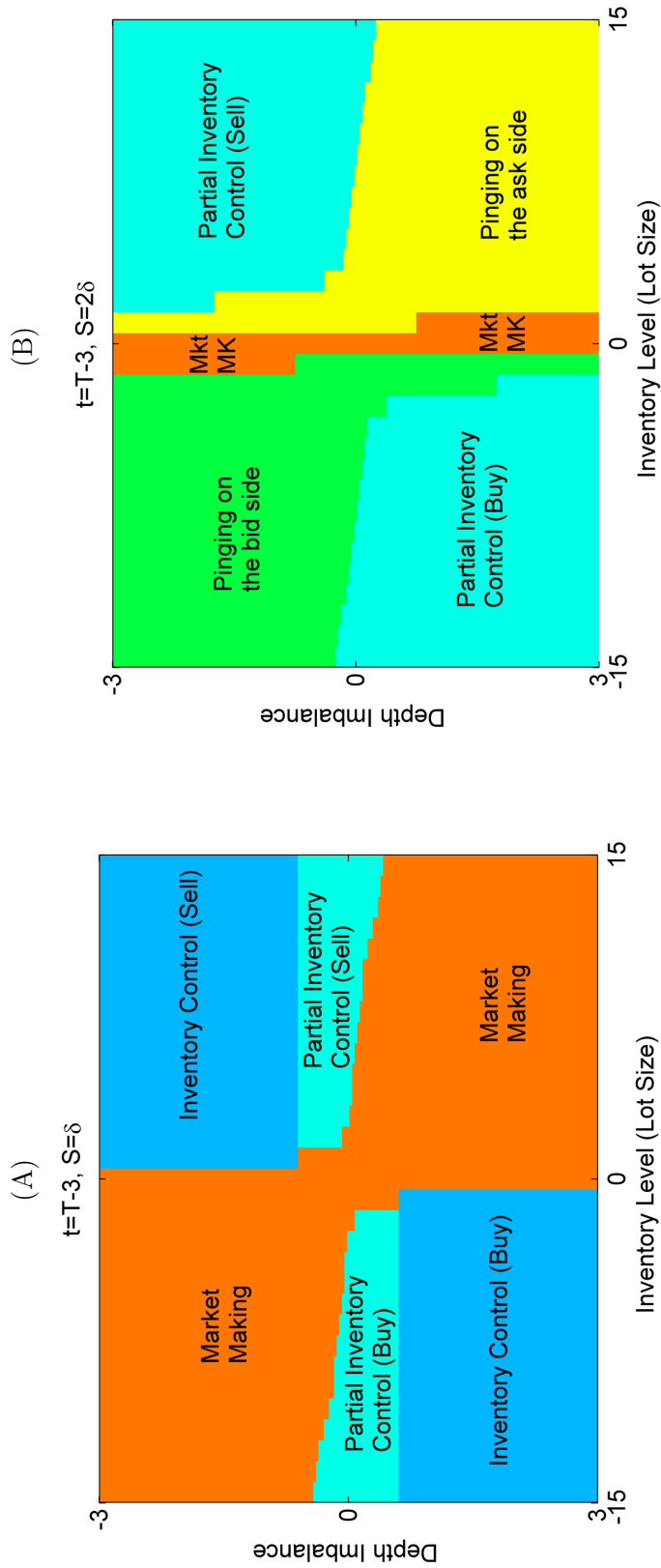


Figure 3: Optimal Strategy, INTC (left) v.s. AMZN (right), $t = 10$, spread = δ

This figure compares the optimal HFT strategy profile for Intel (INTC, left panel) with that for Amazon (AMZN, right panel) at time $t = 10$, with the spread S equal to δ . For given values of the inventory level and the depth imbalance, each colored region represents a particular type of HFT strategy, which is stated in the figure. Specifically, “market making” is the strategy of submitting a limit order at both the best bid and ask prices. “Partial inventory control” means using market orders to unwind the inventory partially, whereas “inventory control” means using market orders to change the inventory precisely back to zero, and buy (sell) indicates using buy (sell) market orders to increase (decrease) the inventory. “Momentum” is the strategy of utilizing market orders to change the inventory and set up a directional position: buy and sell denote establishing positive and negative directional inventory holding positions respectively.

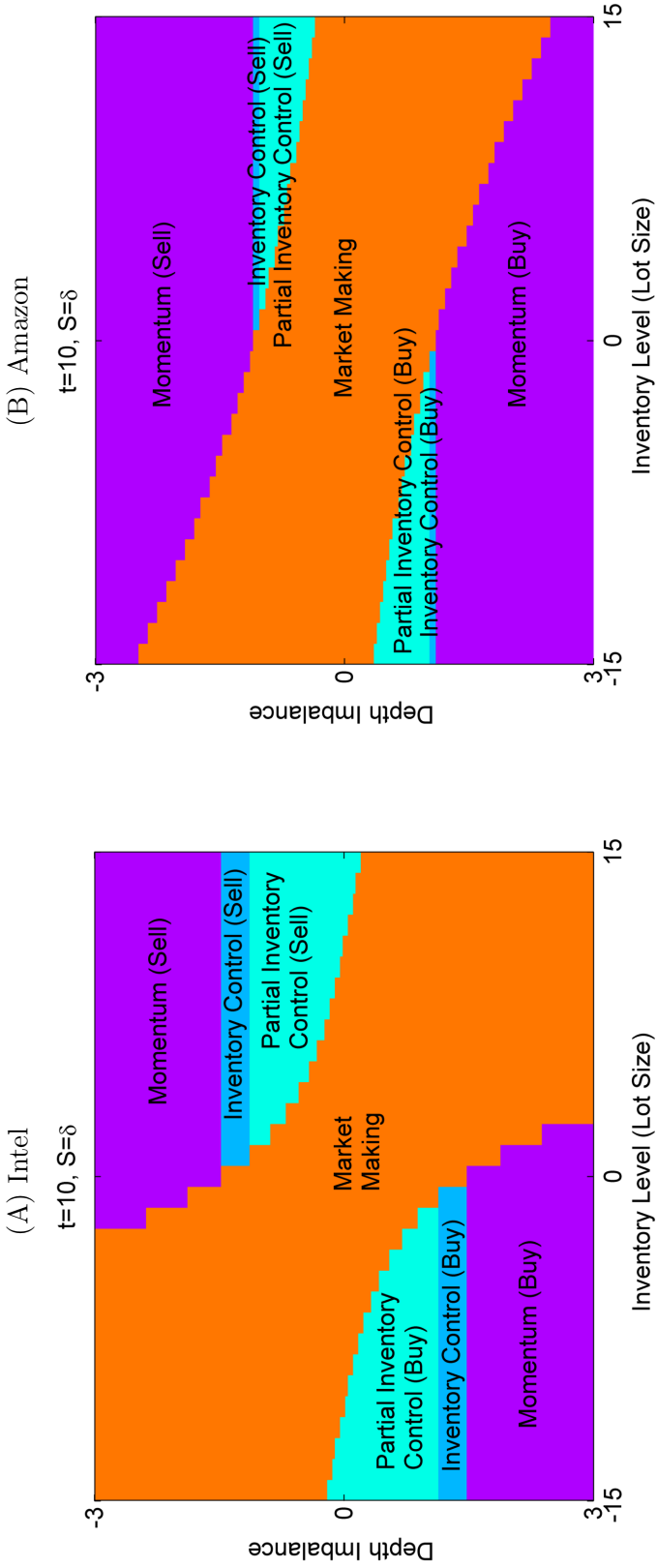


Figure 4: Optimal Strategy, INTC (left) v.s. AMZN (right), $t = 10$, spread = 2δ

This figure compares the optimal HFT strategy profile for Intel (INTC, left panel) with that for Amazon (AMZN, right panel) at time $t = 10$, with the spread S equal to 2δ . For given values of the inventory level and the depth imbalance, each colored region represents a particular type of HFT strategy, which is stated in the figure. Specifically, “market making” is the strategy of submitting a limit order at both the best bid and ask prices. “Partial inventory control” means using market orders to unwind the inventory partially, whereas “inventory control” means using market orders to change the inventory precisely back to zero, and buy (sell) indicates using buy (sell) market orders to increase (decrease) the inventory. “Momentum” is the strategy of utilizing market orders to change the inventory and set up a directional position: buy and sell denote establishing positive and negative directional inventory holding positions respectively. “Pinging on the bid (ask) side” is the strategy of submitting a buy (sell) limit order inside the spread at the best bid plus δ (best ask minus δ), while letting the sell (buy) limit order joining the queue at the best ask (best bid).

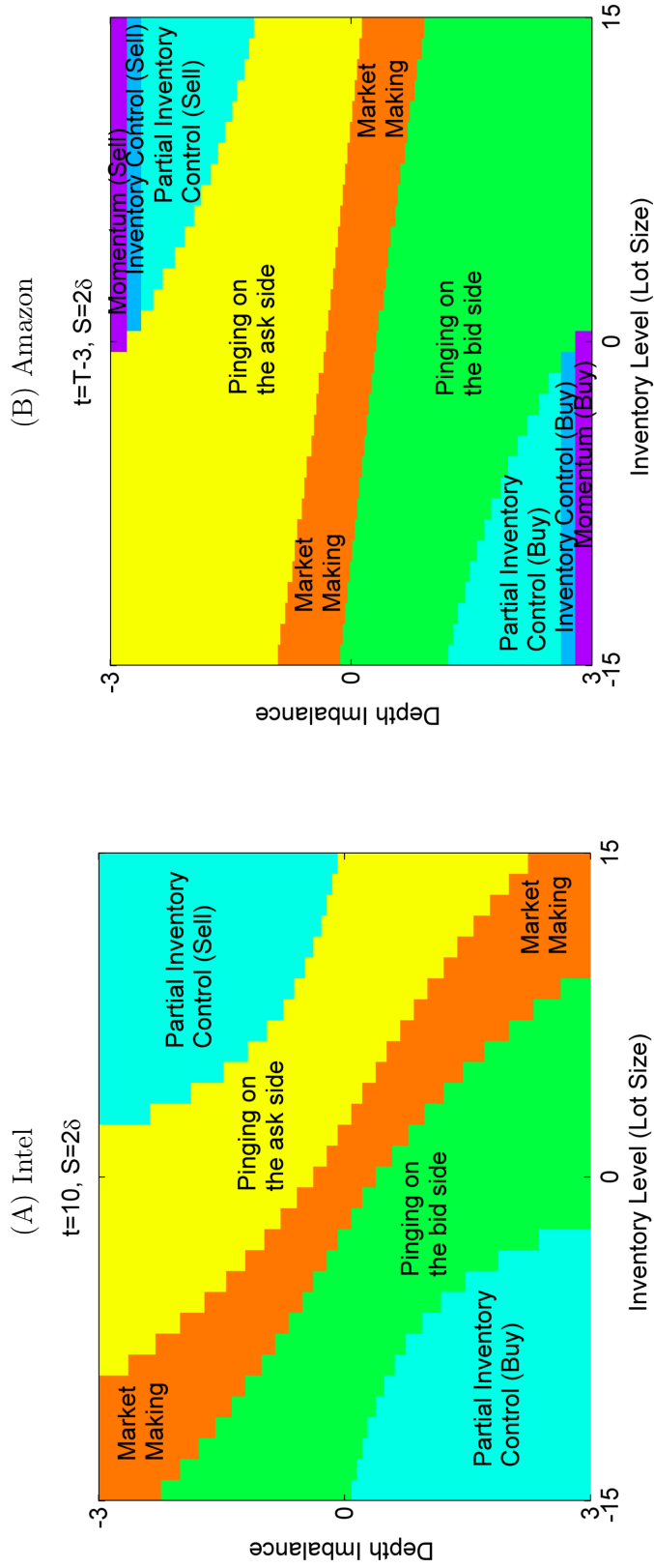


Figure 5: Optimal Strategy, INTC (left) v.s. AMZN (right), $t = T - 3$, spread = δ

This figure compares the optimal HFT strategy profile for Intel (INTC, left panel) with that for Amazon (AMZN, right panel) at time $t = T - 3$, with the spread S equal to δ . For given values of the inventory level and the depth imbalance, each colored region represents a particular type of HFT strategy, which is stated in the figure. Specifically, “market making” is the strategy of submitting a limit order at both the best bid and ask prices. “Partial inventory control” means using market orders to unwind the inventory partially, whereas “inventory control” means using market orders to change the inventory precisely back to zero, and buy (sell) indicates using buy (sell) market orders to increase (decrease) the inventory.

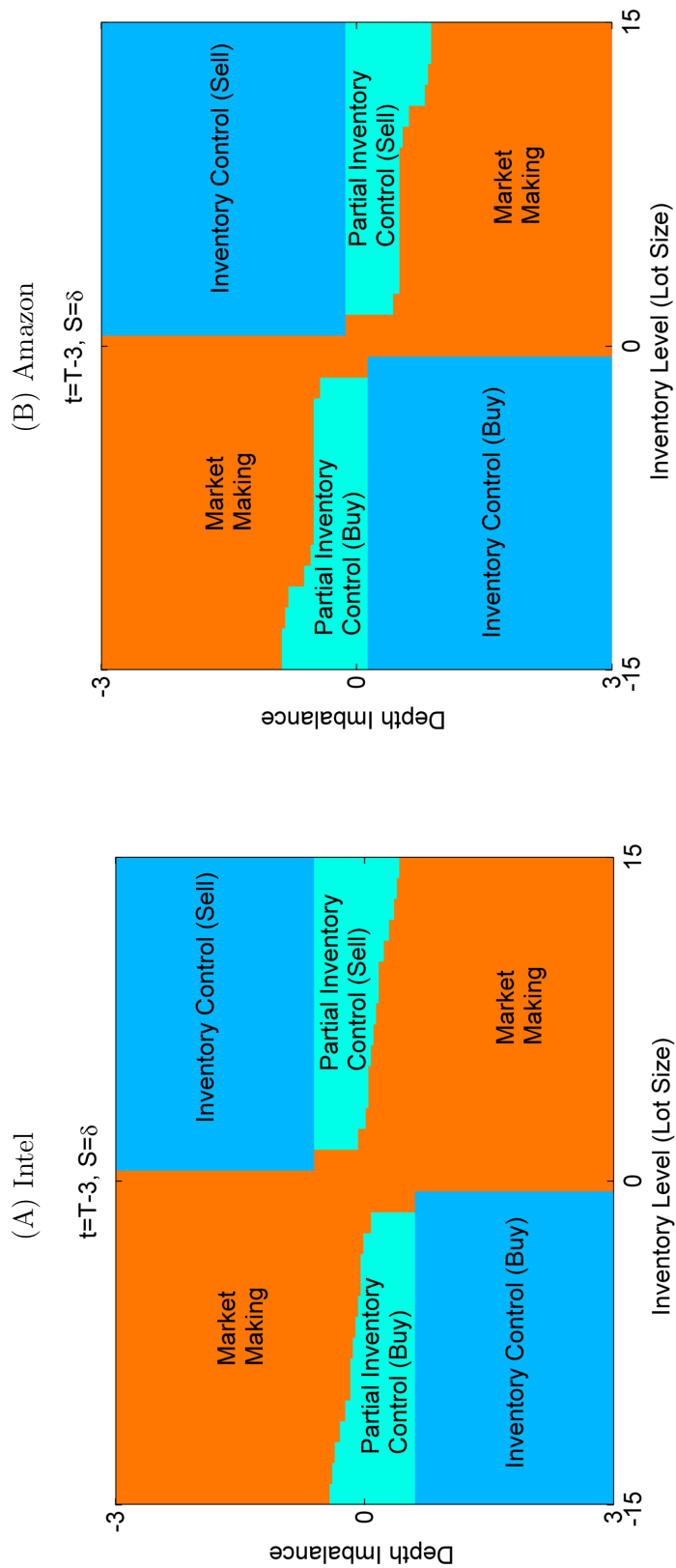


Figure 6: Optimal Strategy, INTC (left) v.s. AMZN (right), $t = T - 3$, $\text{spread} = 2\delta$

This figure compares the optimal HFT strategy profile for Intel (INTC, left panel) with that for Amazon (AMZN, right panel) at time $t = T - 3$, with the spread S equal to 2δ . For given values of the inventory level and the depth imbalance, each colored region represents a particular type of HFT strategy, which is stated in the figure. “Mkt MK” in both panels stands for market making. Specifically, “market making” is the strategy of submitting a limit order at both the best bid and ask prices. “Partial inventory control” means using market orders to unwind the inventory partially, and buy (sell) indicates using buy (sell) limit orders to increase (decrease) the inventory. “Pinging on the bid (ask) side” is the strategy of submitting a buy (sell) limit order inside the spread at the best bid plus δ (best ask minus δ), while letting the sell (buy) limit order joining the queue at the best ask (best bid).

