Annals of Operations Research (2019) 281:315–347
https://doi.org/10.1007/s10479-019-03150-0

S.I. : RISK IN FINANCIAL ECONOMICS

Stylized algorithmic trading: satisfying the predictive near-term demand of liquidity

Edward W. Sun[1] · **Timm Kruse**[2] · **Yi-Ting Chen**[3]

Published online: 21 February 2019

Abstract
Regulatory reform enacted (e.g., the Dodd-Frank Act enforced in the U.S.) requires the financial service industry to consider the "reasonably expected near term demand" (i.e., RENTD) in trading. To manage the price impact and transaction cost associated with orders submitted to an order driven market, market makers or specialists must determine their trading styles (aggressive, neutral, or passive) based on the market liquidity in response to RENTD, particularly for trading a large quantity of some financial instrument. In this article we introduce a model considering different trading styles to satisfy the predictive near-term customer demand of market liquidity in order to find an optimal order submission strategy based on different market situations. We show some analytical properties and numerical performances of our model in search of optimal solutions. We evaluate the performances of our model with simulations run over a set of experiments in comparison with two alternative strategies. Our results suggest that the proposed model illustrates superiority in performance.

Keywords Artificial intelligence · Algorithmic trading · Decision analytics · Discrete optimization · FinTech · Liquidity

JEL Classification C61 · C63 · G10

1 Introduction

Proprietary trading (i.e., "prop trading") refers to an institution that trades in its own account with its own funds for its own direct market gain, whereas flow trading earns commission by trading on behalf of its clients. Institutions engaging in prop and flow trading can boost trading profits by strategically setting the bid-ask spread by accessing the order flows, see

---

This work was supported by the InfoTech research project funded under USt-IdNr. DE320245686.

✉ Edward W. Sun
edward.sun@bem.edu

[1] KEDGE Business School, 680 Cours de la Libèration, 33405 Talance Cedex, France

[2] InfoTech, Frankfurt am Main, Germany

[3] College of Computer Science, National Chiao Tung University, Taiwan, Hsinchu, Taiwan (ROC)

Foucault et al. (2013) for example. Sun et al. (2014) illustrate how a market maker benefits from accessing real-time information in financial market, and Sun et al. (2015) discuss the impact of regulatory intervention on liquidity provision in terms of optimal trading. Kissell (2014) discusses the potential technology that market maker can apply in terms of algorithmic trading.

With fast development of financial technology (i.e., FinTech), liquidity provision in many markets has shifted from market makers to autonomous, computerized multilateral trading facilities. These algorithmic trading systems, electronic trade matching engines, and new connectivity methods that characterize modern financial markets can collect, disseminate, process, and react to market-wide information more comprehensively and efficiently than human traders who will be replaced. Coinciding with the performance of the "representative trader" in Sun et al. (2014) and Sun et al. (2015), Gerig and Michayluk (2017) report the automated market making transacts most orders, quotes price efficiently, decreases transaction costs, and has no effect on volatility. On the other hand, regulatory authorities have proposed a series of risk controls, transparency measures, and other safeguards to enhance the regulatory regime for automated trading. Starting with the Germany's High Frequency Trading Act in 2013, and now continuing with the European MiFID II rules and US Commodity Futures Trading Commission (CFTC) rules on automated trading (known as "Regulation AT"), regulation is put forward to deal with financial market challenges (see Sun et al. 2018, for example).

The Dodd-Frank Wall Street Reform and Consumer Protection Act (referred to as the Dodd-Frank Act)[1] signed into federal law by President Barack Obama on July 21, 2010 brought about most significant regulatory changes for financial services following the crisis erupted in 2007. As a part of the Dodd-Frank Act, the Volcker Rule draws lines between prop trading and other permissible activities and provides a benchmark to decide whether a particular trading strategy is a service, a hedge, or a gamble. It restricts speculations that do not benefit their customers. On December 18, 2014, the Federal Reserve extended the Volcker Rules conformance period until July 21, 2016. Based on the rules, the instruments and related risks that the trading institution carries at any point in time must be designed to meet "reasonably expected near term demand" (i.e., RENTD),[2] which ties the positions of the trading institution to its client activity, but not to prop trading. Therefore, when the trading institutions determine risk limits and position limits on the size, holding periods, and risk exposure of their market making inventory, they must effectively estimate and take into account future customer demand (based predominantly on past activity).

In an order driven market, all buyers and sellers display prices and amounts at which they wish to trade a particular commodity or security. The major advantage of an order driven market is transparency, since the flow of individual orders (i.e., the entire order book) can be displayed for market participants who wish to access this information. In order to trade efficiently, computer-based algorithmic trading is generally applied. When seeking immediacy to guarantee a best execution, strategic order submission and low execution costs are important issues by applying computer based algorithmic trading under different situations of market liquidity.

Many studies on the cost of immediacy assume perfect competition in market making. This leads to the conclusion that the price of immediacy depends on the marginal cost when market makers supply immediacy, see Chacko et al. (2008), Dorn et al. (2015), O'Hara (2015), Sun et al. (2014) and references therein. The cost of holding inventory and the cost of

[1] See http://www.gpo.gov/fdsys/pkg/PLAW-111publ203/html/PLAW-111publ203.htm, for details.

[2] Final rule §4(b)(2)(ii).

adverse selection (that will arise when price does not reflect information) make up the total costs of market making. In an illiquid market, customer demand is weak, large market order submitted by an institutional trader usually has a significant price impact because of the high cost of market making; whereas in a liquid market where customer demand is strong, large market orders might be executed too fast to be adjusted for a better price (i.e., flow toxicity as mentioned by Easley et al. (2012)). In order to reduce the overall price impact of a large market order, it is necessary to manage the order size in an optimal way. Specifically, we investigate how to asynchronously submit orders to grant the transitory price impact that benefits market making with respect to those transactions that demand immediacy. Strategic order submission should determine when to submit orders in an aggressive, neutral, or passive way based on market liquidity (particularly for RENTD). The challenge then is how to optimally choose the trading style (or behavior) - that is, aggressive, neutral, or passive trading - to minimize the price impact and transaction cost. Several studies have investigated this problem of optimal execution - for example, Bertsimas and Lo (1998), Almgren and Chriss (2000), Almgren (2003), Ting et al. (2007), Schöneborn and Schied (2009), Alfonsi et al. (2010), Obizhaeva and Wang (2013), and Sun et al. (2014), among others.

In this article we propose a strategic trading algorithm for placing market orders to optimally minimize the price impact and transaction cost. Our framework is an optimal model of transactions for immediacy and similar to Bertsimas and Lo (1998), Almgren and Chriss (2000), Almgren (2003), Obizhaeva and Wang (2013), and Sun et al. (2014), whereby our optimal trading strategy relies on the price impact function. This function describes the phenomenon that causes a price trajectory to deviate from its naturally intended price path. As Chacko et al. (2008) assume that all agents are symmetrically informed and the inventory risk can be offset due to perfect interdealer markets, the market maker is able to extract some of the difference between a trader's reservation values and fundamental value in exchange for supplying immediacy - that is, to increase the bid-ask spread. Once an order is released, it conveys information. When the information pertains to the immediacy or liquidity demanded by a trader, it will cause an imbalance in supply and demand, then the price disturbance will be short lived. When the information pertains to the change in fundamental value, the price disturbance will persist for a prolonged period. These two types of price disturbance refer to the temporary and permanent price impacts. In this paper, we use the linear price impact function–that is, a linear combination of the temporary and permanent price impacts. Huberman and Stanzel (2004) indicate that the linear price impact function excludes the quasi-arbitrage (i.e., price manipulation) and supports viable market prices. Similar to Chacko et al. (2008), this model setup creates a market for immediacy where the fundamental value is determined in a separate market. Trader submits orders via a stylized strategy, and we assume that processing fundamental value and order flows of other traders are unaffected by the individual's trading decision. The privileged position of the market maker and the asynchronous arrival of immediacy-seeking traders, provide the single market maker with a power of setting the transaction price that is determined by the intensity of opposing orders (that is, market liquidity in perfect competition).

We work herein on a particular case for market makers who occasionally present based on the order flows that have to be placed during the course of trading. Following other studies (see Huberman and Stanzel 2004; Obizhaeva and Wang 2013, for example) we focus on the discrete-time modeling. Obizhaeva and Wang (2013) state that the natural way to address this issue is to take a continuous time limit of the discrete-time formulation. Within the literature, such a limitation might lead to degenerate solutions with certain types of price impact functions. As a consequence, the optimal solutions derived from the discrete-time model are in general sub-optimal (see, for example, Almgren and Chriss (2000)). In practice, the timing of trades is often predetermined before running the computer program. Therefore, in this paper we focus on discrete time modeling for the optimal strategy.

In our model we use the *target participation rate* (which is named by practitioners) to characterize trading styles.[3] Participation rate originally refers to the order size relative to the expected trading volume over a trading period in the market and determines trading aggressiveness. The higher the rate is, the shorter the execution horizon and the greater the price impact will be. We define the target participation rate to characterize the relationship between the actual order-induced depth currently available in response to demand immediacy and the total expected liquidity that the market maker could supply during a specific trading period. For example, under the Volker Rules, if the market maker predicts the near-term demand to be high, then a higher target participation rate will be applied accordingly to meet the demand. As a consequence, the price impact increases and then leads to higher transaction costs. In our model, the target participation rate determines the trade-off between reducing the execution cost and trading aggressiveness. It can be predetermined based on the order execution probability and market liquidity with backward and forward propagation analyses, particularly on trading patterns and turnover in inventory. Our model is more flexible for implementation under different trading styles (i.e., aggressive, neutral, and passive trading). In practice, institutional traders can estimate the target participation rate based on their own primary data, which makes our model more plausible.

When traders searching for viable trading counterparties in capricious markets, they are usually impatient. Consequently, kinetic price is driven by the impatient trader who searches for immediacy. In our study, we build up our model based on two different assumptions for the underlying price dynamics: the Brownian motion suggested by Obizhaeva and Wang (2013) and the geometric Brownian motion as suggested by Ting et al. (2007) and Sun et al. (2014). In order to compare the performance of our optimal trading strategy, we consider two alternative strategies in our simulations. One is the naive trading strategy, which splits a large order equally into small pieces, and the other is the trading strategy proposed by Obizhaeva and Wang (2013). We use the volume weighted average price (VWAP) as a benchmark to measure the average execution cost of these trading strategies (see, for example Werner (2003) and Goldstein et al. (2009)). Only the trading strategy with lowest execution cost is preferred. In order to show the model's implementation for practitioners, we evaluate its performance with simulations run over a set of numerical experiments. The results based on our simulation study indicate that the overall performance of our optimal order submission strategy significantly dominates the alternative trading strategies.

We organized the paper as follows. In Sect. 2, we introduce the fundamental model for our optimal order submission strategy. Section 3 describes analytical solutions of the model–, that is, we consider the target participation rate to satisfy the near-term demand and allow the underlying price dynamics to follow both the Brownian motion and geometric Brownian motion. In Sect. 4, we evaluate the model performance with simulations in comparison with two alternative trading strategies and report the results. We summarize our conclusions in Sect. 5.

[3] See the participation rate in Goldman Sachs Equity Execution Strategies: Street Smart, Issue 39, September 30, 2009, USA.

2 The model

2.1 Trading scenario under the Dodd-Frank Act

The Volcker Rule, as a part of the Dodd-Frank Act, has been proposed to restrict U.S. institutions from trading in certain kinds of speculative investments that might not benefit their customers. It requires that the instruments and related risks that the trading institution carries at any point in time must be designed to meet the "reasonably expected near term demand" (i.e., RENTD). Consequentially, trading institutions must take into account RENTD by effectively estimating the market liquidity when they are determining risk limits and position limits on the size, holding periods, and risk exposure of their market making inventory. Easley et al. (2012) indicate that the order arrival process is informative, and the volume arrival is a metric for it. Passive trading usually leads to a high adverse selection cost, because passive orders are filled quickly when they should be filled slowly and filled slowly when they should be filled quickly, see Sun et al. (2014) and references therein.

RENTD is obviously subject to the trading institution and the rules suggest each trading desk needs to conduct backward and forward looking analyses, which is crucial to determine the historical trading patterns such as turnover in inventory and to predict future possibilities. The action of forward looking means deciding upon the current and expected market conditions and how changes in them will impact the amount of expected demand. In addition, the rule clearly points out that the analysis must be demonstrable, which means it needs to be backed up by evidence and not just by conjecture.

Following Easley et al. (2012), we consider that a passive order does not cross the market and generates adverse selection cost, because the timing of execution is directly out of the trader's control when the market is liquid. We already note that a particular trade or order placed by an impatient customer will generate a high impact on the price and consequentially increase transaction costs (due to the adverse selection effect), see Foucault et al. (2013) and reference therein. On the other hand, Easley et al. (2012) point out that order flow is toxic when it adversely selects market makers who provide liquidity at a loss. In addition, the Volker Rule requires the market maker to consider effective customer demand, which means that the market maker should satisfy customer demand without holding onto positions. Therefore, the objective in our model is to minimize the execution cost by optimally executing orders in response to different market liquidities of the near-term demand by using the proposed target participation rate.

2.2 Model set-up

Similar to Bertsimas and Lo (1998), Almgren and Chriss (2000), Obizhaeva and Wang (2013), and Sun et al. (2014), we assume that a representative trader (e.g., a market maker) is going to execute a trade with total order size $X_0 (X_0 > 0)$ over a fixed time interval $[0, T] (T > 0)$. This trader is allowed to execute the trade partially by submitting market orders at the pre-determined target participation rate π, i.e., $Z_0 := \pi X_0$ with $0 \leq \pi \leq 1$ and $Z_0 \leq X_0$, over the time interval $[0, T_1]$ $(0 < T_1 < T)$. Here, T and T_1 refer to any trading horizon and could be a trading day or an hour. The target participation rate, $\pi = Z_0/X_0$, in our model is the actual order supply (i.e., Z_0 over $[0, T_1]$) in response to customer demand relative to the total expected trading volume (X_0) the trader could supply during a specific trading period ($[0, T]$). It (π) characterizes the trading style (e.g., aggressive trading when π is large and passive trading when π is small) over the interval $[0, T_1]$ in order to meet RENTD. For example, under the Volker Rules, if the market maker predicts the near-term demand to be high, then a higher target participation rate will be applied accordingly to meet the demand. In this article, we focus only on buying orders, since this model can be easily adopted for selling orders. In our model, the trader is only allowed to trade at a discrete time. Here, $N \in \{1, 2, \ldots\}$ stands for the trading frequency, and $t_i, i \in (0, \ldots, N)$ are the time points starting at $t_0 = 0$ and ending in $t_N = T$. We then define $t_i = i\,\tau$, where $\tau = T/N$ as the duration between two successive time points of trading, and t_m is the end of the first trading period T_1 and $0 < m < N$. We define x_{t_n} as the amount of the asset (e.g., number of shares if the underlying asset is a listed stock) the traders wants to buy at time t_n. The objective function, the trader's strategy that minimizes the expected execution cost of the whole trade, is given as follows:

$$\min_{(x_0, \ldots, x_T) \in \Phi} \left(E \sum_{n=0}^{N} x_{t_n} \bar{P}_{t_n} \right),$$

where $\bar{P}_{t_n}$ is the average price of the order at time t_n is executed, Φ is the set of feasible strategies, and

$$\Phi = \left\{ \left\{ x_{t_0}, \ldots, x_{t_N} \right\} : x_{t_n} \geq 0 \; \forall \, n \in \{0, \ldots, N\}; \sum_{n=0}^{m} x_{t_n} = Z_0; \sum_{n=m+1}^{N} x_{t_n} = X_0 - Z_0 \right\}.$$

We shall note that in our study, we assume the representative trader is risk neutral (who cares only about the expected value but not the uncertainty of the total cost) and supplies liquidity in the market. A liquidity provider responds to market information by quickly and frequently updating his quotes and then generates order flows. We impose a non-negativity constraint, $x_{t_n} \geq 0$, to exclude selling an asset before completing the whole trade of buying. We assume the price of underlying asset F_t is independent of t and $E(\Delta\, F_t | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t$ is a set of events at t, and it follows either a Brownian motion or geometric Brownian motion with a drift μ, variance σ, and its initial value $F_0 = A_0$.

In our framework, all traders must search for trading counterparties to complete their positions, and the market maker enables the profit by supplying liquidity (or immediacy) from the difference between the reservation values and fundamental value—that is, the bid-ask spread. We then assume the bid-ask spread is constant with $s > 0$ and symmetric around the asset price following Obizhaeva and Wang (2013). Based on the inventory models, the spread quoted by the dealer is independent of the number of stocks he trades, and so the diversification of the dealer's trading activities does not affect the spread. However, there may not be a single spread from the information-based models. If price varies with trade size, the spread for large trades may be larger than the small trade spread, see O'Hara (1995) and references therein. Gerig and Michayluk (2017) propose that with the automated market maker for providing liquidity, the spread in average transaction price between buys and sells in a security remains the same, but is smaller for liquidity traders and larger for informed investors. One the other hand, the spread can be expressed by a function that are impacted by the market maker's transaction cost, see Sun et al. (2015), for example.

Price impact refers to the relationship between an incoming bid or ask order (to buy or to sell) and the subsequent price change. Letting Δm_t be the change in the midpoint of the spread s at time t (i.e., the mid-price), if the mid-price proportionally changes with respect to buying (bid arrivals) or selling pressure (ask arrivals), then the relationship is expressed as follows:

$$\Delta m_t = \lambda\, q_t + \epsilon_t,$$

where q_t is the order imbalance—that is, the total value of buy less sell market orders executed in the same interval; and ϵ_t is an uncorrelated increment - that is, non-trading driven price variation such that $E(\epsilon_t) = 0$, see Foucault et al. (2013). It states that the net demand during the specific trading interval from market order arrivals puts pressure on the price that is gauged by the coefficient λ. When buy order imbalance (buy orders exceed sell orders) occurs, λ drives price up. On the other hand, when sell order imbalance (sell orders exceed buy orders) occurs, λ pulls price down. The reciprocal of λ is used to measure market liquidity–that is, market depth in our model. A smaller value of λ indicates prices are less sensitive to order imbalance, and vice versa for a large value. A buy trade leads to a price increase that raises the cost, because the second buy trade is on average more expensive than the first caused by self impact, and vice versa for sells, see Kissell (2014), Kyle and Obizhaeva (2016), Kyle et al. (2016), Kyle et al. (2017), and Sun et al. (2015). Therefore, minimizing price impact is the goal of our model. Foucault et al. (2013) review several methods to estimate λ by running a regression model. Hasbrouck (2007) points out that orders do not impact price, but rather forecast price. In addition, orders possess information and price reflects the course of adjustment to this information, see Bessembinder et al. (2016), Easley et al. (2012), and Sun et al. (2014).

In our model we study how the asynchronous arrivals of customer demand grant the market maker a transitory (or temporary) pricing impact when the customer is seeking for immediacy. In this sense, our framework is similar to Chacko et al. (2008), where all traders are simultaneously informed and the market maker does not possess the inventory risk. Given q as the depth of the limit-order book, i.e., the available amount of asset per unit of cost, our trade x_{t_n} has a price impact at time t_n with a magnitude of x_{t_n}/q. The average price impact of the whole trade is then $x_{t_n}/2q$. We decompose the price impact into permanent and temporary price impacts (see Sun et al. (2014) and references therein). The permanent price impact λx_{t_n}, with $0 \leq \lambda \leq 1/q$, is the change in price caused by orders that lead the market to believe the future price will reflect its current expectation or there is a change in the asset's intrinsic value. This in turn brings a jump in price to our model when buying before t_n. The price impact at t_n is $\lambda(X_0 - X_{t_n})$, where we define $X_{t_n} = X_0 - \sum_{i=0}^{n-1} x_{t_i}$ as the amount of asset the trader still has to buy from t_n until the whole trade is completed.

Similar to Chacko et al. (2008), the market maker is enabled to continually trade in the market until the positions are completed. Trading with asynchronous orders for immediacy, the market maker then has a similar price impact as other major market makers. Therefore, the intensity of opposing order arrivals determines the degree of pricing power of the market maker. The market maker can estimate or predict the intensity of order arrival and react accordingly based on such a conditional estimation. The target participation rate π is then to be used to characterize the reaction to the estimated order intensity (i.e., RENTD) by the market maker with respect to reduce the market impact.

Price impact reveals private information, particularly for a fundamental valuation. Such private information causes trades, which make other market agents update their valuations accordingly, therefore incurring a price change. When there is no way to distinguish informed trades from non-informed ones, when all market agents think that a fraction of these trades might contain some news, all trades as a consequence will have a price impact. A 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 the underlying asset. Trades cause temporary increases in price for buy orders and temporary decreases in price for sell orders, subsequently followed by a price reversion back to the initial price trajectory. If private information turns to be public, then the price will surely reflect it and therefore, lead to a permanent impact on price; otherwise, the extent of the temporary price impact of the order vanishes along with time, and we define it as κx_{t_n}, where $\kappa = 1/q - \lambda$. The temporary price impact of an order decays from t_n to t_{n+1} with the speed of $e^{-\rho\tau}$, where $\rho > 0$ is the resilience[4] factor of the limit-order book. The temporary price impact of the order is characterized by $D_{t_n} = \sum_{i=0}^{n-1} x_{t_i} \kappa e^{-\rho(n-i)\tau}$ at t_n, where D_{t_n} is defined by the recursive equation $D_{t_n} = (D_{t_{n-1}} + \kappa x_{t_{n-1}})e^{-\rho\tau}$ with the initial value $D_0 = 0$.

The average price of x_{t_n} executed at t_n is:

$$\bar{P}_{t_n} = F_{t_n} + \frac{s}{2} + \lambda(X_0 - X_{t_n}) + D_{t_n} + \frac{x_{t_n}}{2q}.$$

Asset price rise (decline) is attributable to either (i) a public announcement that causes an increase (decrease) in valuation or (ii) an exogenous buying (selling) pressure from uninformed liquidity traders. In the former case (i.e., the fundamental volatility[5]), there is no reason to expect any further change in prices. In the latter case (i.e., the transitory volatility), the buying (selling) pressure of liquidity traders will cause a temporary rise (fall) in prices that will subsequently be reversed. Thus, price changes generated by non-informational buy (sell) trades tend to be reversed. More generally, informed (uninformed) trades will result in zero (non-zero) autocorrelation in individual stock returns, see Avramov et al. (2006). On the other hand, fundamental volatility results in permanent price impact and transitory volatility leads to temporary price impact. The average execution price can be decomposed as the fundamental value of asset ($F_{t_n} + \frac{s}{2}$) and trade-induced price impact of informed and uninformed traders ($\lambda(X_0 - X_{t_n}) + D_{t_n}$) in previous time periods and ($\frac{x_{t_n}}{2q}$) in current time-of-trade.

In our model we are able to obtain optimal solutions for any given participation rate, which is very practical for traders to choose their own trading style based on different market situations. Keeping other conditions constant, we achieve the same solutions as the model given by Obizhaeva and Wang (2013) if we halve the trading interval and set the participation rate at 50%.

3 Analytical solutions

We define $Y_0 := X_0 - Z_0$, which can be expressed as $X_0 = Z_0 + Y_0$. Here, Z_0 is the amount to be purchased between t_0 and t_m, and Y_0 is the amount to be purchased between t_{m+1} and t_N. The variable Z_t, $Z_t = Z_0 - \sum_{t_n<t} x_{t_n}$, as the remaining amount to be completed before t_m. We define Y_t, which is $Y_t = Y_0 - \sum_{t_m<t_n<t} x_{t_n}$, the remaining amount to be completed between t_{m+1} and t_N.

3.1 Case for the Brownian motion

When the underlying asset price F_T follows a Brownian motion, we obtain the following proposition.

Proposition 1 *Given the model setting described in Sect. 2 with F_T following a Brownian motion, the strategy*

[4] Resiliency refers to how quickly prices revert to former levels after they change in response to large order flow imbalances initialed by uninformed traders, see Harris (2003).

[5] Volatility can be decomposed into fundamental volatility and transitory volatility (see, Harris (2003), for example).

$$x_{t_n} = \begin{cases} Y_{t_N}, \quad \textit{for } t_N, \\ -\frac{1}{2}\delta_{n+1}\left(D_{t_n}(1-\beta_{n+1}e^{-\rho\tau}+\gamma_{n+1}2\kappa e^{-2\rho\tau}) - Y_{t_n}(\lambda+2\alpha_{n+1}-\kappa\beta_{n+1}e^{-\rho\tau})\right), \\ \quad \textit{for } t_n \in \{t_{m+1},\dots,t_{N-1}\}, \\ Z_{t_m}, \quad \textit{for } t_m, \\ -\frac{1}{2}\delta_{n+1}[1-g_{n+1}e^{-\rho\tau}+2\kappa s_{n+1}e^{-\rho\tau}]D_{t_n} + \frac{1}{2}\delta_{n+1}[\lambda+2h_{n+1}-\kappa g_{n+1}e^{-\rho\tau}]Z_{t_n} \\ -\frac{1}{2}\delta_{n+1}[\lambda Z_0 - a_{n+1} + b_{n+1}\kappa e^{-\rho\tau}] \\ \quad \textit{for } t_n \in \{t_0,\dots,t_{m-1}\}, \end{cases} \tag{1}$$

is optimal, if $(x_{t_0},\dots,x_{t_N}) \in \Phi$. *The optimal value function has the form*

$$J_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \begin{cases} J^1_{t_n}(Y_{t_n}, D_{t_n}, F_{t_n}, t_n), & \textit{for } t_n \in \{t_{m+1},\dots,t_N\}, \\ J^2_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n), & \textit{for } t_n \in \{t_0,\dots,t_m\}, \end{cases}$$

where

$$J^1_{t_n}(Y_{t_n}, D_{t_n}, F_{t_n}, t_n) = (F_{t_n} + \frac{s}{2})Y_{t_n} + \lambda(Y_0+Z_0)Y_{t_n} + \alpha_n Y^2_{t_n} + \beta_n Y_{t_n} D_{t_n} + \gamma_n D^2_{t_n}, \tag{2}$$

with the coefficients

$$\begin{aligned} \alpha_n &= \alpha_{n+1} - \frac{1}{4}\delta_{n+1}(\lambda+2\alpha_{n+1}-\beta_{n+1}\kappa e^{-\rho\tau})^2, \\ \beta_n &= \beta_{n+1}e^{-\rho\tau} + \frac{1}{2}\delta_{n+1}(1-\beta_{n+1}e^{-\rho\tau}+2\kappa\gamma_{n+1}e^{-2\rho\tau})(\lambda+2\alpha_{n+1}-\beta_{n+1}\kappa e^{-\rho\tau}), \\ \gamma_n &= \gamma_{n+1}e^{-2\rho\tau} - \frac{1}{4}\delta_{n+1}(1-\beta_{n+1}e^{-\rho\tau}+2\kappa\gamma_{n+1}e^{-2\rho\tau})^2, \\ \delta_{n+1} &= \left[\frac{1}{2q} + \alpha_{n+1} - \beta_{n+1}\kappa e^{-\rho\tau} + \kappa^2\gamma_{n+1}e^{-2\rho\tau}\right]^{-1}, \end{aligned} \tag{3}$$

and terminal conditions

$$\alpha_N = \frac{1}{2q} - \lambda, \;\; \beta_N = 1, \;\; \gamma_N = 0; \tag{4}$$

and

$$\begin{aligned} J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n) &= c_n + (F_{t_n} + \frac{s}{2})Y_0 + (F_{t_n} + \frac{s}{2})Z_{t_n} + a_n Z_{t_n} \\ &\quad + b_n D_{t_n} + h_n Z^2_{t_n} + g_n Z_{t_n} D_{t_n} + s_n D^2_{t_n}, \end{aligned} \tag{5}$$

with the coefficients

$$\begin{aligned} c_n &= c_{n+1} - \frac{1}{4}\delta_{n+1}[\lambda Z_0 - a_{n+1} + b_{n+1}\kappa e^{-\rho\tau}]^2, \\ a_n &= a_{n+1} - \frac{1}{2}\delta_{n+1}(\lambda+2h_{n+1}-\kappa g_{n+1}e^{-\rho\tau})(a_{n+1}-b_{n+1}\kappa e^{-\rho\tau}-\lambda Z_0), \\ b_n &= b_{n+1}e^{-\rho\tau} - \frac{1}{2}\delta_{n+1}(\lambda Z_0 - a_{n+1} + b_{n+1}\kappa e^{-\rho\tau})(-g_{n+1}e^{-\rho\tau}+2\kappa s_{n+1}e^{-2\rho\tau}+1), \\ h_n &= h_{n+1} - \frac{1}{4}\delta_{n+1}(\lambda+2h_{n+1}-\kappa g e^{-\rho\tau})^2, \\ g_n &= g_{n+1}e^{-\rho\tau} + \frac{1}{2}\delta_{n+1}(\lambda+2h_{n+1}-\kappa g e^{-\rho\tau})(1-g e^{-\rho\tau}+2\kappa s e^{-2\rho\tau}), \end{aligned}$$

$$s_n = s_{n+1}e^{-2\rho\tau} - \frac{1}{4}\delta_{n+1}(1 - ge^{-\rho\tau} + 2\kappa s e^{-2\rho\tau})^2,$$

$$\delta_{n+1} = \left[\frac{1}{2q} + h_{n+1} - g_{n+1}\kappa e^{-\rho\tau} + \kappa^2 s_{n+1}e^{-2\rho\tau}\right]^{-1}, \tag{6}$$

and the terminal conditions

$$c_m = \lambda(Y_0 + Z_0)Y_0 + \alpha_{m+1}Y_0^2,\ a_m = \lambda Z_0 + \kappa\beta_{m+1}Y_0e^{-\rho\tau},\ b_m = \beta_{m+1}Y_0e^{-\rho\tau},$$
$$g_m = 1 + 2\gamma_{m+1}\kappa e^{-2\rho\tau},\ h_m = \frac{1}{2q} - \lambda + \gamma_{m+1}\kappa^2 e^{-2\rho\tau},\ s_m = \gamma_{m+1}e^{-2\rho\tau}. \tag{7}$$

Proof 1 See Appendix A. □

3.2 Case for the geometric Brownian motion

In this section we assume the underlying price movement follows a geometric Brownian motion, and then we obtain the following proposition.

Proposition 2 *Given the model setting described in Sect. 2 with F_T following a geometric Brownian motion, the strategy*

$$x_{t_n} = \begin{cases} Y_{t_N}, \quad \text{for } t_N, \\ -\frac{1}{2}\delta_{n+1}[(1 + c_{n+1}2\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})D_{t_n} + (-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa)Y_{t_n} \\ \quad +(1 - a^{N-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau}a)F_{t_n}], \\ \quad \text{for } t_n \in \{t_{m+1}, \ldots, t_{N-1}\}, \\ Z_{t_m}, \quad \text{for } t_m, \\ -\frac{1}{2}\delta_{n+1}[(-\lambda - 2b_{n+1} + g_{n+1}\kappa e^{-\rho\tau})Z_{t_n} + (1 + 2c_{n+1}\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})D_{t_n} \\ \quad +(1 - a^{m-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau}a)F_{t_n} + (\lambda Z_0 - s_{n+1} + u_{n+1}\kappa e^{-\rho\tau})], \\ \quad \text{for } t_n \in \{t_0, \ldots, t_{m-1}\}, \end{cases} \tag{8}$$

is optimal, if $(x_{t_0}, \ldots, x_{t_N}) \in \Phi$. The optimal value function then has the form

$$J_{t_n}(X_{t_n}, D_{t_n}, F_{t_n}, t_n) = \begin{cases} J_{t_n}^1(Y_{t_n}, D_{t_n}, F_{t_n}, t_n), & \text{for } t_n \in \{t_{m+1}, \ldots, t_N\}, \\ J_{t_n}^2(Z_{t_n}, D_{t_n}, F_{t_n}, t_n), & \text{for } t_n \in \{t_0, \ldots, t_m\}, \end{cases}$$

where

$$\begin{aligned} J_{t_n}^1(Y_{t_n}, D_{t_n}, F_{t_n}, t_n) = {} & (a^{N-n}F_n + \frac{s}{2})Y_{t_n} + \lambda(Z_0 + Y_0)Y_{t_n} \\ & + b_nY_{t_n}^2 + c_nD_{t_n}^2 + d_nF_{t_n}^2 + g_nY_{t_n}D_{t_n} \\ & + h_nY_{t_n}F_{t_n} + l_nD_{t_n}F_{t_n}, \end{aligned} \tag{9}$$

with $a = e^{\mu\tau}$ and $m = e^{(2\mu+\sigma^2)\times\tau}$, and the coefficients are given by

$$b_n = b_{n+1} - \frac{1}{4}\delta_{n+1}(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa)^2,$$
$$c_n = c_{n+1}e^{-2\rho\tau} - \frac{1}{4}\delta_{n+1}(1 + c_{n+1}2\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})^2,$$
$$d_n = d_{n+1}m - \frac{1}{4}\delta_{n+1}(1 - a^{N-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau}a)^2,$$
$$g_n = g_{n+1}e^{-\rho\tau} - \frac{1}{2}\delta_{n+1}(1 + c_{n+1}2\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa),$$

$$
\begin{aligned}
h_n &= h_{n+1}a - \frac{1}{2}\delta_{n+1}(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa)(1 - a^{N-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau}a),\\
l_n &= l_{n+1}e^{-\rho\tau}a - \frac{1}{2}\delta_{n+1}(1 - a^{N-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau}a)(1 + c_{n+1}2\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau}),\\
\delta_{n+1} &= (\frac{1}{2q} + b_{n+1} - g_{n+1}\kappa e^{-\rho\tau} + c_{n+1}\kappa^2 e^{-2\rho\tau})^{-1};
\end{aligned}
\tag{10}
$$

and the terminal conditions are given by

$$
b_N = \frac{1}{2q} - \lambda, c_N = 0, d_N = 0, g_N = 1, h_N = 0, l_N = 0. \tag{11}
$$

For $t_n \in \{t_0, \ldots, t_{N-1}\}$ we have

$$
\begin{aligned}
J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n) = (a^{m-n}F_n + \frac{s}{2})Z_{t_n} + (a^{N-n}F_n + \frac{s}{2})Y_0\\
+ b_n Z^2_{t_n} + c_n D^2_{t_n} + d_n F^2_{t_n} + g_n Z_{t_n} D_{t_n}\\
+ h_n Z_{t_n} F_{t_n} + l_n D_{t_n} F_{t_n} + s_n Z_{t_n} + u_n D_{t_n} + w_n F_{t_n} + o_n,
\end{aligned}
\tag{12}
$$

with the coefficients

$$
\begin{aligned}
b_n &= b_{n+1} - \frac{1}{4}(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa)^2,\\
c_n &= c_{n+1}e^{-2\rho\tau} - \frac{1}{4}(1 + 2\kappa c_{n+1}e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})^2,\\
d_n &= d_{n+1}m - \frac{1}{4}\delta_{n+1}(1 - a^{m-n} - h_{n+1}a + l_{n+1}ae^{-\rho\tau}\kappa)^2,\\
g_n &= g_{n+1}e^{-\rho\tau}\kappa - \frac{1}{2}\delta_{n+1}(1 + 2c_{n+1}e^{-2\rho\tau}\kappa - g_{n+1}e^{-\rho\tau})(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa),\\
h_n &= h_{n+1}a - \frac{1}{2}\delta_{n+1}(1 - a^{m-n} - h_{n+1}a + l_{n+1}ae^{-\rho\tau}\kappa)(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa),\\
l_n &= l_{n+1}ae^{-\rho\tau} - \frac{1}{2}\delta_{n+1}(1 - a^{m-n} - h_{n+1}a + l_{n+1}ae^{-\rho\tau}\kappa)(1 + 2c_{n+1}e^{-2\rho\tau}\kappa - g_{n+1}e^{-\rho\tau}),\\
s_n &= s_{n+1} - \frac{1}{2}\delta_{n+1}(-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa)(\lambda Z_0 - s_{n+1} + u_{n+1}e^{-\rho\tau}\kappa),\\
u_n &= u_{n+1}e^{-\rho\tau} - \frac{1}{2}\delta_{n+1}(\lambda Z_0 - s_{n+1} + u_{n+1}e^{-\rho\tau}\kappa)(1 + 2c_{n+1}e^{-2\rho\tau}\kappa - g_{n+1}e^{-\rho\tau}),\\
w_n &= w_{n+1}a - \frac{1}{2}\delta_{n+1}(\lambda Z_0 - s_{n+1} + u_{n+1}e^{-\rho\tau}\kappa)(1 - a^{m-n} - h_{n+1}a + l_{n+1}ae^{-\rho\tau}\kappa),\\
o_n &= o_{n+1} - \frac{1}{4}\delta_{n+1}(\lambda Z_0 - s_{n+1} + u_{n+1}e^{-\rho\tau}\kappa)^2,\\
\delta_{n+1} &= \left[\frac{1}{2q} + b_{n+1} + c_{n+1}\kappa^2 e^{-2\rho\tau} - g_{n+1}\kappa e^{-\rho\tau}\right]^{-1},
\end{aligned}
\tag{13}
$$

and the terminal conditions

$$
\begin{aligned}
&b_m = -\lambda + \frac{1}{2q} + c_{n+1}\kappa^2 e^{-2\rho\tau},\ c_m = c_{n+1}e^{-2\rho\tau},\ d_m = d_{n+1}m,\ g_m = 1 + 2\kappa c_{n+1}e^{-2\rho\tau},\\
&h_m = l_{n+1}a\kappa e^{-\rho\tau},\ l_m = l_{n+1}ae^{-\rho\tau},\ s_m = \lambda Z_0 + g_{n+1}\kappa e^{-\rho\tau}Y_0,\ u_m = g_{n+1}Y_0 e^{-\rho\tau},\\
&w_m = h_{n+1}Y_0 a,\ o_m = \lambda(Z_0 + Y_0)Y_0 + b_{n+1}Y_0^2.
\end{aligned}
\tag{14}
$$

Proof 2 See Appendix B. □

3.3 Considering the "no-selling" constraint

In Proposition 1 the analytical solutions are optimal only if the strategy is in Φ. When $(x_0, \dots, x_N) \notin \Phi$, there exists at least one $x_i < 0$ (i.e., indicating a sell order). This can happen due to different parameter settings of the model, and we cannot predetermine if $(x_0, \dots, x_N) \in \Phi$ or not. We cannot guarantee the trading strategy is still optimal when we sell securities before completing purchases. If we are allowed to sell before completing purchases, then the execution price for sell orders cannot be easily determined. In our model, when $x_i < 0$, it is optimal only if the sell orders are executed at the ask price, but not at the bid price.

When solving the optimal programming problem, we observe there exist negative values for the trading volume at certain time points. Intuitively, these negative values can be considered as sales because in a bearish market with a significant downside trend, traders who lower the long position or conduct a short-selling can profit versus the strategy of holding and waiting for a rally. In practice we have two major concerns. For institutions only dealing with flow trades (i.e., no proprietary position), they profit from commissions, and therefore even though the short-selling is profitable, these institutions have no motivation to do so. For prop trading institutions, many markets apply rules such as a short-selling ban or an up-tick rule during bearish markets that prohibit or reduce profits from short-selling. Particularly, under the Dodd-Frank Act, which originally restricts U.S. banks from making certain kinds of speculative transactions that do not benefit their customers, short-selling does not benefit the near-term demand. Lawmakers argued that such a speculative activity played a key role in the financial crisis of 2007–2010, and new rules are necessary to ban speculation on prop trading, specifically to ban short-selling by institutions, and whereby deposits are used to trade on institutions' own accounts. To make our solution more practical, particularly for executions completed in a very short trading horizon, we thus consider the "no-selling" constraint, $x_i \geq 0$.

As we know, if x_i is not optimal, then the coefficients should be adjusted accordingly. The size of the order x_i depends not only on the coefficients, but also on D_{t_i} and X_{t_i}, which are calculated from the values of x_j for $j < i$. This suggests that we have to use a different approach to find the solution when $x_i < 0$ in order to confirm the solution is feasible for all parameters. We restrict x_i in the interval $[0, X_i]$ and $x_i \leq X_i$ to ensure that there are no forced sells later.

We find that $x_N = X_{t_N}$ is optimal if all x_j for $j \in \{0, \dots, N-1\}$ are $\in [0, X_j]$, which we could ensure. We then can keep this as a part of the solution. For x_{N-1}, we have three cases: (1) $x_{N-1} < 0$, (2) $0 \leq x_{N-1} \leq X_{N-1}$, and (3) $X_{N-1} < x_{N-1}$. For case (1), since the trader is not allowed to sell, $x_{N-1} = 0$ is optimal. We can minimize the optimal value function J in t_{N-1} with the choice of x_{N-1}. In all cases the optimal value function is a quadratic function that opens up. For case (1), the angular point is below 0, and the optimal value function is increasing with $x_{N-1} \in [0, X_{t_{N-1}}]$. This leads to $x_{N-1} = 0$. For this case the coefficients are determined by different recursive equations.

Proposition 3 *We replace $x_{N-1} < 0$ with $x_{N-1} = 0$, and the coefficients to satisfy the following recursive equations for the model:*

$$\alpha_{N-1} = \alpha_N, \;\; \beta_{N-1} = \beta_N e^{-\rho\tau}, \;\; \gamma_{N-1} = \gamma_N e^{-2\rho\tau}, \;\; \delta_N = 0.$$

Proof 3 At $t_n = N-1$, we have

$$\begin{aligned}
&J^1_{t_n}(Y_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \overset{min}{y_{t_n}} \{\left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda(Y_0 + Z_0 - Y_{t_n}) + D_{t_n} + \frac{y_{t_n}}{2q}\right] y_{t_n} + \left(F_{t_n} + \frac{s}{2}\right)(Y_{t_n} - y_{t_n})\\
&\quad + \lambda(Y_0 + Z_0)(Y_{t_n} - y_{t_n}) + \alpha_{n+1}(Y_{t_n} - y_{t_n})^2\\
&\quad + \beta_{n+1}(Y_{t_n} - y_{t_n})(D_{t_n} + \kappa y_{t_n})e^{-\rho\tau} + \gamma_{n+1}(D_{t_n} + \kappa y_{t_n})^2 e^{-2\rho\tau}\}.
\end{aligned}$$

We insert $y_{t_n} = 0$ for y_{t_n} to obtain the minimal value with the constraints we implied for the control variable and get

$$\begin{aligned}
&J^1_{t_n}(Y_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= (F_{t_n} + \frac{s}{2})Y_{t_n} + \lambda(Y_0 + Z_0)Y_{t_n} + \alpha_{n+1}Y^2_{t_n} + \beta_{n+1}Y_{t_n}D_{t_n}e^{-\rho\tau} + \gamma_{n+1}D^2_{t_n}e^{-2\rho\tau}.
\end{aligned}$$

This shows that the optimal value function still has the same form with following coefficients:

$$\alpha_n = \alpha_{n+1},\ \ \beta_n = \beta_{n+1}e^{-\rho\tau},\ \ \gamma_n = \gamma_{n+1}e^{-2\rho\tau},\ \ \delta_{n+1} = 0.$$

□

For case (2), orders do not change and the coefficients satisfy the same recursive equation as given in the propositions.

For case (3), we find that $x_{N-1} = X_{N-1}$. This is optimal for our problem, since it minimizes the optimal value function in t_{N-1} with the choice of x_{N-1}. In all cases the optimal value function is a quadratic function that opens up. For case (3), the angular point is above X_{N-1}, and the optimal value function is increasing with x_{N-1} moving from $X_{t_{N-1}}$ to 0. This leads to $x_{N-1} = X_{N-1}$. The coefficients then follow a different set of recursive equations.

Proposition 4 *When $x_{N-1} > X_{N-1}$, we replace it with $x_{N-1} = X_{N-1}$, and then the coefficients follow the following recursive equations:*

$$\alpha_{N-1} = \frac{1}{2q} - \lambda,\ \ \beta_{N-1} = 1,\ \ \gamma_{N-1} = 0,\ \ \delta_N = 0.$$

Proof 4 At $t_n = N-1$, we encounter the same situation as at t_N, and we have to execute the whole order immediately, which leads to the following optimal value function

$$\begin{aligned}
&J^1_{t_n}(Y_{t_{N-1}}, D_{t_{N-1}}, F_{t_{N-1}}, t_{N-1})\\
&= \left(F_{t_{N-1}} + \frac{s}{2}\right)Y_{t_{N-1}} + \left[\lambda(Y_0 + Z_0 - Y_{t_{N-1}}) + D_{t_{N-1}} + \frac{Y_{t_{N-1}}}{2q}\right]Y_{t_{N-1}}\\
&= \left(F_{t_{N-1}} + \frac{s}{2}\right)Y_{t_{N-1}} + \lambda(Y_0 + Z_0)Y_{t_{N-1}} + Y^2_{t_{N-1}}\left(\frac{1}{2q} - \lambda\right) + D_{t_{N-1}}Y_{t_{N-1}}.
\end{aligned}$$

This shows that the modified optimal value function still has the same form as the previous optimal value function and the coefficients are determined as follows:

$$\alpha_{N-1} = \frac{1}{2q} - \lambda,\ \ \beta_{N-1} = 1,\ \ \gamma_{N-1} = 0,\ \ \delta_N = 0.$$

□

We show that in all cases, the modified optimal value function does not change its original functional form. Only the coefficients are determined differently from their original forms. We can obtain all possible coefficients at each time point before t_{N-1}. We then have 3^{N-1-m} possible combinations of coefficients to complete the trade.

We have the same situation as that between t_0 and t_m. At t_m, we have to complete the first part of the trade. For the coefficients before t_m, we have the following proposition.

Proposition 5 *When $z_{m-1} > Z_{m-1}$, we replace it with $z_{m-1} = Z_{m-1}$, and the coefficients are given by the following recursive equations:*

$$\begin{aligned}
c_{m-1} &= \lambda(Y_0 + Z_0)Y_0 + \alpha_{m+1}Y_0^2, a_{m-1} = \lambda Z_0 + \kappa\beta_{m+1}Y_0 e^{-(m+1-n)\rho\tau},\\
b_{m-1} &= \beta_{m+1}Y_0 e^{-(m+1-n)\rho\tau}, g_{m-1} = 1 + 2\gamma_{m+1}\kappa e^{-2(m+1-n)\rho\tau},\\
h_{m-1} &= \frac{1}{2q} - \lambda + \gamma_{m+1}\kappa^2 e^{-2(m+1-n)\rho\tau}, s_{m-1} = \gamma_{m+1}e^{-2(m+1-n)\rho\tau}.
\end{aligned}$$

When $z_{m-1} < 0$, we replace it with $z_{m-1} = 0$, and then the coefficients follow

$$\begin{aligned}
c_{m-1} &= c_m, a_{m-1} = a_m b_{m-1} = b_m e^{-\rho\tau},\\
h_{m-1} &= h_m, g_{m-1} = g_m e^{-\rho\tau}, s_{m-1} = s_m e^{-2\rho\tau},\\
\delta_{n+1} &= 0.
\end{aligned}$$

Proof 5 For the first case we have to complete the whole order. We have the modified optimal value function for $t_n = t_{m-1}$ as follows:

$$\begin{aligned}
&J_{t_n}^2(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n} + \left[\lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{Z_{t_n}}{2q}\right] Z_{t_n}\\
&\quad + E_{t_n} J^1(Y_0, (D_{t_n} + \kappa Z_{t_n})e^{-(m+1-n)\rho\tau}, F_{t_n}, t_n)\\
&= \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n} + \left[\lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{Z_{t_n}}{2q}\right] Z_{t_n} + \left(F_{t_n} + \frac{s}{2}\right) Y_0 + \lambda(Y_0 + Z_0)Y_0\\
&\quad + \ \ \alpha_{m+1}Y_0^2 + \beta_{m+1}Y_0(D_{t_n} + \kappa Z_{t_n})e^{-(m+1-n)\rho\tau} + \gamma_{m+1}(D_{t_n} + \kappa Z_{t_n})^2 e^{-2(m+1-n)\rho\tau}\\
&= (\lambda(Y_0 + Z_0)Y_0 + \alpha_{m+1}Y_0^2) + \left(F_{t_n} + \frac{s}{2}\right) Y_0 + \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n}\\
&\quad + Z_{t_n}[\lambda Z_0 + \beta_{m+1}Y_0\kappa e^{-(m+1-n)\rho\tau}]\\
&\quad + D_{t_n}[\beta_{m+1}Y_0 e^{-(m+1-n)\rho\tau}] + D_{t_n}Z_{t_n}[1 + 2\gamma_{m+1}\kappa e^{-2(m+1-n)\rho\tau}]\\
&\quad + Z_{t_n}^2\left[-\lambda + \frac{1}{2q} + \gamma_{m+1}\kappa^2 e^{-2(m+1-n)\rho\tau}\right] + D_{t_n}^2[\gamma_{m+1}e^{-2(m+1-n)\rho\tau}],
\end{aligned}$$

with which we can determine the coefficients. For the second case we find that the modified optimal function for $t_n = t_{m-1}$ has the following form

$$J_{t_n}^2(Z_{t_n}, D_{t_n}, F_{t_n}, t_n) = E_{t_n} J^2(Z_{t_n}, (D_{t_n})e^{-\rho\tau}, F_{t_{n+1}}, t_{n+1})$$

$$= c_{n+1} + (F_{t_n} + \frac{s}{2})Y_0 + (F_{t_n} + \frac{s}{2})Z_{t_n} + a_{n+1}Z_{t_n} + b_{n+1}D_{t_n}e^{-\rho\tau} + h_{n+1}Z_{t_n}^2$$
$$+ g_{n+1}Z_{t_n}D_{t_n}e^{-\rho\tau} + s_{n+1}D_{t_n}^2 e^{-2\rho\tau},$$

where we can derive the recursive equations for the coefficients. □

For completing the first part of the trade, we have 3^{m-1} possible combinations of coefficients. With the 3^{N-1-m} sets of coefficients to complete the second part of order, we totally have 3^{N-2} possible strategies. After calculating each of them, we remove the strategies that are not in Φ and keep the one with the lowest expected cost to ensure the optimal solution.

When the underlying price follows the geometric Brownian motion, we obtain the following propositions.

Proposition 6 *When $x_{t_n} = 0$, we find that the optimal value function has the same form with the following coefficients*

$$b_n = b_{n+1},\ c_n = c_{n+1}e^{-2\rho\tau},\ d_n = d_{n+1}m,\ g_n = g_{n+1}e^{-\rho\tau},\ h_n = h_{n+1}a,\ l_n = l_{n+1}e^{-\rho\tau}a,$$

and

$$b_n = b_{n+1},\ c_n = c_{n+1}e^{-2\rho\tau},\ d_n = d_{n+1}m,\ g_n = g_{n+1}e^{-\rho\tau},\ h_n = h_{n+1}a,\ l_n = l_{n+1}e^{-\rho\tau}a,$$
$$s_n = s_{n+1},\ u_n = u_{n+1}e^{-\rho\tau},\ w_n = w_{n+1}a,\ o_n = o_{n+1}.$$

Proof 6 When $n > m$:

$$J_{t_n}^1(Y_{t_n}, D_{t_n}, F_{t_n}, t_n) = (a \times a^{N-(n+1)}F_n + \frac{s}{2})Y_{t_n} + \lambda(Z_0 + Y_0)Y_{t_n} + b_{n+1}Y_{t_n}^2$$
$$+ c_{n+1}D_{t_n}^2 e^{-2\rho\tau} + d_{n+1}mF_{t_n}^2 + g_{n+1}Y_{t_n}D_{t_n}e^{-\rho\tau}$$
$$+ h_{n+1}Y_{t_n}aF_{t_n} + l_{n+1}D_{t_n}e^{-\rho\tau}aF_{t_n},$$

which leads to the coefficients. When $n < m$:

$$J_{t_n}^2(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)$$
$$= (a^{m-n-1}aF_n + \frac{s}{2})Z_{t_n} + (a^{N-n}F_n + \frac{s}{2})Y_0 + b_{n+1}Z_{t_n}^2 + c_{n+1}D_{t_n}^2 e^{-2\rho\tau}$$
$$+ d_{n+1}mF_{t_n}^2 + g_{n+1}Z_{t_n}D_{t_n}e^{-\rho\tau}$$
$$+ h_{n+1}Z_{t_n}aF_{t_n} + l_{n+1}D_{t_n}e^{-\rho\tau}aF_{t_n} + s_{n+1}Z_{t_n} + u_{n+1}D_{t_n}e^{-\rho\tau}$$
$$+ w_{n+1}aF_{t_n} + o_{n+1}.$$

□

Proposition 7 *When $x_{t_n} = X_{t_n}$, we find that the optimal value function has the same form with the following coefficients*

$$b_N = \frac{1}{2q} - \lambda, c_N = 0, d_N = 0, g_N = 1, h_N = 0, l_N = 0,$$

and

$$b_m = -\lambda + \frac{1}{2q} + c_{m+1}\kappa^2 e^{-2(m+1-n)\rho\tau},\ c_m = c_{m+1}e^{-2(m+1-n)\rho\tau},\ d_m = d_{m+1}m^{(m+1-n)},$$
$$g_m = 1 + 2\kappa c_{m+1}e^{-2(m+1-n)\rho\tau},\ h_m = l_{m+1}a^{(m+1-n)}\kappa e^{-(m+1-n)\rho\tau},$$
$$l_m = l_{m+1}a^{(m+1-n)}e^{-(m+1-n)\rho\tau},\ s_m = \lambda Z_0 + g_{m+1}\kappa e^{-(m+1-n)\rho\tau}Y_0,\ u_m = g_{m+1}Y_0e^{-(m+1-n)\rho\tau},$$
$$w_m = h_{m+1}Y_0a^{(m+1-n)},\ o_m = \lambda(Z_0 + Y_0)Y_0 + b_{m+1}Y_0^2.$$

Proof 7 When $n > m$:

$$J^1_{t_n}(Y_{t_n}, D_{t_n}, F_{t_n}, t_n) = [(F_{t_n} + \frac{s}{2}) + \lambda(Z_0 + Y_0 - Y_{t_n}) + D_{t_n} + \frac{Y_{t_n}}{2q}]Y_{t_n},$$

which leads to the coefficients. When $n < m$:

$$\begin{aligned}
&J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{Z_{t_n}}{2q}\right] Z_{t_n} + E_{t_n} J^1_{m+1}(Y_0, (D_{t_n} + \kappa Z_{t_n})e^{-(m+1-n)\rho\tau}, F_{t_n}, t_n)\\
&= \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n} + (a^{N-n} F_n + \frac{s}{2})Y_0 + Z^2_{t_n}\left(-\lambda + \frac{1}{2q} + c_{m+1}\kappa^2 e^{-2(m+1-n)\rho\tau}\right)\\
&\quad + D^2_{t_n}(c_{m+1}e^{-2(m+1-n)\rho\tau}) + F^2_{t_n}(d_{m+1}m^{m+1-n}) + F_{t_n} D_{t_n}(l_{m+1}e^{-(m+1-n)\rho\tau} a^{m+1-n})\\
&\quad + Z_{t_n} D_{t_n}(1 + c_{m+1}\kappa e^{-2(m+1-n)\rho\tau}) + Z_{t_n} F_{t_n}(l_{m+1}\kappa e^{-(m+1-n)\rho\tau} a^{m+1-n})\\
&\quad + Z_{t_n}(\lambda Z_0 + g_{m+1} Y_0 \kappa e^{-(m+1-n)\rho\tau}) + F_{t_n}(h_{m+1} Y_0 a^{m+1-n})\\
&\quad + D_{t_n}(g_{m+1} Y_0 e^{-(m+1-n)\rho\tau}) + (\lambda(Z_0 + Y_0)Y_0 + b_{m+1} Y_0^2).
\end{aligned}$$

□

4 Simulation study

In this section we evaluate the model performance by comparing simulations run over a set of numerical experiments. In order to show the importance of introducing the participation rate π for an optimal order submission strategy, we compare the performance of our strategy with two alternative strategies: the strategy proposed by Obizhaeva and Wang (2013) and the naive trading strategy that splits the large order into equally small pieces. We use the volume weighted average price (VWAP) as a benchmark to measure the average execution cost of these trading strategies (see, for example Werner (2003) and Goldstein et al. (2009)). Only the trading strategy with the lowest average execution cost is preferred.

4.1 Numerical examples

We separate the whole trading period into two parts and use the following parameters: $q = 5000$, $\lambda = 1/2q$, $\kappa = 1/q - \lambda$, $\rho = 2.2$, $N = 9$, $X_0 = 100{,}000$, $A_0 = 100$, and $T = 1$, which are used in Obizhaeva and Wang (2013). Table 1 illustrates one example of order behavior for our optimal order submission strategy with different participation rates when the underlying asset price follows the Brownian motion. We also show the graphic presentation in Fig. 1. In Table 2 we show the optimal order behavior of our model when the underlying asset price follows a geometric Brownian motion with $\mu = 3\%$. Figure 2 provides the corresponding graphic presentation. As we have shown that σ does not influence the optimal solution, when the underlying asset price follows the geometric Brownian motion, only μ can influence the trading strategy. Table 3 shows the change of optimal order behavior along with the change of μ when we set the participation rate at 75%. We show it graphically in Fig. 3.

4.2 Results

In Fig. 1, the trader splits the trading period into two parts (i.e., sub-time intervals) when the market price follows the Brownian motion. When the target participation rate is 50% (trade

Table 1 Optimal order sizes of the S-K strategy for different participation rates when the underlying asset price F_T follows a Brownian motion

Participation rate π (%)	0	1	2	3	4	5	6	7	8	9
10	10000	0	0	0	0	32120	7604	7604	7604	35066
20	19758	242	0	0	0	26501	7029	7029	7029	32412
30	23054	5000	1946	0	0	20427	6513	6513	6513	30034
40	25421	5513	5513	3553	0	13676	6086	6086	6086	28065
50	26775	5806	5806	5806	5806	5806	5806	5806	5806	26775
60	28065	6086	6086	6086	13676	0	3553	5513	5513	25421
70	30034	6513	6513	6513	20427	0	0	1946	5000	23054
80	32412	7029	7029	7029	26501	0	0	0	242	19758
90	35066	7604	7604	7604	32120	0	0	0	0	10000

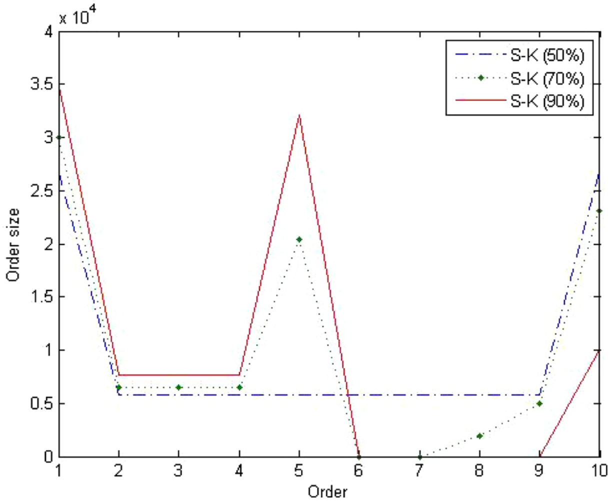

Fig. 1 Optimal order behavior of the S-K strategy ($t_n \in \{t_0, \ldots, t_m\}$) for different participation rate when the underlying asset price F_T follows a Brownian motion

the same volume in the two consecutive trading periods) which coincides with Obizhaeva and Wang (2013), the trading strategy illustrates a U-shape. When the target participation rate is 70% or 90% (trade more first, then trade less, i.e., 70% or 90% in the first interval and then 30% or 10% in the second interval), the solution is a U-shaped strategy in the first interval and then a J-shaped strategy.

In Fig. 2, the trader splits the trading period into two sub-time intervals and the market price follows a geometric Brownian motion with drift $\mu = 3\%$, where the Obizhaeva and Wang (2013) model does not work. When the target participation rate is 50% (trade equally in the two consecutive trading periods), the solution is an L-shaped trading strategy in the first interval then a U-shaped strategy in the second interval. When the target participation rate is 70%, we have an L-shaped strategy in the first interval and a J-shaped strategy in the second, which looks like a U-shaped strategy over the two trading periods. When the target participation rate is 90% the solution is a U-shaped strategy in the first interval and then a J-shaped strategy.

In Fig. 3, the trader splits the trading period into two sub-time intervals as well and the market price follows the geometric Brownian motion with different drift μ, where the Obizhaeva and Wang (2013) model does not work either. When the target participation rate is fixed to 75% (trade more in the first interval and less in the second), the optimal solution depends on the drift, i.e., the volatility. When the drift $\mu = 1\%$, we observe a U-shaped strategy in the first sub-time interval and a J-shaped strategy in the second. When the drift $\mu = 3\%$, we observe an L-shaped strategy in the first sub-time interval and a J-shaped in the second, which is a U-shaped strategy for the whole trading period. When the drift $\mu = 5\%$, we observe an L-shaped strategy in the first sub-time interval and a transposed I-shaped strategy in the second, which is overall an L-shaped strategy for the whole trading period.

The Volker Rule requires that the trading positions of market makers should react to the near-term demand. In our model, this is determined by the target participation rate in order to optimize the trading strategy. We can see that in each sub-time interval, the trading strategy remains smooth under either a U-, L-, or J-shaped strategy. Actually, the temporary

Table 2 Optimal order sizes of the S-K strategy for different participation rates when the underlying asset price F_T follows a geometric Brownian motion with $\mu = 3\%$

Participation rate π (%)	0	1	2	3	4	5	6	7	8	9
10	10000	0	0	0	0	42660	8037	7629	7220	24454
20	20000	0	0	0	0	37053	7460	7052	6642	21793
30	30000	0	0	0	0	31446	6883	6475	6065	19132
40	35693	4307	0	0	0	25620	6334	5926	5517	16603
50	38662	6561	4776	0	0	19346	5845	5437	5027	14345
60	40814	7028	6627	5532	0	12403	5443	5035	4626	12493
70	42003	7286	6885	6482	7345	4330	5190	4782	4373	11326
80	43412	7591	7190	6787	15019	0	1686	4454	4045	9815
90	45522	8049	7648	7245	21537	0	0	0	2840	7160

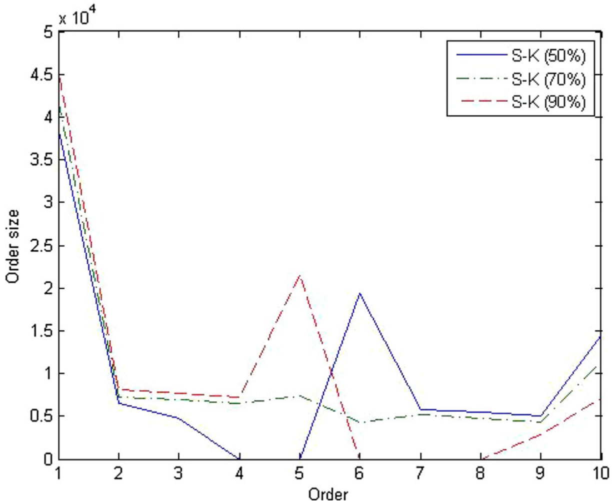

Fig. 2 Optimal order behavior of the S-K strategy ($t_n \in \{t_0, \ldots, t_m\}$) for different participation rate when the underlying asset price F_T follows a geometric Brownian motion with $\mu = 3\%$

Table 3 Optimal order sizes of the S-K strategy when the participation rate π=75% and underlying asset price F_T follows the geometric Brownian motion with different μ

Drift μ (%)	0	1	2	3	4	5	6	7	8	9
− 10	0	969	6932	8312	58787	0	0	0	0	25000
− 5	13237	6123	6810	7493	41337	0	0	0	0	25000
− 3	20459	6355	6764	7172	34251	0	0	0	0	25000
− 1	27579	6609	6745	6880	27187	0	0	0	497	24503
1	34848	6939	6805	6670	19738	0	0	2145	4590	18266
3	42553	7405	7004	6601	11437	238	5071	4663	4253	10776
5	51120	8097	7433	6765	1584	10198	5010	4327	3641	1823
10	66748	8252	0	0	0	25000	0	0	0	0

price impact factor determines this trading behavior in our model and reflects the order intensity that our model considers to reduce the transaction cost. The O-W model assumes the temporary price impact remains constant. In our model, when the target participation rate is high, the temporary price impact turns out to be higher in the first sub-time interval. The trader then acts more passively in the beginning of the second sub-time interval such that the limit order book can be replenished and the temporary price impact declines. In addition, when the target participation rate is fixed at certain level, the drift μ which infers the market volatility in our model, determines the optimal trading strategy.

We use the parameters mentioned above and run the simulation for 10,000 and 100,000 times. We report our results in Tables 4 and 5 respectively. Figures 4, 5 and 6 illustrate the respective individual performance of these optimal trading strategies under the average execution cost (i.e., the VWAP) measure. For different participation rates, we see that our strategy is significantly superior to both alternative strategies when the underlying price movement follows the Brownian motion (i.e., $\mu = 0$). We illustrate this in Fig. 7. When the

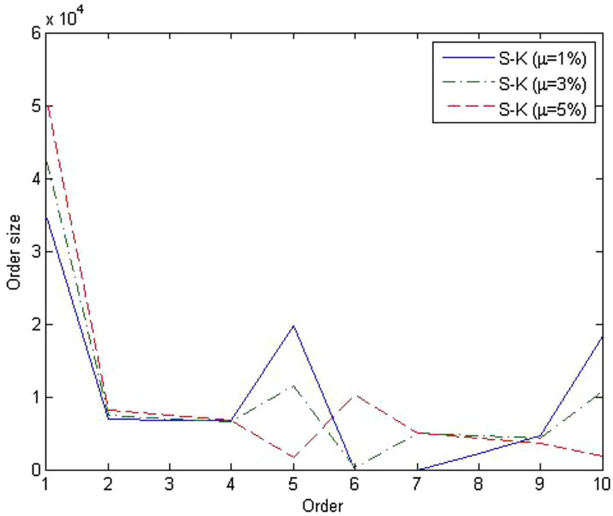

Fig. 3 Optimal order behavior of the S-K strategy ($t_n \in \{t_0, \ldots, t_m\}$) for the participation rate of 75% when the underlying price follows the geometric Brownian motion with different μ

underlying price movement follows the geometric Brownian motion, from Tables 4 and 5 we can see that our strategy is still significantly superior to the alternatives. Figure 8 illustrates this comparison when the underlying price follows the geometric Brownian motion (when $\mu = 3\%$).

We have shown that the performance of our optimal order submission strategy is influenced by the participation rate π and underlying asset price dynamics characterized by the drift μ. We now compare the performance of our trading strategy with two alternatives under the average execution cost measure. As we have already learned from Tables 4 and 5, our optimal strategy is significantly better than the alternatives, and we thus focus on how exactly our strategy can reduce the execution cost. We compute the difference (i.e., the reduced execution cost) between our strategy and the alternatives. The reduced execution cost is the average transaction cost of the alternative strategy minus that of our strategy. The more the reduced execution cost is, the better the performance of our strategy is versus the alternatives. When the reduced execution cost is equal to zero, there is no difference between our strategy and the alternatives.

We illustrate the results in Figs. 9 and 10. From Fig. 9, we can easily identify that our trading strategy can reduce the execution cost to a greater extent in comparison with the naive trading strategy, which agrees with our previous conclusion. From Fig. 10, we can see that, when the participation rate π=50% and drift μ=0 (i.e., the underlying asset price follows the Brownian motion), there is no difference between our optimal strategy and the O-W strategy (which we mentioned in Sect. 2). Except for this, our optimal strategy performs significantly better than the O-W strategy for reducing the execution cost.

The O-W strategy implies a U-shaped trading pattern, but our solutions have shown that a U-shaped pattern turns out not to be always optimal. In practice, traders are inclined towards completing of their positions before the market closes without holding an open position overnight. Active traders typically trade at known and preferred prices throughout the trading period, as we shown in our model, rather than wait until approaching to the end of the day when prices are still unknown. Kissell (2014) mentions that there are dramatic increases in

Table 4 Comparison of strategies using the average execution cost (VWAP) and its variance (shown in parenthesis) for 10,000 runs for different values of participation rate (π) and drift (μ)

Drift	Participation rate	$\pi = 50\%$	$\pi = 60\%$	$\pi = 70\%$	$\pi = 80\%$	$\pi = 90\%$
$\mu = -10\%$	S-K (GBM)	100.9370 (1.8289)	101.5190 (1.5908)	102.2697 (1.4781)	103.1296 (1.3034)	104.1049 (1.2583)
	O-W	102.4957 (0.9826)	103.0990 (0.8356)	103.7300 (0.6860)	104.4586 (0.5726)	105.2298 (0.4726)
	Naive	102.5154 (1.0203)	103.1185 (0.8696)	103.7715 (0.7394)	104.5217 (0.6300)	105.3346 (0.5373)
$\mu = -5\%$	S-K (GBM)	104.4489 (1.3890)	104.6930 (1.2374)	105.1192 (1.0242)	105.6181 (0.8720)	106.2527 (0.7471)
	O-W	105.0454 (0.9596)	105.3416 (0.7944)	105.7333 (0.6490)	106.1744 (0.5403)	106.6799 (0.4490)
	Naive	105.0737 (0.9863)	105.3820 (0.8310)	105.7733 (0.6798)	106.2335 (0.5926)	106.7702 (0.4861)
$\mu = -1\%$	S-K (GBM)	106.8852 (0.9582)	106.8958 (0.8295)	107.0534 (0.6959)	107.2933 (0.5803)	107.6448 (0.4754)
	O-W	107.0730 (0.9382)	107.1552 (0.7766)	107.3162 (0.6385)	107.5195 (0.5097)	107.7942 (0.4194)
	Naive	107.0973 (0.9406)	107.1933 (0.8191)	107.3593 (0.6558)	107.5863 (0.5470)	107.9070 (0.4738)
$\mu = 0\%$	S-K (GBM)	107.3889 (0.8037)	107.4066 (0.7254)	107.4934 (0.6139)	107.6645 (0.5138)	107.9519 (0.4047)
	O-W	107.5654 (0.9397)	107.6064 (0.7630)	107.6931 (0.6227)	107.8473 (0.5177)	108.0817 (0.4151)
	Naive	107.5973 (0.9456)	107.6129 (0.7895)	107.7494 (0.6687)	107.9374 (0.5570)	108.1841 (0.4621)
$\mu = 1\%$	S-K (GBM)	107.8541 (0.7008)	107.8685 (0.6285)	107.8886 (0.5413)	108.0126 (0.4481)	108.2448 (0.3628)
	O-W	108.0720 (0.9015)	108.0436 (0.7600)	108.0849 (0.6297)	108.1940 (0.5052)	108.3516 (0.4081)
	Naive	108.0966 (0.9410)	108.0793 (0.7675)	108.1187 (0.6550)	108.2632 (0.5323)	108.4686 (0.4610)
$\mu = 5\%$	S-K (BM)	109.4454 (0.4463)	109.2735 (0.3215)	109.1858 (0.2514)	109.1713 (0.2039)	109.1837 (0.1719)
	O-W	110.0689 (0.9119)	109.8114 (0.7334)	109.6332 (0.6028)	109.5086 (0.4927)	109.4492 (0.3894)
	Naive	110.0632 (0.9370)	109.8413 (0.7620)	109.6574 (0.6381)	109.5890 (0.5254)	109.5461 (0.4381)
$\mu = 10\%$	S-K (GBM)	110.9511 (0.3828)	110.4792 (0.2417)	110.1468 (0.1478)	109.9229 (0.0837)	109.8126 (0.0492)
	O-W	112.4987 (0.8620)	111.9782 (0.7007)	111.5139 (0.5836)	111.1257 (0.4688)	110.7988 (0.3778)
	Naive	112.5150 (0.8836)	112.0028 (0.7338)	111.5549 (0.6045)	111.1893 (0.5013)	110.9003 (0.4043)

Table 5 Comparison of strategies using the average execution cost (VWAP) and its variance (shown in parenthesis) for 100,000 runs for different values of participation rate (π) and drift (μ)

Drift	Participation rate	$\pi = 50\%$	$\pi = 60\%$	$\pi = 70\%$	$\pi = 80\%$	$\pi = 90\%$
$\mu = -10\%$	S-K (GBM)	100.9382 (1.7988)	101.5355 (1.6066)	102.2718 (1.4573)	103.1341 (1.3218)	104.1218 (1.2285)
	O-W	102.4958 (0.9919)	103.0878 (0.8304)	103.7325 (0.6845)	104.4492 (0.5679)	105.2295 (0.4691)
	Naive	102.5148 (1.0181)	103.1070 (0.8599)	103.7717 (0.7259)	104.5143 (0.6209)	105.3330 (0.5294)
$\mu = -5\%$	S-K (GBM)	104.4498 (1.4195)	104.7173 (1.2050)	105.1076 (1.0186)	105.6265 (0.8604)	106.2615 (0.7475)
	O-W	105.0504 (0.9572)	105.3628 (0.7927)	105.7329 (0.6536)	106.1671 (0.5420)	106.6705 (0.4437)
	Naive	105.0698 (0.9835)	105.3858 (0.8253)	105.7785 (0.6950)	106.2396 (0.5825)	106.7795 (0.4951)
$\mu = -1\%$	S-K (GBM)	106.8689 (0.9533)	106.9173 (0.8377)	107.0515 (0.6981)	107.2939 (0.5666)	107.6489 (0.4610)
	O-W	107.0679 (0.9246)	107.1604 (0.7722)	107.3068 (0.6374)	107.5176 (0.5147)	107.7971 (0.4257)
	Naive	107.0866 (0.9642)	107.1808 (0.7982)	107.3500 (0.6703)	107.5879 (0.5578)	107.9139 (0.4723)
μ =0%	S-K (BM)	107.3843 (0.8243)	107.4067 (0.7329)	107.4944 (0.6257)	107.6619 (0.5085)	107.9476 (0.4072)
	O-W	107.5689 (0.9312)	107.5963 (0.7656)	107.6916 (0.6339)	107.8554 (0.5138)	108.0757 (0.4182)
	Naive	107.5879 (0.9578)	107.6272 (0.7953)	107.7425 (0.6586)	107.9283 (0.5542)	108.1897 (0.4694)
μ =1%	S-K (GBM)	107.8670 (0.6924)	107.8607 (0.6356)	107.9049 (0.5480)	108.0203 (0.4472)	108.2341 (0.3565)
	O-W	108.0645 (0.9138)	108.0383 (0.7617)	108.0827 (0.6271)	108.1846 (0.5090)	108.3510 (0.4139)
	Naive	108.0873 (0.9429)	108.0756 (0.7939)	108.1304 (0.6542)	108.2589 (0.5488)	108.4652 (0.4578)
μ =5%	S-K (GBM)	109.4454 (0.4516)	109.2752 (0.3249)	109.1888 (0.2435)	109.1675 (0.2025)	109.1831 (0.1735)
	O-W	110.0460 (0.8939)	109.8036 (0.7357)	109.6234 (0.6070)	109.5065 (0.4927)	109.4462 (0.3986)
	Naive	110.0765 (0.9209)	109.8285 (0.7646)	109.6700 (0.6369)	109.5747 (0.5218)	109.5543 (0.4402)
μ =10%	S-K (GBM)	110.9439 (0.3828)	110.4777 (0.2454)	110.1422 (0.1468)	109.9278 (0.0836)	109.8110 (0.0487)
	O-W	112.5001 (0.8730)	111.9763 (0.7190)	111.5232 (0.5824)	111.1227 (0.4734)	110.7935 (0.3803)
	Naive	112.5281 (0.8928)	112.0031 (0.7388)	111.5596 (0.6109)	111.1918 (0.4998)	110.8979 (0.4170)

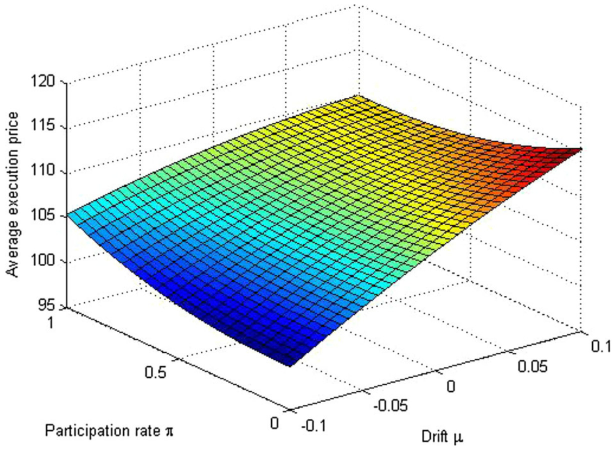

Fig. 4 Performance of the S-K strategy under an average execution cost (VWAP) measure for different values of the drift (μ) and participation rate (π)

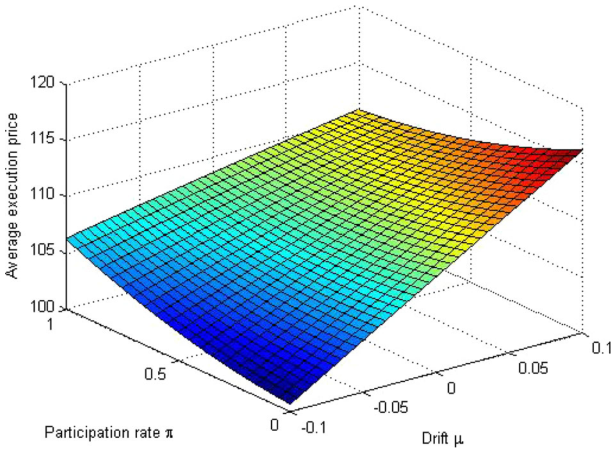

Fig. 5 Performance of the Naive strategy under an average execution cost (VWAP) measure for different values of the drift (μ) and participation rate (π)

exchange traded funds (ETFs) that gain certain market exposure or hedge very short-term risk, thus shift their trading volume profile.

Our results from the proposed constrained model can also significantly improve the trading performance compared with Obizhaeva and Wang (2013), and our model reduces the transaction cost particularly when the market is volatile. It could perform much better when we conduct an L-shaped trading (i.e., trade more, then trade less) or a J-shaped trading (trade less, then trade more) strategy when market volatility is high e.g., the impact of μ in geometric Brownian motion.

4.3 Discussion and limitations

Risk-neutral probabilities are used in our study to describe objective fair prices for the underlying asset. We treat price dynamics exogenously by taking out the risk. The risk-neutral

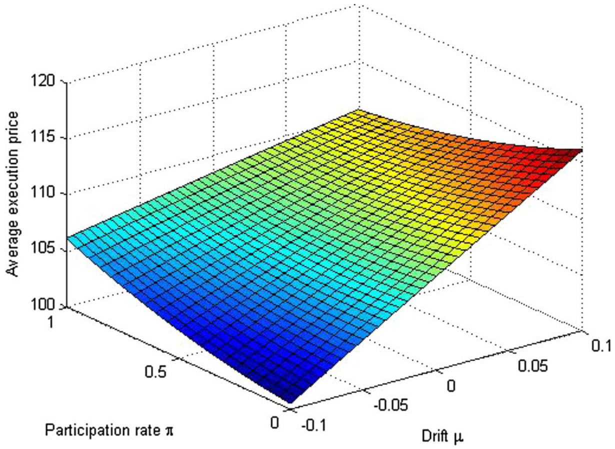

Fig. 6 Performance of the O-W strategy under an average execution cost (VWAP) measure for different values of the drift (μ) and participation rate (π)

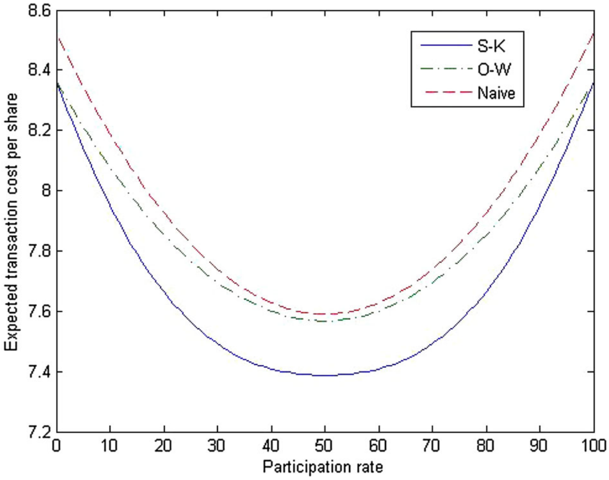

Fig. 7 Expected transaction cost of different trading strategies with different participation rates when the underlying price follows a Brownian motion

trader cares only about the expected value but not the uncertainty of the total cost. The two price dynamics are Brownian motion and geometric Brownian motion, which makes the model feasibly easy. By contrast, if one attempts to estimate the anticipated value of a particular asset based on its trend (i.e., how likely the market goes up or down), considering unique factors or market conditions that influence the underlying asset, risk should be included and thus actual real-world or physical probability shall be considered. The benefit with risk-neutrality is that once the risk-neutral probabilities are calculated, they can be used to price every asset based on its expected payoff. These theoretical risk-neutral probabilities differ from real probabilities. When using the real-world probabilities, expected values of each security would need to be adjusted for its individual risk profile. As a consequence, the trading strategy might be influenced by the individual risk profile.

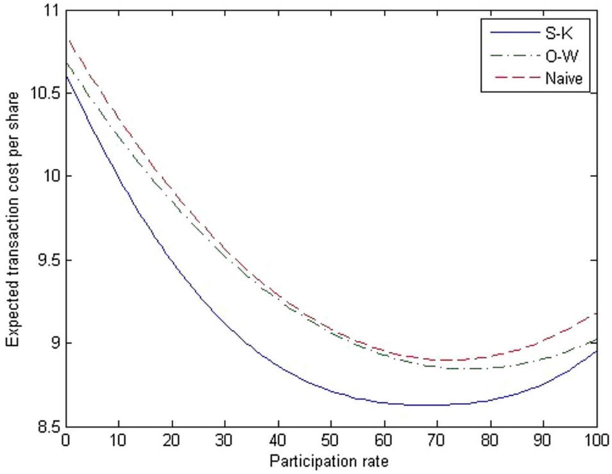

Fig. 8 Expected transaction cost of different trading strategies with different participation rates when the underlying price follows a geometric Brownian motion with $\mu = 3\%$

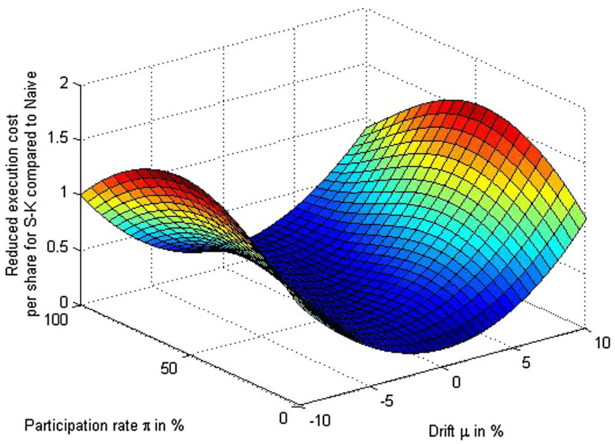

Fig. 9 Reduced execution cost of the S-K strategy comparing with the Naive strategy under an average execution cost (VWAP) measure for different values of (μ) and participation rates (π)

Brownian motion and geometric Brownian motion have been used for the underlying price dynamics. Our model works well with both in comparison with alternative trading strategy by Obizhaeva and Wang (2013). Fluctuations vary in real asset prices over time, but in geometric Brownian motion, fluctuations do not change over time. In our model, μ characterizes the fluctuation which leads to different trading strategies. Risk-neutrality enervates the influence of σ for both Brownian motion and geometric Brownian motion when considering the optimal execution problem. The assumption of risk-neutrality can be released to risk-aversion by introducing a risk-aversion coefficient to the tractable mean-variance objective function. Obizhaeva and Wang (2013) pointed out that the nature of the optimal execution strategy remains qualitatively similar between a risk-averse trader and a risk-neutral one. The risk-aversion coefficient is positively correlated to the aggressiveness of trades. A risk-averse trader will place discrete trades at the beginning and at the end of

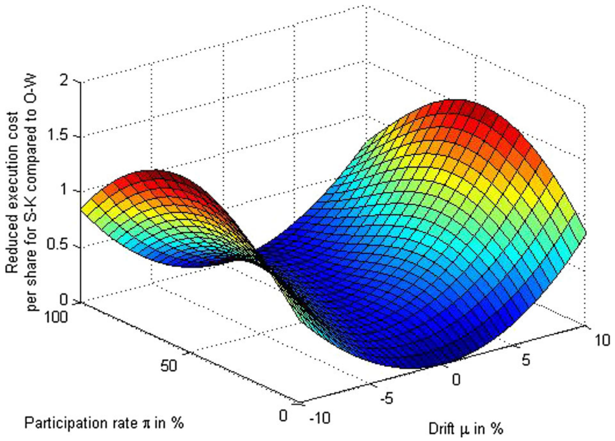

Fig. 10 Reduced execution cost of the S-K strategy comparing with the O-W strategy under an average execution cost (VWAP) measure for different values of (μ) and participation rates (π)

trading period and trade continuously in between. The initial and final discrete trades are, however, of different magnitude. The more risk averse the trader is, the faster he wants to execute his order to avoid future uncertainty and the more aggressive orders he submits in the beginning.

In our work, we only consider buy orders. This model can be extended for selling. To enable both buying and selling for the same asset we can simply remove the condition of $x_{t_n} \geq 0$ for all n from the set of feasible strategies to enable purchases and sales in the same strategy. If the trader's objective changes from purchasing a total amount of X_0 shares to selling X_0 shares, we have to change the restriction from $X_0 > 0$ to $X_0 < 0$ and replace the condition of $x_{t_n} \geq 0$ for the set of feasible strategies with $x_{t_n} \leq 0$.

We have worked only on trading one asset. This model can be modified to trade a portfolio of several assets instead of a single one by replacing x and X with vectors. For our model similar ways can be worked out as discussed above (i.e. buy/sell in the same strategy and either buy or sell) for the vectors. When the price dynamics for each asset would be measured with risk-neutrality, a same solution will be obtained for each component in the vector. In case $x(i)$ has the price impact on $x(j)\ \forall\, i \neq j;\, i, j \in T$, the decision of the optimal strategy becomes more complicated. This is an interesting aspect and shall be addressed in another study.

5 Conclusions

In this paper we consider an optimal order submission strategy under new regulatory requirement of the near-term demand when supplying liquidity. We build a model based on the discrete time optimal control theory, like other studies documented in the literature, in order to fix the number of trades at a predetermined horizon. In our model, we introduce the target participation rate as a variable to characterize the trading style (aggressive, neutral, or passive) in response to customers' near-term demand for liquidity. We assume the underlying asset price follows a Brownian motion and geometric Brownian motion, which makes our model much more flexible in search of optimal solutions under different market situations. We adopt the linear price impact function in our model–that is, characterizing a linear combination of the permanent and temporary price impacts. We derive the analytical solutions for the optimal order submission strategy and assess its performance with simulations run over a set of experiments in comparison with two alternative strategies (i.e., the naive trading strategy and strategy suggested by Obizhaeva and Wang (2013)). We show that under the measure of average execution cost (i.e., VWAP), our optimal trading strategy performs significantly better than the alternative trading strategies. In practice, our model can be easily extended and implemented in any order driven market for trading a large amount of assets.

Both models, i.e., the Obizhaeva and Wang (2013) model and our model, are under the same framework originally from Bertsimas and Lo (1998), such that the optimal trading with respect to price impact is considered. Keeping other conditions constant, we obtain the same solutions as the model given by Obizhaeva and Wang (2013) if we halve the trading interval and set the participation rate at 50%. This means that we can set the constraints for our solution to coincide with the solution by Obizhaeva and Wang (2013) in the unconstrained case. The motivation for the enhancement of the Obizhaeva and Wang (2013) model is to consider the intensity of order flows in order to meet the expected near-term demand as required by the Volker Rules of the Dodd-Frank Act. Therefore, Obizhaeva and Wang (2013) is a special case of our model. To show the significance of this modification in deriving the optimal solution, we have compared the performances of the trading strategies under various settings (in Sect. 4.2). In the case when the trader tries to implement a (predetermined) target participation rate based on his expected volume intensity, for example, he wants to execute 70% of his total position in the morning and 30% in the afternoon. The performance based on the unconstrained solution from Obizhaeva and Wang (2013) is worse than the performance based on the constrained solution given by our model. Particularly, our results from the proposed constrained model can significantly improve the trading performance versus that of Obizhaeva and Wang (2013), and our model reduces the transaction costs, especially when the market is volatile.

Further extensions of our proposed model can be considered in several ways. We consider a multi-period trading strategy with two or more target participation rates. For example, if a trader intends to trade in three consecutive trading periods, he will then determine the participation rate to meet the near-term customers' demand as π_1, π_2, and $1 - \pi_1 - \pi_2$ in the first, second, and third trading intervals respectively. Another direction to extend our model is to consider a functional variable for bid-ask spread and market impact. In addition, the participation rate in our model is a subject probability, one can extend it to a function. This model could be applied in different markets. For example, when trading electricity, the participate rate could be considered as a conditional probability given weather information.

Appendix

A. Proof of Proposition 1

For $t_n = t_N$, the optimal value function has the form

$$J^1_{t_N}(Y_T, D_T, F_T, T) = (F_T + \frac{s}{2})Y_T + [\lambda(Y_0 + Z_0 - Y_T) + D_T + \frac{Y_T}{2q}]Y_T,$$

which satisfies Eq. 2 with the coefficients given by Eq. 4.

Given Eqs. 1– 4 hold for $t_{n+1} \in \{t_{m+2}, \ldots, t_N\}$, we have the optimal value function for t_n

$$
\begin{aligned}
&J^1_{t_n}(Y_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \min_{y_{t_n}} \left\{ \left[\left(F_{t_n} + \frac{s}{2} \right) + \lambda(Y_0 + Z_0 - Y_{t_n}) + D_{t_n} + \frac{y_{t_n}}{2q} \right] y_{t_n} \right.\\
&\quad \left. + E_{t_n} J^1_{t_{n+1}}(Y_{t_n} - y_{t_n}, (D_{t_n} + \kappa y_{t_n})e^{-\rho\tau}, F_{t_{n+1}}, t_{n+1}) \right\}\\
&= \min_{y_{t_n}} \left\{ \left[\left(F_{t_n} + \frac{s}{2} \right) + \lambda(Y_0 + Z_0 - Y_{t_n}) + D_{t_n} + \frac{y_{t_n}}{2q} \right] y_{t_n} + (F_{t_n} + \frac{s}{2})(Y_{t_n} - y_{t_n}) \right.\\
&\quad + \lambda(Y_0 + Z_0)(Y_{t_n} - y_{t_n}) + \alpha_{n+1}(Y_{t_n} - y_{t_n})^2 + \beta_{n+1}(Y_{t_n} - y_{t_n})(D_{t_n} + \kappa y_{t_n})e^{-\rho\tau}\\
&\quad \left. + \gamma_{n+1}(D_{t_n} + \kappa y_{t_n})^2 e^{-2\rho\tau} \right\}.
\end{aligned} \tag{15}
$$

We obtain the optimal value for y_{t_n} from Equation 15, which is

$$
y_{t_n} = -\frac{1}{2}\delta_{n+1}(D_{t_n}(1 - \beta_{n+1}e^{-\rho\tau} + \gamma_{n+1}2\kappa e^{-2\rho\tau}) - Y_{t_n}(\lambda + 2\alpha_{n+1} - \kappa\beta_{n+1}e^{-\rho\tau})), \tag{16}
$$

with

$$
\delta_{n+1} = \left(\frac{1}{2q} + \alpha_{n+1} - \beta_{n+1}\kappa e^{-\rho\tau} + \gamma_{n+1}\kappa^2 e^{-2\rho\tau} \right)^{-1}.
$$

Inserting Eq. 16 into Eq. 15, we obtain the results given by Eqs. 1– 4 which are also valid for t_n, and for all $t_n \in \{t_{m+1}, \ldots, t_N\}$. Then the optimal value function for $t_n = t_m$, called J^2, has the following form

$$
\begin{aligned}
&J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \left(F_{t_n} + \frac{s}{2} \right) Z_{t_n} + \left[\lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{Z_{t_n}}{2q} \right] Z_{t_n} + E_{t_n} J^1_{t_{n+1}}(Y_0, (D_{t_n} + \kappa Z_{t_n})e^{-\rho\tau}, F_{t_n}, t_n)\\
&= \left(F_{t_n} + \frac{s}{2} \right) Z_{t_n} + \left[\lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{Z_{t_n}}{2q} \right] Z_{t_n} + \left(F_{t_n} + \frac{s}{2} \right) Y_0 + \lambda(Y_0 + Z_0)Y_0\\
&\quad + \alpha_{n+1}Y_0^2 + \beta_{n+1}Y_0(D_{t_n} + \kappa Z_{t_n})e^{-\rho\tau} + \gamma_{n+1}(D_{t_n} + \kappa Z_{t_n})^2 e^{-2\rho\tau}\\
&= \left(\lambda(Y_0 + Z_0)Y_0 + \alpha_{n+1}Y_0^2 \right) + \left(F_{t_n} + \frac{s}{2} \right) Y_0 + \left(F_{t_n} + \frac{s}{2} \right) Z_{t_n} + Z_{t_n}[\lambda Z_0 + \beta_{n+1}Y_0\kappa e^{-\rho\tau}]\\
&\quad + D_{t_n}[\beta_{n+1}Y_0 e^{-\rho\tau}] + D_{t_n}Z_{t_n}[1 + 2\gamma_{n+1}\kappa e^{-2\rho\tau}] + Z^2_{t_n}[-\lambda + \frac{1}{2q} + \gamma_{n+1}\kappa^2 e^{-2\rho\tau}]\\
&\quad + D^2_{t_n}[\gamma_{n+1}e^{-2\rho\tau}].
\end{aligned} \tag{17}
$$

We then show that for all $t_n \in \{t_0, \ldots, t_m\}$ the optimal value function has the form of Eq. 5 with coefficients given by Eqs. 6 and 7.

We have shown that the Eq. 5 is true for $t_n = t_m$. Keeping same assumption for t_{n+1}, we will show that it is still valid for each $t_n \in \{t_0, \ldots, t_{m-1}\}$. We find

$$
\begin{aligned}
&J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \min_{z_{t_n}} \left\{ \left[\left(F_{t_n} + \frac{s}{2} \right) + \lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{z_{t_n}}{2q} \right] z_{t_n} \right.\\
&\quad \left. + E_{t_n} J^2_{t_{n+1}}(Z_{t_n} - z_{t_n}, (D_{t_n} + \kappa z_{t_n})e^{-\rho\tau}, F_{t_{n+1}}, t_{n+1}) \right\}\\
&= \min_{z_{t_n}} \left\{ \left[\left(F_{t_n} + \frac{s}{2} \right) + \lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{z_{t_n}}{2q} \right] z_{t_n} + c_{n+1} + \left(F_{t_n} + \frac{s}{2} \right) Y_0 \right.
\end{aligned}
$$

$$+ \left(F_{t_n} + \frac{s}{2}\right)(Z_{t_n} - z_{t-n}) + a_{n+1}(Z_{t_n} - z_{t-n}) + b_{n+1}(D_{t_n} + \kappa z_{t-n})e^{-\rho\tau}$$
$$+ h_{n+1}(Z_{t_n} - z_{t-n})^2 + g_{n+1}(Z_{t_n} - z_{t-n})(D_{t_n} + \kappa z_{t-n})e^{-\rho\tau}$$
$$+ s_{n+1}(D_{t_n} + \kappa z_{t-n})^2 e^{-2\rho\tau} \Bigg\}. \tag{18}$$

We can obtain the optimal value for z_{t_n} from Eq. 18

$$z_{t_n} = -\frac{1}{2}\delta_{n+1}[1 - g_{n+1}e^{-\rho\tau} + 2\kappa s_{n+1}e^{-\rho\tau}]D_{t_n} - -\frac{1}{2}\delta_{n+1}[\lambda + 2h_{n+1} - \kappa g_{n+1}e^{-\rho\tau}]Z_{t_n}$$
$$-\frac{1}{2}\delta_{n+1}[\lambda Z_0 - a_{n+1} + b_{n+1}\kappa e^{-\rho\tau}]. \tag{19}$$

Inserting Eqs. 19 into 18, we obtain the coefficients given by Eq. 6. This concludes the proof.

B. Proof of Proposition 2

We apply induction to prove Proposition 2. For $t_n = t_N$ the optimal value function takes the form

$$J_T^1(Y_T, D_T, F_T, T) = (F_T + \frac{s}{2})Y_T + [\lambda(Z_0 + Y_0 - Y_T) + D_T + \frac{Y_T}{2q}]Y_T$$

which satisfies Eq. 9 with the coefficients given by Eq. 11. For the induction step for some $t_n \in \{t_{m+1}, \ldots, t_{N-1}\}$, if Eqs. 8– 11 hold for t_{n+1}, we get

$$J_{t_n}^1(Y_{t_n}, D_{t_n}, F_{t_n}, t_n)$$
$$= \overset{min}{y_{t_n}} \left\{ \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda(Z_0 + Y_0 - Y_{t_n}) + D_{t_n} + \frac{y_{t_n}}{2q} \right] y_{t_n} \right.$$
$$\left. + E_{t_n} J_{t_{n+1}}^1(Y_{t_n} - y_{t_n}, (D_{t_n} + \kappa y_{t_n})e^{-\rho\tau}, F_{t_{n+1}}, t_{n+1}) \right\}$$
$$= \overset{min}{y_{t_n}} \left\{ \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda(Z_0 + Y_0 - Y_{t_n}) + D_{t_n} + \frac{y_{t_n}}{2q} \right] y_{t_n} \right.$$
$$+ (a \times a^{N-(n+1)} F_n + \frac{s}{2})(Y_{t_n} - y_{t_n}) + \lambda(Z_0 + Y_0)(Y_{t_n} - y_{t_n}) + b_{n+1}(Y_{t_n} - y_{t_n})^2$$
$$+ c_{n+1}(D_{t_n} + \kappa y_{t_n})^2 e^{-2\rho\tau} + d_{n+1} m F_{t_n}^2 + g_{n+1}(Y_{t_n} - y_{t_n})(D_{t_n} + \kappa y_{t_n})e^{-\rho\tau}$$
$$\left. + h_{n+1}(Y_{t_n} - y_{t_n}) a F_{t_n} + l_{n+1}(D_{t_n} + \kappa y_{t_n})e^{-\rho\tau} a F_{t_n} \right\}. \tag{20}$$

From Eq. 20 we obtain the optimal value for y_{t_n}, which satisfies

$$y_{t_n} = -\frac{1}{2}\delta_{n+1}((1 + c_{n+1}2\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})D_{t_n} + (-\lambda - 2b_{n+1} + g_{n+1}e^{-\rho\tau}\kappa)Y_{t_n}$$
$$+ (1 - a^{N-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau} a)F_{t_n}), \tag{21}$$

with

$$\delta_{n+1} = \left(\frac{1}{2q} + b_{n+1} - g_{n+1}\kappa e^{-\rho\tau} + c_{n+1}\kappa^2 e^{-2\rho\tau}\right)^{-1}.$$

Inserting Eq. 21 into Eq. 20 we obtain the coefficients given by Eq. 10. This concludes the induction for the first part of the Proposition 2. For the optimal value function J^2 at $t_n = t_m$, we find

$$
\begin{aligned}
&J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n} + \left[\lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{Z_{t_n}}{2q}\right] Z_{t_n}\\
&\quad + E_{t_n} J^1_{t_{n+1}}\left(Y_0, \left(D_{t_n} + \kappa Z_{t_n}\right) e^{-\rho\tau}, F_{t_n}, t_n\right)\\
&= \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n} + \left[\lambda\left(Z_0 - Z_{t_n}\right) + D_{t_n} + \frac{Z_{t_n}}{2q}\right] Z_{t_n} + \left(a^{N-n} F_n + \frac{s}{2}\right) Y_0\\
&\quad + \lambda\left(Z_0 + Y_0\right) Y_0 + b_{n+1} Y_0^2 + c_{n+1}\left(D_{t_n} + \kappa Z_{t_n}\right)^2 e^{-2\rho\tau} + d_{n+1} m F^2_{t_n}\\
&\quad + \ g_{n+1} Y_0 \left(D_{t_n} + \kappa Z_{t_n}\right) e^{-\rho\tau}\\
&\quad + \ h_{n+1} Y_0 a F_{t_n} + l_{n+1}\left(D_{t_n} + \kappa Z_{t_n}\right) e^{-\rho\tau} a F_{t_n}\\
&= \left(F_{t_n} + \frac{s}{2}\right) Z_{t_n} + \left(a^{N-n} F_n + \frac{s}{2}\right) Y_0 + (\lambda(Z_0 + Y_0) Y_0 + b_{n+1} Y_0^2)\\
&\quad + \left(\lambda Z_0 + g_{n+1}\kappa e^{-\rho\tau} Y_0\right) Z_{t_n} + \left(g_{n+1} Y_0 e^{-\rho\tau}\right) D_{t_n} + (h_{n+1} Y_0 a) F_{t_n}\\
&\quad + \left(-\lambda + \frac{1}{2q} + c_{n+1}\kappa^2 e^{-2\rho\tau}\right) Z^2_{t_n} + \left(c_{n+1} e^{-2\rho\tau}\right) D^2_{t_n} + (d_{n+1} m) F^2_{t_n}\\
&\quad + \left(1 + 2\kappa c_{n+1} e^{-2\rho\tau}\right) D_{t_n} Z_{t_n} + \left(l_{n+1} a e^{-\rho\tau}\right) D_{t_n} F_{t_n} + \left(l_{n+1} a\kappa e^{-\rho\tau}\right) F_{t_n} Z_{t_n}.
\end{aligned} \tag{22}
$$

We then show that for all $t_n \in \{t_0, \ldots, t_m\}$, the optimal value function has the form of Eq. 12 with coefficients given by Eqs. 13 and 14.

We have already shown that Eqs. 12– 14 hold for both $t_n = t_m$ and $t_n \in \{t_0, \ldots, t_{m-1}\}$, with the same assumption for t_{n+1}. We find

$$
\begin{aligned}
&J^2_{t_n}(Z_{t_n}, D_{t_n}, F_{t_n}, t_n)\\
&= \min_{z_{t_n}} \Bigg\{ \left[\left(F_{t_n} + \frac{s}{2}\right) + \lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{z_{t_n}}{2q}\right] z_{t_n}\\
&\quad + E_{t_n} J^2_{t_{n+1}}(Z_{t_n} - z_{t_n}, (D_{t_n} + \kappa z_{t_n}) e^{-\rho\tau}, F_{t_{n+1}}, t_{n+1}) \Bigg\}\\
&= \min_{z_{t_n}} \Bigg\{ \left[(F_{t_n} + \frac{s}{2}) + \lambda(Z_0 - Z_{t_n}) + D_{t_n} + \frac{z_{t_n}}{2q}\right] z_{t_n}\\
&\quad + \left(a^{m-n-1} a F_n + \frac{s}{2}\right)(Z_{t_n} - z_{t_n}) + \left(a^{N-n-1} a F_n + \frac{s}{2}\right) Y_0 + b_{n+1}(Z_{t_n} - z_{t_n})^2\\
&\quad + c_{n+1}(D_{t_n} + \kappa z_{t_n})^2 e^{-2\rho\tau}\\
&\quad + d_{n+1} m F^2_{t_n} + g_{n+1}(Z_{t_n} - z_{t_n})(D_{t_n} + \kappa z_{t_n}) e^{-\rho\tau}\\
&\quad + h_{n+1}(Z_{t_n} - z_{t_n}) a F_{t_n} + l_{n+1}(D_{t_n} + \kappa z_{t_n}) e^{-\rho\tau} a F_{t_n} + s_{n+1}(Z_{t_n} - z_{t_n})\\
&\quad + u_{n+1}(D_{t_n} + \kappa z_{t_n}) e^{-\rho\tau}\\
&\quad + w_{n+1} a F_{t_n} + o_{n+1} \Bigg\}.
\end{aligned} \tag{23}
$$

To obtain the optimal value for z_{t_n}, we set $\frac{\partial J^2}{\partial z_{t_n}} \stackrel{!}{=} 0$ and obtain

$$z_{t_n} = -\frac{1}{2}\delta_{n+1}((-\lambda - 2b_{n+1} + g_{n+1}\kappa e^{-\rho\tau})Z_{t_n} + (1 + 2c_{n+1}\kappa e^{-2\rho\tau} - g_{n+1}e^{-\rho\tau})D_{t_n} + (1 - a^{m-n} - h_{n+1}a + l_{n+1}\kappa e^{-\rho\tau}a)F_{t_n} + (\lambda Z_0 - s_{n+1} + u_{n+1}\kappa e^{-\rho\tau})), \quad (24)$$

with

$$\delta_{n+1} = \left(\frac{1}{2q} + b_{n+1} + c_{n+1}\kappa^2 e^{-2\rho\tau} - g_{n+1}\kappa e^{-\rho\tau}\right)^{-1}.$$

Inserting Eqs. 24 into 23 we get the optimal value function given by Eq. 12 and the coefficients given by Eq. 13.

References

Alfonsi, A., Fruth, A., & Schied, A. (2010). Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, *10*(2), 143–157.

Almgren, R. (2003). Optimal execution with nonlinear impact functions and trading enhanced risk. *Applied Mathematical Finance*, *10*, 1–18.

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

Avramov, D., Chordia, T., & Goyal, A. (2006). The impact of trades on daily volatility. *Review of Financial Studies*, *19*(4), 1241–1277.

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

Bessembinder, H., Carrion, A., Tuttle, L., & Venkataraman, K. (2016). Liquidity, resiliency and market quality around predictable trades: Theory and evidence. *Journal of Financial Economics*, *121*, 142–166.

Chacko, G., Jurek, J., & Stafford, E. (2008). The price of immediacy. *Journal of Finance*, *63*(3), 1253–1290.

Dorn, A. J., Dorn, D., & Sengmueller, P. (2015). Trading as gambling. *Management Science*, *61*(10), 2281–2547.

Easley, D., Prado, M., & O'Hara, M. (2012). Flow toxicity and liquidity in a high frequency world. *Review of Financial Studies*, *25*(5), 14571493.

Foucault, T., Pagano, M., & Roell, A. (2013). *Market liquidity: Theory, evidence, and policy*. Oxford: Oxford University Press.

Gerig, A., & Michayluk, D. (2017). Automated liquidity provision. *Pacific-Basin Finance Journal*, *45*, 1–13.

Goldstein, M. A., Irvine, P., Kandel, E., & Wienern, Z. (2009). Brokerage commissions and institutional trading patterns. *Review of Financial Studies*, *22*(12), 5175–5212.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. Oxford: Oxford University Press.

Hasbrouck, J. (2007). *Empirical market microstructure*. Oxford: Oxford University Press.

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

Kissell, R. (2014). The Science of Algorithmic trading and Portfolio Management.

Kyle, A., Kritzman, M., & Obizhaeva, A. (2016). A practitioner's guide for market microstructure invariance. *Journal of Portfolio Management*, *43*(1), 43–53.

Kyle, A., & Obizhaeva, A. (2016). Market microstructure invariance: Empirical hypotheses. *Econometrica*, *84*(4), 1345–1404.

Kyle, A., Obizhaeva, A., & Wang, J. (2017). Smooth trading with overconfidence and market power. *The Review of Economic Studies*, *85*(1), 611–662.

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

O'Hara, M. (1995). *Market microstructure theory*. Hoboken: Blackwell Publishing.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, *116*(2), 257–270.

Schöneborn, T., & Schied, A. (2009). Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance and Stochastics*, *13*, 181–204.

Sun, E. W., Kruse, T., & Yu, M. T. (2014). High frequency trading, liquidity, and execution cost. *Annals of Operations Research*, *223*, 403–432.

Sun, E. W., Kruse, T., & Yu, M. T. (2015). Financial transaction tax: Policy analytics based on optimal trading. *Computational Economics*, *46*, 103–141.
Sun, E. W., Kruse, T., & Yu, M. T. (2018). Systemic risk, financial markets, and performance of financial institutions. *Annals of Operations Research*, *262*, 579–603.
Ting, C., Warachka, M., & Zhao, Y. (2007). Optimal liquidation strategies and their implications. *Journal of Economic Dynamics and Control*, *31*, 1431–1450.
Werner, I. M. (2003). Nyse order flow, spreads, and information. *Journal of Financial Markets*, *6*, 309–335.