Ann Oper Res (2014) 223:403–432
DOI 10.1007/s10479-013-1382-8

High frequency trading, liquidity, and execution cost

Edward W. Sun · Timm Kruse · Min-Teh Yu

Published online: 25 April 2013

Abstract We build a model under the framework of discrete optimization to explain how high frequency trading (HFT) can be applied to supply liquidity and reduce execution cost. We derive the analytical properties of our model in finding the optimal solution to minimize the overall execution cost of HFT. We show that the execution cost can be reduced after increasing trading frequency (i.e., the higher the trading frequency, the lower the execution cost) with a simulation study. In addition, we conduct an empirical investigation with tick level data from US equity market through January 2008 to October 2010 to verify our conclusion drawn from the simulation study. Based on the simulation and empirical results we collected, we show that the HFT can reduce the execution cost when supplying liquidity.

Keywords Discrete optimization · High frequency trading · Liquidity · Price impact · Optimal execution

1 Introduction

Technological innovation has improved the trading capacity of financial assets. Many market participants are now employing algorithmic trading (AT) that uses computer algorithms to automatically complete the trading process which involves making trade decision, order

The authors thank the editor, G. Arivarignan, and two referees for their helpful comments on an earlier version of this paper. The authors thank Cheng-Few Lee and Quentin Chu for their valuable comments.

E.W. Sun
KEDGE Business School & BEM Management School Bordeaux, 680 Cours de la Libèration, 33405 Talance Cedex, France
e-mail: edward.sun@bem.edu

T. Kruse
d-fine GmbH, Frankfurt am Main, Germany

M.-T. Yu (✉)
Institute of Finance, National Chiao Tung University, Hsinchu 30010, Taiwan
e-mail: mtyu@nctu.edu.tw

submission, and order management after submission. High frequency trading (HFT) is a special class of algorithmic trading (AT) that uses computer algorithms making elaborate trading decision based on the electronically received information before human traders are capable of processing same information they observe. Much attention has, and continues to be, focused on the impact of high frequency trading, particularly due to the tremendous increases in volumes of HFT and the U.S. May 6, 2010 flash crash.

Many market participants are claiming that HFT leads to flickering quotes and disappearing liquidity, and others believe that HFT benefits the market by adding liquidity resulting in reduced spreads and lower volatility. It is important to understand the mechanism of HFT. Well understanding of HFT can help the regulator or policy maker to apply more effective tools to regulate the market, for example, to introduce financial transaction tax or to restrain certain kind of trading activities. Market participants will benefit from understanding the mechanism of HFT by improving the efficiency of their trading algorithms. Exchanges will provide better service for HFT in order to maintain market quality and control transaction cost. Several studies have investigated the empirical results to support that AT improves liquidity and enhances the informativeness of quotes, see Hendershott et al. (2011), for example. Most documented studies are focusing on the empirical consideration for HFT assuming that the HFT tasks can be successfully realized. Biais et al. (2011) propose an equilibrium analysis of HFT and conclude that HFT can increase gains from trade and generate adverse selection. Jovanovic and Menkveld (2012) suggest a model and show that HFT entry can indeed increase welfare, but it might also reduce it.

In this paper, we are going to investigate how HF trader to achieve the trading goal (i.e., providing liquidity and reduce execution cost) based on his optimal policy. Market maker (sometimes called designated market maker or specialist) serves as liquidity provider to guarantee that every trader who wants to trade can find a counterparty willing to trade the same amount at that time. One of the market maker's privilege is to observe the order flow. By observing order flow, HFT algorithm is capable of extracting information that has not yet crossed the news screens. Since all quote and volume information is public, such behavior is fully compliant with all the applicable laws. Typically, when conducting HFT, trader quotes a pair of bid-ask prices to trade with two simultaneously incoming orders and takes the spread as his profit.

In this paper we propose a model considering the motivation of HFT for liquidity supply, that is, HF traders act as market makers providing liquidity by placing passive orders. Currently, the regulation shows a tendency to require the HFT firms post firm quotes at competitive prices, thus provide a regular source of liquidity and effectively turn HFT firms into market makers.[1] Following Easley et al. (2012), we consider the passive order does not cross the market and generates adverse selection cost[2] for HFT since we cannot directly control the timing of its execution. We focus on the unique characteristics of HFT, that is, low latency such that the computerized algorithms will elaborate the trading strategy based on the information electronically received before human traders process on the same information they observed. We define the information electronically received as the "news" on the limit order book. We have already learned that a particular trade or order (for example, large passive orders) placed by HF traders will generate high impact on asset price and increase transaction costs (due to adverse selection). Easley et al. (2012) point out that order flow is toxic when

[1] For example, see Article 51 of The Markets in Financial Instruments Directive (MiFID II) proposed by the European Commission in 2011.

[2] The passive orders are filled quickly when they should be filled slowly and filled slowly when they should be filled quickly, see Jeria and Sofianos (2008).

it adversely selects market makers who provide liquidity at a loss. Therefore, the objective of HFT in our model is to minimize the execution cost by optimally executing orders.

Similar as Bertsimas and Lo (1998), Almgren and Chriss (1999, 2000), Obizhaeva and Wang (2013), and Sun and Kruse (2012) the model we presented to show the optimal execution are relying on the price impact functions that characterize the impact of a sequence of trades on prices when the trades are executed. In our model we decompose the price impact resulting from HFT into two parts: a permanent impact when the price moves to a new direction (upward or downward), and a temporary impact which instantaneously affects the trade that has been triggered. Huberman and Stanzel (2004) point out that the linear price impact function excludes the quasi-arbitrage (i.e., price manipulation) and support viable market prices. Therefore, we adopt the linear price impact function in our model, that is, a linear combination of the permanent and temporary price impact is characterized by the price impact function. In order to reduce the adverse selection, HFT algorithms will predict the liquidity change after the news (i.e., the information electronically received by HFT algorithms) disseminated to all traders. HFT algorithms will then decide the optimal order submission strategy accordingly in order to fill the large order undertaken.

Easley et al. (2012) argue that in the high frequency world, the order arrival process is informative and volume arrival is a metric for it. In this paper, we model the effect of disseminated news on liquidity (captured by volume arrival) with depth and the resilience effect of order flow on the limit order book for the underlying asset (see Patell and Wolfson 1984 and Lee et al. 1993). Based on the resilience model proposed by Obizhaeva and Wang (2013), we allow the variables that model the limit order book to be changed at the predetermined time points within the trading period. Based on this setting we can characterize different order flows measured by volume arrival (as liquidity increasing, decreasing, or remaining constant) to the "news" by assigning a probability to different consequences after the disseminated news. This makes our model more flexible to capture different market situations since we allow the permanent and temporary market impact to be adjusted accordingly to the order flow changes. In practice, HFT algorithms can define the probability based on their own estimations or other indicators (for example, VPIN suggested by Easley et al. 2012) so that our model can be implemented easily.

We show the importance of incorporating drift of market when the HFT algorithms looking for the optimal execution. The higher the altitude market moves to one direction (upward or downward), the more important the drift variable considered in our model turns to be. We show that the market volatility (measured by the variance of underlying price changes) has no significant influence on the optimal HFT strategy, although it depends on the fair price dynamics. Our model suggests that (i) it is better not to provide liquidity immediately after detecting the disseminated order flow, and (ii) market makers only start to supply liquidity after adopting the trading strategy (i.e., size of the passive order) accordingly to the volume arrival inferred from order flow.

In this paper we compare the performance of our HFT trading strategy with two alternative trading strategies: the optimal trading without considering HFT originally proposed by Obizhaeva and Wang (2013) and the non-optimal trading (i.e., naive trading by submitting orders with equalized size) mentioned in Bertsimas and Lo (1998). We run a simulation to examine the analytical properties derived from our model, and show that the HFT performs better in providing liquidity and reducing execution cost than other trading strategies we investigated. In addition, we empirically verify the performance of HFT by using tick level data of 105 stocks from S&P 100, S&P MidCap 400, and S&P SmallCap 600 component companies in US market through January 2008 to October 2010. We found that (i) the empirical results coincide with those from the simulation study, that is, HFT can significantly reduce the execution cost; and (ii) when increasing the trading frequency, the overall execution cost of HFT is significantly reduced comparing with other non-HFT strategies investigated, particularly in our sample, the execution cost will be reduced between 25 bps and 33 bps per trade per day when fixed the size of position for providing liquidity.

The rest of paper is organized as follows. An introduction of the model setup is provided in Sect. 2. Section 3 describes our contributions and introduces the analytical solutions for solving the optimization problem of minimizing the market impact of HFT. Section 4 investigates the performance of our model by running simulations. In order to show the model implementation for practitioners, we provide numerical examples to show the performance of trading strategies based on the proposed model comparing with alternative non-HFT trading strategies and discuss the results. In Sect. 5, we then examine the performance of our model with real American market data to show the practical implementation. Section 6 concludes.

2 The model setup

High-frequency trading (HFT) utilizes computers to transact within a finer time interval at incomprehensible speed. High frequency (HF) trading firms represent approximately 2 % of the nearly 20,000 trading firms operating in the U.S. markets, but since 2009 they have accounted for over 70 % of the volume in U.S. equity markets and are fast approaching 50 % of the volume in futures markets.[3] How does this sweeping market change affect retail investors? There are two very different answers to that question. Supporters claim that HFT is a net-positive market force because it provides liquidity and tightens bid-ask spreads. On the other side, detractors claim that HFTs is a net-negative force on the market and should be reined in because HF traders regularly manipulate unaware investors and otherwise destabilize markets. The answer surely lies somewhere in between. But which is closer to the truth? To find out, we build a model to investigate. In order to make our model remains neutral (i.e., towards neither supporter nor detractor's viewpoint to build the model), we have several assumptions that take into account both supporter and detractor's consideration for HFT.

Suppose there is a representative HF trader (for example, a trader from proprietary trading desks) who can observe the whole order flow. He should provide liquidity to accumulate (or liquidate) a large position of securities with size X_0 ($X_0 > 0$) during a given time interval $[0, T]$ ($T > 0$). T refers to any trading period, for example, it could be one second or millisecond under HFT. The major assumptions are (i) the representative HF trader has direct market access (DMA), i.e., electronic trading facilities that give HF trader a low-latency way to interact with the order book of an exchange, (ii) HF trader is able to transact the incoming orders before other non-HF traders do, and (iii) HFT leads to three possible consequences, i.e., liquidity increase, decrease, and unaffected, therefore, HF trader can decide to do or not to do HFT based on the subsequent liquidity situation they estimated.

2.1 A trading scenario

We use following example to describe the trading scenario (e.g., DMA) investigated by our model. Assume that a seller wants to sell 100,000 shares of XXXX. The market price of an XXXX share is \$26.40, but the seller's limit price is \$26.10. In other words, the seller is willing to receive at least \$26.10 for each share of XXXX or \$0.30 less than its current price. With flash orders from the exchange (e.g., NASDAQ), HF trader gets a peek at these

[3] CFTC (2010). Proposed rules, *Federal Register 75* (112), 33198–33202.

orders for a finer time (e.g., 30 milliseconds) before they are disseminated to everyone else. Having detected an incoming order flow of bid for XXXX shares (e.g., limit bid at $26.30), the computers of HF trader start issuing small orders, for example, limit orders with prices higher than $26.10 but below $26.30 or immediate or cancel (IOC) orders at specific levels below the current price ($26.30) of XXXX shares. If the first buy order at $26.15 is accepted, another buy order at $26.20 is issued, and so on. This continues until a buy order at $26.09 is issued. Because the seller's limit price is $26.10, the buy order at $26.09 cannot be executed. Meantime, HF trader could transact the accumulated position to match the incoming bid order for the immediate profit. At this stage, HF trader floods the seller with buy orders at $26.29, causing most of the company's order of 100,000 XXXX shares to be filled at $0.11 below the market price. If HF trader observes that an incoming order flow of bid for XXXX shares limited at $26.00, the computers of HF trader start continuously issuing small orders at limit prices below $26.10 of XXXX shares and immediately cancel them to make the seller misinterpret and drop the limit price. Under normal circumstances, a seller would see the buy order at $26.30 and might subsequently raise the limit price on his/her order. However, HF trader's has such a direct access to see the whole order flow before the seller that unless the seller can observe the whole order flow, he/she would have no chance to do this.

In above example, HFT either involves front-running or manipulating, which are the major accusation of HFT. In order to investigate HFT under a neutral consideration, when we build our model with the above HFT framework (e.g., DMA), we impose following trading constrains (that have also been considered by regulators) to eliminate front-running and manipulating: (1) HF trader is obliged to submit market order (i.e., firm order requirement), and (2) HF trader cannot simultaneously buy and sell (i.e., trade only on one side).

In this paper, we also consider other two trading strategies. One is the naive trading strategy which splits the large order equally to small pieces and the other is the non-HF trading strategy proposed by Obizhaeva and Wang (2013) (O-W). Naive trading focuses on the minimal order size of each submission and non-HF trading (O-W) focuses on the optimal choice following the price changes.

2.2 Market order

Market order will be executed at the best price currently available in the market and it is used to realize immediacy of a trade. In our model, as we are going to investigate HFT neutrally without considering its suspicion of manipulating, we allow HFT trade only with market orders. The advantage of this setting for our model is twofold. First, it can reduce the difficulty of the algorithm run by HF trader to estimate the execution uncertainty of limit orders, particularly when the market bid-ask spread is relatively large. Second, it can limit HF trader from flickering quotes and make them bring liquidity to the market. Easley et al. (2012) pointed out that order flow is toxic when it adversely selects market makers who may be unaware they are providing liquidity on a loss. It is necessary to consider optimal trading when there exists flow toxicity in the market. Following the literature, we assume that liquidity does not replenish immediately after it is taken but only gradually over time. The HF trader is trying to passively accumulate or liquidate his large position, that is, break up the order (see Keim and Madhavan 1995 and Chan and Lakonishok 1995). In our paper, we will focus only on accumulating a position (i.e., buying), since our model can be easily adopted for liquidation a position (i.e., selling).

In our model, HF trader is only allowed to submit market orders at discrete time points which are equidistantly distributed, this means HF trader trades at time point $N+1$ during the whole trading period. $N \in \{1, 2, \dots\}$ stands for the trading frequency. t_i $(i \in 0, \dots, N)$

are time points starting at $t_0 = 0$ and ending in $t_N = T$. We then write $t_i = i\tau$, where $\tau = T/N$ is the duration between two successive time points we are able to trade. We define x_{t_n} as the size of the order HF trader has submitted at time point t_n, then $X_0 = \sum_{n=0}^{N} x_{t_n}$. We define $X_{t_n} = X_0 - \sum_{i=0}^{n-1} x_{t_i}$ as the position HF trader still needs to accumulate before t_n. We also assume $x_{t_n} \geq 0$. The space of feasible strategies is then defined as follows:

$$\Phi = \left\{ \{x_{t_0}, \dots, x_{t_N}\} : x_{t_n} \geq 0 \; \forall n \in \{0, \dots, N\}; \; \sum_{n=0}^{N} x_{t_n} = X_0 \right\}.$$

2.3 Price impact

We assume the underlying transaction price follows a geometric Brownian motion F_t with drift μ and volatility σ and an initial value $F_0 = V_0$. In addition, part of the exogenous price movement (i.e., without trading influence) is the size of the spread $s > 0$ which is assumed to be constant. We then call the exogenous price movement the unaffected best ask price A_t, where $A_t = F_t + s/2$. The limit-order book is modeled by using a constant depth q, which means when executing a buy order of the size q the price will be increased by 1 unit. In general, this translates into the price impact of an order x_{t_n} is x_{t_n}/q. The average price impact of the whole order is then $x_{t_n}/2q$. The price jump is due to the influence of order submitted to the market and consists of the permanent and temporary price impact.

We decompose the price impact of an order x_{t_n} into two parts $x_{t_n}/q = \lambda x_{t_n} + \kappa x_{t_n}$, where $0 \leq \lambda \leq 1/q$ is the percentage of the permanent price impact and $\kappa = 1/q - \lambda$ is the percentage of the temporary price impact contributed respectively to the total price impact. We call the term λx_{t_n} with $0 \leq \lambda \leq 1/q$, the permanent price impact of the trade x_{t_n} and κx_{t_n} with $\kappa = 1/q - \lambda$, the temporary price impact. Permanent price impact is the change in price caused by HF trader's order that leads the market to believe that future prices will be different than originally expected or there is a change in the asset's intrinsic value. Temporary price impact occurs whenever an order is released to the market but does not provide fundamental news or information that changes the market current valuation or long-run outlook of underlying asset. Trades cause temporary increases in price for buy orders and temporary decrease for sell orders subsequently followed by a price reversion back to the initial price trajectory. In order to model the way the temporary price impact of an order vanishes along with time we use a resilience factor ρ following Obizhaeva and Wang (2013). The part of the temporary price impact of the order x_{t_n} that remains until $t > t_n$ is $\kappa x_{t_n} e^{-\rho(t-t_n)}$, where the resilience factor $\rho > 0$. The temporary price impact at time point t_n before we make a trade is defined as $D_{t_n} = \sum_{i=0}^{n-1} x_{t_i} \kappa e^{-\rho(t_n - t_i)}$. In fact, the temporary price impact D_{t_n} at t_n before we submit the order, satisfies the recursive equation $D_{t_n} = (D_{t_{n-1}} + \kappa x_{t_{n-1}}) e^{-\rho\tau}$ with the initial condition $D_0 = 0$.

2.4 Order flow

We assume that HF trader has the direct access to the order flow and can observe the incoming order before other traders at one of the equidistant distributed time points during the trading period. In this model, we focus on trading one security and assume HF trader can observe the incoming order (or order flow) of that security during the trading period. We define $t_m \in \{t_1, \dots, t_{N-1}\}$ as equidistant distributed time points. HF trader cannot immediately identify how other traders react to the incoming order flow since there might exist simultaneous orders. We assume that the market only has r different possible reactions to the event. To each of the possible market reaction we assign a set of variables $\omega_i = (q_i, \rho_i, \kappa_i, \lambda_i, \mu_i, \sigma_i)$ for $i \in [1, r]$, which characterize the influence of news once it is disseminated to the market and observed by other traders. By doing so we are able to model the changes of the depth in the limit order book, the temporary and permanent price impact, and the resilience speed. We also assign a probability $0 \le p_i \le 1$ for $i \in [1, r]$ to each possible market reaction. We impose the condition that $\sum_{i=1}^{r} p_i = 1$. We assume that these probabilities are independent from the price development before the news is publicly available and the probability can be estimated or predefined based on the order flow we can see under HFT. $\omega_0 = (q_0, \rho_0, \kappa_0, \lambda_0, \mu_0, \sigma_0)$ are the variables describe the market before the news is publicly available.

2.5 Execution cost and the objective function

Then the objective function is to find the strategy that minimizes the execution cost of accumulating the whole position, that is,

$$\min_{x_0,\dots,x_T} \left(E \sum_{n=0}^{m-1} x_{t_n} \left(F_{t_n} + \frac{s}{2} + \lambda_0 (X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_0} \right) + \sum_{n=m}^{N} x_{t_n} \left(F_{t_n} + \frac{s}{2} + \lambda_j (X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_j} \right) \right).$$

3 Analytical solutions

The "incoming order" we considered in this paper will significantly influence the market by stimulating different order flows. Several empirical studies have tried to determine the effect of the news, see, for example, Patell and Wolfson (1984), Lee et al. (1993), and Sun et al. (2011). These studies show that the consequences introduced by the news are absorbed by the market mostly within the first 10–15 minutes after it is disseminated. The control variable in our model is size of the market order submitted by HFT algorithm.

Proposition 1 *Given the model setting described in Sect. 2, with* $\omega_0 = (q_0, \rho_0, \kappa_0, \lambda_0, \mu_0, \sigma_0)$ *for* $[0, t_{m-1}]$, *and* $\omega_i = (q_i, \rho_i, \kappa_i, \lambda_i, \mu_i, \sigma_i)$ *for* $i \in [1, r]$ *from* $[t_m, t_N]$ *with probability* p_i *and the underlying asset price following a geometric Brownian motion, the strategy* $x_{t_N} = X_{t_N}$, *and*

$$x_{t_n} = \begin{cases} -\frac{1}{2}\delta^i_{n+1}((1 + c^i_{n+1} 2\kappa_i e^{-2\rho_i \tau} - g^i_{n+1} e^{-\rho_i \tau}) D_{t_n} + (-\lambda_i - 2b^i_{n+1} + g^i_{n+1} e^{-\rho_i \tau} \kappa_i) X_{t_n} \\ \quad + (1 - a_i^{N-n} - h^i_{n+1} a_i + l^i_{n+1} \kappa_i e^{-\rho_i \tau} a_i) F_{t_n}), \\ \quad \textit{for } t_n \in \{t_{m+1}, \dots, t_{N-1}\}, \\ -\frac{1}{2}\delta^0_{m+1}((1 + 2\sum_{i=1}^r p_i \times c^i_{n+1} e^{-2\rho_i \tau} \kappa_i - \sum_{i=1}^r p_i \times g^i_{n+1} e^{-\rho_i \tau}) D_{t_m} \\ \quad + (-\sum_{i=1}^r p_i \times \lambda_i - 2\sum_{i=1}^r p_i \times b^i_{n+1} + \sum_{i=1}^r p_i \times g^i_{n+1} e^{-\rho_i \tau} \kappa_i) X_{t_m} \\ \quad + (1 - \sum_{i=1}^r p_i \times (a_i^{N+1-(m+1)}) - \sum_{i=1}^r p_i \times h^i_{n+1} a_i \\ \quad + \sum_{i=1}^r p_i \times l^i_{n+1} a_i e^{-\rho_i \tau} \kappa_i) F_{t_m}), \\ \quad \textit{for } t_n = t_m, \\ -\frac{1}{2}\delta^0_{n+1}((1 + c^0_{n+1} 2\kappa_0 e^{-2\rho_0 \tau} - g^0_{n+1} e^{-\rho_0 \tau}) D_{t_n} + (-\lambda_0 - 2b^0_{n+1} + g^0_{n+1} e^{-\rho_0 \tau} \kappa_0) X_{t_n} \\ \quad + (1 - a_0^{m-n} \times \overline{a} - h^0_{n+1} a_0 + l^0_{n+1} \kappa_0 e^{-\rho_0 \tau} a_0) F_{t_n}), \\ \quad \textit{for } t_n \in \{t_0, \dots, t_{m-1}\}, \end{cases} \tag{1}$$

is optimal, if $(x_{t_0}, \dots, x_{t_N}) \in \Phi$.

The optimal value function has following form

$$J_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \begin{cases} J^i_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) & \text{for } t_n \in \{t_{m+1}, \ldots, t_N\}, \\ J^0_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) & \text{for } t_n \in \{t_0, \ldots, t_m\}, \end{cases}$$

where

$$J^i_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \left(a_i^{N-n} F_n + \frac{s}{2}\right) X_{t_n} + \lambda_i \left(\frac{\lambda_0}{\lambda_i} X_0 - \frac{\lambda_0}{\lambda_i} X_{t_m} + X_{t_m}\right) X_{t_n} + b^i_n X^2_{t_n} \\ + c^i_n D^2_{t_n} + d^i_n F^2_{t_n} + g^i_n X_{t_n} D_{t_n} + h^i_n X_{t_n} F_{t_n} + l^i_n D_{t_n} F_{t_n}, \quad (2)$$

and

$$J^0_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \left(a_0^{m-n} \times \overline{a} F_n + \frac{s}{2}\right) X_{t_n} + \lambda_0 X_0 X_{t_n} + b^0_n X^2_{t_n} + c^0_n D^2_{t_n} + d^0_n F^2_{t_n} \\ + g^0_n X_{t_n} D_{t_n} + h^0_n X_{t_n} F_{t_n} + l^0_n D_{t_n} F_{t_n}, \quad (3)$$

with $a_i = e^{\mu_i \tau}$, $\upsilon_i = e^{(2\mu_i + \sigma_i^2)\times\tau}$, $\overline{a} = \sum_{i=1}^r p_i \times (a_i^{N+1-(m+1)})$, *and the coefficients given as follows*:

$$b_n = \begin{cases} b^i_N = \frac{1}{2q_i} - \lambda_i, \\ \quad \text{for } n = N, \\ b^i_n = b^i_{n+1} - \frac{1}{4}\delta^i_{n+1}(-\lambda_i - 2b^i_{n+1} + g^i_{n+1} e^{-\rho_i \tau}\kappa_i)^2, \\ \quad \text{for } n \in \{m+1, \ldots, N-1\}, \\ b^0_n = \sum_{i=1}^r p_i \times \lambda_i - \lambda_0 + \sum_{i=1}^r p_i \times b^i_{m+1} \\ \qquad - \frac{1}{4}\delta^0_{m+1}(-\sum_{i=1}^r p_i \times \lambda_i - 2\sum_{i=1}^r p_i \times b^i_{m+1} + \sum_{i=1}^r p_i \times g^i_{m+1} e^{-\rho_i \tau}\kappa_i)^2, \\ \quad \text{for } n = m, \\ b^0_n = b^0_{n+1} - \frac{1}{4}\delta^0_{n+1}(-\lambda_0 - 2b^0_{n+1} + g^0_{n+1} e^{-\rho_0 \tau}\kappa_0)^2, \\ \quad \text{for } n \in \{0, \ldots, m-1\}; \end{cases} \quad (4)$$

$$c_n = \begin{cases} c^i_N = 0, \\ \quad \text{for } n = N, \\ c^i_n = c^i_{n+1} e^{-2\rho_i \tau} - \frac{1}{4}\delta^i_{n+1}(1 + c^i_{n+1} 2\kappa_i e^{-2\rho_i \tau} - g^i_{n+1} e^{-\rho_i \tau})^2, \\ \quad \text{for } n \in \{m+1, \ldots, N-1\}, \\ c^0_n = \sum_{i=1}^r p_i \times c^i_{m+1} e^{-2\rho_i \tau} - \frac{1}{4}\delta^0_{m+1}(1 + 2\sum_{i=1}^r p_i \times c^i_{m+1} e^{-2\rho_i \tau}\kappa_i \\ \qquad - \sum_{i=1}^r p_i \times g^i_{m+1} e^{-\rho_i \tau})^2, \\ \quad \text{for } n = m, \\ c^0_n = c^0_{n+1} e^{-2\rho_0 \tau} - \frac{1}{4}\delta^0_{n+1}(1 + c^0_{n+1} 2\kappa_0 e^{-2\rho_0 \tau} - g^0_{n+1} e^{-\rho_0 \tau})^2, \\ \quad \text{for } n \in \{0, \ldots, m-1\}; \end{cases} \quad (5)$$

$$d_n = \begin{cases} d_N^i = 0, \\ \quad \mathit{for}\ n = N, \\ d_n^i = d_{n+1}^i v_i - \frac{1}{4}\delta_{n+1}^i (1 - a_i^{N-n} - h_{n+1}^i a_i + l_{n+1}^i \kappa_i e^{-\rho_i \tau} a_i)^2, \\ \quad \mathit{for}\ n \in \{m+1, \dots, N-1\}, \\ d_n^0 = \sum_{i=1}^r p_i \times d_{n+1}^i v_i \\ \qquad - \frac{1}{4}\delta_{m+1}^0 (1 - \sum_{i=1}^r p_i \times (a_i^{N+1-(m+1)}) - \sum_{i=1}^r p_i \times h_{m+1}^i a_i \\ \qquad + \sum_{i=1}^r p_i \times l_{m+1}^i a_i e^{-\rho_i \tau} \kappa_i)^2, \\ \quad \mathit{for}\ n = m, \\ d_n^0 = d_{n+1}^0 v_0 - \frac{1}{4}\delta_{n+1}^0 (1 - a_0^{m-n} \times \overline{a} - h_{n+1}^0 a_0 + l_{n+1}^0 \kappa_0 e^{-\rho_0 \tau} a_0)^2, \\ \quad \mathit{for}\ n \in \{0, \dots, m-1\}; \end{cases} \tag{6}$$

$$g_n = \begin{cases} g_N^i = 1, \\ \quad \mathit{for}\ n = N, \\ g_n^i = g_{n+1}^i e^{-\rho_i \tau} - \frac{1}{2}\delta_{n+1}^i (1 + c_{n+1}^i 2\kappa_i e^{-2\rho_i \tau} - g_{n+1}^i e^{-\rho_i \tau}) \\ \qquad \times (-\lambda_i - 2b_{n+1}^i + g_{n+1}^i e^{-\rho_i \tau} \kappa_i), \\ \quad \mathit{for}\ n \in \{m+1, \dots, N-1\}, \\ g_n^0 = \sum_{i=1}^r p_i \times g_{n+1}^i e^{-\rho_i \tau} \\ \qquad - \frac{1}{2}\delta_{m+1}^0 (1 + 2\sum_{i=1}^r p_i \times c_{m+1}^i e^{-2\rho_i \tau} \kappa_i - \sum_{i=1}^r p_i \times g_{m+1}^i e^{-\rho_i \tau}) \\ \qquad \times (-\sum_{i=1}^r p_i \times \lambda_i - 2\sum_{i=1}^r p_i \times b_{m+1}^i + \sum_{i=1}^r p_i \times g_{m+1}^i e^{-\rho_i \tau} \kappa_i), \\ \quad \mathit{for}\ n = m, \\ g_n^0 = g_{n+1}^0 e^{-\rho_0 \tau} - \frac{1}{2}\delta_{n+1}^0 (1 + c_{n+1}^0 2\kappa_0 e^{-2\rho_0 \tau} - g_{n+1}^0 e^{-\rho_0 \tau}) \\ \qquad \times (-\lambda_0 - 2b_{n+1}^0 + g_{n+1}^0 e^{-\rho_0 \tau} \kappa_0), \\ \quad \mathit{for}\ n \in \{0, \dots, m-1\}; \end{cases} \tag{7}$$

$$h_n = \begin{cases} h_N^i = 0, \\ \quad \mathit{for}\ n = N, \\ h_n^i = h_{n+1}^i a_i - \frac{1}{2}\delta_{n+1}^i (-\lambda_i - 2b_{n+1}^i + g_{n+1}^i e^{-\rho_i \tau} \kappa_i) \\ \qquad \times (1 - a_i^{N-n} - h_{n+1}^i a_i + l_{n+1}^i \kappa_i e^{-\rho_i \tau} a_i), \\ \quad \mathit{for}\ n \in \{m+1, \dots, N-1\}, \\ h_n^0 = \sum_{i=1}^r p_i \times h_{m+1}^i a_i \\ \qquad - \frac{1}{2}\delta_{m+1}^0 (-\sum_{i=1}^r p_i \times \lambda_i - 2\sum_{i=1}^r p_i \times b_{m+1}^i \\ \qquad + \sum_{i=1}^r p_i \times g_{m+1}^i e^{-\rho_i \tau} \kappa_i)(1 - \sum_{i=1}^r p_i \times (a_i^{N+1-(m+1)}) \\ \qquad - \sum_{i=1}^r p_i \times h_{m+1}^i a_i + \sum_{i=1}^r p_i \times l_{m+1}^i a_i e^{-\rho_i \tau} \kappa_i), \\ \quad \mathit{for}\ n = m, \\ h_n^0 = h_{n+1}^0 a_0 - \frac{1}{2}\delta_{n+1}^0 (-\lambda_0 - 2b_{n+1}^0 + g_{n+1}^0 e^{-\rho_0 \tau} \kappa_0) \\ \qquad \times (1 - a_0^{m-n} \times \overline{a} - h_{n+1}^0 a_0 + l_{n+1}^0 \kappa_0 e^{-\rho_0 \tau} a_0), \\ \quad \mathit{for}\ n \in \{0, \dots, m-1\}; \end{cases} \tag{8}$$

$$
l_n = \begin{cases}
l_N^i = 0, \\
\quad \mathit{for}\ n = N, \\
l_n^i = l_{n+1}^i e^{-\rho_i \tau} a_i - \frac{1}{2}\delta_{n+1}^i (1 - a_i^{N-n} - h_{n+1}^i a_i + l_{n+1}^i \kappa_i e^{-\rho_i \tau} a_i) \\
\qquad \times (1 + c_{n+1}^i 2\kappa_i e^{-2\rho_i \tau} - g_{n+1}^i e^{-\rho_i \tau}), \\
\quad \mathit{for}\ n \in \{m+1, \dots, N-1\}, \\
l_n^0 = \sum_{i=1}^r p_i \times l_{m+1}^i a_i e^{-\rho_i \tau} \\
\qquad - \frac{1}{2}\delta_{m+1}^0 (1 + 2\sum_{i=1}^r p_i \times c_{m+1}^i e^{-2\rho_i \tau}\kappa_i - \sum_{i=1}^r p_i \times g_{m+1}^i e^{-\rho_i \tau}) \\
\qquad \times (1 - \sum_{i=1}^r p_i \times (a_i^{N+1-(m+1)}) - \sum_{i=1}^r p_i \times h_{m+1}^i a_i \\
\qquad + \sum_{i=1}^r p_i \times l_{m+1}^i a_i e^{-\rho_i \tau}\kappa_i), \\
\quad \mathit{for}\ n = m, \\
l_n^0 = l_{n+1}^0 e^{-\rho_0 \tau} a_0 - \frac{1}{2}\delta_{n+1}^0 (1 - a_0^{m-n} \times \overline{a} - h_{n+1}^0 a_0 + l_{n+1}^0 \kappa_0 e^{-\rho_0 \tau} a_0) \\
\qquad \times (1 + c_{n+1}^0 2\kappa_0 e^{-2\rho_0 \tau} - g_{n+1}^0 e^{-\rho_0 \tau}), \\
\quad \mathit{for}\ n \in \{0, \dots, m-1\};
\end{cases} \tag{9}
$$

and

$$
\delta_{n+1} = \begin{cases}
\delta_{n+1}^i = (\frac{1}{2q_i} + b_{n+1}^i - g_{n+1}^i \kappa_i e^{-\rho_i \tau} + c_{n+1}^i \kappa_i^2 e^{-2\rho_i \tau})^{-1}, \\
\quad \mathit{for}\ n \in \{m+1, \dots, N-1\}, \\
\delta_{n+1}^0 = (\sum_{i=1}^r p_i \times \frac{1}{2q_i} + \sum_{i=1}^r p_i \times b_{n+1}^i - \sum_{i=1}^r p_i \times g_{n+1}^i e^{-\rho_i \tau}\kappa_i \\
\qquad + \sum_{i=1}^r p_i \times c_{n+1}^i e^{-2\rho_i \tau}\kappa_i^2)^{-1}, \\
\quad \mathit{for}\ n = m, \\
\delta_{n+1}^0 = (\frac{1}{2q_0} + b_{n+1}^0 - g_{n+1}^0 \kappa_0 e^{-\rho_0 \tau} + c_{n+1}^0 \kappa_0^2 e^{-2\rho_0 \tau})^{-1}, \\
\quad \mathit{for}\ n \in \{0, \dots, m-1\}.
\end{cases} \tag{10}
$$

Proof See Appendix. □

3.1 Case for "one-side trading" constraint

"One-side trading" refers to the situation that the HF trader is not allowed to make an opposite trade, i.e., cannot simultaneously buy and sell. In our model, the optimal choice of HFT may include orders for opposite trade, i.e., $x_{t_n} < 0$ given different combinations of parameters which means the strategy is not in Φ, then the solution is not optimal. HF trader then faces three choices: (1) to follow what the optimal solution suggests, (2) not to submit order, or (3) to submit an order that completes the whole trade.

Not to submit order, that is, $x_{t_n} = 0$, which is the next best solution if the optimal solution suggests an order with $x_{t_n} < 0$, because of the parabolic structure of the optimal value function. Submitting an order which completes the whole trade, that is, $x_{t_n} = X_{t_n}$, which is the next best solution if the optimal solution suggests an order with $x_{t_n} > X_{t_n}$, because of the parabolic structure of the optimal value function. An order with $x_{t_n} > X_{t_n}$ is not allowed, because this forces an order with $x_{t_n} < 0$ later on. If we choose beforehand which order we will be using, the coefficients for this time point will follow different equations. We can do this for every time point, except for the last one when we are forced to submit the order $x_{t_N} = X_{t_N}$ to complete the whole trade. We then end up with 3^N different sets of recursive equations for the parameters. The only way we can be sure to find the best feasible strategy, is to calculate all possible strategies then select the best strategy, which is in Φ.

When the underlying asset price follows a geometric Brownian motion, we can derive the following propositions:

Proposition 2 *For the case $x_{t_n} = X_{t_n}$, the coefficients follow the equations: when $n \in \{m+1, \dots, N-1\}$:*

$$b_n^i = \frac{1}{2q_i} - \lambda_i, \qquad c_n^i = 0, \qquad d_n^i = 0, \qquad g_n^i = 1, \qquad h_n^i = 0, \qquad l_n^i = 0;$$

when $n = m$:

$$b_n^0 = \sum_{i=1}^{r} p_i \times \frac{1}{2q_i} - \lambda_0, \qquad c_n^0 = 0, \qquad d_n^0 = 0, \qquad g_n^0 = 1, \qquad h_n^0 = 0, \qquad l_n^0 = 0;$$

and when $n \in \{0, \dots, m-1\}$:

$$b_n^0 = \frac{1}{2q_0} - \lambda_0, \qquad c_n^0 = 0, \qquad d_n^0 = 0, \qquad g_n^0 = 1, \qquad h_n^0 = 0, \qquad l_n^0 = 0.$$

Proof When $n \in \{m+1, \dots, N-1\}$, we have the optimal value function

$$J_{t_n}^i(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_m}) + \lambda_i(X_{t_m} - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_i}\right]x_{t_n},$$

where we find the coefficients. When $n = m$, the optimal value function has the form

$$J_{t_n}^0(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_u}\right]x_{t_n},$$

from which we find the coefficients. When $n \in \{0, \dots, m-1\}$ we have the optimal value function

$$J_{t_n}^0(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_0}\right]x_{t_n},$$

that gives the coefficients. □

Proposition 3 *For the case $x_{t_n} = 0$, the coefficients follow the equations: when $n \in \{m+1, \dots, N-1\}$:*

$$b_n^i = b_{n+1}^i, \qquad c_n^i = c_{n+1}^i e^{-2\rho_i \tau}, \qquad d_n^i = d_{n+1}^i v_i,$$
$$g_n^i = g_{n+1}^i e^{-\rho_i \tau}, \qquad h_n^i = h_{n+1}^i a_i, \qquad l_n^i = l_{n+1}^i e^{-\rho_i \tau} a_i;$$

when $n = m$:

$$b_n^0 = \sum_{i=1}^{r} p_i \times b_{n+1}^i, \qquad c_n^0 = \sum_{i=1}^{r} p_i \times c_{n+1}^i e^{-2\rho_i \tau},$$
$$d_n^0 = \sum_{i=1}^{r} p_i \times d_{n+1}^i v_i, \qquad g_n^0 = \sum_{i=1}^{r} p_i \times g_{n+1}^i e^{-\rho_i \tau},$$
$$h_n^0 = \sum_{i=1}^{r} p_i \times h_{n+1}^i a_i, \qquad l_n^0 = \sum_{i=1}^{r} p_i \times l_{n+1}^i a_i e^{-\rho_i \tau};$$

and when $n \in \{0, \dots, m-1\}$:

$$b_n^0 = b_{n+1}^0, \qquad c_n^0 = c_{n+1}^0 e^{-2\rho_0 \tau}, \qquad d_n^0 = d_{n+1}^0 v_0,$$
$$g_n^0 = g_{n+1}^0 e^{-\rho_0 \tau}, \qquad h_n^0 = h_{n+1}^0 a_0, \qquad l_n^0 = l_{n+1}^0 e^{-\rho_0 \tau} a_0.$$

Proof When $n \in \{m+1, \dots, N-1\}$, we have the optimal value function

$$\begin{aligned} J^i_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = &\left(a_i \times a_i^{N-(n+1)} F_n + \frac{s}{2}\right) X_{t_n} + \lambda_i \left(\frac{\lambda_0}{\lambda_i} X_0 - \frac{\lambda_0}{\lambda_i} X_{t_m} + X_{t_m}\right) X_{t_n} \\ &+ b^i_{n+1} X^2_{t_n} + c^i_{n+1} D^2_{t_n} e^{-2\rho_i \tau} + d^i_{n+1} \upsilon_i F^2_{t_n} + g^i_{n+1} X_{t_n} D_{t_n} e^{-\rho_i \tau} \\ &+ h^i_{n+1} X_{t_n} a_i F_{t_n} + l^i_{n+1} D_{t_n} e^{-\rho_i \tau} a_i F_{t_n}, \end{aligned}$$

and find the coefficients from it. When $n = m$, the optimal value function is

$$\begin{aligned} J^0_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = &\left(F_n \overline{a} + \frac{s}{2}\right) X_{t_n} + \overline{\lambda}\left(\frac{\lambda_0}{\overline{\lambda}} X_0 - \frac{\lambda_0}{\overline{\lambda}} X_{t_n} + X_{t_n}\right) X_{t_n} \\ &+ \left(\sum_{i=1}^r p_i \times b^i_{n+1}\right) X^2_{t_n} + D^2_{t_n}\left(\sum_{i=1}^r p_i \times c^i_{n+1} e^{-2\rho_i \tau}\right) \\ &+ \left(\sum_{i=1}^r p_i \times d^i_{n+1} \upsilon_i\right) F^2_{t_n} \\ &+ X_{t_n} D_{t_n}\left(\sum_{i=1}^r p_i \times g^i_{n+1} e^{-\rho_i \tau}\right) + \left(\sum_{i=1}^r p_i \times h^i_{n+1} a_i\right) X_{t_n} F_{t_n} \\ &+ D_{t_n} F_{t_n}\left(\sum_{i=1}^r p_i \times l^i_{n+1} a_i e^{-\rho_i \tau}\right), \end{aligned}$$

which gives the coefficients. When $n \in \{0, \dots, m-1\}$, we can obtain the coefficients from the optimal value function

$$\begin{aligned} J^0_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = &\left(a_0 \times a_0^{m-(n+1)} \times \sum_{i=1}^r p_i \times \left(a_i^{N+1-(m+1)}\right) F_n + \frac{s}{2}\right) X_{t_n} + \lambda_0 X_0 X_{t_n} \\ &+ b^0_{n+1} X^2_{t_n} + c^0_{n+1} D^2_{t_n} e^{-2\rho_0 \tau} + d^0_{n+1} \upsilon_0 F^2_{t_n} + g^0_{n+1} X_{t_n} D_{t_n} e^{-\rho_0 \tau} \\ &+ h^0_{n+1} X_{t_n} a_0 F_{t_n} + l^0_{n+1} D_{t_n} e^{-\rho_0 \tau} a_0 F_{t_n}. \end{aligned}$$

□

4 Simulation study

In this section we conduct a simulation study to compare the performance (i.e., execution cost reduction) of HFT, non-HF optimal trading proposed by Obizhaeva and Wang (2013) (O-W in short), and the naive trading (that equally splits the large order into small pieces), which is popular for practitioners. In order to calibrate the magnitude of the execution cost reduction provided by these candidate strategies, we run the simulation for 100,000 times to ensure the statistical significance. We set the initial value of underlying asset price at 100. In our simulation study, the lower the execution cost, the better the performance of the trading strategy we investigated.

4.1 Simulation results

In this simulation, we examine performance of HFT based on the liquidity change measured by volume arrival. In our HFT model, we describe the volume arrival with q, ρ, and p.

We consider three different market reactions inferred from order flow for liquidity changes under HFT, that are, liquidity remains same, increasing liquidity, and decreasing liquidity. These three situations are uniformly distributed with probability of 1/3. For practitioners, they can modify the probability based on their own estimation. Following Obizhaeva and Wang (2013), we assign $\omega_0 = (5000, 2.2, 1/10000, 1/10000)$, $\omega_1 = (4000, 2.0, 1/8000, 1/8000)$, $\omega_2 = (5000, 2.2, 1/10000, 1/10000)$, and $\omega_3 = (6000, 2.4, 1/12000, 1/12000)$. We can consider ω_1 as a situation when the liquidity decreases in the market. In consequence, the depth of the limit-order book turns out to be smaller, which leads to a higher price impact for the orders we submitted to the market. The resilience speed ρ is smaller according to the liquidity decreasing. For ω_3, it describes an opposite situation as ω_1 does, which is the situation when the liquidity increases. For ω_2, the HFT order has no impact to the market. For the case the underlying value follows the geometric Brownian motion, we use the same setting as described above and keep the parameters of the geometric Brownian motion unchanged, regardless of the order flow.

We consider different scenarios to describe the change of liquidity after we originate HFT. Table 1 shows the result of comparing the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading when ρ and q move symmetrically. Table 2 shows the result of comparing the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading when ρ and q move asymmetrically. Table 3 shows the result of comparing the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading when p changes. We can conclude that the expected execution cost of HFT is lower than non-HF optimal trading (O-W) and Naive trading. Figure 1 illustrates the expected execution cost of HFT is lower than non-HF trading and Naive trading.

In order to show the influence of trading frequency on the execution cost, we compute the analytical value of execution cost for different trading frequencies under a scenario given by our model. Table 4 shows the result of comparing the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading at different trading frequencies. We can see that when increase trading frequency, the execution cost of HFT is reduced comparing with that of O-W and Naive trading. From Table 4, we can find that when increasing trading frequency, the execution cost of Naive trading is accordingly increased while that of O-W does not change significantly. Figures 2 and 3 illustrate the pattern of execution cost reduction when we increase trading frequency.

5 Empirical study

In Sect. 4, we investigate the performance of our trading strategy based on the simulated prices which follow the geometric Brownian motion. In this section, we assess the performance of HFT with the real market data at different trading frequencies.

5.1 Data

We apply the HFT strategy derived from our theoretical model for a sample of US stocks from January 2008 to October 2010. We start with a sample of S&P 100, S&P MidCap 400, and S&P SmallCap 600. In our sample, we have 105 stocks. 89 stocks from S&P 100, 9 stocks from S&P MidCap 400, and 7 stocks from S&P SmallCap 600 component stocks at tick level from the Trades and Quotes (TAQ). We apply the trade classification rules suggested by Ellis et al. (2000), that is, trades executed at the ask quote are buys, executed at the bid quote are sells, and all other trades are categorized by the tick rule since the HFT strategy investigated can only submit market orders.

Table 1 Comparison of the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading (100,000 runs) for different combinations of the drift of the underlying value and the size of the liquidity change after the HFT order. A liquidity change of, for example, 65 % means that the parameters ω_1 (which model the situation after the HFT order) in case of liquidity decreasing, are obtained from ω_0 (which models the situation before the HFT order) by $q_1 = q_0 \cdot (1 - 0.65) = q_0 \cdot 0.35$, $\rho_1 = \rho_0 \cdot 0.35$. The remaining parameters that model the market after HFT order remain the same. The parameters ω_3, which model the situation after the HFT order in case of liquidity increasing can be obtained by $q_3 = q_0 \cdot (1 + 0.65) = q_0 \cdot 1.65$ and $\rho_3 = \rho_0 \cdot 1.65$

Drift	Liquidity change	25 %	35 %	45 %	55 %	65 %	75 %
$\mu = -10$ %	HFT	100.2240 (0.8659)	100.7193 (0.8499)	101.5055 (0.8470)	102.4649 (0.6284)	103.9148 (0.5533)	105.9156 (0.4382)
	O-W	102.6064 (0.2823)	102.8912 (0.2840)	103.3332 (0.2835)	104.0317 (0.2825)	105.1835 (0.2822)	107.3203 (0.2846)
	Naive	102.7731 (0.3331)	103.0809 (0.3338)	103.5629 (0.3375)	104.3141 (0.3337)	105.5649 (0.3378)	107.8815 (0.3374)
$\mu = -5$ %	HFT	104.5536 (0.5687)	104.9440 (0.5086)	105.4848 (0.4468)	106.1860 (0.3792)	107.0544 (0.3186)	107.9710 (0.1328)
	O-W	105.1406 (0.2779)	105.4267 (0.2774)	105.8670 (0.2756)	106.5638 (0.2780)	107.7141 (0.2788)	109.8538 (0.2785)
	Naive	105.3298 (0.3263)	105.6422 (0.3224)	106.1189 (0.3230)	106.8754 (0.3242)	108.1227 (0.3247)	110.4423 (0.3250)
$\mu = -1$ %	HFT	107.1420 (0.2937)	107.4038 (0.2630)	107.7597 (0.2307)	108.1901 (0.2016)	108.6643 (0.0830)	108.6637 (0.0828)
	O-W	107.1501 (0.2723)	107.4340 (0.2732)	107.8760 (0.2729)	108.5759 (0.2735)	109.7269 (0.2754)	111.8659 (0.2726)
	Naive	107.3499 (0.3135)	107.6577 (0.3146)	108.1380 (0.3152)	108.8906 (0.3150)	110.1403 (0.3146)	112.4572 (0.3147)
$\mu = 0$ %	HFT	107.6350 (0.2414)	107.8707 (0.2141)	108.1821 (0.1881)	108.5617 (0.1647)	108.8116 (0.0719)	108.8125 (0.0729)
	O-W	107.6518 (0.2734)	107.9370 (0.2732)	108.3803 (0.2740)	109.0783 (0.2745)	110.2304 (0.2741)	112.3658 (0.2723)
	Naive	107.8485 (0.3113)	108.1574 (0.3143)	108.6332 (0.3128)	109.3940 (0.3125)	110.6373 (0.3153)	112.9575 (0.3140)
$\mu = 1$ %	HFT	108.0858 (0.1922)	108.2869 (0.1727)	108.5586 (0.1507)	108.7989 (0.1332)	108.9494 (0.0629)	108.9497 (0.0631)
	O-W	108.1540 (0.2713)	108.4339 (0.2692)	108.8806 (0.2720)	109.5772 (0.2705)	110.7280 (0.2706)	112.8636 (0.2708)
	Naive	108.3486 (0.3094)	108.6529 (0.3103)	109.1329 (0.3069)	109.8881 (0.3077)	111.1357 (0.3090)	113.4570 (0.3091)

Table 1 (*Continued*)

Drift	Liquidity change	25 %	35 %	45 %	55 %	65 %	75 %
$\mu = 5$ %	HFT	108.6553 (0.0916)	109.3516 (0.0537)	109.4078 (0.0335)	109.4080 (0.0333)	109.4082 (0.0337)	109.4072 (0.0336)
	O-W	110.1425 (0.2654)	110.4274 (0.2666)	110.8688 (0.2676)	111.5641 (0.2659)	112.7127 (0.2671)	114.8533 (0.2670)
	Naive	110.3308 (0.3020)	110.6372 (0.3046)	111.1215 (0.3032)	111.8773 (0.3032)	113.1243 (0.3023)	115.4410 (0.3046)
$\mu = 10$ %	HFT	109.7827 (0.0112)	109.7831 (0.0112)	109.7827 (0.0113)	109.7826 (0.0111)	109.7827 (0.0112)	109.7829 (0.0112)
	O-W	112.6113 (0.2619)	112.8971 (0.2615)	113.3371 (0.2639)	114.0346 (0.2631)	115.1856 (0.2624)	117.3222 (0.2595)
	Naive	112.7813 (0.2921)	113.0876 (0.2924)	113.5712 (0.2937)	114.3231 (0.2921)	115.5729 (0.2930)	117.8856 (0.2925)

Table 2 Comparison of the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading (100,000 runs) for different values of liquidity change. A liquidity decrease, for example, 65 % means that the parameters ω_1 (which model the situation after the HFT order) in case of liquidity decreasing can be obtained from ω_0 (which models the situation before the HFT order) by $q_1 = q_0 \cdot (1 - 0.65) = q_0 \cdot 0.35$, $\rho_1 = \rho_0 \cdot 0.35$, and the remaining parameters that model the market after the HFT order remain the same as before the HFT order. A liquidity increase, for example, 20 % means that the parameters ω_3 (which model the situation after the HFT order) in case of liquidity increasing can be obtained from ω_0 (which models the situation before the HFT order) by $q_3 = q_0 \cdot (1 + 0.20) = q_0 \cdot 1.2$, $\rho_3 = \rho_0 \cdot 1.2$, and the remaining parameters that model the market after the HFT order remain the same as before the HFT order

ω_3		ω_1				
		20 %	35 %	50 %	65 %	80 %
20 %	HFT	108.7237 (0.1208)	108.9588 (0.1022)	109.0477 (0.0601)	109.0485 (0.0602)	109.0483 (0.0603)
	O-W	109.0368 (0.2699)	109.5611 (0.2705)	110.3933 (0.2709)	111.9375 (0.2708)	115.7646 (0.2676)
	Naive	109.2283 (0.3082)	109.7748 (0.3069)	110.6436 (0.3103)	112.2501 (0.3099)	116.2354 (0.3091)
35 %	HFT	108.6156 (0.1290)	108.8962 (0.1063)	109.0481 (0.0597)	109.0474 (0.0604)	109.0476 (0.0602)
	O-W	108.8665 (0.2700)	109.3900 (0.2696)	110.2245 (0.2689)	111.7646 (0.2682)	115.5934 (0.2684)
	Naive	109.0538 (0.3092)	109.5950 (0.3080)	110.4671 (0.3055)	112.0697 (0.3092)	116.0554 (0.3088)
50 %	HFT	108.5174 (0.1367)	108.8276 (0.1120)	109.0468 (0.0599)	109.0483 (0.0598)	109.0464 (0.0606)
	O-W	108.7304 (0.2687)	109.2494 (0.2700)	110.0856 (0.2699)	111.6286 (0.2678)	115.4534 (0.2698)
	Naive	108.9090 (0.3099)	109.4554 (0.3100)	110.3241 (0.3079)	111.9260 (0.3074)	115.9089 (0.3080)
65 %	HFT	108.4299 (0.1447)	108.7709 (0.1148)	109.0480 (0.0604)	109.0490 (0.0598)	109.0466 (0.0606)
	O-W	108.6137 (0.2683)	109.1369 (0.2683)	109.9749 (0.2686)	111.5168 (0.2688)	115.3412 (0.2679)
	Naive	108.7930 (0.3097)	109.3370 (0.3082)	110.2080 (0.3091)	111.8079 (0.3072)	115.7956 (0.3092)
80 %	HFT	108.3508 (0.1528)	108.7191 (0.1195)	109.0481 (0.0600)	109.0477 (0.0601)	109.0485 (0.0602)
	O-W	108.5223 (0.2686)	109.0459 (0.2677)	109.8872 (0.2698)	111.4230 (0.2681)	115.2491 (0.2698)
	Naive	108.6932 (0.3094)	109.2352 (0.3089)	110.1086 (0.3062)	111.7137 (0.3102)	115.6979 (0.3071)

Table 3 Comparison of the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading (100,000 runs) for different values of probabilities of liquidity change. p_1 is the probability that the outcome of the HFT order leads to a decrease in liquidity, and p_3 is the probability that to a increase in liquidity. The probability, that the liquidity remains unchanged after the HFT order is calculated by $p_2 = 1 - p_1 - p_3$

p_3		p_1				
		1/2	1/3	1/4	1/5	1/6
1/2	HFT	108.8287 (0.1677)	108.6646 (0.1459)	108.5634 (0.1490)	108.4971 (0.1585)	108.4469 (0.1669)
	O-W	109.2501 (0.4026)	108.9426 (0.3128)	108.7850 (0.3028)	108.6983 (0.3070)	108.6376 (0.3138)
	Naive	109.4524 (0.4632)	109.1288 (0.3585)	108.9724 (0.3460)	108.8761 (0.3497)	108.8145 (0.3592)
1/3	HFT	108.9151 (0.1243)	108.7702 (0.1167)	108.6867 (0.1284)	108.6269 (0.1401)	108.5869 (0.1517)
	O-W	109.4318 (0.3167)	109.1269 (0.2715)	108.9750 (0.2802)	108.8791 (0.2982)	108.8264 (0.3144)
	Naive	109.6416 (0.3615)	109.3230 (0.3094)	109.1667 (0.3192)	109.0718 (0.3416)	109.0083 (0.3600)
1/4	HFT	108.9502 (0.1188)	108.8173 (0.1165)	108.7378 (0.1352)	108.6885 (0.1489)	108.6491 (0.1620)
	O-W	109.5265 (0.3021)	109.2172 (0.2807)	109.0668 (0.3033)	108.9711 (0.3294)	108.9156 (0.3498)
	Naive	109.7355 (0.3471)	109.4218 (0.3223)	109.2618 (0.3477)	109.1649 (0.3731)	109.0992 (0.3965)
1/5	HFT	108.9675 (0.1158)	108.8464 (0.1223)	108.7712 (0.1412)	108.7177 (0.1584)	108.6820 (0.1732)
	O-W	109.5788 (0.3054)	109.2743 (0.2987)	109.1208 (0.3275)	109.0262 (0.3541)	108.9707 (0.3767)
	Naive	109.7942 (0.3511)	109.4749 (0.3427)	109.3155 (0.3764)	109.2250 (0.4080)	109.1602 (0.4366)
1/6	HFT	108.9835 (0.1185)	108.8610 (0.1269)	108.7884 (0.1483)	108.7398 (0.1670)	108.7056 (0.1812)
	O-W	109.6182 (0.3132)	109.3091 (0.3144)	109.1573 (0.3467)	109.0673 (0.3785)	109.0046 (0.3995)
	Naive	109.8357 (0.3577)	109.5144 (0.3602)	109.3590 (0.4006)	109.2627 (0.4325)	109.1994 (0.4603)

Fig. 1 Expected execution cost for different trading strategies ($\mu = 3$ %)

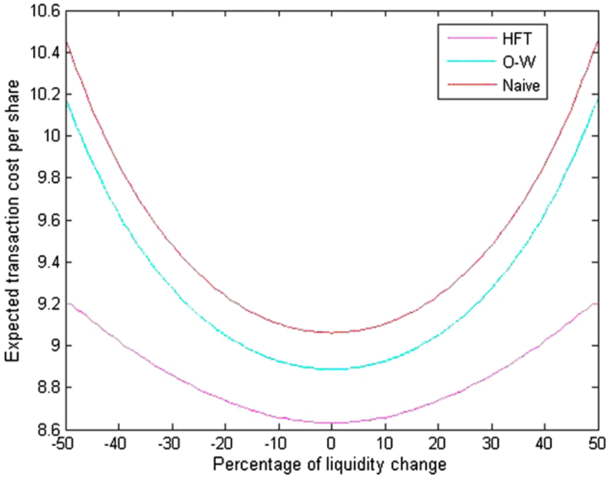


Table 4 Comparison of the expected execution cost for HFT, non-HF optimal trading (O-W), and Naive trading at different trading frequencies

Drift	Trading frequency N	60 min	30 min	15 min	10 min	5 min	1 min
$\mu = -5$ %	HFT	104.5402	104.5244	104.4895	104.4580	104.3804	104.1756
	O-W	105.1298	105.1305	105.1273	105.1268	105.1264	105.1262
	Naive	105.2932	105.3723	105.4163	105.4324	105.4491	105.4629
$\mu = -3$ %	HFT	105.9459	105.9311	105.8971	105.8681	105.7996	105.6270
	O-W	106.1368	106.1376	106.1345	106.1340	106.1336	106.1334
	Naive	106.3047	106.3849	106.4295	106.4458	106.4626	106.4766
$\mu = 0$ %	HFT	107.6235	107.6101	107.5791	107.5544	107.5000	107.3723
	O-W	107.6406	107.6416	107.6385	107.6379	107.6376	107.6374
	Naive	107.8110	107.8919	107.9368	107.9532	107.9702	107.9842
$\mu = 3$ %	HFT	108.7840	108.7722	108.7458	108.7265	108.6866	108.5998
	O-W	109.1369	109.1377	109.1346	109.1341	109.1337	109.1335
	Naive	109.3049	109.3851	109.4297	109.4460	109.4629	109.4768
$\mu = 5$ %	HFT	109.2706	109.2597	109.2375	109.2223	109.1922	109.1305
	O-W	110.1303	110.1310	110.1279	110.1273	110.1269	110.1267
	Naive	110.2941	110.3732	110.4173	110.4334	110.4501	110.4639

The S&P 100 stocks are large cap companies (i.e., blue chip companies across multiple industry groups) in the U.S. The S&P MidCap 400 stocks are mid-sized companies which over 7 % of the U.S. equity market. The S&P SmallCap 600 covers approximately 3 % of

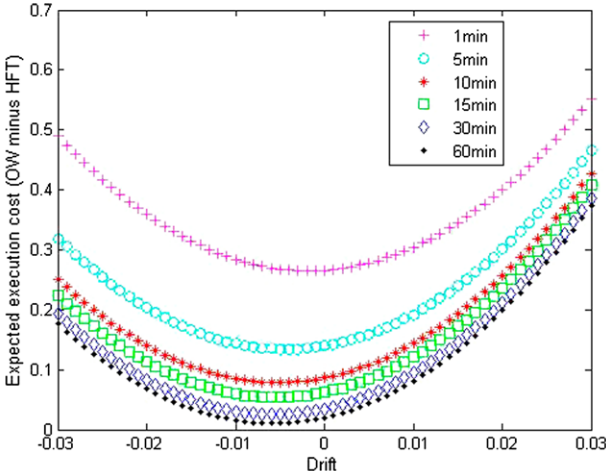


Fig. 2 The simulation result: expected execution cost reduction of HFT comparing with the non-HF optimal trading (O-W)

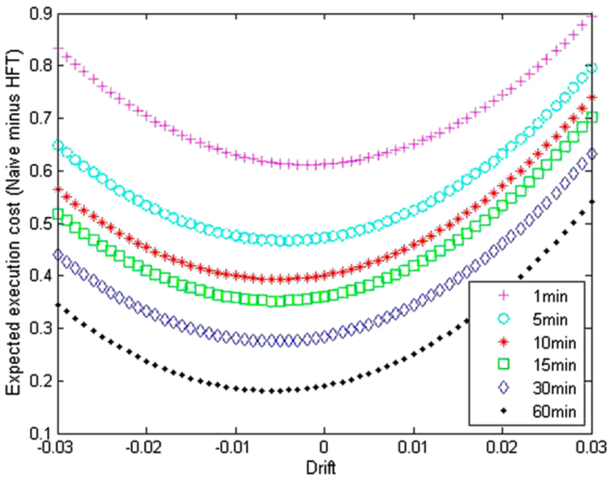


Fig. 3 The simulation result: expected execution cost reduction of HFT comparing with Naive trading

the U.S. equities market.[4] Therefore, in this empirical study we are able to assess our HFT strategy for different stocks based on their cap segment in the market which is typically renowned for trading liquidity and financial instability.

5.2 Empirical results

We use the tick level data to verify the conclusion we have drawn based on the simulation study. In this empirical test, we assume that we only provide liquidity once per trading day by submitting market orders for accumulating a position of 100,000 shares. The order size is decided by our HFT algorithm derived in Sect. 3. Transaction prices are determined by aggregating the data with the trade classification rules suggested by Ellis et al. (2000) mentioned above. For each trading day through our sample, we compute the average execution cost of HFT, non-HF optimal trading (O-W), and Naive trading at different trading

[4]See S&P, http://www.standardandpoors.com.

Fig. 4 Expected execution cost reduction of HFT comparing with the non-HF optimal trading (O-W) for Procter and Gamble (P&G)

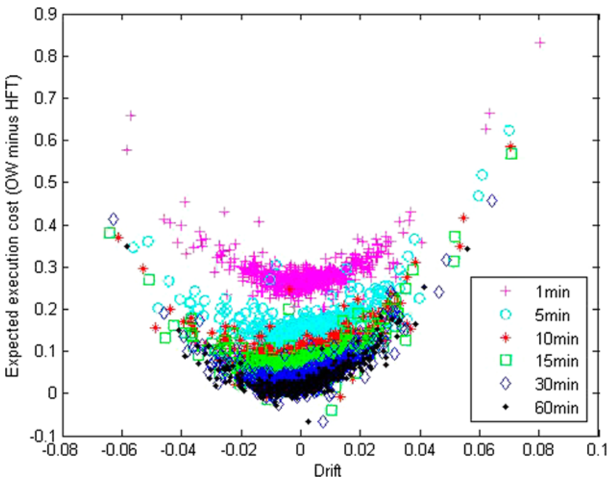
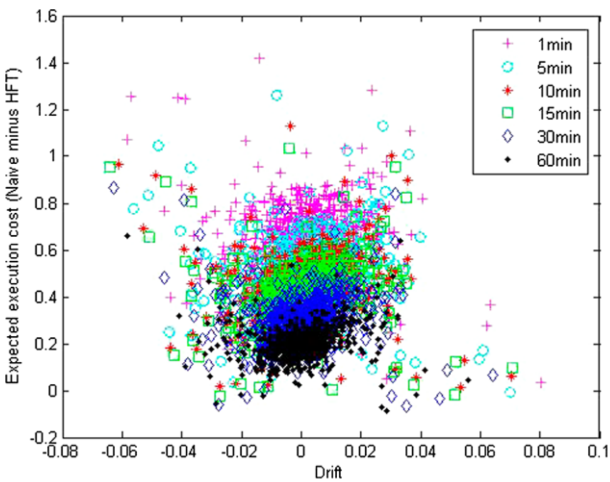


Fig. 5 Expected execution cost reduction of HFT comparing with Naive trading for Procter and Gamble (P&G)



frequencies (i.e., from 1 minute to 60 minutes). Figure 4 to Fig. 9 illustrate some representative patterns when we comparing the expected execution cost reduction of HFT with non-HF optimal trading (O-W) and Naive trading. We can see that the empirical observation for comparing the expected execution cost reduction of HFT with non-HF optimal trading (O-W) is relatively close to the patterns we theoretically shown (see Fig. 2). But the empirical results for comparing the expected execution cost reduction of HFT with Naive trading is not close to the patterns we theoretically shown (see Fig. 3).

In order to quantitatively compare the execution cost reduction of HFT at different trading frequencies with non-HF optimal trading (O-W) and Naive trading, we regress the cost of non-HF optimal trading (O-W) and Naive trading on that of HFT. The specification (regression 1) is

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_{it}, \tag{11}$$

where X_{it} stands for the measure of execution cost of stock i on day t based on HFT and Y_{it} measures the execution cost of stock i on day t without considering HFT under the model

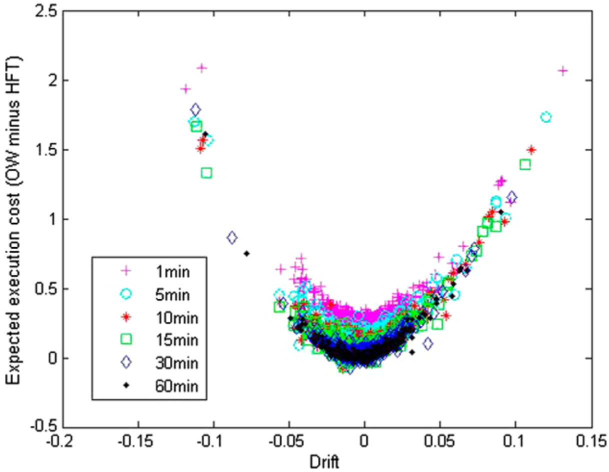


Fig. 6 Expected execution cost reduction of HFT comparing with the non-HF optimal trading (O-W) for Exxon

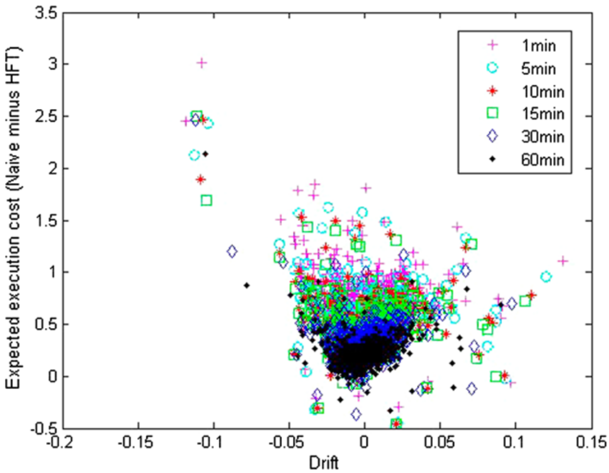


Fig. 7 Expected execution cost reduction of HFT comparing with Naive trading for Exxon

setting, i.e., non-HF optimal trading (O-W) and Naive trading. Based on our simulation result, we can infer that $\alpha_i + \gamma_t$ in Eq. (11) should be greater than zero and β should be greater than 1 if HFT can significantly reduce the execution cost comparing with the non-HF optimal trading (O-W) and Naive trading.

In order to remove the individual and time effects that might influence the execution costs, we then run the regression with following specification (regression 2):

$$Y_{it}^* = \bar{\beta} X_{it}^* + \epsilon_{it}. \tag{12}$$

Based on our simulation result, we can infer that $\bar{\beta}$ in Eq. (12) should be greater than 1 if HFT can significantly reduce the execution cost comparing with the non-HF optimal trading (O-W) and Naive trading.

We report the results of expected execution cost reduction of HFT with non-HF optimal trading (O-W) in Table 5 and with Naive trading in Table 6 respectively. In our regression, the term of $\alpha_i + \gamma_t$ measures the expected execution cost reduction in dollar and the term of $\bar{\beta}$ measures that in bps. From Tables 5 and 6, we can see (1) the execution cost of HFT given by

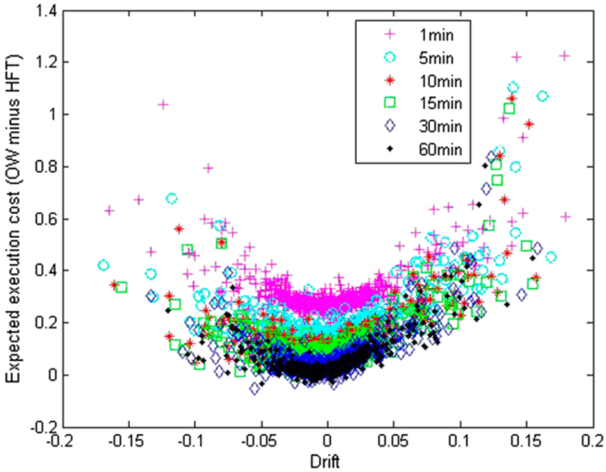


Fig. 8 Expected execution cost reduction of HFT comparing with the non-HF optimal trading (O-W) for JP Morgan

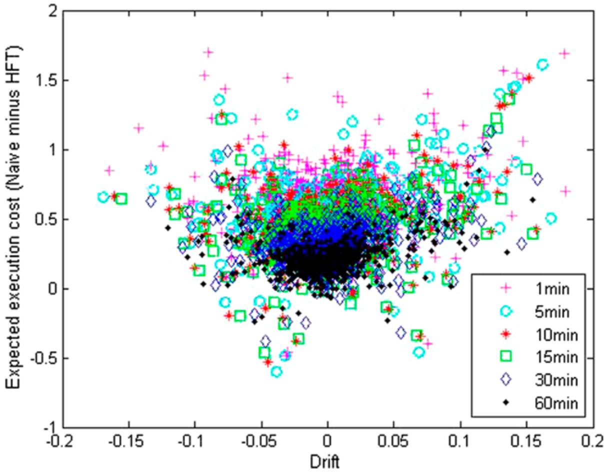


Fig. 9 Expected execution cost reduction of HFT comparing with Naive trading for JP Morgan

our model is lower than that of the non-HFT no matter if it is optimal trading (O-W strategy) or not (Naive strategy) and (2) increasing trading frequency will further reduce the execution cost, that is, the higher the trading frequency, the lower the execution cost. Particularly, in the sample we investigated, under the empirical setting we made (i.e., only one fixed position for supplying liquidity each trading day), when increasing trading frequency from 60 minutes to 1 minute, the corresponding expected execution cost reduction comparing with non-HF optimal trading (O-W) and Naive trading is 16 bps and 23 bps respectively.

We compute average daily trading volume (i.e., daily trading volume divided by the number of daily transactions) and average daily transaction for each stock. We then sort stocks into five quintiles based on the average trading volume per trade (i.e., average daily trading volume divided by the average daily transaction). Quintile 1 refers to stocks with largest average trading volume per trade and quintile 5 corresponds to the smallest. We investigate

Table 5 Comparison of execution cost between HFT and O-W strategy. Regression 1 is specified as $Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_{it}$ and regression 2 is specified as $Y^*_{it} = \bar{\beta} X^*_{it} + \epsilon_{it}$, where X stands for the execution cost of HFT and Y for the O-W strategy. Fixed effects are included and t-values in parentheses are based on standard errors that are robust to general cross-sectional and time-series heteroskedasticity and within-group autocorrelation

		60 min	30 min	15 min	10 min	5 min	1 min
Regression 1	β	1.0009 (2.7e+04)	1.0010 (2.6e+04)	1.0012 (2.4e+04)	1.0013 (2.2e+04)	1.0015 (2.1e+04)	1.0025 (1.9e+04)
	$\alpha_i + \gamma_t$	0.1990 (87.84)	0.3334 (140.28)	0.5284 (202.80)	0.6512 (236.96)	0.8369 (285.24)	1.0273 (325.73)
	F(104, 150784)	43.87	42.00	43.21	42.29	41.17	36.04
Regression 2	$\bar{\beta}$	1.0011 (2.6e+04)	1.0011 (2.5e+04)	1.0014 (2.2e+04)	1.0015 (2.1e+04)	1.0017 (2.0e+04)	1.0026 (1.8e+04)

Table 6 Comparison of execution cost between HFT and Naive strategy. Regression 1 is specified as $Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_{it}$ and regression 2 is specified as $Y^*_{it} = \bar{\beta} X^*_{it} + \epsilon_{it}$, where X stands for the execution cost of HFT and Y for the Naive strategy. Fixed effects are included and t-values in parentheses are based on standard errors that are robust to general cross-sectional and time-series heteroskedasticity and within-group autocorrelation

		60 min	30 min	15 min	10 min	5 min	1 min
Regression 1	β	1.0010 (2.5e+04)	1.0011 (2.3e+04)	1.0014 (2.1e+04)	1.0016 (1.9e+04)	1.0019 (1.8e+04)	1.0033 (1.7e+04)
	$\alpha_i + \gamma_t$	0.3927 (158.73)	0.6298 (233.27)	0.8708 (292.86)	1.0096 (317.47)	1.2098 (362.33)	1.3927 (384.24)
	F(104, 150784)	35.68	31.18	31.40	29.70	30.11	29.66
Regression 2	$\bar{\beta}$	1.0012 (2.4e+04)	1.0013 (2.2e+04)	1.0016 (2.0e+04)	1.0018 (1.8e+04)	1.0021 (1.8e+04)	1.0033 (1.6e+04)

if the performance of our trading strategies is different with respect to different quintiles. We use the performance dummy $\delta_j(i)$, which occurs in the following fixed effect model:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{j=1}^{5} \beta_j \delta_j(i) X_{it} + \epsilon_{it}, \tag{13}$$

where $\delta_j(i) = 1$ when the i-th observation belongs to the j-th group and $\delta_j(i) = 0$ otherwise. We report the results based on Eq. (13) in Tables 7 and 8. We can see that the performance of HFT strategy is better than the O-W and Naive strategy for each quintile. We can verify that the HFT strategy can reduce transaction cost comparing with non-HFT strategy and such superior performance is independent of the characteristics (trading volume or turnover) of underlying stock.

5.3 Comparison with simulation results

In our simulation study, following the geometric Brownian motion assumption, we generate the price dynamics and assess the performance of HFT together with non-HF trading and Naive trading. We find that (1) for a given trading frequency, the HFT strategy performs better than the non-HF (O-W) trading and Naive trading; and (2) when we increase trading frequency, our HFT strategy performs better than other two trading strategies. Our empirical results confirm the simulation results. We can conclude this based on Tables 4, 5 and 6. For example, in Table 4, we can see that when increase trading frequency, the expected execution cost of HFT turns to decrease while the cost of Naive trading increases and we conclude the same from Table 6.

Figure 2 illustrates the expected execution cost reduction of HFT comparing with the non-HF (O-W) trading. From the empirical results, we could observe a similar pattern for the comparison between HFT and non-HF (O-W) trading, see Figs. 4, 6, and 8 for different stocks. Figure 3 illustrates the expected execution cost reduction of HFT comparing with the Naive trading, the corresponding empirical results illustrate different patterns, see Figs. 5, 7, and 9 for different stocks, but these different patterns do not violate our conclusion.

6 Conclusions

This paper presents a model under discrete optimization to explain how high frequency trading (HFT) can reduce the expected execution cost and supply liquidity. In this model, we consider only liquidity influence by HFT and treat market volatility as an exogenous variable. The HF traders in our model act as market makers providing liquidity by submitting passive orders. They neither make directional bets nor strive to earn tiny margins on the trades. They try to limit their position risk by controlling adverse selection in the execution of their passive orders, that is, to predict order flow or volume arrival after submitting the HFT order and decide how to break up the large order optimally to reduce market impact.

In this paper, we show the numerical solution and analyze the properties derived from the model. In order to show the rigidity of the model, we conduct both simulation and empirical investigation. With the simulation study, we illustrate the theoretical performance of HFT given the model setting. This study shows that HFT can reduce the expected execution cost when comparing with non-HF optimal trading and Naive trading. This study investigates the expected execution cost reduction of HFT with tick level data of US stocks sampled from S&P 100, S&P MidCap 400, and S&P SmallCap 600 through January 2008 to October

Table 7 Comparison of execution cost between HFT and O-W strategy. We use $Y_{it} = \alpha_i + \gamma_t + \sum_{j=1}^{5} \beta_j \delta_j(i) X_{it} + \epsilon_{it}$, to check if there exists inter-quantile difference, where X stands for the execution cost of HFT and Y for the O-W strategy. Fixed effects are included and t-values in parentheses are based on standard errors that are robust to general cross-sectional and time-series heteroskedasticity and within-group autocorrelation

Quintile	60 min	30 min	15 min	10 min	5 min	1 min
1	1.0201 (1166.99)	1.0172 (1097.54)	1.0137 (1027.53)	1.0154 (1012.04)	1.0142 (986.94)	1.0161 (979.85)
2	1.0202 (1622.65)	1.0193 (1527.64)	1.0177 (1434.07)	1.0166 (1413.08)	1.0174 (1377.52)	1.0185 (1375.79)
3	1.0123 (2220.90)	1.0113 (2092.72)	1.0079 (1970.71)	1.0100 (1937.24)	1.0103 (1887.33)	1.0105 (1871.25)
4	1.0103 (2814.57)	1.0094 (2665.78)	1.0078 (2507.74)	1.0085 (2454.22)	1.0079 (2389.19)	1.0078 (2376.50)
5	1.0098 (4594.20)	1.0097 (4366.68)	1.0095 (4227.59)	1.0089 (4217.77)	1.0089 (4135.98)	1.0095 (4049.72)
cons.	1.3884 (82.55)	1.5802 (88.74)	1.8083 (95.30)	1.8262 (94.70)	1.8911 (96.24)	1.8793 (96.23)

Table 8 Comparison of execution cost between HFT and Naive strategy. We use $Y_{it} = \alpha_i + \gamma_t + \sum_{j=1}^{5} \beta_j \delta_j(i) X_{it} + \epsilon_{it}$, to check if there exists inter-quintile difference, where X stands for the execution cost of HFT and Y for the Naive strategy. Fixed effects are included and t-values in parentheses are based on standard errors that are robust to general cross-sectional and time-series heteroskedasticity and within-group autocorrelation

Quintile	60 min	30 min	15 min	10 min	5 min	1 min
1	1.0197 (1167.05)	1.0166 (1098.26)	1.0136 (1027.53)	1.0150 (1011.01)	1.0140 (986.36)	1.0160 (979.27)
2	1.0199 (1622.16)	1.0189 (1528.20)	1.0177 (1434.30)	1.0164 (1411.72)	1.0174 (1376.26)	1.0185 (1374.64)
3	1.0121 (2220.57)	1.0112 (2092.42)	1.0080 (1970.72)	1.0099 (1935.98)	1.0103 (1886.02)	1.0106 (1870.29)
4	1.0101 (2814.17)	1.0093 (2664.58)	1.0079 (2508.00)	1.0084 (2452.51)	1.0079 (2385.06)	1.0079 (2374.39)
5	1.0097 (4583.79)	1.0098 (4348.63)	1.0097 (4225.25)	1.0088 (4195.84)	1.0090 (4115.43)	1.0097 (4029.63)
cons.	1.3993 (83.18)	1.5888 (89.18)	1.8098 (95.40)	1.8386 (95.26)	1.8968 (96.46)	1.8826 (96.33)

2010. We show that in the sample we investigated, under the empirical setting (i.e., only one fixed position for supplying liquidity each trading day), HFT can significantly reduce the expected execution cost. Particularly, when increasing trading frequency from 60 minutes to 1 minute, the corresponding expected execution cost reduction comparing with non-HF optimal trading (O-W) and Naive trading is 16 bps and 23 bps respectively.

Our finding coincides with other empirical papers stating that HFT improve market quality and reduce price impact (see for example, Hendershott et al. 2011). In our model, HFT has neither information on the fundamental values of the asset traded nor increases microstructure noise since these variables are all exogenous. Since the volatility is also treated as exogenous variable, with our model we cannot conclude if HFT increases stock price volatility.

Appendix

A.1 Proof of Proposition 1

For the induction basis at time $t_N = T$ we have

$$\begin{aligned} J_T^i(X_T, D_T, F_T, T) &= \left(F_T + \frac{s}{2}\right)X_T + \left[\lambda_0(X_0 - X_{t_m}) + \lambda_i(X_{t_m} - X_T) + D_T + \frac{X_T}{2q_i}\right]X_T \\ &= \left(F_T + \frac{s}{2}\right)X_T + \lambda_i\left(\frac{\lambda_0}{\lambda_i}X_0 - \frac{\lambda_0}{\lambda_i}X_{t_m} + X_{t_m}\right)X_T \\ &\quad + \left(\frac{1}{2q_i} - \lambda_i\right)X_T^2 + X_T D_T. \end{aligned}$$

For the induction step for some $t_n \in \{t_{m+1}, \dots, t_{N-1}\}$ we get

$$\begin{aligned} J_{t_n}^i(X_{t_n}, D_{t_n}, F_{t_n}, t_n) &= \min_{x_{t_n}}\left\{\left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_m}) + \lambda_i(X_{t_m} - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_i}\right]x_{t_n}\right. \\ &\quad \left. + E_{t_n} J_{t_{n+1}}^i\left(X_{t_n} - x_{t_n}, (D_{t_n} + \kappa_i x_{t_n})e^{-\rho_i\tau}, F_{t_{n+1}}, t_{n+1}\right)\right\} \\ &= \min_{x_{t_n}}\left\{\left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_m}) + \lambda_i(X_{t_m} - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_i}\right]x_{t_n}\right. \\ &\quad + \left(a_i \times a_i^{N-(n+1)}F_n + \frac{s}{2}\right)(X_{t_n} - x_{t_n}) \\ &\quad + \lambda_i\left(\frac{\lambda_0}{\lambda_i}X_0 - \frac{\lambda_0}{\lambda_i}X_{t_m} + X_{t_m}\right)(X_{t_n} - x_{t_n}) + b_{n+1}^i(X_{t_n} - x_{t_n})^2 \\ &\quad + c_{n+1}^i(D_{t_n} + \kappa_i x_{t_n})^2 e^{-2\rho_i\tau} + d_{n+1}^i v_i F_{t_n}^2 \\ &\quad + g_{n+1}^i(X_{t_n} - x_{t_n})(D_{t_n} + \kappa_i x_{t_n})e^{-\rho_i\tau} \\ &\quad \left. + h_{n+1}^i(X_{t_n} - x_{t_n})a_i F_{t_n} + l_{n+1}^i(D_{t_n} + \kappa_i x_{t_n})e^{-\rho_i\tau}a_i F_{t_n}\right\}. \end{aligned} \tag{14}$$

To obtain the minimum, we differentiate Eq. (14) with respect to x_{t_n}

$$\begin{aligned} \frac{\partial J}{\partial x_{t_n}} &= \left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_m}) + \lambda_i(X_{t_m} - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{q_i} \\ &\quad - \left(a_i^{N-n}F_n + \frac{s}{2}\right) - \lambda_i\left(\frac{\lambda_0}{\lambda_i}X_0 - \frac{\lambda_0}{\lambda_i}X_{t_m} + X_{t_m}\right) - 2b_{n+1}^i(X_{t_n} - x_{t_n}) \end{aligned}$$

$$
\begin{aligned}
&+ 2\kappa_i c^i_{n+1}(D_{t_n} + \kappa_i x_{t_n})e^{-2\rho_i\tau} + g^i_{n+1}e^{-\rho_i\tau}\big[\kappa_i(X_{t_n} - x_{t_n}) - (D_{t_n} + \kappa_i x_{t_n})\big] \\
&- h^i_{n+1}a_i F_{t_n} + l^i_{n+1}\kappa_i e^{-\rho_i\tau}a_i F_{t_n} \\
= {}& x_{t_n}\left(\frac{1}{q_i} + 2b^i_{n+1} - 2g^i_{n+1}\kappa_i e^{-\rho_i\tau} + c^i_{n+1}2\kappa_i^2 e^{-2\rho_i\tau}\right) \\
&+ X_{t_n}\big(-\lambda_i - 2b^i_{n+1} + g^i_{n+1}e^{-\rho_i\tau}\kappa_i\big) + D_{t_n}\big(1 + c^i_{n+1}2\kappa_i e^{-2\rho_i\tau} - g^i_{n+1}e^{-\rho_i\tau}\big) \\
&+ F_{t_n}\big(1 - a_i^{N-n} - h^i_{n+1}a_i + l^i_{n+1}\kappa_i e^{-\rho_i\tau}a_i\big).
\end{aligned} \tag{15}
$$

Setting $\frac{\partial J}{\partial x_{t_n}} \stackrel{!}{=} 0$ for Eq. (15) to obtain the optimal choice

$$
x_{t_n} = oD_{t_n} + wX_{t_n} + uF_{t_n}, \tag{16}
$$

where

$$
\begin{aligned}
o &= -\frac{1}{2}\delta^i_{n+1}\big(1 + c^i_{n+1}2\kappa_i e^{-2\rho_i\tau} - g^i_{n+1}e^{-\rho_i\tau}\big), \\
w &= -\frac{1}{2}\delta^i_{n+1}\big(-\lambda_i - 2b^i_{n+1} + g^i_{n+1}e^{-\rho_i\tau}\kappa_i\big), \\
u &= -\frac{1}{2}\delta^i_{n+1}\big(1 - a_i^{N-n} - h^i_{n+1}a_i + l^i_{n+1}\kappa_i e^{-\rho_i\tau}a_i\big), \\
\delta^i_{n+1} &= \left(\frac{1}{2q_i} + b^i_{n+1} - g^i_{n+1}\kappa_i e^{-\rho_i\tau} + c^i_{n+1}\kappa_i^2 e^{-2\rho_i\tau}\right)^{-1}.
\end{aligned}
$$

Putting Eq. (16) into Eq. (14) we obtain the optimal value function given by Eq. (2) and find the coefficients given by Eqs. (4)–(10). This completes the induction for $t_n \in \{t_{m+1}, \dots, t_N\}$. We are unsure about market reaction to the event and the following change of the parameters that describe the market. At t_m we face the following problem

$$
\begin{aligned}
J^0_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = {}& \overset{min}{x_{t_n}} E_{t_n}\left\{\left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q_u}\right]x_{t_n}\right. \\
&\left. + J_{t_{n+1}}\big(X_{t_n} - x_{t_n}, (D_{t_n} + \kappa_u x_{t_n})e^{-\rho_u\tau}, F_{t_{n+1}}, t_{n+1}\big)\right\}
\end{aligned} \tag{17}
$$

where q_u, κ_u, a_u and ρ_u should indicate that the current value of q and the future value of κ, ρ, and a_u are unknown.

Because the event is modeled as a discrete random variable, we obtain

$$
\begin{aligned}
&E_{t_n} J_{t_{n+1}}\big(X_{t_n} - x_{t_n}, (D_{t_n} + \kappa_u x_{t_n})e^{-\rho_u\tau}, a_u F_{t_{n+1}}, t_{n+1}\big) \\
&\quad = \sum_{i=0}^{r} p_i \times J^i_{t_{n+1}}\big(X_{t_n} - x_{t_n}, (D_{t_n} + \kappa_i x_{t_n})e^{-\rho_i\tau}, a_i F_{t_n}, t_{n+1}\big).
\end{aligned} \tag{18}
$$

We use $a_i = e^{\mu_i\tau}$ and $v_i = e^{(2\mu_i+\sigma_i^2)\times\tau}$ and define

$$
\frac{1}{\overline{q}} = \sum_{i=1}^{r} p_i \times \frac{1}{q_i}, \qquad \overline{a} = \sum_{i=1}^{r} p_i \times \big(a_i^{N+1-(m+1)}\big), \qquad \overline{\lambda} = \sum_{i=1}^{r} p_i \times \lambda_i.
$$

Combining Eqs. (17) and (18) with this definitions, we find that

$$
\begin{aligned}
&J^0_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) \\
&\quad = \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2\overline{q}}\right]x_{t_n}
\end{aligned}
$$

$$+\left(F_n\overline{a}+\frac{s}{2}\right)(X_{t_n}-x_{t_n})+\overline{\lambda}\left(\frac{\lambda_0}{\overline{\lambda}}X_0-\frac{\lambda_0}{\overline{\lambda}}X_{t_n}+X_{t_n}\right)(X_{t_n}-x_{t_n})$$
$$+\left(\sum_{i=1}^{r}p_i\times b_{n+1}^i\right)(X_{t_n}-x_{t_n})^2$$
$$+D_{t_n}^2\left(\sum_{i=1}^{r}p_i\times c_{n+1}^i e^{-2\rho_i\tau}\right)+2D_{t_n}x_{t_n}\left(\sum_{i=1}^{r}p_i\times c_{n+1}^i e^{-2\rho_i\tau}\kappa_i\right)$$
$$+x_{t_n}^2\left(\sum_{i=1}^{r}p_i\times c_{n+1}^i e^{-2\rho_i\tau}\kappa_i^2\right)+\left(\sum_{i=1}^{r}p_i\times d_{n+1}^i\upsilon_i\right)F_{t_n}^2$$
$$+(X_{t_n}-x_{t_n})\left(D_{t_n}\left(\sum_{i=1}^{r}p_i\times g_{n+1}^i e^{-\rho_i\tau}\right)+\left(\sum_{i=1}^{r}p_i\times g_{n+1}^i e^{-\rho_i\tau}\kappa_i\right)x_{t_n}\right)$$
$$+\left(\sum_{i=1}^{r}p_i\times h_{n+1}^i a_i\right)(X_{t_n}-x_{t_n})F_{t_n}+D_{t_n}F_{t_n}\left(\sum_{i=1}^{r}p_i\times l_{n+1}^i a_i e^{-\rho_i\tau}\right)$$
$$+x_{t_n}F_{t_n}\left(\sum_{i=1}^{r}p_i\times l_{n+1}^i a_i e^{-\rho_i\tau}\kappa_i\right). \tag{19}$$

We then obtain the solution that minimizes Eq. (19) is

$$x_m=-\frac{1}{2}\delta_{m+1}\left(\left(1+2\sum_{i=1}^{r}p_i\times c_{n+1}^i e^{-2\rho_i\tau}\kappa_i-\sum_{i=1}^{r}p_i\times g_{n+1}^i e^{-\rho_i\tau}\right)D_{t_m}\right.$$
$$+\left(-\overline{\lambda}-2\sum_{i=1}^{r}p_i\times b_{n+1}^i+\sum_{i=1}^{r}p_i\times g_{n+1}^i e^{-\rho_i\tau}\kappa_i\right)X_{t_m}$$
$$\left.+\left(1-\overline{a}-\sum_{i=1}^{r}p_i\times h_{n+1}^i a_i+\sum_{i=1}^{r}p_i\times l_{n+1}^i a_i e^{-\rho_i\tau}\kappa_i\right)F_{t_m}\right), \tag{20}$$

with

$$\delta_{m+1}=\left(\frac{1}{2\overline{q}}+\sum_{i=1}^{r}p_i\times b_{n+1}^i-\sum_{i=1}^{r}p_i\times g_{n+1}^i e^{-\rho_i\tau}\kappa_i+\sum_{i=1}^{r}p_i\times c_{n+1}^i e^{-2\rho_i\tau}\kappa_i^2\right)^{-1}.$$

Inserting Eq. (20) into Eq. (19), we find the optimal value function given by Eq. (2) and the coefficients given by Eqs. (4)–(10). For the induction step for some $t_n\in\{t_0,\dots,t_{m-1}\}$ we get

$$J_{t_n}^0(X_{t_n},D_{t_n},F_{t_n},t_n)=\min_{x_{t_n}}\left\{\left[\left(F_{t_n}+\frac{s}{2}\right)+\lambda_0(X_0-X_{t_n})+D_{t_n}+\frac{x_{t_n}}{2q_0}\right]x_{t_n}\right.$$
$$\left.+E_{t_n}J_{t_{n+1}}^0\left(X_{t_n}-x_{t_n},(D_{t_n}+\kappa_0x_{t_n})e^{-\rho_0\tau},F_{t_{n+1}},t_{n+1}\right)\right\}$$
$$=\min_{x_{t_n}}\left\{\left[\left(F_{t_n}+\frac{s}{2}\right)+\lambda_0(X_0-X_{t_n})+D_{t_n}+\frac{x_{t_n}}{2q_0}\right]x_{t_n}\right.$$
$$+\left(a_0\times a_0^{m-(n+1)}\times\overline{a}F_n+\frac{s}{2}\right)(X_{t_n}-x_{t_n})+\lambda_0X_0(X_{t_n}-x_{t_n})$$
$$+b_{n+1}^0(X_{t_n}-x_{t_n})^2+c_{n+1}^0(D_{t_n}+\kappa_0x_{t_n})^2e^{-2\rho_0\tau}+d_{n+1}^0\upsilon_0F_{t_n}^2$$

$$+ g^0_{n+1}(X_{t_n} - x_{t_n})(D_{t_n} + \kappa_0 x_{t_n})e^{-\rho_0\tau}$$
$$+ h^0_{n+1}(X_{t_n} - x_{t_n})a_0 F_{t_n} + l^0_{n+1}(D_{t_n} + \kappa_0 x_{t_n})e^{-\rho_0\tau} a_0 F_{t_n}\Bigg\}. \quad (21)$$

To obtain the minimum we differentiate Eq. (21) with respect to x_{t_n}

$$\begin{aligned}
\frac{\partial J}{\partial x_{t_n}} &= \left(F_{t_n} + \frac{s}{2}\right) + \lambda_0(X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{q_0} \\
&\quad - \left(a_0^{m-n} \times \overline{a} F_n + \frac{s}{2}\right) - \lambda_0 X_0 - 2b^0_{n+1}(X_{t_n} - x_{t_n}) \\
&\quad + 2\kappa_0 c^0_{n+1}(D_{t_n} + \kappa_0 x_{t_n})e^{-2\rho_0\tau} + g^0_{n+1}e^{-\rho_0\tau}\left[\kappa_0(X_{t_n} - x_{t_n}) - (D_{t_n} + \kappa_0 x_{t_n})\right] \\
&\quad - h^0_{n+1}a_0 F_{t_n} + l^0_{n+1}\kappa_0 e^{-\rho_0\tau} a_0 F_{t_n} \\
&= x_{t_n}\left(\frac{1}{q_0} + 2b^0_{n+1} - 2g^0_{n+1}\kappa_0 e^{-\rho_0\tau} + c^0_{n+1}2\kappa_0^2 e^{-2\rho_0\tau}\right) \\
&\quad + X_{t_n}\left(-\lambda_0 - 2b^0_{n+1} + g^0_{n+1}e^{-\rho_0\tau}\kappa_0\right) + D_{t_n}\left(1 + c^0_{n+1}2\kappa_0 e^{-2\rho_0\tau} - g^0_{n+1}e^{-\rho_0\tau}\right) \\
&\quad + F_{t_n}\left(1 - a_0^{m-n} \times \overline{a} - h^0_{n+1}a_0 + l^0_{n+1}\kappa_0 e^{-\rho_0\tau} a_0\right). \qquad (22)
\end{aligned}$$

Setting $\frac{\partial J}{\partial x_{t_n}} \stackrel{!}{=} 0$ for Eq. (22) to obtain the optimal choice

$$x_{t_n} = oD_{t_n} + wX_{t_n} + uF_{t_n}, \quad (23)$$

where

$$\begin{aligned}
o &= -\frac{1}{2}\delta^0_{n+1}\left(1 + c^0_{n+1}2\kappa_0 e^{-2\rho_0\tau} - g^0_{n+1}e^{-\rho_0\tau}\right), \\
w &= -\frac{1}{2}\delta^0_{n+1}\left(-\lambda_0 - 2b^0_{n+1} + g^0_{n+1}e^{-\rho_0\tau}\kappa_0\right), \\
u &= -\frac{1}{2}\delta^0_{n+1}\left(1 - a_0^{m-n} \times \overline{a} - h^0_{n+1}a_0 + l^0_{n+1}\kappa_0 e^{-\rho_0\tau} a_0\right), \\
\delta^0_{n+1} &= \left(\frac{1}{2q_0} + b^0_{n+1} - g^0_{n+1}\kappa_0 e^{-\rho_0\tau} + c^0_{n+1}\kappa_0^2 e^{-2\rho_0\tau}\right)^{-1}.
\end{aligned}$$

Putting Eq. (23) into Eq. (21) we obtain the optimal value function given by Eq. (2) and the coefficients given by Eqs. (4)–(10). This concludes the induction.

References

Almgren, R., & Chriss, N. (1999). Value under liquidation. *Risk*, *12*, 61–63.

Almgren, R., & Chriss, N. (2000). Optimal execution of portfolios. *Journal of Risk*, *3*(2), 5–39.

Biais, B., Foucault, T., & Moinas, S. (2011). *Equilibrium high frequency trading*. Working Paper.

Bertsimas, D., & Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, *1*, 1–50.

Chan, L. K. C., & Lakonishok, J. (1995). The behavior of stock prices around institutional trades. *The Journal of Finance*, *50*(4), 1147–1174.

Easley, D., Prado, M., & O'Hara, M. (2012). Flow toxicity and liquidity in a high frequency world. *Review of Financial Studies*, *25*(5), 1457–1493.

Ellis, K., Michaely, R., & O'Hara, M. (2000). The accuracy of trade classification rules: evidence from NASDAQ. *Journal of Financial and Quantitative Analysis*, *35*(4), 529–551.

Hendershott, T., Jones, C., & Menkveld, A. (2011). Does algorithmic trading improve liquidity? *Journal of Finance*, *66*(1), 1–33.

Huberman, G., & Stanzel, W. (2004). Arbitrage-free price update and price-impact functions. *Econometrica*, *72*(4), 1247–1275.

Jeria, D., & Sofianos, G. (2008). Passive orders and natural adverse selection. *Street Smart*, *33*(September 4).

Jovanovic, B., & Menkveld, A. J. (2012). *Middlemen in limit-order markets*. Working Paper.

Keim, D., & Madhavan, A. (1995). Anatomy of the trading process: empirical evidence on the behavior of institutional traders. *Journal of Financial Economics*, *37*, 371–398.

Lee, C., Mucklow, B., & Ready, M. (1993). Spreads, depths and the impact of earnings information: an intraday Analysis. *The Review of Financial Studies*, *6*(2), 345–374.

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

Patell, J. M., & Wolfson, M. A. (1984). The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics*, *13*, 223–252.

Sun, E. W., Rezania, O., Rachev, S., & Fabozzi, F. (2011). Analysis of the intraday effects of economic releases on the currency market. *Journal of International Money and Finance*, *30*(4), 692–707.

Sun, E. W., & Kruse, T. (2012). *Optimal order submission strategy when market liquidity is uncertain*. Working Paper.