Finance and Stochastics
https://doi.org/10.1007/s00780-026-00602-x

Optimal execution and speculation with trade signals

Peter Bank[1] · **Álvaro Cartea**[2] · **Laura Körber**[1]

Received: 21 July 2023 / Accepted: 2 June 2025

Abstract
We propose a price impact model where changes in prices are purely driven by the order flow in the market. The stochastic price impact of market orders and the arrival rates of limit and market orders are functions of the market liquidity process which reflects the balance of the demand and supply of liquidity. Limit and market orders mutually excite each other so that the liquidity is mean-reverting. We use the theory of Meyer σ-fields to introduce a short-term signal process from which a trader learns about imminent changes in order flow. Her trades impact the market through the same mechanism as other orders. With a novel version of Marcus-type SDEs, we efficiently describe the intricate timing of market dynamics at moments when her orders coincide with orders issued by others. In this setting, we examine an optimal execution problem and derive the Hamilton–Jacobi–Bellman (HJB) equation for the value function of the trader. The HJB equation is solved numerically, and we illustrate how the trader uses the signals to enhance the performance of execution problems and to execute speculative strategies.

Keywords Optimal execution · Speculation · Trade signal · Meyer σ-field · Stochastic optimal control · Marcus SDEs

Mathematics Subject Classification 91B70 · 93E20

JEL Classification C02 · C61

P. Bank and L. Körber acknowledge support by Deutsche Forschungsgemeinschaft through IRTG 2544. P. Bank also acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC/TRR 388 “Rough Analysis, Stochastic Dynamics and Related Fields”, Project ID 516748464.

✉ Peter Bank
bank@math.tu-berlin.de

Álvaro Cartea
alvaro.cartea@maths.ox.ac.uk

Laura Körber
koerber@math.tu-berlin.de

[1] Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

[2] Oxford-Man Institute of Quantitative Finance and Mathematical Institute, University of Oxford, Oxford, UK

Published online: 10 September 2026

1 Introduction

In this paper, we propose a reduced-form model where the evolution of prices is determined by the order flow in the market. In our model, changes in the asset's price are caused by market order flow, which is modelled as a pure jump process. Thus we do not require exogenous dynamics for the evolution of some unaffected fundamental price as is generally proposed in standard models including Almgren and Chriss [2] and Obizhaeva and Wang [28]. In the present paper, the price impact of a market order depends on the market's liquidity which is given by the difference between the volume of posted limit orders and the volume of market orders and cancellations. When this difference is high, i.e., when the market is liquid, the price impact of the next market order is low, and when this difference is low, the price impact is high. Our model may be viewed as a permanent price impact model with finite *market depth* similar to that in Huberman and Stanzl [19]. However, our impact specification depends on liquidity and we do not require an exogenous fundamental price.

For simplicity, we do not distinguish between the volumes posted on the bid or ask side of the book, a model extension we leave for future research. This allows us to specify the *market's tightness* with a constant spread. The *market's resilience* is captured by letting the arrival rates of market orders, cancellations and limit orders depend on the liquidity in the market: when the level of liquidity is low, the arrival rate of further market orders is low and the arrival rate of limit orders is high, and vice versa when the liquidity is high; see Cartea et al. [10] who provide empirical evidence of this effect based on data from the Nasdaq stock exchange. As a consequence, resilience is an endogenous feature in our model as opposed to models such as the one by [28] where resilience is an exponential relaxation of price impact towards zero.

We introduce a trader who uses market orders to complete an execution programme, where her objective is to maximise expected utility of terminal wealth and she does not provide liquidity to the book. The trader receives a private signal about the arrival of other limit and market orders and uses it to execute informed trades before other orders arrive in the book. Delicate timing issues arise when in reaction to a signal, the trader pre-emptively sends orders just before other orders arrive in the market, while having the option to trade after other orders arrive. As a result, our continuous-time market dynamics have to accommodate three simultaneous jump-shocks in a timely manner. For this, we present a novel Marcus-type SDE system with which we compute the dynamics of the trader's portfolio value. The system of SDEs highlights the positive effect of the arrival of liquidity and reveals the adverse effects of the trader's price impact and transaction costs. Interestingly, the wealth dynamics also reveal how for a trader with a long position, the arrival of market buy orders is beneficial if their upwards pressure on market prices is not offset by their detrimental effect on market liquidity. An analogous effect is found for short positions and market sell orders.

As mentioned, the trader's market orders affect the price of the asset and the liquidity in the market in the same way as the *external* orders sent by other market participants. Due to the resilience of the book in our model, the trader's orders trigger more liquidity provision, an indirect market impact which may incentivise pump-and-dump schemes. In such cases, to ensure a well-functioning market, we introduce a trading halt mechanism that imposes a minimum liquidity requirement for the market to continue operating so that the profitability of pump-and-dump schemes is limited. When a liquidity-taking order exhausts liquidity below the minimum liquidity requirement, trading is halted and positions are executed in an "auction" where the price is randomly drawn from a Gaussian distribution.

In a market with trading halts, the trader's optimal execution problem is well posed and the value function is non-degenerate. The optimisation problem is a singular stochastic control problem when trading is continuous, and it results in an impulse control problem when trading at a minimum lot size. Under some regularity assumptions, we prove continuity of the value function. This is non-trivial because traders may be confronted with abruptly differing liquidity evolutions even when starting with very similar initial liquidity. This is due to the jumps corresponding to an external order flow. As a result, pathwise closeness cannot be expected and the dynamics of liquidity have to be controlled carefully, which we do with the Marcus-SDE dynamics that provides a remarkably convenient tool for this stability analysis.

We apply the dynamic programming approach to derive the Hamilton–Jacobi–Bellman quasi-variational inequality (HJBQVI) for both continuous and discrete trading. As a novelty, our HJBQVI contains a double integral straddling a sup-operator that yields the optimal signal-based trade.

The trader's private signal means that strategies should include more than just predictable processes. This can be conveniently captured by requiring them to be measurable with respect to a Meyer σ-field that contains the usual predictable σ-field. We refer to Lenglart [24] for the introduction of this concept. In stochastic optimisation, the technique of Meyer σ-fields was first applied by El Karoui [17, pp. 117–149] in the context of optimal stopping. Bank and Besslich [4] apply the theory of Meyer σ-fields in a singular stochastic control problem which is solved with convex analysis tools. In Merton's optimal investment problem, Bank and Körber [5] use a Meyer σ-field to incorporate a short-term signal about jumps in the price of the asset and use dynamic programming methods to solve for the optimal investment strategy.

We use simulations to study the trader's execution strategies. The trader uses signals about liquidity-taking orders to optimise the times of execution and the trading volume. Specifically, upon receiving the signal of an imminent arrival of a liquidity-taking order, the trader may submit a market order before the external order arrives and impacts the price. As the bid–ask spread widens, a signal about liquidity provision becomes irrelevant for the execution programme of a trader because there is no incentive to execute a market order before the liquidity increases in the book; so the price impact of market orders decreases.

For a narrow bid–ask spread, we find that the trader uses the signal about liquidity provision to start speculative round-trip trades when the liquidity is low and completes the round-trip after the liquidity has recovered upon receiving a signal about a liquidity-taking order.

Trading on information from signals increases the average terminal wealth, and it also increases the variance of the distribution of the terminal wealth. In other words, to extract value from the signal, the trader uses strategies that increase the expected gain and that increase the risk of the financial performance of the execution programme.

There exists a broad range of literature on informed trading. Important seminal works are those by Kyle [20] and Back [3]. More recent work on optimal trading with market signals is in Cartea and Jaimungal [8] who examine optimal execution with a general Markovian signal and derive closed-form optimal strategies; see also Casgrain and Jaimungal [13] who incorporate latent factors. Similarly, Lehalle and Neuman [23] and Neuman and Voß [27] study optimal trading with signals when there is transient price impact, and Belak et al. [6] use non-Markovian finite-variation signals. For a market maker, Cartea and Wang [12] show how to use a signal of the trend of the price of an asset in optimal liquidity-provision strategies; similarly, Lehalle and Mounjid [22] study optimal market-maker strategies with a signal on liquidity imbalance. More recently, Cartea et al. [7] use signatures of the market to generate signals, and Cartea and Sánchez-Betancourt [11] study how a broker provides liquidity to informed and uninformed traders.

The feature that the trader's market orders influence the arrival rate of other limit and market orders in the same way as external orders is a novelty compared with the existing literature on stochastic optimal control with Hawkes processes. For instance, Alfonsi and Blanc [1] solve an optimal execution problem in a modified version of the model in [28] where the external order flow is modelled as a Hawkes process that is not influenced by the trader. Similarly, Cartea et al. [9] study a framework where the fill rates of the trader's controlled limit order flow are driven by an external, uncontrolled Hawkes process. The work of Cayé and Muhle-Karbe [14] proposes a modification of the Almgren and Chriss [2] model where the past orders of the trader have a self-exciting effect on the price impact. In Fu et al. [18], the trader influences the base intensity of the mutually exciting external market order dynamics so that the order flow of the trader influences the intensity process in a different way from the external order flow.

The remainder of this paper is organised as follows. Section 2 presents the market model and introduces a trader who receives signals on the order flow. Section 3 examines the problem of optimal investment and execution in a market with trading halts and illustrates the performance of the optimal strategy with trading signals. Results and proofs are collected in the Appendix.

2 The model

In this section, we present a model of stock prices where innovations in prices are driven by the flow of market orders. First, we introduce a market model where the arrival of market orders and limit orders is driven by a marked Poisson point process. The arrival rate of orders is a function of the liquidity in the market. At times of high liquidity, market orders arrive more frequently, while at times of low liquidity, the arrival rate of limit orders increases. This feature of the model introduces resilience in the supply and demand of liquidity. Next, we introduce a strategic trader who receives signals about order flow and executes market orders. In our model, all orders arriving in the market, external or those of the trader, have the same effect on the dynamics of liquidity and prices.

2.1 The uncontrolled model

In the following, we present a model for the price dynamics $(P_t)_{t\geq 0}$ of a single asset where price changes are driven by the market order flow $(M_t)_{t\geq 0}$. The market liquidity λ_t measures the capacity of the market to fill an incoming market order at time $t \geq 0$. Liquidity increases with incoming limit orders, but is diminished by cancellations and by the volume of market orders. Incoming limit orders correspond to the increasing part of the process $(L_t)_{t\geq 0}$, while its decreasing part is interpreted as cancellations. The volume of market orders is described by the total variation $|dM_t|$ of the market order flow at time t. For simplicity, we assume that the liquidity on the buy and sell side is the same. So the liquidity process $(\lambda_t)_{t\geq 0}$ satisfies the dynamics

$$d\lambda_t = dL_t - |dM_t|. \tag{2.1}$$

Many specifications for the order flows L and M are conceivable. As a tractable possibility, we let them be driven by a marked Poisson point process and put

$$dL_t = \int_{E\times\mathbb{R}_+} 1_{\{y\leq f(\lambda_{t-})\}}\rho(e)N(dt, de, dy), \tag{2.2}$$

$$dM_t = \int_{E\times\mathbb{R}_+} 1_{\{y\leq g(\lambda_{t-})\}}\eta(e)N(dt, de, dy). \tag{2.3}$$

Here, N is a marked Poisson point process on $[0,\infty) \times E \times \mathbb{R}_+$ with compensator $dt \otimes \nu(de) \otimes dy$, where we assume $\nu(E) = 1$. The mappings $\eta, \rho \in L^1(\nu)$ determine the volume of the order associated with a mark e that the point process N sets in the Polish mark space E with its Borel σ-field $\mathcal{E}$, and $\nu(de)$ is the probability for this mark. When $\rho(e) > 0$, a new limit order of size $\rho(e)$ is posted, and $\rho(e) < 0$ corresponds to a cancellation of limit orders of size $|\rho(e)|$. Similarly, $\eta(e) > 0$ corresponds to a buy market order of size $\eta(e)$, and $\eta(e) < 0$ is a sell market order of size $|\eta(e)|$. Modelling the dynamics of limit and market orders with the same marked Poisson point process allows us to specify the dependence structure between the arrival of orders in a tractable way. More complex specifications are conceivable to address momentum in order flow or time-of-day effects, but are beyond the scope of this paper. It is worth pointing out, however, that the SDE dynamics of our model to be derived below are fully versatile and able to accommodate more involved order flow dynamics.

To exclude the simultaneous arrival of limit orders and market orders (which would be cumbersome for bookkeeping), we assume

$$\eta(e)\rho(e) = 0 \qquad \text{for all } e \in E. \tag{2.4}$$

The functions f and g specify how the liquidity affects the posting frequency of limit and market orders. We assume that

$$\begin{aligned}&f \text{ and } g \text{ are nonnegative, Lipschitz-continuous}\\ &\text{and decreasing and increasing, respectively.}\end{aligned} \tag{2.5}$$

Lipschitz-continuity of f and g yields in particular that the Hawkes-like liquidity dynamics (2.1)–(2.3) have a unique solution; see Lemma A.1 in the Appendix. The monotonicity introduces resilience of liquidity provision. Indeed, it ensures that as the liquidity decreases, fewer market orders arrive while limit orders arrive more frequently, and vice versa when the liquidity increases, imposing some degree of self-stabilisation for the liquidity balance λ.

Next, we specify the price dynamics of the asset as the pure-jump càdlàg process P with jumps

$$\Delta_t P = I(\Delta_t M, \lambda_{t-}), \tag{2.6}$$

where $\Delta_t M = M_t - M_{t-}$ for the càdlàg process M solving the dynamics (2.3), and with the price impact function $I(\cdot, \cdot)$ defined by

$$I(\Delta, \lambda) := \operatorname{sgn}(\Delta) \int_0^{|\Delta|} \iota(\lambda - z)dz = \int_0^{\Delta} \iota(\lambda - |z|)dz, \qquad \Delta, \lambda \in \mathbb{R}, \tag{2.7}$$

where $\operatorname{sgn} := 1_{(0,\infty)} - 1_{(-\infty,0)}$ denotes the sign function and $\int_0^{\Delta} = -\int_0^{|\Delta|}$ for $\Delta < 0$. Here, the function $I(\Delta, \lambda)$ describes the price impact of a market order of size $\Delta \in \mathbb{R}$ that arrives when the liquidity of the market is λ. The sign of $I(\Delta, \lambda)$ is determined by that of Δ: buy orders increase the price and sell orders decrease it. We assume that the function

$$\iota : \mathbb{R} \to \mathbb{R}_+ \text{ is Lipschitz-continuous and decreasing}$$

to ensure that market orders have less impact when the liquidity is high. Similarly, the absolute value of the function I is increasing in the size $|\Delta|$ because, everything else being equal, as the volume of market orders increases, so does the impact of the order on the price of the asset. Finally, the definition (2.7) makes the price dynamics (2.6) consistent when orders are split, i.e., $I(\Delta, \lambda) = I(\Delta_1, \lambda) + I(\Delta_2, \lambda - |\Delta_1|)$ for $\Delta_1 + \Delta_2 = \Delta$ with $\operatorname{sgn}(\Delta_1) = \operatorname{sgn}(\Delta_2)$. In fact, when taking order splitting to the limit, this observation reveals that an equivalent specification of the price dynamics can be given in the form of the Marcus-type SDE system (cf. Marcus [26])

$$\diamond\, d\lambda_t = \diamond\,(dL_t - |dM_t|) \quad \text{and} \quad \diamond\, dP_t = \iota(\lambda_t) \diamond dM_t. \tag{2.8}$$

Remark 2.1 We briefly explain the Marcus-type SDE dynamics in (2.8). The symbol $\diamond$ indicates that at any time t, a jump by the drivers (here M, its total variation $V[0, \cdot](M) = \int_0^{\cdot} |dM_t|$ and L) makes the system (here (λ, P)) jump, too. The system's jump is determined by the endpoint at $s = 1$ of a "fictitious" flow that runs in line with the given dynamics on a separate time line $s \in [0, 1]$ and starts with the pre-jump values. In our example, this flow is given, for $s \in [0, 1]$, by

$$\begin{aligned} (\ell_0, p_0) := (\lambda_{t-}, P_{t-}), \qquad d\ell_s &= \Delta_t L\, ds - |\Delta_t M|\, ds, \\ dp_s &= \iota(\ell_s)\Delta_t M ds. \end{aligned}$$

Its solution is easily found to be

$$\ell_s = \ell_0 + (\Delta_t L - |\Delta_t M|)s, \qquad p_s = p_0 + \int_0^s \iota(\ell_r)\Delta_t M dr, \qquad s \in [0,1],$$

and so, recalling that $\Delta_t L = 0$ whenever $\Delta_t M \neq 0$ by (2.4), the post-jump values become

$$\begin{aligned}\lambda_t &:= \ell_1 = \lambda_{t-} + \Delta_t L - |\Delta_t M|,\\ P_t &:= p_1 = P_{t-} + \int_0^1 \iota(\ell_0 - |\Delta_t M| r)\Delta_t M dr = P_{t-} + I(\Delta_t M, \lambda_{t-}),\end{aligned}$$

in exact accordance with (2.1) and (2.6).

Our model exhibits a direct link between price volatility and market liquidity because price volatility is determined by the arrival rate and the size of price changes, both of which depend on the liquidity in the market. More precisely, the market liquidity affects the arrival rate of market orders, i.e., the arrival rate of price changes, through the dynamics in (2.3) and the size of price changes in (2.6) through the price impact function I in (2.7). For instance, when the liquidity in the market decreases, the arrival rate of market orders decreases and the size of the price impact of market orders increases. The next lemma makes this link between volatility and liquidity explicit and states an elasticity condition under which the volatility of prices decreases with the market liquidity; see Pástor and Stambaugh [29] for empirical evidence of this negative relation between price volatility and liquidity.

Lemma 2.2 *The predictable quadratic variation of the price P is given by*

$$d\langle P\rangle_t = \sigma^2(\lambda_{t-})dt, \tag{2.9}$$

where

$$\sigma(\lambda) := \left(g(\lambda)\int_E I\big(\eta(e),\lambda\big)^2 \nu(de)\right)^{1/2}. \tag{2.10}$$

The function in (2.10) is strictly decreasing in λ under the elasticity condition

$$0 < \frac{\partial_\lambda g(\lambda)}{g(\lambda)} \Bigg/ \left(-\frac{\partial_\lambda \mathbb{I}^2(\lambda)}{\mathbb{I}^2(\lambda)}\right) < 1, \tag{2.11}$$

where $\mathbb{I}(\lambda) := (\int_E I(\eta(e),\lambda)^2\nu(de))^{1/2}$ is the L^2-norm of the price impact size from market orders at the liquidity level λ.

For a proof, see Appendix A.1.

The elasticity condition (2.11) compares the relative change in the arrival rate g of market orders with the absolute value of the relative change in the impact size of market orders under the L^2-norm. Thus condition (2.11) says that when the liquidity decreases, the price impact increases at a higher rate than the arrival rate of market orders decreases. Hence price volatility rises with low liquidity because the increase of the price impact of market orders at low liquidity more than compensates for the decrease in the arrival rate of market orders.

2.2 The controlled model with trading signal

Next, we introduce a trader who follows an optimal execution programme. For simplicity, we assume that the trader sends market orders and does not provide liquidity to the market.

To describe the trader's information structure, we introduce the (right-continuous) filtration $\mathbb{F} = (\mathcal{F}_t)$ with

$$\mathcal{F}_t = \sigma\big(N([0,s] \times A \times [0,y]), s \in [0,t], A \in \mathcal{E}, y \in \mathbb{R}_+\big), \qquad t \geq 0,$$

generated by the point process N. The usual information framework of the predictable information flow $\mathcal{P}(\mathbb{F})$ contains only past information about the order flow, and a trader with information $\mathcal{P}(\mathbb{F})$ can only trade after the arrival of external limit and market orders are revealed. Thus she can only react to market events after they have happened.

By contrast, in our setting, the trader receives a short-term signal about imminent changes in the order flow. She uses this information to anticipate jumps in liquidity and in prices, and executes signal-based trades before the external order arrives in the market. The signal is given by the process

$$Z_t = \int_{\{t\} \times E \times \mathbb{R}_+} z(e,y) N(ds, de, dy), \qquad t \geq 0, \tag{2.12}$$

where N is the same marked Poisson point process that also drives the external order flows (2.2) and (2.3). The function $z \in L^0(\nu(de) \otimes dy)$ determines what the trader learns about any mark (e, y) set by N, with $z(e, y) = 0$ amounting to the same as receiving no signal. Signals $\bar{z}$ different from zero are received with frequency $\mu(d\bar{z}) := (\nu \otimes \mathrm{Leb}) \circ z^{-1}(d\bar{z})$, which we assume to be a σ-finite measure on $z(E \times \mathbb{R}_+) \setminus \{0\}$. We introduce the Meyer σ-field (cf. Lenglart [24, Def. 2])

$$\Lambda := \mathcal{P}(\mathbb{F}) \vee \sigma(Z)$$

which adds the information $\sigma(Z)$ of the signal Z to the predictable information $\mathcal{P}(\mathbb{F})$. It is this σ-field which formally describes the information flow based on which the investor chooses her strategy.

Remark 2.3 A key competitive edge of high-frequency traders is strategies that use their speed advantage and that employ trading signals, which include those that use limit order book information to predict order flow; see Lewis [25]. Other signals are designed to anticipate large meta-orders split and routed to different exchanges at slightly different points in time. In our case study below, we consider a signal of whether an impending next order provides or consumes liquidity.

The trader's strategy is described by a process $(C_t)_{t\geq 0}$ of locally bounded variation starting at $C_{0-} = 0$. Its changes represent changes in the trader's inventory so that at any time t, the net number of shares sold or bought up to this moment is given by C_t. We assume admissible strategies C to induce uniformly bounded inventories

$$Q^C_t := q_0 + C_t \qquad \text{with } |Q^C_t| \leq \bar{q} \tag{2.13}$$

for some constant $\bar{q} \in [0, \infty)$ (which could be for instance the total number of shares outstanding). Trades can be of any size in $D = \mathbb{R}$, or – in line with market practice – a multiple of a minimal lot size $\delta > 0$ so that $D = \{\dots, -2\delta, -\delta, 0, \delta, 2\delta, \dots\}$. The set of admissible controls is therefore

$$\begin{aligned}\mathcal{C}(q) = \{C : D\text{-valued, } \Lambda\text{-measurable process of bounded variation}\\ \text{with (2.13) and } C_{0-} = 0\}.\end{aligned}$$

Notice that we allow C to be merely làdlàg; so there may be "left jumps" $\Delta^\ell_t C := C_t - C_{t-}$ and "right jumps" $\Delta^r_t C := C_{t+} - C_t$ at the same time $t \geq 0$. They correspond to the proactive signal-based trades and reactive state-based trades we introduce shortly.

The following lemma provides a representation of the trader's strategy and clarifies how she uses the signal Z; its proof is given in Appendix A.3.

Lemma 2.4 *A process $C = (C_t)_{t\in[0,T]}$ of bounded variation is Λ-measurable if and only if it its decomposition*

$$C_t = C^c_t + \sum_{0\leq s\leq t} \Delta^\ell_s C + \sum_{0\leq s<t} \Delta^r_s C, \qquad t \in [0, T], \tag{2.14}$$

into its continuous part C^c, left jumps $\Delta^\ell C$ and right jumps $\Delta^r C$ is such that C^c is adapted, $\Delta^r C$ is optional and $\Delta^\ell C$ is of the form

$$\Delta^\ell_t C = \int_{\{t\}\times\{z\neq 0\}} \Gamma_s\big(z(e, y)\big) N(ds, de, dy) + \Delta^{\ell,p}_t C, \qquad t \in [0, T], \tag{2.15}$$

almost surely for some pure-jump predictable càdlàg process $C^{\ell,p} = \sum_{t\leq\cdot} \Delta_t C^{\ell,p}$ and some $(\mathcal{P}(\mathbb{F}) \otimes \mathcal{B}(\mathbb{R}))$-measurable field Γ satisfying

$$\int_{[0,T]\times\{z\neq 0\}} \big|\Gamma_s\big(z(e, y)\big)\big| N(ds, de, dy) < \infty \qquad \textit{almost surely.}$$

Putting $\Delta^{\ell,p}_t C := \Delta_t C^{\ell,p}$, we note that almost surely, predictable left jumps of such a control C only happen when there are no marks, i.e.,

$$\begin{aligned}&\mathbb{P}\Big[\{t \in [0, T] : \Delta^{\ell,p}_t C \neq 0\} \subseteq \Big\{t \in [0, T] : N\big(\{t\} \times E \times [0, \infty)\big) = 0\Big\}\Big]\\ &= 1.\end{aligned} \tag{2.16}$$

Thus upon observing a signal $Z_t \neq 0$, the trader sends the order $\Delta_t^{\ell} C = \Gamma_t(Z_t)$ before the full information about external orders becomes known to all market participants and its effect materialises, i.e., liquidity and prices change as a result. Hence we call such a trade *signal-based*.

In contrast, the right jump $\Delta_t^r C$ is the trader's action after the arrival of an external order in the market. It incorporates the full information about the post-shock market state, i.e., the market state after the arrival of external orders. In addition, $\Delta_t^r C$ can also represent an order sent out of the trader's own volition, typically motivated by the state of her execution programme. The latter also applies to the predictable left trades $\Delta_t^{\ell,p} C$. Therefore, we call these trades *state-based*.

Remark 2.5 In our model, the controlled orders $\Delta_t^{\ell} C$, $\Delta_t^r C$, dC_t^c affect market liquidity and with it the arrival rate of external shocks in the same way as the external limit orders, market orders and cancellations. This is a key difference between our model and those proposed in the existing literature on stochastic control for Hawkes processes; see e.g. Alfonsi and Blanc [1], Cartea et al. [9] and Fu et al. [18].

Let us now introduce the controlled dynamics of the state process

$$S_t^C := (\lambda_t^C, Q_t^C, P_t^C, X_t^C), \qquad t \geq 0,$$

starting at $S_{0-}^C := (\lambda, q, p, x)$. Here, λ_t^C is the current level of market liquidity, Q_t^C is the trader's current inventory, P_t^C is the current asset price, and X_t^C is the trader's cash process resulting from her strategy C.

When no external market or limits orders arrive and when there are no jumps in the trader's strategy, the state process satisfies the continuous dynamics

$$\begin{aligned} dS_t^C = \big(-|dC_t^c|, dC_t^c, \iota(\lambda_t^C) dC_t^c, -P_t^C dC_t^c - \zeta |dC_t^c| \big) \\ \text{when } \Delta_t^{\ell} C = \Delta_t^r C = 0, N(\{t\} \times E \times \mathbb{R}_+) = 0, \end{aligned}$$

where $\zeta \geq 0$ is the half-spread quoted in the book.

Next, we describe how the state changes when there are jumps in the order flow. A jump can result for instance from an order of size $\Delta > 0$ sent by the trader. Thus the asset price changes by $I(\Delta, \lambda)$, where λ denotes the pre-trade liquidity level; see (2.6). When the trader sends an order of size $\Delta > 0$, the price impact from her trade affects the trader's post-trade cash position in a nonlinear way. To describe this, it is convenient to introduce the function

$$\Xi(\Delta, \lambda) := \int_0^{|\Delta|} I(z, \lambda) dz \tag{2.17}$$

which reflects the impact costs of the trade Δ when the liquidity is λ. So the total costs of buying $\Delta \geq 0$ shares when the pre-transaction mid-price is p amount to $(p + \zeta)\Delta + \Xi(\Delta, \lambda)$; similarly, the revenue from selling $\Delta \geq 0$ shares is $(p - \zeta)\Delta - \Xi(\Delta, \lambda)$. In the definition of I in (2.7), the function Ξ is consistent with order splitting, i.e., $\Xi(\Delta, \lambda) = \Xi(\Delta_1, \lambda) + \Xi(\Delta_2, \lambda - |\Delta_1|) + \Delta_2 I(\Delta_1, \lambda)$ for $\Delta_1 + \Delta_2 = \Delta$ with $\mathrm{sgn}(\Delta_1) = \mathrm{sgn}(\Delta_2)$. Similarly to the infinitesimal version

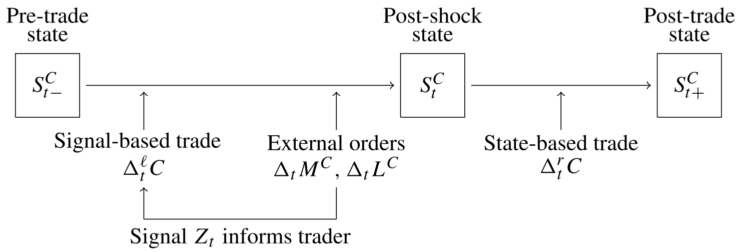

Fig. 1 Evolution of our state process S^C

of the price and liquidity dynamics (2.8), this splitting yields a Marcus-style SDE description also for the investor's cash position; cf. (2.22) below.

Finally, the precise timing when market orders, limit orders and cancellations arrive is important and not interchangeable. This is because the different orders arrive when the respective levels of liquidity are λ^C_{t-}, $\lambda^C_{t-} - |\Gamma_t(Z_t)|$ and $\lambda^C_t := \lambda^C_{t-} - |\Gamma_t(Z_t)| - |\Delta_t M^C| + \Delta_t L^C$, and so the impact on the price of the asset varies and the effect on the trader's cash process is different, too. We explain this step by step; see Fig. 1 for an illustration.

2.2.1 From pre-trade state S^C_{t-} to post-shock state S^C_t

We assume the pre-trade state $\mathbf{s}_- := S^C_{t-}$ is

$$\mathbf{s}_- = (\boldsymbol{\lambda}_-, \mathbf{q}_-, \mathbf{p}_-, \mathbf{x}_-) := (\lambda^C_{t-}, Q^C_{t-}, P^C_{t-}, X^C_{t-}).$$

The impending external limit and market orders at time t are

$$\rho := \Delta_t L^C = \int_{\{t\}\times E\times\mathbb{R}_+} 1_{\{y\le f(\lambda^C_{s-})\}}\rho(e)N(ds, de, dy), \tag{2.18}$$

$$\eta := \Delta_t M^C = \int_{\{t\}\times E\times\mathbb{R}_+} 1_{\{y\le g(\lambda^C_{s-})\}}\eta(e)N(ds, de, dy). \tag{2.19}$$

Remark 2.6 Equations (2.18) and (2.19) specify the dynamics of incoming limit and market orders analogously to (2.2) and (2.3). However, their jump frequencies are now determined by the liquidity λ^C_{t-}. This is another way by which the investor affects market dynamics because her market orders attract more limit orders, an effect that may be exploited in a pump-and-dump scheme.

As derived in Lemma 2.4, when the trader receives a signal Z_t, she trades the quantity $\gamma := \Gamma_t(z)$. Without a signal, she can still opt to send the predictable order $\gamma := \Delta^{\ell,p}_t C$. In either case, the market liquidity updates to

$$\boldsymbol{\lambda}(\gamma, \eta, \rho; \mathbf{s}_-) := \boldsymbol{\lambda}_- - |\gamma| - |\eta| + \rho$$

and the trader's inventory becomes

$$\mathbf{q}(\gamma, \eta, \rho; \mathbf{s}_-) := \mathbf{q}_- + \gamma.$$

Due to price impact, the price changes to

$$\mathbf{p}(\gamma, \eta, \rho; \mathbf{s}_-) := \mathbf{p}_- + I(\gamma, \boldsymbol{\lambda}_-) + I(\eta, \boldsymbol{\lambda}_- - |\gamma|),$$

where the market order of size η arrives when the liquidity is $\boldsymbol{\lambda}_- - |\gamma|$. The trader's cash position becomes

$$\mathbf{x}(\gamma, \eta, \rho; \mathbf{s}_-) := \mathbf{x}_- - \mathbf{p}_-\gamma - \zeta|\gamma| - \Xi(\gamma, \boldsymbol{\lambda}_-).$$

Thus we define the state-update function $\mathbb{s}$ as

$$\mathbb{s}(\gamma, \eta, \rho; \mathbf{s}_-) := (\boldsymbol{\lambda}, \mathbf{q}, \mathbf{p}, \mathbf{x})(\gamma, \eta, \rho; \mathbf{s}_-) \tag{2.20}$$

to write the post-shock state as

$$S_t^C := \mathbb{s}\big(\Gamma_t(Z_t)1_{\{Z_t \neq 0\}} + \Delta_t^{\ell,p} C, \Delta_t M^C, \Delta_t L^C; S_{t-}^C\big).$$

2.2.2 From post-shock state S_t^C to post-trade state S_{t+}^C

After the realised external shock and the post-shock state S_t^C become fully known to the trader, she executes the state-based trade $\Delta_t^r C$ and the post-trade state is defined as

$$S_{t+}^C := \mathbb{s}(\Delta_t^r C, 0, 0; S_t^C),$$

where $\mathbb{s}$ is given in (2.20).

It is convenient to have an equivalent alternative description of the above state dynamics in the form of an SDE system. For this, one can use the framework of Chevyrev et al. [16] which introduces a fictitious time interval [0, 1] at every jump time to specify a "manual" describing the SDE system's reaction to jumps. Our idea is to use this fictitious time interval to coordinate the three jumps that can affect the market at a given time: a (typically signal-based) trade dC_t^ℓ, either a limit order/cancellation dL_t^C or an exogenous market order dM_t^C (recall (2.4)), and a state-based trade $dC_t^{r,c} = dC_t^r + dC_t^c$ (to which we have added the continuous part of the control for the sake of simplicity and completeness; cf. also Lemma 2.4).

Recall that the exogenous orders only depend on the pre-jump liquidity λ_{t-}^C, i.e.,

$$dL_t^C = \int_{E\times\mathbb{R}_+} 1_{\{y \leq f(\lambda_{t-}^C)\}}\rho(e)N(dt, de, dy),$$

$$dM_t^C = \int_{E\times\mathbb{R}_+} 1_{\{y \leq g(\lambda_{t-}^C)\}}\eta(e)N(dt, de, dy).$$

In line with the order-splitting impact specification, we let a Marcus-style fictitious flow (cf. Remark 2.1) continuously drive the liquidity λ^C, the cash position X^C, the stock position Q^C and the price P^C by these shocks via

$$\lambda^C_{0-} = \lambda, \quad \diamond\, d\lambda^C_t = \diamond\big((-|dC^\ell_t|)\vec{+}(dL^C_t - |dM^C_t|)\vec{+}(-|dC^{r,c}_t|)\big), \tag{2.21}$$
$$X^C_{0-} = x, \quad \diamond\, dX^C_t = -P^C_t \diamond (dC^\ell_t \vec{+} 0 \vec{+} dC^{r,c}_t) - \zeta \diamond (|dC^\ell_t| \vec{+} 0 \vec{+} |dC^{r,c}_t|), \tag{2.22}$$
$$Q^C_{0-} = q, \quad \diamond\, dQ^C_t = \diamond(dC^\ell_t \vec{+} 0 \vec{+} dC^{r,c}_t), \tag{2.23}$$
$$P^C_{0-} = p, \quad \diamond\, dP^C_t = \iota(\lambda^C_t) \diamond (dC^\ell_t \vec{+} dM^C_t \vec{+} dC^{r,c}_t). \tag{2.24}$$

Here, the $\vec{+}$-operator means that the jumps it connects are to be run through in the order indicated (with 0 signifying a pause) when following the fictitious Marcus flow to determine the resulting jumps; cf. Remark 2.1. This is done with the understanding that the first of the three $\vec{+}$-summands is taken care of over the first third of the fictitious time interval [0, 1], the second over the second third, and the third in the remaining third. In this way, the liquidity λ^C impacts the prices P^C in sync with all incoming orders, while the external orders have no impact on the investor's cash or stock positions as indicated by the $\vec{+}0$-pause terms in the middle of the drivers of both X^C and Q^C. The "intermediate" values λ^C_t, Q^C_t, X^C_t, P^C_t at time t (rather than their right limits at time $t+$) can be obtained from these dynamics by evaluating the Marcus-style dynamics when its fictitious flow at time t has run two thirds of its course (i.e., when at time $s = 2/3$, two of the three jump contributions at time t have been taken care of). In this way, we get the SDE system to fully describe also the dynamics of such làdlàg processes.

The advantages of the SDE-system description are borne out by the following result which describes the dynamics of the realisable portfolio value

$$W^C_t := \mathrm{w}(X^C_t, Q^C_t, P^C_t, \lambda^C_t), \qquad t \in [0, T],$$

where the function

$$\mathrm{w}(x, q, p, \lambda) := x + qp - \big(\zeta|q| + \Xi(q, \lambda)\big) \tag{2.25}$$

adjusts the mark-to-market or book value $x + qp$ of the agent's cash and stock position (x, q) by the transaction costs $\zeta|q|$ and the impact costs $\Xi(q, \lambda)$ that would be incurred if the stock position was liquidated at the present liquidity λ.

Proposition 2.7 *The realisable portfolio value W^C associated with a strategy C follows for $t \in [0,T]$ the dynamics*

$$
\begin{aligned}
W_t^C &= \mathrm{w}(x,q,p,\lambda)\\
&\quad + \int_0^t \big(I(|Q^C|,\lambda^C) - |Q^C|\iota(\lambda^C)\big) \diamond (0\vec{+}dL^C\vec{+}0)\\
&\quad - \int_0^t \big(I(|Q^C|,\lambda^C) - 1_{\{Q^C>0\}}2|Q^C|\iota(\lambda^C)\big) \diamond (0\vec{+}dM^{C,+}\vec{+}0)\\
&\quad - \int_0^t \big(I(|Q^C|,\lambda^C) - 1_{\{Q^C<0\}}2|Q^C|\iota(\lambda^C)\big) \diamond (0\vec{+}dM^{C,-}\vec{+}0)\\
&\quad - \int_0^t 2\big(I(|Q^C|,\lambda^C) - |Q^C|\iota(\lambda^C) + \zeta\big)\\
&\qquad\qquad \times (1_{\{Q^C>0\}} \diamond dQ^{C,+} + 1_{\{Q^C<0\}} \diamond dQ^{C,-}),
\end{aligned} \tag{2.26}
$$

where we use the notation $dM^C = dM^{C,+} - dM^{C,-}$ and $dQ^C = dQ^{C,+} - dQ^{C,-}$ for the Hahn decompositions of M^C and Q^C.

The proof of this result is deferred to Appendix A.4.

The above result does not depend on our choice for the dynamics of L^C and M^C, but only on the market mechanics we have specified above. As a result, the following observations hold true generically. For instance, we have $I(|Q^C|,\lambda^C) - |Q^C|\iota(\lambda^C) = \int_0^{|Q^C|} \iota(\lambda^C - z) - \iota(\lambda^C) \geq 0$ because ι is decreasing, and so added liquidity ("$dL^{C,+} > 0$") boosts the realisable portfolio value while any cancellation ("$dL^{C,-} > 0$") has the opposite effect. For the same reason, expanding an existing position (e.g. "$dQ^{C,+} > 0$" when $Q^C > 0$) also reduces the realisable portfolio value, an effect amplified by the spread ζ. External orders in opposition to the investor's position (e.g. "$dM^{C,+} > 0$" when $Q^C < 0$) penalise the liquidation value in proportion to the liquidation price impact $I(|Q^C|,\lambda^C)$. Only external market orders in the direction of the investor's current position can result in both positive or negative consequences. For instance, an external buy order "$dM^{C,+} > 0$" is welcome for a trader with a long position $Q^C > 0$, but only if its favourable price impact dominates its detrimental effect on the market liquidity in the sense that $2|Q^C|\iota(\lambda^C) > I(|Q^C|,\lambda^C) + 2\zeta$.

The above dynamics show furthermore that direct market manipulation is not profitable. In particular, immediate round-trips result in a loss. Note that this finding does not contradict Huberman and Stanzl [19] since they consider impact factors that are independent of the trader's orders.

While there is no way to turn a profit by one's own trading, the trader can try to game the system through her short- to medium-term influence on the market liquidity. Indeed, this is made possible by our choice for the dynamics (2.18) and (2.19) of incoming limit and market orders: In this specification, the jump frequencies of these external orders are determined by the liquidity λ^C which the trader can influence. This is an indirect way in which the trader affects market dynamics because the market orders issued by her attract more limit orders, an effect that could be exploited in a pump-and-dump scheme; see Sect. 3.1 below for further considerations in that respect.

As a final observation, the investor's influence on liquidity does not allow her to directly trigger or preempt impending exogenous orders by making λ^C jump. This is because the dynamics of L^C and M^C are solely governed by the pre-jump liquidity level λ^C_- that cannot be changed in an ad hoc manner.

3 The optimal investment and execution problem

In this section, we show how a trader who receives the private signal Z executes a position with market orders over some finite time interval $[0, T]$. The trader's performance criterion is the expected utility of terminal wealth.

The self-exciting nature of our system requires special care to avoid blow-ups. Thus we introduce a lower bound on the liquidity so that liquidity-taking orders (i.e., market orders and limit order cancellations) are limited by the supply of liquidity in the market. The lower bound on the liquidity can be interpreted as a threshold that triggers a trading halt imposed by the exchange to ensure a well-functioning market. As we show in this section's main result in Theorem 3.2, the trading halt ensures that the value function of the optimisation problem is non-degenerate. Finally, we illustrate the market model with signals and showcase the performance of the optimal strategy.

3.1 Trading halts for sufficient market liquidity

In practice, we observe that trading in a stock is temporarily halted if its price dynamics become too volatile or undergo an abrupt change exceeding a predefined level; see e.g. Chen et al. [15]. Such a mechanism is called a (single-stock) circuit breaker or more simply a trading halt. According to Kyle [21], the purpose of such mechanisms is in particular to reduce price volatility and to protect customers from illiquid, disorderly markets. In our reduced-form model, the liquidity level λ can be viewed as a measure for market quality. Consequently, a trading halt is triggered when a market participant's order lets market quality deteriorate beyond a given threshold $\underline{\lambda}$, the liquidity trigger, i.e., at the time

$$\tau^C := \inf\big\{t \geq 0 : \min\{\lambda^C_{t-} - |\Delta^\ell_t C|, \lambda^C_t\} < \underline{\lambda}\big\}, \tag{3.1}$$

with $\inf\emptyset = +\infty$. In this case, the market order or the cancellation of a limit order that reduces the liquidity to the level $\underline{\lambda}$ is partially executed with the liquidity $\lambda_t - \underline{\lambda}$ available in the market, and trading is halted temporarily. Any orders that are sent thereafter cannot be executed and thus have no effect on the state process.

Remark 3.1 We briefly comment on how our stylised trading halt mechanism relates to the ones used in exchanges such as NYSE or Nasdaq to prevent excessive market movements. At the time of writing, trading in a stock is halted in Nasdaq when within five minutes, its price moves by more than 10%.

Lemma 2.2 shows how volatility is linked to our market liquidity process λ and clarifies how our liquidity-based trigger uses a bound on volatility to tame market fluctuations. Moreover, a well-functioning market requires a healthy amount of liquidity, and the price band trigger of Nasdaq can be viewed as a proxy indicator for poor liquidity levels. In our reduced-form model, such a dire market condition can be directly related to its liquidity process λ^C; so a passage time like τ^C is a natural time to halt trading. Of course, constructing this liquidity indicator from real market data is challenging (for example, what levels of the limit order books to use? How to address the difference of liquidity posted on bid vs. ask side?), and so the price band rule of Nasdaq can be viewed as a more readily implementable substitute for a liquidity-oriented trigger of trading halts such as ours. Also, while not directly price-oriented, our liquidity-based trading halts often coincide with big price changes that breach a price band.

We also observe that monitoring a rolling five-minute price window leads to a path-dependency that severely complicates any control problem in this context because the state process becomes infinite-dimensional. Thus our choice of liquidity-based trading halt trigger significantly improves mathematical tractability.

To formalise the dynamics of the state process with trading halts, denote it by

$$\tilde{S}_t^C := (\tilde{\lambda}_t^C, \tilde{Q}_t^C, \tilde{P}_t^C, \tilde{X}_t^C).$$

Clearly, while the trading halt has not been triggered, the state process in the market with and without trading halts coincides so that

$$\tilde{S}^C = S^C \qquad \text{on } [\![0, \tau^C[\![\cap [\![0, T]\!].$$

When a trading halt is triggered at time $\tau^C \leq T$, we proceed analogously to Sect. 2.2 and put, if $\tau^C \leq T$,

$$\begin{aligned}\tilde{S}_{\tau^C}^C &:= \tilde{\mathrm{s}}(\Delta_{\tau^C}^{\ell} C, \Delta_{\tau^C} M^C, \Delta_{\tau^C} L^C; \tilde{S}_{\tau^C-}^C),\\ \tilde{S}_{\tau^C+}^C &:= \tilde{\mathrm{s}}(\Delta_{\tau^C}^{r} C, 0, 0; \tilde{S}_{\tau^C}^C),\end{aligned}$$

where $\tilde{\mathrm{s}}$ is a suitably amended state-update function that caps orders when they trigger the trading halt. Specifically, this capping is accomplished by

$$\Upsilon(\Delta, \lambda) := \begin{cases} \mathrm{sgn}(\Delta)(\lambda - \underline{\lambda})^+, & \text{if } \lambda - |\Delta| < \underline{\lambda},\\ \Delta, & \text{if } \lambda - |\Delta| \geq \underline{\lambda},\end{cases}$$

and consequently

$$\tilde{\mathrm{s}}(\Delta, \eta, \rho; \mathbf{s}_-) := (\tilde{\boldsymbol{\lambda}}, \tilde{\mathbf{q}}, \tilde{\mathbf{p}}, \tilde{\mathbf{x}})$$

with

$$\begin{aligned}
\tilde{\boldsymbol{\lambda}} &= \big(\boldsymbol{\lambda}_- - |\Delta| + 1_{\{\boldsymbol{\lambda}_- - |\Delta| \geq \underline{\lambda}\}}(-|\eta| + \rho)\big) \vee \underline{\lambda},\\
\tilde{\mathbf{q}} &= \mathbf{q}_- + \Upsilon(\Delta, \boldsymbol{\lambda}_-),\\
\tilde{\mathbf{p}} &= \mathbf{p}_- + I\big(\Upsilon(\Delta, \boldsymbol{\lambda}_-), \boldsymbol{\lambda}_-\big) + 1_{\{\boldsymbol{\lambda}_- - |\Delta| \geq \underline{\lambda}\}} I\big(\Upsilon(\eta, \boldsymbol{\lambda}_- - |\Delta|), \boldsymbol{\lambda}_- - |\Delta|\big),\\
\tilde{\mathbf{x}} &= \mathbf{x}_- - \mathbf{p}_- \Upsilon(\Delta, \boldsymbol{\lambda}_-) - \zeta |\Upsilon(\Delta, \boldsymbol{\lambda}_-)| - \Xi\big(\Upsilon(\Delta, \boldsymbol{\lambda}_-), \boldsymbol{\lambda}_-\big).
\end{aligned}$$

Trading is halted when the market liquidity becomes too low; thus

$$\tilde{S}^C := \tilde{S}^C_{\tau^C+} \qquad \text{on } \rrbracket \tau^C, T \rrbracket.$$

For the sake of completeness, let us also define the external market order flow $\tilde{M}^C$ in the market with trading halts as

$$\tilde{M}^C_t := \int_{[0, t\wedge\tau^C]\times E\times\mathbb{R}_+} \Upsilon\Big(\eta(e), \tilde{\lambda}^C_{s-} - \big|\Gamma_s\big(z(e,y)\big)\big|\Big) 1_{\{y \leq g(\tilde{\lambda}^C_{s-})\}} N(ds, de, dy), \tag{3.2}$$

and proceed similarly for the liquidity-providing limit orders

$$\tilde{L}^{C,+}_t := \int_{[0, t\wedge\tau^C]\times E\times\mathbb{R}_+} \rho^+(e) 1_{\{\tilde{\lambda}^C_{s-} - |\Gamma_s(z(e,y))| \geq \underline{\lambda}\}\cap\{y \leq f(\tilde{\lambda}^C_{s-})\}} N(ds, de, dy) \tag{3.3}$$

and cancellations

$$\tilde{L}^{C,-}_t := \int_{[0, t\wedge\tau^C]\times E\times\mathbb{R}_+} \Upsilon\Big(\rho^-(e), \tilde{\lambda}^C_{s-} - \big|\Gamma_s\big(z(e,y)\big)\big|\Big) 1_{\{y \leq f(\tilde{\lambda}^C_{s-})\}} N(ds, de, dy). \tag{3.4}$$

The order flow is frozen after τ^C, and at time τ^C, the function Υ in (3.2) and (3.4) and the indicator in (3.3) ensure that trading is halted after the liquidity has been depleted beyond $\underline{\lambda}$.

Let us finally record that $(\lambda^C, \tilde{\lambda}^C, \tilde{Q}^C, \tilde{X}^C, \tilde{P}^C, \tilde{L}^C, \tilde{M}^C)$ jointly solve the SDE system

$$\begin{aligned}
\lambda^C_{0-} &= \lambda, & \diamond\, d\lambda^C &= \diamond\, \big((-|dC^\ell|) \vec{+} (d\tilde{L}^C - |d\tilde{M}^C|) \vec{+} (-|dC^{r,c}|)\big),\\
\tilde{\lambda}^C_{0-} &= \lambda', & \diamond\, d\tilde{\lambda}^C &= -1_{\{\lambda^C \geq \underline{\lambda}\}} \diamond\, \big((-|dC^\ell|) \vec{+} (d\tilde{L}^C - |d\tilde{M}^C|) \vec{+} (-|dC^{r,c}|)\big),\\
\tilde{Q}^C_{0-} &= q, & \diamond\, d\tilde{Q}^C &= 1_{\{\lambda^C \geq \underline{\lambda}\}} \diamond dC,\\
\tilde{X}^C_{0-} &= x, & \diamond\, d\tilde{X}^C &= -\tilde{P}^C \diamond d\tilde{Q}^C - \zeta \diamond |d\tilde{Q}^C|,\\
\tilde{P}^C_{0-} &= p, & \diamond\, d\tilde{P}^C &= \iota(\tilde{\lambda}^C) 1_{\{\lambda^C \geq \underline{\lambda}\}} \diamond (dC^\ell \vec{+} d\tilde{M}^C \vec{+} dC^{r,c}),\\
\tilde{L}^C_{0-} &= 0, & d\tilde{L}^C &= 1_{\{y \leq f(\lambda^C_-),\ \lambda^C_- \geq \underline{\lambda}\}} \rho(e) N(dt, de, dy),\\
\tilde{M}^C_{0-} &= 0, & d\tilde{M}^C &= 1_{\{y \leq g(\lambda^C_-),\ \lambda^C_- \geq \underline{\lambda}\}} \eta(e) N(dt, de, dy).
\end{aligned}$$

Also here the "intermediate" values λ_t^C, $\tilde{\lambda}_t^C$, $\tilde{Q}_t^C$, $\tilde{X}_t^C$, $\tilde{P}_t^C$ at time t (rather than $t+$) can be obtained from these dynamics by considering the Marcus-style dynamics at the moment when its fictitious flow at time t has run two thirds of its course (i.e., when, in the notation of Remark 2.1, at the fictitious time $s = 2/3$, two of the three jump contributions at time t have been taken care of).

At the terminal time T, any remaining inventory $\tilde{Q}_{T+}^C$ has to be liquidated. If trading has been halted or will be during this liquidation, an auction determines the new mid-price from which the remaining liquidation is executed with price impact as before. For simplicity, we assume that this new mid-price deviates from the pre-halt mid-price by a shock

$$\begin{aligned}&Y \text{ independent of } \mathcal{F}_T \text{ with}\\&\mathcal{L}_Y(a) := \log \mathbb{E}[\exp(aY)] < \infty \text{ symmetric in } a \in \mathbb{R}.\end{aligned} \tag{3.5}$$

For instance, any centered and independent Gaussian Y will do. As a result, our trader's terminal wealth is

$$\tilde{W}_{T+}^C := \mathtt{w}(\tilde{X}_{T+}^C, \tilde{Q}_{T+}^C, \tilde{P}_{T+}^C, \tilde{\lambda}_{T+}^C) + Y\mathtt{r}(\tilde{Q}_{T+}^C, \tilde{\lambda}_{T+}^C). \tag{3.6}$$

Here, the function $\mathtt{w}$ of (2.25) provides the cost-adjusted liquidation value of the investor's position. The Y-term adjusts this further to account for the part of the investor's position that goes into auction and is thus additionally exposed to the mid-price shock Y. This part is given by the function

$$\mathtt{r}(q, \lambda) := \big(|q| - (\lambda - \underline{\lambda})^+\big)^+.$$

Indeed, if at the liquidity level λ, a trade of size q has to be executed, this is possible without triggering a halt only if $|q|$ does not exceed the available liquidity $(\lambda - \underline{\lambda})^+$. Otherwise, this available liquidity is exhausted and the remaining bits of the order $\mathtt{r}(q, \lambda)$ are auctioned.

3.2 The utility maximisation problem

Let us consider a trader who maximises her expected utility from terminal wealth over the trading interval $[0, T]$ as given by the performance criterion

$$J(C) := \mathbb{E}[U(\tilde{W}_{T+}^C)] \qquad \text{with } U(w) = -\exp(-\alpha w),$$

where $\tilde{W}_{T+}^C$ in (3.6) specifies the investor's terminal wealth and $\alpha > 0$ measures her (constant) absolute risk aversion.

For continuity of the resulting value function, we need to assume that at any time, the trader can take the remaining liquidity without triggering a trading halt and that she can also trade exactly to the extreme inventory levels $-\bar{q}$ and $\bar{q}$. Obviously, this is not a problem when any order size is allowed; but when trading is in lots of size δ, this may not be possible. Thus we let

$$D(\lambda, q) := \mathbb{R}$$

or

$$D(\lambda, q) := \{\dots, -2\delta, -\delta, 0, \delta, 2\delta, \dots\} \cup \{\lambda - \underline{\lambda}, \underline{\lambda} - \lambda, \bar{q} - q, -\bar{q} - q\}$$

and define the set of admissible strategies when starting with liquidity λ and in position q as

$$\begin{aligned}\tilde{\mathcal{C}}(\lambda, q) := \{C : \;&\Lambda\text{-measurable process of bounded variation with } C_{0-} = 0\\ &\text{satisfying (2.13) and } \Delta_t^\ell C \in D(\lambda_{t-}^C, Q_{t-}^C), \Delta_t^r C \in D(\lambda_t^C, Q_t^C)\\ &\text{for } t \in [0, T]\}.\end{aligned}$$

Now the value function can be defined as

$$v(T, \mathbf{s}) = v(T, \lambda, q, p, x) := \sup_{C \in \tilde{\mathcal{C}}(\lambda, q)} \mathbb{E}[U(\tilde{W}_{T+}^C)]. \tag{3.7}$$

Theorem 3.2 *The value function v in (3.7) for our utility maximisation problem with trading halts has the multiplicative structure*

$$v(T, \lambda, q, p, x) = \exp\big(-\alpha(x + pq)\big) v_0(T, \lambda, q), \tag{3.8}$$

where $v_0 = v(\cdot, \cdot, \cdot, 0, 0)$ is the value function for initial data $x = 0$, $p = 0$. Moreover, the value function is non-degenerate in the sense that it is strictly less than $U(\infty) = 0$ everywhere.

Suppose that the distributions $\nu \circ (\rho^-)^{-1}$ of the cancellations and $\nu \circ |\eta|^{-1}$ of the external market orders are continuous with a possible atom at 0 and both have a finite second moment. Assume in addition that the posting of market orders is controlled in the sense that their maximum arrival rate is finite, i.e., $g(\infty) < \infty$. Finally, let the size of limit orders be controlled in the sense that

$$\int_E \exp\big(c\rho^+(e)\big)\nu(de) < \infty \qquad \text{for all } c \in [0, \infty). \tag{3.9}$$

Then the value function $(T, \lambda, q, p, x) \mapsto v(T, \lambda, q, p, x)$ is continuous on the set $[0, \infty) \times [\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}] \times \mathbb{R}^2$.

We defer the proof of this result to Appendix B. One cannot follow standard arguments in our setting because the external order flow, at least pathwise, does not depend continuously on the control. For instance, small differences in liquidity levels can allow external orders to affect one strategy, but not another one which has almost the same liquidity level. This may further widen the difference in liquidity for these strategies, making it ever more difficult for one of them to for instance copy the other's trades and do so under similar market conditions and thus with similar profitability. For reasons like this, the relevant analysis is rather challenging and requires some care; see in particular Lemma B.2 below.

Let us briefly comment on the extra assumptions we need for our continuity result for the value function. Continuity of the distributions of the cancellation and market order sizes is needed because with an atom, say at 1 lot, a trader with liquidity $\underline{\lambda} + 1$ has no risk of seeing a halt triggered by an external order, whereas there is a strictly positive frequency of this disruption happening when the liquidity is just below this level. The integrability and frequency assumptions on the external order flow are of little practical concern and are made for technical reasons as we need to bound the order flow fluctuations; see in particular the proof of Corollary B.4.

In Sect. 4, we show how one can use standard techniques from dynamic programming and the theory of Hamilton–Jacobi–Bellman (HJB) equations to derive numerical approximations of solutions to the trader's problem which we describe in the next section.

3.3 Performance of signal-based trading strategies

In the following, we report the results from a numerical scheme to solve the HJB equation (4.6) from Sect. 4.1 below and illustrate the trader's signal-based strategy in an optimal acquisition problem. Section 4.2 provides details of the numerical scheme we used.

In our benchmark, the trader receives, with probability $\hat{p}$, a private signal about limit orders, market orders and cancellations of limit orders. Compared to a trader without the signal, the trader with the signal optimises her execution times and trading volumes by splitting the parent order into fewer and larger child orders than those of the trader without signal. She executes her orders when a signal about a liquidity-taking order arrives. Signals about liquidity provision are not relevant for the execution programme of a risk-averse trader because there is no incentive to execute before the liquidity-providing order increases the liquidity in the book and the price impact of market orders decreases.

Speculative trades are not profitable in the benchmark due to the size of the bid–ask spread. In contrast, when the bid–ask spread is narrow, speculation based on signals can become profitable and, in fact, we see the trader initiates speculative trades upon receiving a signal about liquidity provision at times when the level of liquidity is low. After waiting for the liquidity to improve and the cost of price impact to decline, she unwinds the speculative position upon receiving a signal about the imminent arrival of a liquidity-taking order.

3.3.1 Parameter specification: benchmark case

In the simulations, we consider six types of external market orders: buy orders of one, two and three lots, and sell orders of one, two and three lots. Similarly, limit orders and their cancellations are of size one, two and three lots.

We choose the arrival rates f and g as

$$f(\lambda) = \theta_f \exp(-\kappa_f \lambda), \qquad g(\lambda) = \theta_g \exp(\kappa_g \lambda), \tag{3.10}$$

where $\theta_f = 40$, $\theta_g = 20$ and $\kappa_f = \kappa_g = 0.01$, i.e., if the liquidity λ is at the level 0, the market expects to receive 20 external market orders, 30 limit orders and 10 cancellations of limit orders over a trading interval of length $T = 1$.

The price impact function is given by

$$\iota(\lambda) = \theta_\iota + \kappa_\iota \lambda, \tag{3.11}$$

where $\theta_\iota = 0.01$ and $\kappa_\iota = -0.0002$.

For numerical reasons, we introduce an upper bound $\overline{\lambda}$ for the liquidity parameter λ so that we obtain a bounded domain $[\underline{\lambda}, \overline{\lambda}]$ for λ, where $\underline{\lambda}$ is the lower bound from the trading halt mechanism as in (3.1). We set the lower bound at $\underline{\lambda} = -40$ and the upper bound at $\overline{\lambda} = 40$ so that $\iota(\lambda) \geq 0$ for every $\lambda \in [\underline{\lambda}, \overline{\lambda}]$. Moreover, we consider the inventory q in the bounded domain $[-\overline{q}, \overline{q}]$ with $\overline{q} := 10$.

With these parameter values, the elasticity condition (2.11) holds. More precisely, for the specification of ι given in (3.11), the denominator $-\frac{\partial_\lambda \mathbb{I}^2(\lambda)}{\mathbb{I}^2(\lambda)}$ in the elasticity condition (2.11) lies in the range $\approx [0.022, 0.177]$ for $\lambda \in [-40, 40]$. Hence (2.11) requires $\kappa_f = \frac{\partial_\lambda g(\lambda)}{g(\lambda)} < 0.022$ and thus yields an upper bound for this exponential arrival rate of market orders.

The independent auction mark-up Y on the mid-price in (3.6) is $\mathcal{N}(0, \sigma^2)$-distributed with $\sigma = 0.3$. The bid–ask spread is set to one cent, i.e., $2\zeta = 0.01$.

Finally, the trader's risk aversion is $\alpha = 0.1$.

3.3.2 Signal design

The signal is given by the variable Z_t. A value of $Z_t = 1$ signals an incoming limit order, and a value of $Z_t = -1$ signals liquidity taking, i.e., an incoming market order or the cancellation of a limit order. With fixed probability $\hat{p} \in [0, 1]$, the signal alerts the trader of an imminent order, and so an external order has a $1 - \hat{p}$ chance of taking the trader by surprise.

Thus the signal informs the trader about the sign of the imminent change in liquidity, but does not provide any information about the size of an order, i.e., the trader anticipates the direction, but not the magnitude of changes in liquidity. In particular, in the case of a signal $Z_t = -1$ about liquidity taking, the trader does not know whether a buy market order, a sell market order or the cancellation of a limit order will be posted.

Note that the definition (2.12) allows a huge variety of conceivable signals, and the above specification is chosen to simplify the implementation.

3.3.3 Formal model specification as in Sect. 3.3

In this part, we follow the notations in Chap. 2. For simplicity, we only treat the benchmark; the specification of the remaining settings is similar.

First, the mark space is given by

$$E = \{-3, -2, -1, 1, 2, 3\} \times \{-3, -2, -1, 1, 2, 3\} \times \{-1, 1\}.$$

The components e_1 and e_2 of a mark $e = (e_1, e_2, e_3) \in E$ describe the volume of a possible market and of a limit order, respectively; component e_3 decides which of these two possibilities materialises. The mappings η and ρ are

$$\eta(e) = e_1 1_{\{-1\}}(e_3) \quad \text{and} \quad \rho(e) = e_2 1_{\{1\}}(e_3) \qquad \text{for } e = (e_1, e_2, e_3) \in E.$$

The model is driven by the homogeneous Poisson point process $N(dt, de, dy)$ with compensator $dt \otimes \nu(de) \otimes dy$. Here, we have $\nu(de) = P^1(de_1) \otimes P^2(de_2) \otimes P^3(de_3)$, where P^1 is the distribution of the mark that determines the market order size, P^2 is the distribution of the mark that determines the limit order size and P^3 is a coin flip deciding whether a market or a limit order arrives in the book.

In our case study, we use the specifications

$$
\begin{aligned}
&P^1(\{-3\}) = P^1(\{3\}) = 0.15,\\
&\quad P^1(\{2\}) = P^1(\{-2\}) = P^1(\{-1\}) = P^1(\{1\}) = 0.175,\\
&P^2(\{-3\}) = 0.1, \quad P^2(\{-2\}) = P^2(\{-1\}) = 0.075,\\
&\quad P^2(\{1\}) = P^2(\{2\}) = P^2(\{3\}) = 0.25,\\
&P^3(\{-1\}) = P^3(\{1\}) = 0.5.
\end{aligned}
$$

As the arrival rates of market and limit orders are halved by the coin flip, we double f and g in (3.10), i.e., we take $\hat{f} = 2f$ and $\hat{g} = 2g$, to ensure that in expectation, the market still receives 20 external market orders, 30 limit orders and 10 cancellations of limit orders over the trading interval [0, 1].

The signal process Z is

$$
Z_t := \int_{\{t\}\times E\times\mathbb{R}_+} z(e, y) N(ds, de, dy) \qquad \text{with } z(e, y) := (e_3, y).
$$

We disintegrate for $z = z(e, y) \in \{-1, 1\} \times \mathbb{R}_+$ the jump measure

$$
\nu(de) \otimes dy = \int_{\{-1,1\}\times[0,\infty)} K(z; de, dy)\mu(dz)
$$

into

$$
\begin{aligned}
\mu(dz) &= \big(P^3(\{-1\})\mathrm{Dirac}_{-1} + P^3(\{1\})\mathrm{Dirac}_1\big)(dz_1) \otimes dy,\\
K(z_1, y; \cdot, \cdot) &= P^1 \otimes P^2 \otimes \mathrm{Dirac}_{z_1} \otimes \mathrm{Dirac}_y,
\end{aligned}
$$

where we generically denote $z = (z_1, y)$.

3.3.4 Value function and certainty equivalent

We implement the numerical scheme from Sect. 4.2 to solve the HJB equation (4.7). Figure 2 shows the value function v in (3.7) for $p = x = 0$ and when the probability of receiving a signal is $\hat{p} = 0.2$.

The value function is symmetric about $q = 0$ due to the symmetry of (2.7) and (2.17) with respect to buy and sell orders. Hence the expected utility of terminal wealth for an acquisition or liquidation programme is the same, everything else being equal.

Moreover, the value function is increasing in the liquidity λ because execution costs decrease as the liquidity in the market increases. When the liquidity approaches the liquidity trigger $\underline{\lambda}$, the value function is steepest because of the additional risk of

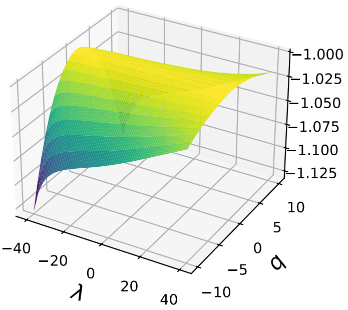

Fig. 2 Value function v in (3.7) for $p = x = 0$

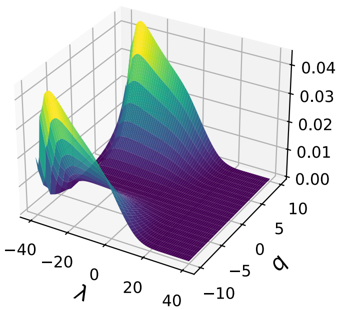

Fig. 3 Certainty equivalent for a trader who receives a signal with probability $\hat{p} = 0.2$ compared to a trader without signal, for spread 0.01

entering an auction through a trading halt. Similarly, the value function is particularly steep for large values of $|q|$ because a large execution programme is linked to higher risk in price and liquidity changes.

Figure 3 illustrates the signal-specific certainty equivalent of the value function when the probability of receiving a signal is $\hat{p} = 0.2$. This certainty equivalent corresponds to the additional amount of initial wealth that is needed when the trader does not receive a signal to achieve the same expected utility as in the case with signal, i.e., the certainty equivalent CE is such that

$$v_{\hat{p}=0}(T, q, \lambda, p, x + \mathrm{CE}) = v_{\hat{p}=0.2}(T, q, \lambda, p, x). \tag{3.12}$$

We use (3.8) and rearrange (3.12) to write

$$\mathrm{CE} = -\frac{1}{\alpha} \log \frac{v_{\hat{p}=0.2}(T, q, \lambda, p, x)}{v_{\hat{p}=0}(T, q, \lambda, p, x)}. \tag{3.13}$$

The certainty equivalent is symmetric about q and does not depend on p or x. Moreover, the signal is worth most for large values of the inventory, long or short. At most, the trader can make up to four cents from the information of the signal, which corresponds to four times the bid–ask spread; see Fig. 3.

The certainty equivalent is decreasing in the liquidity component λ because as the liquidity decreases, the execution times become more relevant because price impact is more costly. Moreover, the certainty equivalent increases slightly towards the liquidity trigger $\underline{\lambda} = -40$. Here, the signal warns the trader of potentially reaching the lower bound $\underline{\lambda}$ and reduces the risk of executing a final trade in an auction.

For liquidity greater than 20, the certainty equivalent is zero, i.e., the signal does not add any value because with and without a signal, a trader immediately completes the execution programme.

3.3.5 Benchmark: the signal optimises times of execution and trading volume

In the benchmark with spread 0.01, we look at the optimal acquisition problem where the trader acquires 8 lots of the asset over the time interval $[0, T]$; so the initial inventory is $q = -8$. The trader can buy and sell any multiple of lots and can execute speculative trades by trading away from the target of buying only 8 lots over the trading window.

When there is no private signal, the trader sends child orders of at most two lots when the liquidity is sufficiently high; see Fig. 4. However, when the trader receives private signals, she times her executions with those of the signal with information about liquidity-taking orders and optimises the trading volume of her trades; see Fig. 5 (simulated with the same seed as that for the results in Fig. 4). Here she executes her acquisition programme by sending two child orders of sizes 3 and 5 lots upon receiving a signal about liquidity taking when the level of liquidity is high. When the liquidity is low during $t \in [0.15, 0.7]$, the trader does not act on the signal and prefers to wait for the liquidity to increase. Note that only the signals about liquidity-taking orders are relevant in her execution programme.

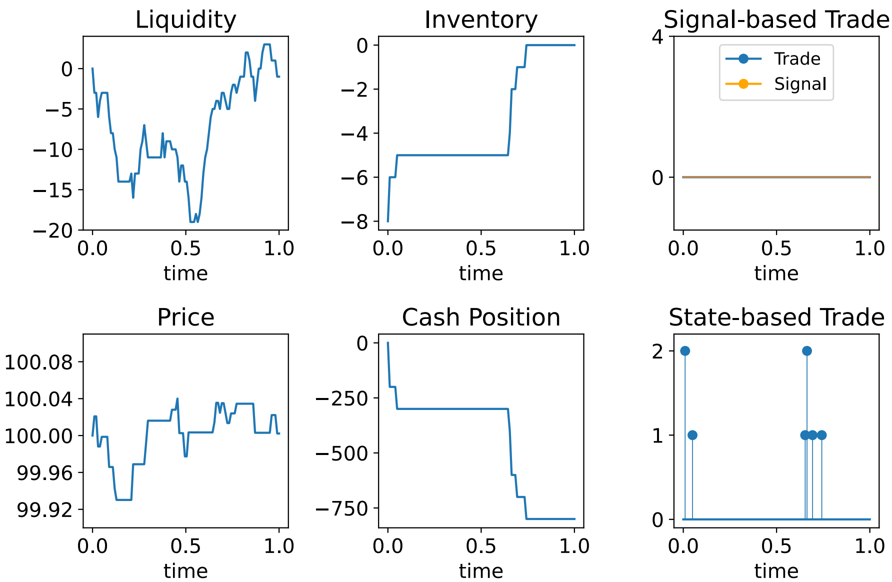

Fig. 4 Optimal acquisition of 8 lots. Pathwise plot when the trader does not receive a signal and with spread 0.01

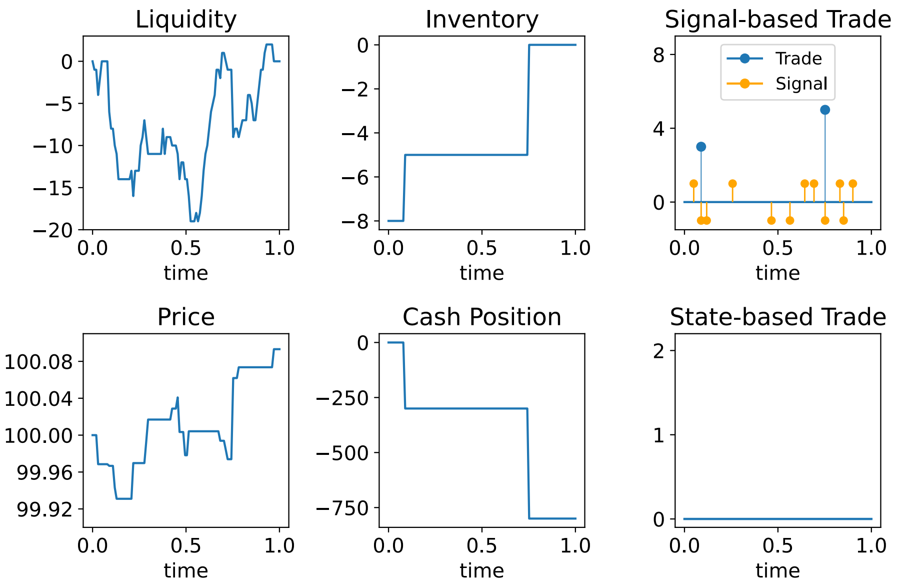

Fig. 5 Optimal acquisition of 8 lots. Pathwise plot when the trader receives a signal with probability $\hat{p} = 0.2$ and with spread 0.01

3.3.6 More signals lead to lower execution costs and riskier strategies

Figure 6 illustrates the performance of signal-based strategies depending on the probability of receiving a signal. Figure 6 (left) shows the distribution of the terminal wealth for different probabilities of receiving a signal compared to the distribution without a signal. When the probability of receiving a signal increases, the densities shift to the right so that the signal increases the expected terminal wealth by optimising the execution. On the other hand, as the probability of receiving a signal increases, also the variance of terminal wealth increases because the trader takes more risks due to the additional information through the private signal. More precisely, knowing that

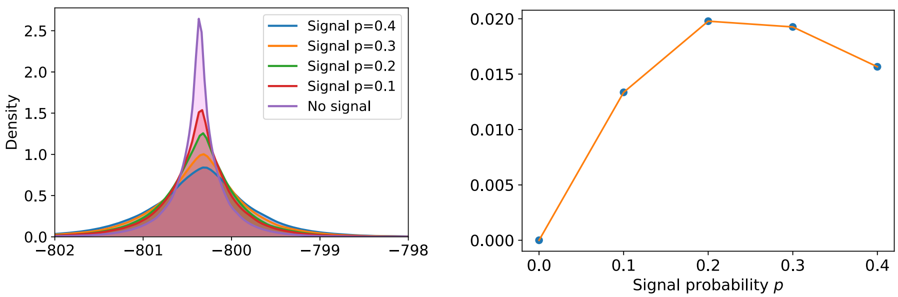

Fig. 6 Optimal acquisition of 8 lots. Performance of a trader without a signal compared to a trader who receives a signal with probability $\hat{p} \in \{0, 1, 0.2, 0.3, 0.4\}$ and with spread 0.01. Left: Distribution of terminal wealth. Right: Signal Sharpe ratio

she will be informed about liquidity shocks through the signal, the trader waits longer for the liquidity to recover from shocks before trading to minimise price impact costs. However, this is linked to more risk because unsignaled external orders can still arrive which increases the variance of the trader's terminal wealth.

To quantify the value of the private signal Z, we introduce the signal Sharpe ratio (SSR)

$$\mathrm{SSR}(Z) := \frac{\bar{W}(Z) - \bar{W}(0)}{\sigma(Z)},$$

where

$$\bar{W}(Z) := \frac{1}{n_{\mathrm{sim}}} \sum_{j=1}^{n_{\mathrm{sim}}} \tilde{W}^j_{T+}(Z), \qquad \bar{W}(0) := \frac{1}{n_{\mathrm{sim}}} \sum_{j=1}^{n_{\mathrm{sim}}} \tilde{W}^j_{T+}(0)$$

and

$$\sigma(Z) := \sqrt{\frac{1}{n_{\mathrm{sim}}} \sum_{j=1}^{n_{\mathrm{sim}}} \big(\tilde{W}^j_{T+}(Z) - \bar{W}(Z)\big)^2}.$$

Here $\tilde{W}^j_{T+}(Z)$ and $\tilde{W}^j_{T+}(0)$ denote the terminal wealth for scenario $j \in \{1, \dots, n_{\mathrm{sim}}\}$ with signal Z and without signal, respectively. Precisely, the SSR is the excess return $\bar{W}(Z) - \bar{W}(0)$ for a trader with signal Z compared to a trader without signal, weighted by the risk $\sigma(Z)$ of a trader with signal Z.

Figure 6 (right) shows the SSR as a function of the probability $\hat{p}$ of receiving a signal. The SSR increases up to $\hat{p} = 0.2$ and slowly decreases thereafter because the increase in variance dominates the increase in expected terminal wealth, which leads to a non-monotonicity in Fig. 6 (left). Note that mean–variance optimisation is not the objective of the trader who maximises expected utility of terminal wealth; see (3.7).

3.3.7 Signal is more valuable for speculation as bid–ask spread narrows

Next, consider a bid–ask spread of size 0.002, i.e., the bid–ask spread is one fifth of that in the benchmark. In this case, the signal incentivises speculative trades because everything else being equal, the costs to execute round-trip trades are lower.

The speculative round-trip trades start after receiving a signal on liquidity provision, i.e., when the trader knows through the signal that the speculation can be unwound at better liquidity; see Fig. 7. After triggering liquidity provision through her own market order, the trader waits for liquidity to arrive in the market to unwind the speculative position with better liquidity upon receiving a signal about a liquidity-taking order.

Next, we study the number of paths with speculative trades among $n_{\mathrm{sim}} = 100'000$ simulated paths with bid–ask spread 0.01 and 0.002 in the acquisition ($q = -8$) and pure speculation ($q = 0$) scenarios, where the paths for wide and narrow spreads are simulated with the same seed. We say that a path contains a speculative trade if the trader trades away from the target of buying 8 lots over the time interval [0, 1] by trading with both buy and sell market orders.

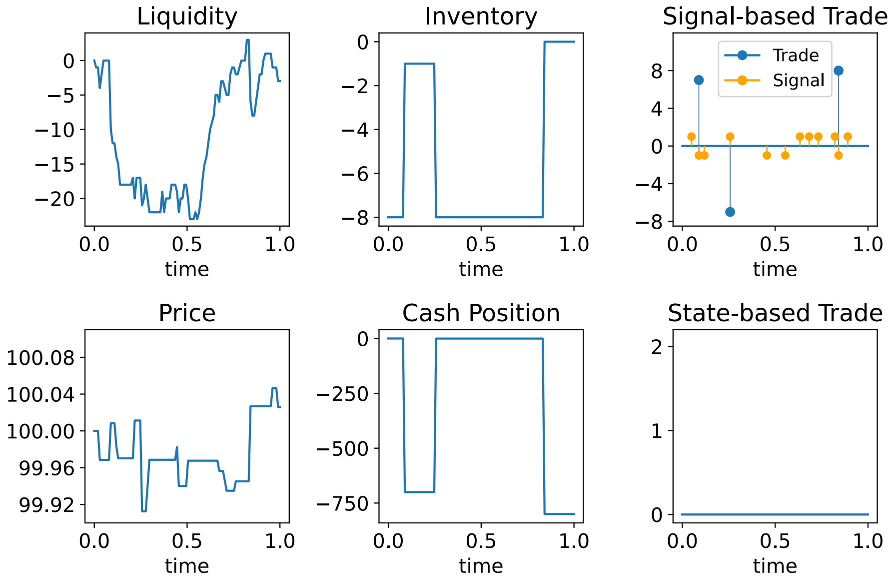

Fig. 7 Optimal acquisition of 8 lots. Pathwise plot when the trader receives a signal with probability $\hat{p} = 0.2$ and with spread 0.002

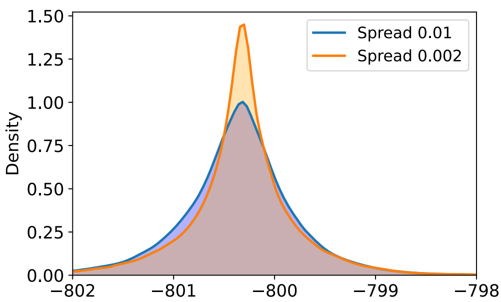

Fig. 8 Optimal acquisition of 8 lots. Distribution of terminal wealth when the trader receives a signal with probability $\hat{p} = 0.2$ with a spread of 0.01 and 0.002

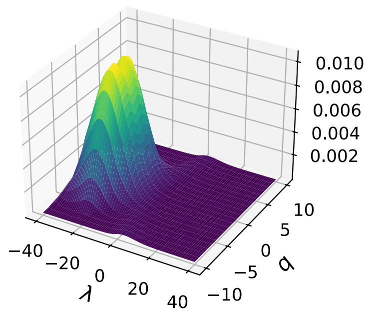

Fig. 9 Certainty equivalent for a small spread of 0.002 minus the certainty equivalent for a spread of 0.01

When the spread is wide (at 0.01), there is no speculation because round-trip trades are too costly. On the other hand, when the spread is narrow (at 0.002), there is speculation in about 22% of paths in the acquisition problem ($q = -8$) and in about 11% of paths in the pure speculation scenario ($q = 0$). There are more speculative paths in the execution example because the trader's execution of trades triggers liquidity provision and speculative trades become profitable. Especially, when the trader trades towards a position close to zero very early in the trading window, i.e., completes the acquisition problem early on, the remaining time interval is long enough so that a speculative round-trip starting close to zero is profitable, and vice versa if the trader completes the acquisition at a later point in the trading window. Finally, without signal, there is no speculation for both the wide and the narrow bid–ask spread.

Figure 8 shows the distribution of the terminal wealth in the optimal acquisition problem of $q = -8$ when the trader receives a private signal with probability $\hat{p} = 0.2$ and the bid–ask spread is 0.01 and 0.002. With a narrow spread, the terminal wealth is on average higher because the trader pays less spread and profits from speculative trades. The terminal wealth increases most in the interval $[-801, -800.5]$, see Fig. 8, because low levels of liquidity, which cause smaller values of the terminal wealth for a spread of 0.01 due to price impact costs, can lead to profitable round-trips when the spread is at 0.002. Similarly, the variance of the terminal wealth decreases because in general, the trader executes more orders to bring her inventory to zero and only deviates from this target in about 22% of cases.

Similarly, Fig. 9 illustrates the certainty equivalent in (3.13) for spread 0.002 minus the certainty equivalent in (3.13) for spread 0.01. The certainty equivalent improves when the liquidity is low and when the trader's inventory is close to zero which is when speculative trades are executed.

3.3.8 Signal on liquidity-taking orders versus signal on liquidity-providing orders

Finally, we compare the performance of the strategy of a trader who receives a signal about the arrival of liquidity-taking orders against the strategy of a trader who receives a signal about the arrival of liquidity provision. The trader with a signal on liquidity taking uses the signal to optimise the execution times and trading volumes, which increases the terminal wealth and its variance; see Fig. 10. In contrast, similarly to Fig. 5, the trader does not use the signal about liquidity provision for her execution so that her terminal wealth coincides with that of the trader who receives no signal; see Fig. 10.

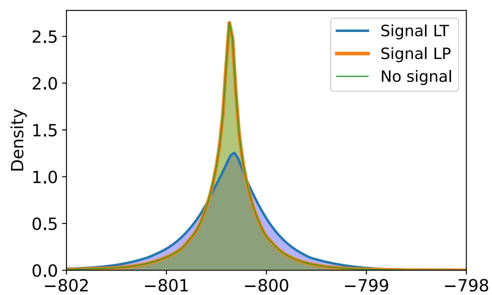

Fig. 10 Optimal acquisition of 8 lots. Distribution of terminal wealth for a trader who receives a signal on liquidity taking (LT) with probability $\hat{p} = 0.2$ versus a trader who receives a signal on liquidity provision (LP) with probability $\hat{p} = 0.2$; spread is 0.01

4 The Hamilton–Jacobi–Bellman equation and its numerical solution

In this section, we present the derivation of the HJB equation and the numerical scheme we used for our illustrations in the preceding section.

4.1 The Hamilton–Jacobi–Bellman equation

In the following, we derive the HJB equation that is satisfied by the value function v in (3.7). In particular, we investigate for which conditions the value process $V_t^C := v(\bar{T}-t, \tilde{S}_{t+}^C)$, $t \in [0, \bar{T}]$ with arbitrary but fixed time horizon $\bar{T}$, satisfies supermartingale dynamics for every admissible $C \in \tilde{\mathcal{C}}(\lambda, q)$ and martingale dynamics for some strategy $C^* \in \tilde{\mathcal{C}}(\lambda, q)$ (which is then optimal).

We adopt the notation from the representation of C in Lemma 2.4 and suppose that the value function $v = v(T, \lambda, q, p, x)$ is sufficiently smooth to apply Itô's formula. Thus we formally write on $[\![0, \tau^C \wedge \bar{T}]\!]$ the dynamics of V^C as

$$
\begin{aligned}
& dV_t^C \\
& = -\frac{\partial v}{\partial T}(\bar{T}-t, \tilde{S}_{t-}^C)dt \\
& \quad + \bigg(\frac{\partial v}{\partial p}(\bar{T}-t, \tilde{S}_{t-}^C)\iota(\tilde{\lambda}_{t-}^C) - \frac{\partial v}{\partial x}(\bar{T}-t, \tilde{S}_{t-}^C)\tilde{P}_{t-}^C + \frac{\partial v}{\partial q}(\bar{T}-t, \tilde{S}_{t-}^C)\bigg)dC_t^c \\
& \quad + \bigg(\frac{\partial v}{\partial x}(\bar{T}-t, \tilde{S}_{t-}^C)\zeta - \frac{\partial v}{\partial \lambda}(\bar{T}-t, \tilde{S}_{t-}^C)\bigg)|dC_t^c| \\
& \quad + \int_{E\times\{z\neq 0\}} \bigg(v\Big(\bar{T}-t, \tilde{\mathfrak{s}}\Big(\Gamma_t\big(z(e,y)\big), \eta(e)1_{\{y\le f(\tilde{\lambda}_{t-}^C)\}}, \rho(e)1_{\{y\le f(\tilde{\lambda}_{t-}^C)\}}; \tilde{S}_{t-}^C\Big)\Big) \\
& \qquad\qquad - v(\bar{T}-t, \tilde{S}_{t-}^C)\bigg)N(dt, de, dy) \\
& \quad + \int_{E\times\{z=0\}} \bigg(v\Big(\bar{T}-t, \tilde{\mathfrak{s}}\big(0, \eta(e)1_{\{y\le g(\tilde{\lambda}_{t-}^C)\}}, \rho(e)1_{\{y\le f(\tilde{\lambda}_{t-}^C)\}}; \tilde{S}_{t-}^C\big)\Big) \\
& \qquad\qquad - v(\bar{T}-t, \tilde{S}_{t-}^C)\bigg)N(dt, de, dy) \\
& \quad + d\bigg(\sum_{s\le t}\Big(v\big(\bar{T}-s, \tilde{\mathfrak{s}}(\Delta_s^{\ell,p}C, 0, 0; \tilde{S}_{s-}^C)\big) - v(\bar{T}-s, \tilde{S}_{s-}^C)\Big)\bigg) \\
& \quad + d\bigg(\sum_{s\le t}\Big(v\big(\bar{T}-s, \tilde{\mathfrak{s}}(\Delta_s^{r}C, 0, 0; \tilde{S}_{s}^C)\big) - v(\bar{T}-s, \tilde{S}_{t}^C)\Big)\bigg),
\end{aligned}
\tag{4.1}
$$

where we used that the predictable left jumps $\Delta_t^{\ell,p}C$ cannot coincide with mark times; cf. (2.16).

For V^C to be a supermartingale for *any* admissible C, we see from the last two lines that

$$
\sup_{\Delta\in D(\lambda,q)} \Big(v\big(T, \tilde{\mathfrak{s}}(\Delta, 0, 0; s)\big) - v(T, s)\Big) \le 0 \tag{4.2}
$$

must hold for $(T, s) \in [0, \infty) \times [\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}] \times \mathbb{R}^2$. When $D(\lambda, q) = D = \mathbb{R}$ (i.e., continuous trading is admissible), the above condition ensures that contributions from the dC_t^c terms in dV_t^C lead to supermartingale dynamics. Indeed, upon formal differentiation of $v(T, \tilde{\mathrm{s}}(\Delta, 0, 0; s)) - v(T, s)$ from above and from below at $\Delta = 0$, the condition entails in particular $\frac{\partial v}{\partial x}\zeta - \frac{\partial v}{\partial \lambda} \pm (\frac{\partial v}{\partial p}\iota(\lambda) - \frac{\partial v}{\partial x}p + \frac{\partial v}{\partial q}) \leq 0$. As a result, the netted integrands of $dC^{c,\pm}$ in (4.1) are nonpositive as required.

Next, we compensate the dN-integrals which give local martingales and drifts accumulating with rates

$$\begin{aligned}
&\int_{(E\times\mathbb{R}_+)\cap\{z\neq 0\}} \bigg(v\Big(\bar{T}-t, \tilde{\mathrm{s}}\Big(\Gamma_t\big(z(e,y)\big), \eta(e)1_{\{y\leq g(\tilde{\lambda}^C_{t-})\}}, \rho(e)1_{\{y\leq f(\tilde{\lambda}^C_{t-})\}}; \tilde{S}^C_{t-}\Big)\Big) \\
&\qquad\qquad - v(\bar{T}-, \tilde{S}^C_{t-})\bigg)\big(\nu(de)\otimes dy\big) \qquad (4.3) \\
&+\int_{(E\times\mathbb{R}_+)\cap\{z=0\}} \bigg(v\Big(\bar{T}-t, \tilde{\mathrm{s}}\big(0, \eta(e)1_{\{y\leq g(\tilde{\lambda}^C_{t-})\}}, \rho(e)1_{\{y\leq f(\tilde{\lambda}^C_{t-})\}}; \tilde{S}^C_{t-}\big)\Big) \\
&\qquad\qquad - v(\bar{T}-t, \tilde{S}^C_{t-})\bigg)\big(\nu(de)\otimes dy\big).
\end{aligned}$$

We use the disintegration

$$\nu(de)\otimes dy = \int_{z(E\times\mathbb{R}_+)\setminus\{0\}} K(\bar{z}; de, dy)\mu(d\bar{z}) \qquad (4.4)$$

with $\mu = (\nu \otimes \mathrm{Leb}) \circ (z)^{-1}$ to rewrite and estimate the first integral in (4.3) as

$$\begin{aligned}
&\int_{(E\times\mathbb{R}_+)\cap\{z\neq 0\}} \bigg(v\Big(\bar{T}- \ \tilde{\mathrm{s}}\Big(\Gamma_t\big(z(e,y)\big), \eta(e)1_{\{y\leq g(\tilde{\lambda}^C_{t-})\}}, \rho(e)1_{\{y\leq f(\tilde{\lambda}^C_{t-})\}}; \tilde{S}^C_{t-}\Big)\Big) \\
&\qquad\qquad - v(\bar{T}-t, \tilde{S}^C_{t-})\bigg)\big(\nu(de)\otimes dy\big) \\
&= \int_{z(E\times\mathbb{R}_+)\setminus\{0\}} \int_{(E\times\mathbb{R}_+)\cap\{z=\bar{z}\}} \bigg(v\Big(\bar{T}-t, \tilde{\mathrm{s}}\big(\Gamma_t(\bar{z}), \eta(e)1_{\{y\leq g(\tilde{\lambda}^C_{t-})\}}, \\
&\qquad\qquad \rho(e)1_{\{y\leq f(\tilde{\lambda}^C_{t-})\}}; \tilde{S}^C_{t-}\big)\Big) \\
&\qquad\qquad - v(\bar{T}-t, \tilde{S}^C_{t-})\bigg) K(\bar{z}; de, dy)\mu(d\bar{z}) \\
&\leq \int_{z(E\times\mathbb{R}_+)\setminus\{0\}} \sup_{\gamma\in D(\tilde{\lambda}^C_{t-}, \tilde{Q}^C_{t-})} \int_{(E\times\mathbb{R}_+)\cap\{z=\bar{z}\}} \bigg(v\Big(\bar{T}-t, \tilde{\mathrm{s}}\big(\gamma, \eta(e)1_{\{y\leq g(\tilde{\lambda}^C_{t-})\}}, \\
&\qquad\qquad \rho(e)1_{\{y\leq f(\tilde{\lambda}^C_{t-})\}}; \tilde{S}^C_{t-}\big)\Big) \\
&\qquad\qquad - v(\bar{T}-t, \tilde{S}^C_{t-})\bigg) K(\bar{z}; de, dy)\mu(d\bar{z}).
\end{aligned}$$

Finally, we collect all the dt-terms from the dynamics of dV_t^C and have the tight upper bound

$$\begin{aligned}
&-\frac{\partial v}{\partial T}(\bar{T}-t,\tilde{S}^C_{t-})\\
&+\int_{(E\times\mathbb{R}_+)\cap\{z=0\}}\Big(v\Big(\bar{T}-t,\tilde{\mathrm{s}}\big(0,\eta(e)1_{\{y\le g(\tilde{\lambda}^C_{t-})\}},\rho(e)1_{\{y\le f(\tilde{\lambda}^C_{t-})\}};\tilde{S}^C_{t-}\big)\Big)\\
&\qquad\qquad -v(\bar{T}-t,\tilde{S}^C_{t-})\Big)\big(\nu(de)\otimes dy\big) \qquad (4.5)\\
&+\int_{z(E\times\mathbb{R}_+)\setminus\{0\}}\sup_{\gamma\in D(\tilde{\lambda}^C_{t-},\tilde{Q}^C_{t-})}\int_{(E\times\mathbb{R}_+)\cap\{z=\bar{z}\}}\Big(v\Big(\bar{T}-t,\tilde{\mathrm{s}}\big(\gamma,\eta(e)1_{\{y\le g(\tilde{\lambda}^C_{t-})\}},\\
&\qquad\qquad \rho(e)1_{\{y\le f(\tilde{\lambda}^C_{t-})\}};\tilde{S}^C_{t-}\big)\Big)\\
&\qquad\qquad -v(\bar{T}-t,\tilde{S}^C_{t-})\Big)K(\bar{z};de,dy)\mu(d\bar{z}).
\end{aligned}$$

To satisfy the martingale optimality principle, the terms in (4.5) must be smaller than or equal to zero for any admissible strategy $C\in\tilde{\mathcal{C}}(\lambda,q)$, and zero for some (optimal) strategy C^*. With (4.2), we obtain the HJBQVI

$$\begin{aligned}
\max\Big\{&-\frac{\partial v}{\partial T}(T,s)\\
&+\int_{(E\times\mathbb{R}_+)\cap\{z=0\}}\Big(v\Big(T,\tilde{\mathrm{s}}\big(0,\eta(e)1_{\{y\le g(\lambda)\}},\rho(e)1_{\{y\le f(\lambda)\}};s\big)\Big)\\
&\qquad\qquad -v(T,s)\Big)\big(\nu(de)\otimes dy\big)\\
&+\int_{z(E\times\mathbb{R}_+)\setminus\{0\}}\sup_{\gamma\in D(\lambda,q)}\int_{(E\times\mathbb{R}_+)\cap\{z=\bar{z}\}}\Big(v\Big(T,\tilde{\mathrm{s}}\big(\gamma,\eta(e)1_{\{y\le g(\lambda)\}},\\
&\qquad\qquad \rho(e)1_{\{y\le f(\lambda)\}};s\big)\Big)\\
&\qquad\qquad -v(T,s)\Big)K(\bar{z};de,dy)\mu(d\bar{z}),\\
&\sup_{\Delta\in D(\lambda,q)}\Big(v\big(T,\tilde{\mathrm{s}}(\Delta,0,0;s)\big)-v(T,s)\Big)\Big\}=0 \qquad (4.6)
\end{aligned}$$

for $(T,s)=(T,\lambda,q,p,x)\in[0,\infty)\times[\underline{\lambda},\infty)\times[-\bar{q},\bar{q}]\times\mathbb{R}\times\mathbb{R}$. Due to trading being halted when the liquidity falls below $\underline{\lambda}$, the value function evaluated in $\lambda<\underline{\lambda}$ is

$$v(T,\lambda,q,p,x)=U\big(\mathrm{w}(x,q,p,\underline{\lambda})\big)\exp\big(\mathcal{L}_Y(\alpha q)\big)$$

for $(T,q,p,x)\in[0,\infty)\times[-\bar{q},\bar{q}]\times\mathbb{R}\times\mathbb{R}$, and the initial condition for $s=(\lambda,q,p,x)\in[\underline{\lambda},\infty)\times[-\bar{q},\bar{q}]\times\mathbb{R}^2$ with zero time to go ($T=0$) is

$$v(0,\lambda,q,p,x)=\begin{cases}U(\mathrm{w}(x,q,p,\lambda))\exp(\mathcal{L}_Y(\alpha\mathtt{r}(q,\lambda))), & \text{if }\lambda\ge\underline{\lambda},\\ U(\mathrm{w}(x,q,p,\underline{\lambda}))\exp(\mathcal{L}_Y(\alpha q)), & \text{if }\lambda<\underline{\lambda}.\end{cases}$$

To reduce the dimension of the HJB equation, we recall the multiplicative structure (3.8) of the value function. For notational simplicity, we introduce

$$\begin{aligned}
\gamma^{\lambda} &:= \Upsilon(\gamma, \lambda),\\
\eta^{\lambda,\gamma}(e, y) &:= \Upsilon\big(\eta(e), \lambda - |\gamma^{\lambda}|\big)1_{\{y \le g(\lambda)\}},\\
\rho^{\lambda,\gamma}(e, y) &:= \Big(\rho^{+}(e)1_{\{\lambda - |\gamma^{\lambda}| \ge \underline{\lambda}\}} + \Upsilon\big(\rho^{-}(e), \lambda - |\gamma^{\lambda}|\big)\Big)1_{\{y \le f(\lambda)\}},\\
\lambda^{\lambda,\gamma}(e, y) &:= \lambda - |\gamma| - |\eta(e)|1_{\{y \le g(\lambda)\}} + \rho(e)1_{\{y \le f(\lambda)\}},\\
\Delta^{\lambda} &:= \Upsilon(\Delta, \lambda),\\
\mathrm{w}_0(q, \lambda) &:= \mathrm{w}(0, q, 0, \lambda) = -\zeta|q| - \Xi(q, \lambda)
\end{aligned}$$

and write, with $(T, \lambda, q) \in [0, \infty) \times [\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}]$, the reduced HJB equation for $v_0(T, \lambda, q)$ when $\alpha > 0$ as

$$\begin{aligned}
\max\Bigg\{ &-\frac{\partial v_0}{\partial T}(T, \lambda, q)\\
&+ \int_{(E\times\mathbb{R}_+)\cap\{z=0\}} \bigg(v_0\big(T, \lambda^{\lambda,0}(e, y), q\big)\Big|U\Big(I\big(\eta^{\lambda,0}(e, y), \lambda\big)q\Big)\Big|\\
&\qquad\qquad - v_0(T, \lambda, q)\bigg)\big(\nu(de) \otimes dy\big)\\
&+ \int_{z(E\times\mathbb{R}_+)\setminus\{0\}}\\
&\sup_{\gamma\in D(\lambda,q)} \bigg(\int_{(E\times\mathbb{R}_+)\cap\{z=\bar{z}\}} \bigg(v_0\big(T, \lambda^{\lambda,\gamma}(e, y), q + \gamma^{\lambda}\big)\\
&\qquad\qquad \times \bigg|U\bigg(\mathrm{w}_0(\gamma^{\lambda}, \lambda)\\
&\qquad\qquad\qquad + \Big(I(\gamma^{\lambda}, \lambda) + I\big(\eta^{\lambda,\gamma}(e, y), \lambda - |\gamma^{\lambda}|\big)\Big)\\
&\qquad\qquad\qquad \times (q + \gamma^{\lambda})\bigg)\bigg|\\
&\qquad\qquad - v_0(T, \lambda, q)\bigg)K(\bar{z}; de, dy)\bigg)\mu(d\bar{z}),\\
&\sup_{\Delta\in D(\lambda,q)} \bigg(v_0(T, \lambda - |\Delta|, q + \Delta^{\lambda})\Big|U\Big(\mathrm{w}_0(\Delta^{\lambda}, \lambda) + I(\Delta^{\lambda}, \lambda)(q + \Delta^{\lambda})\Big)\Big|\\
&\qquad - v_0(T, \lambda, q)\bigg)\Bigg\} = 0.
\end{aligned}$$

For $\lambda < \underline{\lambda}$, the function $v_0(T, \lambda, q)$ takes the value

$$v_0(T, \lambda, q) = U\big(\mathrm{w}_0(q, \lambda)\big) \exp\big(\mathcal{L}_Y(\alpha q)\big), \qquad (T, q) \in [0, \infty) \times [-\bar{q}, \bar{q}],$$

and the initial condition for $(\lambda, q) \in \mathbb{R} \times [-\bar{q}, \bar{q}]$ with zero time to go is

$$v_0(0, \lambda, q) = \begin{cases} U(\mathrm{w}_0(q, \lambda)) \exp(\mathcal{L}_Y(\alpha \mathrm{r}(q, \lambda)^+)), & \text{if } \lambda \geq \underline{\lambda}, \\ U(\mathrm{w}_0(q, \underline{\lambda})) \exp(\mathcal{L}_Y(\alpha q)), & \text{if } \lambda < \underline{\lambda}. \end{cases}$$

Similarly, one obtains the reduced HJB equation for w when $\alpha = 0$.

The suggested optimal strategy can as usual be constructed in feedback form by determining the maximiser in the suprema of the HJB equation (4.7).

4.2 The numerical scheme

In the following, we present a numerical finite-difference scheme to approximate the solution of the dimension-reduced HJB equation in (4.7) for $\alpha > 0$; a similar scheme can be derived for $\alpha = 0$.

First, we discretise the value function's domain $[0, T] \times [\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}]$. For the time dimension $T' \in [0, T]$, we take $\mathbb{T}_{\delta T}$ as the grid on $[0, T]$ with step size $\delta T > 0$, where $T = N_1 \delta T$ for some $N_1 \in \mathbb{N}$.

For the liquidity dimension $\lambda \in [\underline{\lambda}, \infty)$, we choose the grid $\mathbb{R}^{\overline{\lambda}}_{\delta\lambda}$ discretising $[\underline{\lambda} - \delta\lambda, \overline{\lambda}]$ with stepsize $\delta\lambda > 0$, where $\overline{\lambda} = \underline{\lambda} + N_2 \delta\lambda$ for some $N_2 \in \mathbb{N}$ is a conveniently chosen maximum liquidity level. To remain within a bounded domain, we assume λ equals $\underline{\lambda} - \delta\lambda$ after trading was halted and define the numerical scheme below accordingly.

For the inventory levels $q \in \mathbb{R}$, we introduce the bounded grid $\mathbb{R}^{\underline{Q}, \overline{Q}}_{\delta q}$ on $[\underline{Q}, \overline{Q}]$ with step size $\delta q > 0$, where $\underline{Q} = N_3 \delta q$ and $\overline{Q} = N_4 \delta q$ for some constants $N_3, N_4 \in \mathbb{Z}$, $N_3 < N_4$ and $[\underline{Q}, \overline{Q}] \subseteq [-\overline{q}, \overline{q}]$. For the inventory q to remain within the grid $\mathbb{R}^{\underline{Q}, \overline{Q}}_{\delta q}$ and λ to remain above $\underline{\lambda} - \delta\lambda$, we define the set of actions for $q \in \mathbb{R}^{\underline{Q}, \overline{Q}}_{\delta q}$, $\lambda \in \mathbb{R}^{\overline{\lambda}}_{\delta\lambda}$ as

$$D^{\underline{Q}, \overline{Q}}(q, \lambda) := \{n\delta q \text{ with } n \in \mathbb{Z} : q + n\delta q \in \mathbb{R}^{\underline{Q}, \overline{Q}}_{\delta q} \text{ and } \lambda - |n\delta| \geq \underline{\lambda} - \delta\lambda\}.$$

To propagate the value function, we introduce on functions $h^{T'} : \mathbb{R}^{\overline{\lambda}}_{\delta\lambda} \times \mathbb{R}^{\underline{Q}, \overline{Q}}_{\delta q} \to \mathbb{R}$ with $T' \in \mathbb{T}_{\delta T}$ the operator

$$\begin{aligned} &\mathcal{L}^{\delta T, \delta\lambda}(q, \lambda, h^{T'}) \\ &:= h^{T'}(\lambda, q) + \delta T \bigg(\Delta_{z=0} h^{T'}(\lambda, q) \\ &\qquad\qquad + \int_{z(E \times \mathbb{R}_+) \setminus \{0\}} \sup_{\gamma \in D^{\underline{Q}, \overline{Q}}(q, \lambda)} \Delta_{\mathrm{z} = \bar{z}} h^{T'}(\lambda, q, \gamma, \psi) \;\; \mu(d\bar{z}) \bigg), \end{aligned}$$

where

$$\begin{aligned}
&\Delta_{\mathrm{z}=0} h^{T'}(\lambda, q) \\
&:= \int_{(E\times\mathbb{R}_+)\cap\{z=0\}} \Big(\bar{h}^{T'}\big(\lambda^{\lambda,0}(e,y) \vee (\underline{\lambda}-\delta\lambda), q\big)\big|U\big(I(\eta^{\lambda,0}(e,y),\lambda)q\big)\big| \\
&\qquad - h^{T'}(\lambda, q)\Big)\big(\nu(de)\otimes dy\big)
\end{aligned}$$

describes the propagation while there is no signal to act upon, and

$$\begin{aligned}
&\Delta_{\mathrm{z}=\bar{z}} h^{T'}(\lambda, q, \gamma) \\
&:= \int_{(E\times\mathbb{R}_+)\cap\{z=\bar{z}\}} \Bigg(\bar{h}^{T'}\big(\lambda^{\lambda,\gamma}(e,y) \vee (\underline{\lambda}-\delta\lambda), q+\gamma^{\lambda}\big) \\
&\qquad \times \Bigg|U\Bigg(\mathrm{w}_0(\gamma^{\lambda},\lambda) \\
&\qquad\qquad - \Big(I(\gamma^{\lambda},\lambda) + I\big(\eta^{\lambda,\gamma}(e,y), \lambda-|\gamma^{\lambda}|\big)\Big)(q+\gamma^{\lambda})\Bigg)\Bigg| \\
&\qquad - h^{T'}(\lambda, q)\Bigg) K(\bar{z}; de, dy)
\end{aligned}$$

describes the propagation when an order γ is placed upon observing a signal $\bar{z}$. For values $\lambda \notin \mathbb{R}^{\overline{\lambda}}_{\delta\lambda}$, we linearly interpolate to obtain

$$\begin{aligned}
&\bar{h}^{T'}(q, \lambda) \\
&:= h^{T'}\left(q, \left\lceil \frac{\lambda}{\delta\lambda} \right\rceil \delta\lambda\right) - \left(\left\lceil \frac{\lambda}{\delta\lambda} \right\rceil \delta\lambda - \lambda\right)\left(h^{T'}\left(q, \left\lceil \frac{\lambda}{\delta\lambda} \right\rceil \delta\lambda\right) - h^{T'}\left(q, \left\lfloor \frac{\lambda}{\delta\lambda} \right\rfloor \delta\lambda\right)\right).
\end{aligned}$$

Here $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote rounding to the closest integer below and above. Similarly, we define the operator

$$\begin{aligned}
\mathcal{M}^{\delta T,\delta\lambda}(\lambda, q, h^{T'}) := \sup_{\Delta\in D^{\underline{Q},\overline{Q}}(q,\lambda)} &\Big(\bar{h}^{T'}\big((\lambda-|\Delta|)\vee(\lambda-\delta\lambda), q+\Delta\big) \\
&\times \big|U\big(\mathrm{w}_0(\Delta^{\lambda},\lambda) - I(\Delta^{\lambda},\lambda)(q+\Delta^{\lambda})\big)\big|\Big)
\end{aligned}$$

to ensure that state-based trades are executed optimally.

With this notation in place, we specify the following numerical scheme to compute an approximate solution w to (4.7). First, initialise with time $T = 0$ to go with the boundary values

$$v_0^0(q, \lambda) := \begin{cases} U(\mathrm{w}_0(q, \lambda)) \exp(\mathcal{L}_Y(\alpha \mathrm{r}(q, \lambda))), & \text{if } \lambda \geq \underline{\lambda}, \\ U(\mathrm{w}_0(q, \underline{\lambda})) \exp(\mathcal{L}_Y(\alpha q)), & \text{if } \lambda = \underline{\lambda} - \delta\lambda. \end{cases}$$

Next, after computing the value function up to a time horizon T', we propagate it a step further to $T' + \delta T$ by calculating the value when one can play for an extra period to obtain

$$\tilde{v_0}^{T'+\delta T}(q,\lambda) := \mathcal{L}^{\delta T,\delta\lambda}(q,\lambda,v_0^{T'}), \qquad \lambda \geq \underline{\lambda}.$$

This value is updated to account for possible state-based trades to obtain the next instance

$$v_0^{T'+\delta T}(q,\lambda) := \max\{\tilde{v_0}^{T'+\delta T}(q,\lambda), \mathcal{M}^{\delta t,\delta\lambda}(q,\lambda,\tilde{v_0}^{T'+\delta T})\}, \qquad \lambda \geq \underline{\lambda},$$

while maintaining the boundary condition

$$v_0^{T'+\delta T}(q,\underline{\lambda}-\delta\lambda) := U\big(\mathrm{w}_0(q,\underline{\lambda})\big)\exp\big(\mathcal{L}_Y(\alpha q)\big)$$

for $T' \in \{0, \delta t, \ldots, T-2\delta t, T-\delta t\}$ when trading halts.

Appendix A: Proofs for Sect. 2

A.1 Model dynamics admit a unique solution

Lemma A.1 *Under condition* (2.5), *the system dynamics* (2.1)–(2.3) *for* (λ, L, M) *allow a unique solution.*

Proof It suffices to rule out that the mutually exciting dynamics lead to blow-ups. To this end, we prove that the expectation of the total variation $V_{[0,t]}(\lambda)$ of the liquidity process (λ_s) over $[0,t]$ is bounded for each $t \in [0,T]$. For this, we use that f and g have at most linear growth to estimate

$$\begin{aligned}
&\mathbb{E}[V_{[0,t]}(\lambda)] \\
&= \mathbb{E}\Big[\int_{[0,t]\times E\times\mathbb{R}_+} 1_{\{y\leq f(\lambda_{s-})\}}|\rho(e)|N(ds,de,dy) \\
&\qquad + \int_{[0,t]\times E\times\mathbb{R}_+} 1_{\{y\leq g(\lambda_{s-})\}}|\eta(e)|N(ds,de,dy)\Big] \\
&= \int_E |\rho(e)|\nu(de)\mathbb{E}\Big[\int_0^t f(\lambda_s)ds\Big] + \int_E |\eta(e)|\nu(de)\mathbb{E}\Big[\int_0^t g(\lambda_s)ds\Big] \\
&\leq \Big(\int_E |\rho(e)|\nu(de) + \int_E |\eta(e)|\nu(de)\Big)cT\Big(1 + |\lambda_{0-}| + \mathbb{E}\Big[\int_0^t V_{[0,s]}(\lambda)ds\Big]\Big).
\end{aligned}$$

The claim follows by Gronwall's inequality. □

A.2 Link between price volatility and liquidity of the market — proof of Lemma 2.2

The quadratic variation of the price process has dynamics

$$d[P]_t = \int_E 1_{\{y \le g(\lambda_{t-})\}} I\big(\eta(e), \lambda_{t-}\big)^2 N(dt, de, dy), \qquad [P]_0 = 0.$$

Hence the dynamics of its predictable compensator coincide with (2.9). The derivative of the expression in (2.10) with respect to λ is strictly negative if condition (2.11) holds; this proves the monotonicity claim.

A.3 Decomposition of C — proof of Lemma 2.4

It is easy to see that C with C^c, $\Delta^\ell C$, and $\Delta^r C$ of the described form is Λ-measurable.

To prove the converse, we have to show for a Λ-measurable C of bounded variation that its left jumps $\Delta^\ell C$ are of the form (2.15) with Γ and $C^{\ell,p}$ as described in the lemma. Without loss of generality, we can assume that all these left jumps are negative and that their sum C^ℓ is bounded. Then C^ℓ is a Λ-supermartingale and thus can be decomposed into $C^\ell = M - A - B_-$ for some local Λ-martingale, some predictable, increasing, right-continuous A with $A_0 = 0$ and some Λ-measurable, increasing, right-continuous pure-jump process B with $B_0 = 0$; cf. Lenglart [24, Theorem 4]. By localisation, M is even a true martingale and by the martingale representation theorem for Poisson point processes, there is a $(\mathcal{P} \otimes \mathcal{E} \otimes \mathcal{B}([0, T]))$-measurable $\xi \in L^1(\mathbb{P} \otimes ds \otimes \nu \otimes dy)$ such that

$$M_T = \int_{[0,T]\times E\times[0,\infty)} \xi_s(e, y)\bar{N}(ds, de, dy),$$

where we adopt the notation from Sect. 4. Taking the Λ-projection gives

$$M_t = ({}^{\Lambda}M)_t = \int_{[0,t)\times E\times[0,\infty)} \xi_s(e, y)\bar{N}(ds, de, dy) + \hat{\xi}_t(Z_t), \qquad t \in [0, T],$$

where $\hat{\xi}_t(0) := 0$ and $\hat{\xi}_t(\bar{z}) := \int_{\{z=\bar{z}\}} \xi_t(e, y)K(\bar{z}; de, dy)$ for $\bar{z} \in z(E \times [0, \infty)) \setminus \{0\}$, with K given in (4.4). Since A is predictable and N places marks only at $\mathcal{P}$-totally inaccessible stopping times, the jumps $\Delta A =: -\Delta C^{\ell,p} =: -\Delta^{\ell,p} C$ almost surely do not coincide with mark times, i.e., $\{t \in [0, T] : \Delta_t A \neq 0\} \subseteq \{t \in [0, T] : N(\{t\} \times E \times [0, \infty)) = 0\}$. It follows that $\Delta^\ell_t C = \Delta^\ell_t M - \Delta^\ell_t A - \Delta^\ell_t B_- = \hat{\xi}_t(Z_t) + \Delta^{\ell,p}_t C + 0$. This is (2.15) with $\Gamma := \hat{\xi}$.

A.4 Dynamics of realisable portfolio value — proof of Proposition 2.7

We use that the Marcus-style dynamics (2.21)–(2.24) of λ^C, X^C, Q^C and P^C are governed by classical calculus to compute the dynamics of W^C as

$$\begin{aligned}
&W_t^C - W_{0-}^C \\
&= X_t^C + P_t^C Q_t^C - (x + pq) - \zeta(|Q_t^C| - |q|) - \big(\Xi(Q_t^C, \lambda_t^C) - \Xi(q, \lambda)\big) \\
&= \int_0^t Q^C \diamond dP^C - \zeta \int_0^t \diamond\, |dQ^C| - \zeta \int_0^t \operatorname{sgn}(Q^C) \diamond dQ^C \\
&\quad - \bigg(\int_0^t \partial_\Delta \Xi(Q^C, \lambda^C) \diamond dQ^C + \int_0^t \partial_\lambda \Xi(-Q^C, \lambda^C) \diamond d\lambda^C \bigg).
\end{aligned}$$

With

$$\begin{aligned}
\partial_\Delta \Xi(\Delta, \lambda) &= \operatorname{sgn}(\Delta) I(|\Delta|, \lambda), \\
\partial_\lambda \Xi(\Delta, \lambda) &= \int_0^{|\Delta|} \partial_\lambda I(z, \lambda) dz = \int_0^{|\Delta|} \int_0^z \iota'(\lambda - \tilde{z}) d\tilde{z} dz \\
&\qquad\qquad\qquad\qquad\quad = \int_0^{|\Delta|} \big(\iota(\lambda) - \iota(\lambda - z)\big) dz \\
&\qquad\qquad\qquad\qquad\quad = |\Delta|\iota(\lambda) - I(|\Delta|, \lambda),
\end{aligned}$$

we find

$$\begin{aligned}
&W_t^C - W_{0-}^C \\
&= \int_0^t Q^C \iota(\lambda^C) \diamond \big(dQ^C + (0 \vec{+} dM^C \vec{+} 0)\big) \\
&\quad - \zeta \int_0^t \big(\diamond\ |dQ^C| + \operatorname{sgn}(Q^C) \diamond dQ^C\big) \\
&\quad - \bigg(\int_0^t \operatorname{sgn}(Q^C) I(|Q^C|, \lambda^C) \diamond dQ^C + \int_0^t \big(|Q^C|\iota(\lambda^C) - I(|Q^C|, \lambda^C)\big) \diamond d\lambda^C \bigg) \\
&= \int_0^t \big(|Q^C|\iota(\lambda^C) - I(|Q^C|, \lambda^C) - \zeta\big)\big(\operatorname{sgn}(Q^C) \diamond dQ^C + \diamond\, |dQ^C|\big) \\
&\quad + \int_0^t |Q^C|\iota(\lambda^C) \operatorname{sgn}(Q^C) \diamond (0 \vec{+} dM^C \vec{+} 0) \\
&\quad - \int_0^t \big(|Q^C|\iota(\lambda^C) - I(|Q^C|, \lambda^C)\big) \diamond \big(0 \vec{+} (dL^C - |dM^C|) \vec{+} 0\big).
\end{aligned}$$

Decomposing

$$dQ^C = dQ^{C,+} - dQ^{C,-} = (dC^{\ell,+} - dC^{\ell,-})\vec{+}0\vec{+}(dC^{r,c,+} - dC^{r,c,-}),$$
$$|dQ^C| = dQ^{C,+} + dQ^{C,-} = (dC^{\ell,+} + dC^{\ell,-})\vec{+}0\vec{+}(dC^{r,c,+} + dC^{r,c,-})$$

and decomposing similarly the dM^C-terms leads to

$$\begin{aligned}
&W_t^C - W_{0-}^C \\
&= \int_0^t (|Q^C|\iota(\lambda^C) - I(|Q^C|, \lambda^C) - \zeta)2(1_{\{Q^C>0\}} \diamond dQ^{C,+} + 1_{\{Q^C<0\}} dQ^{C,-}) \\
&\quad + \int_0^t \big(1_{\{Q^C>0\}} 2|Q^C|\iota(\lambda^C) - I(|Q^C|, \lambda^C)\big) \diamond (0\vec{+}dM^{C,+}\vec{+}0) \\
&\quad + \int_0^t \big(1_{\{Q^C<0\}} 2|Q^C|\iota(\lambda^C) - I(|Q^C|, \lambda^C)\big) \diamond (0\vec{+}dM^{C,-}\vec{+}0) \\
&\quad + \int_0^t \big(I(|Q^C|, \lambda^C) - |Q^C|\iota(\lambda^C)\big) \diamond (0\vec{+}dL^C\vec{+}0).
\end{aligned}$$

Rearranging terms gives (2.26).

Appendix B: Regularity of the value function

This section gives the proof of our regularity result in Theorem 3.2 for the value function.

B.1 Total order activity bounded by provided liquidity

We can use that market orders and cancellations are executed only as long as the liquidity remains greater than $\underline{\lambda}$ to estimate

$$\underline{\lambda} \le \tilde{\lambda}_t^C = \lambda - V_{[0,t]}(\tilde{Q}^C) - V_{[0,t]}(\tilde{M}^C) - \tilde{L}_t^{C,-} + \tilde{L}_t^{C,+}, \qquad t \in [0,T], \quad \text{(B.1)}$$

with $\tilde{M}^C$, $\tilde{L}^{C,+}$, $\tilde{L}^{C,-}$ as in (3.2)–(3.4). By rearranging (B.1), we find

$$V_{[0,t]}(\tilde{Q}^C) + V_{[0,t]}(\tilde{M}^C) + \tilde{L}_t^{C,-} \le \lambda - \underline{\lambda} + \tilde{L}_t^{C,+}, \qquad t \in [0,T].$$

Since the overall liquidity supplied in a market with minimum liquidity $\underline{\lambda}$ can be bounded by

$$\tilde{L}_t^{C,+} \le \int_{[0,t]\times E\times[0,f(\underline{\lambda})]} \rho^+(e) N(ds, de, dy) =: \bar{L}_t^+,$$

we find the total amount of market orders and cancellations controlled by

$$V_{[0,t]}(\tilde{Q}^C) + V_{[0,t]}(\tilde{M}^C) + \tilde{L}_t^{C,-} \le \lambda - \underline{\lambda} + \bar{L}_t^+, \qquad t \in [0,T]. \quad \text{(B.2)}$$

B.2 Multiplicative structure and nondegeneracy of the value function

By Proposition 2.7,

$$
\begin{aligned}
&\tilde{W}^C_{T+}\\
&= \mathrm{w}(x,q,p,\lambda)\\
&\quad + \int_0^{T+} \big(I(|\tilde{Q}^C|,\tilde{\lambda}^C) - |\tilde{Q}^C|\iota(\tilde{\lambda}^C)\big) \diamond (0\vec{+}d\tilde{L}^C\vec{+}0)\\
&\quad - \int_0^{T+} \big(I(|\tilde{Q}^C|,\tilde{\lambda}^C) - 1_{\{\tilde{Q}^C>0\}}2|\tilde{Q}^C|\iota(\tilde{\lambda}^C)\big) \diamond (0\vec{+}d\tilde{M}^{C,+}\vec{+}0)\\
&\quad - \int_0^{T+} \big(I(|\tilde{Q}^C|,\tilde{\lambda}^C) - 1_{\{\tilde{Q}^C<0\}}2|\tilde{Q}^C|\iota(\tilde{\lambda}^C)\big) \diamond (0\vec{+}d\tilde{M}^{C,-}\vec{+}0)\\
&\quad - \int_0^{T+} 2\big(I(|\tilde{Q}^C|,\tilde{\lambda}^C) - |\tilde{Q}^C|\iota(\tilde{\lambda}^C) + \zeta\big) \diamond (1_{\{\tilde{Q}^C>0\}}d\tilde{Q}^{C,+} + 1_{\{\tilde{Q}^C<0\}}d\tilde{Q}^{C,-})\\
&\quad + Y\mathrm{r}(\tilde{Q}^C_{T+},\tilde{\lambda}^C_{T+}).
\end{aligned}
\tag{B.3}
$$

We notice that neither x nor p plays a role in the above integrals or the Y-term. As $\mathrm{w}(x,q,p,\lambda) = x + pq + \mathrm{w}(0,q,0,\lambda)$, the exponential utility structure allows us to separate $x + pq$ and we conclude that

$$
v(T,\lambda,q,p,x) = -\exp\big(-\alpha(x+pq)\big)v(T,\lambda,q,0,0),
$$

which is the claimed multiplicative form (3.8) of the value function.

As a further consequence of the wealth representation (B.3), we can use the position limit $\bar{q}$ of (2.13) and the liquidity bound $\underline{\lambda}$ to estimate

$$
\tilde{W}^C_{T+} \le \mathrm{w}(x,q,p,\lambda) + \bar{q}\iota(\underline{\lambda})\big(\tilde{L}^{C,+}_T + 2V_{[0,T]}(\tilde{M}^C)\big) + Y\mathrm{r}(\tilde{Q}^C_{T+},\tilde{\lambda}^C_{T+}). \tag{B.4}
$$

Therefore, we find with (B.2) that

$$
\begin{aligned}
\mathbb{E}[U_\alpha(\tilde{W}^C_{T+})] \le{}& -\exp\big(-\alpha\mathrm{w}(x,q,p,\lambda)\big)\mathbb{E}\big[\exp\big(-3\alpha\bar{q}\iota(\underline{\lambda})(\lambda-\underline{\lambda}+\bar{L}^+_T)\big)\\
&\times \mathbb{E}[\exp(aY)\,|\,\mathcal{F}_T]\big|_{a=-\alpha\mathrm{r}(\tilde{Q}^C_{T+},\tilde{\lambda}^C_{T+})}\big].
\end{aligned}
$$

Next, due to (3.5), Y is centered and independent of $\mathcal{F}_T$; so the conditional expectation is at least 1 by Jensen's inequality, and hence

$$
\mathbb{E}[U_\alpha(\tilde{W}^C_{T+})] \le -\exp\big(-\alpha\mathrm{w}(x,q,p,\lambda)\big)\mathbb{E}\big[\exp\big(-3\alpha\bar{q}\iota(\underline{\lambda})(\lambda-\underline{\lambda}+\bar{L}^+_T)\big)\big] < 0.
$$

As this bound does not depend on the choice of $C \in \tilde{\mathcal{C}}(\lambda,q)$, the value function v of (3.7) is non-degenerate in the sense that its value is less than $U(\infty) = 0$.

B.3 Continuity of the value function with respect to the time horizon

Let us first show that our value functions depend continuously and monotonically on the time horizon.

Lemma B.1 *For any $T' \le T$, we have*

$$v(T', q, p, x) \le v(T, \lambda, q, p, x) \le v(T', \lambda, q, p, x)\tau(T - T') < 0 \tag{B.5}$$

for some decreasing continuous function $\tau : [0, \infty) \to (0, 1]$ with $\tau(0) = 1$.

Proof As the investor can liquidate and wait until the time available for investment has elapsed, it is obvious that the value function is monotonic in T.

For the second inequality, let us consider a control $C \in \tilde{C}$ for the time horizon T and use the properties of Y from (3.5) to write its expected utility as

$$\begin{aligned}&\mathbb{E}[U(\tilde{W}^C_{T+})]\\&= \mathbb{E}\Big[- \exp\Big(- \alpha \mathrm{w}(\tilde{X}^C_{T+}, \tilde{Q}^C_{T+}, \tilde{P}^C_{T+}, \tilde{\lambda}^C_{T+}) + \mathcal{L}_Y\big(\alpha \mathrm{r}(\tilde{Q}^C_{T+}, \tilde{\lambda}^C_{T+})\big)\Big)\Big].\end{aligned} \tag{B.6}$$

Analogously to (B.4), we conclude that

$$\begin{aligned}&\mathrm{w}(\tilde{X}^C_{T+}, \tilde{Q}^C_{T+}, \tilde{P}^C_{+}, \tilde{\lambda}^C_{T+}) - \mathrm{w}(\tilde{X}^C_{T'+}, \tilde{Q}^C_{T'+}, \tilde{P}^C_{T'+}, \tilde{\lambda}^C_{T'+})\\&\le \bar{q}\iota(\underline{\lambda})\big(\tilde{L}^{C,+}_T - \tilde{L}^{C,+}_{T'} + 2V_{(T',T]}(\tilde{M}^C)\big)\\&\le \bar{q}\iota(\underline{\lambda}) \int_{(T',T]\times E\times[0,\infty)} \big(1_{[0,f(\underline{\lambda})]}(y)\rho^+(e) + 1_{[0,g(\infty)]}(y)2|\eta(e)|\big)N(ds, de, dy).\end{aligned}$$

Similarly,

$$\begin{aligned}&\mathcal{L}_Y\big(\alpha \mathrm{r}(\tilde{Q}^C_{T+}, \tilde{\lambda}^C_{T+})\big) - \mathcal{L}_Y\big(\alpha \mathrm{r}(\tilde{Q}^C_{T'+}, \tilde{\lambda}^C_{T'+})\big)\\&= \int_{T'}^{T+} \mathcal{L}'_Y\big(\alpha \mathrm{r}(\tilde{Q}^C, \tilde{\lambda}^C)\big)\alpha 1_{\{\mathrm{r}(\tilde{Q}^C,\tilde{\lambda}^C)^C>0\}}\\&\qquad\times \Big(\mathrm{sgn}(\tilde{Q}^C) \diamond d\tilde{Q}^C - \diamond\Big(- |d\tilde{Q}^C| + \big(0\vec{+}(d\tilde{L}^C - |d\tilde{M}^C|)\vec{+}0\big)\Big)\Big)\\&= \int_{T'}^{T+} \mathcal{L}'_Y\big(\alpha \mathrm{r}(\tilde{Q}^C, \tilde{\lambda}^C)\big)\alpha 1_{\{\mathrm{r}(\tilde{Q}^C,\tilde{\lambda}^C)>0\}}\\&\qquad\times \Big(1_{\{\tilde{Q}^C>0\}}2 \diamond d\tilde{Q}^{C,+} + 1_{\{\tilde{Q}^C<0\}}2 \diamond d\tilde{Q}^{C,-} - \diamond\big(0\vec{+}(d\tilde{L}^C - |d\tilde{M}^C|)\vec{+}0\big)\Big)\\&\ge -\mathcal{L}'_Y(\alpha\bar{q})\alpha(\tilde{L}^{C,+}_T - \tilde{L}^{C,+}_{T'})\\&= -\mathcal{L}'_Y(\alpha\bar{q})\alpha \int_{(T',T]\times E\times[0,\infty)} 1_{[0,f(\underline{\lambda})]}(y)\rho^+(e)N(ds, de, dy),\end{aligned}$$

where for the last estimate we used that $\mathcal{L}'_Y$ is increasing from $\mathcal{L}'_Y(0) = 0$ because Y is centered with symmetric, convex cumulant generating function $\mathcal{L}_Y$; cf. (3.5).

With these estimates, we get from (B.6) that

$$
\begin{aligned}
&\mathbb{E}[U(\tilde{W}^C_{T+})] \\
&= \mathbb{E}\Big[-\exp\Big(-\alpha \mathtt{w}(\tilde{X}^C_{T'+}, \tilde{Q}^C_{T'+}, \tilde{P}^C_{+}, \tilde{\lambda}^C_{T'+}) + \mathcal{L}_Y\big(\alpha \mathtt{r}(\tilde{Q}^C_{T'+}, \tilde{\lambda}^C_{T'+})\big)\Big) \\
&\qquad \times \exp\Big(-\alpha\big(\mathtt{w}(\tilde{X}^C_{T+}, \tilde{Q}^C_{T+}, \tilde{P}^C_{+}, \tilde{\lambda}^C_{T+}) - \mathtt{w}(\tilde{X}^C_{T'+}, \tilde{Q}^C_{T'+}, \tilde{P}^C_{T'+}, \tilde{\lambda}^C_{T'+})\big)\Big) \\
&\qquad \times \exp\Big(\mathcal{L}_Y\big(\alpha \mathtt{r}(\tilde{Q}^C_{T+}, \tilde{\lambda}^C_{T+})\big) - \mathcal{L}_Y\big(\alpha \mathtt{r}(\tilde{Q}^C_{T'+}, \tilde{\lambda}^C_{T'+})\big)\Big)\Big] \\
&\le \mathbb{E}\Big[-\exp\Big(-\alpha \mathtt{w}(\tilde{X}^C_{T'+}, \tilde{Q}^C_{T'+}, \tilde{P}^C_{T'+}, \tilde{\lambda}^C_{T'+}) + \mathcal{L}_Y\big(\alpha \mathtt{r}(\tilde{Q}^C_{T'+}, \tilde{\lambda}^C_{T'+})\big)\Big) \\
&\qquad \times \exp\Big(-\alpha \bar{q}\iota(\underline{\lambda}) \int_{(T',T]\times E\times[0,\infty)} \big(1_{[0,f(\underline{\lambda})]}(y)\rho^+(e) \\
&\qquad\qquad + 1_{[0,g(\infty)]}(y) 2|\eta(e)|\big) N(ds,de,dy)\Big) \\
&\qquad \times \exp\Big(-\mathcal{L}'_Y(\alpha\bar{q})\alpha \int_{(T',T]\times E\times[0,\infty)} 1_{[0,f(\underline{\lambda})]}(y)\rho^+(e) N(ds,de,dy)\Big)\Big].
\end{aligned}
$$

The above integrals with respect to N are independent of $\mathcal{F}_{T'}$. Hence their contribution can be worked out separately and by the Lévy–Khintchine formula, there is a constant $\theta > 0$ such that it is of the form $\tau(T-T') = \exp(-\theta(T-T'))$. With another application of (B.6), but now for T' instead of T, it follows that

$$
\mathbb{E}[U(\tilde{W}^C_{T+})] \le \mathbb{E}[U(\tilde{W}^C_{T'+})]\tau(T-T').
$$

With $C \in \tilde{\mathcal{C}}$ chosen arbitrarily, this yields the second estimate in (B.5). $\square$

B.4 Continuity of the value function with respect to the initial liquidity level and inventory

Let us next investigate continuity of v with respect to $(\lambda, q) \in [\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}]$. By the multiplicative structure (3.8) of the value function, it suffices to do this for v_0, i.e., for $x = 0$ and $p = 0$.

Our task is to construct from a given strategy $C \in \tilde{\mathcal{C}}(\lambda, q)$ a strategy C' which is admissible when starting from $(\lambda', q') \in [\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}]$ and performs at least almost as well as C does. As we show later, such a C' can be obtained by executing the same orders as C, at least to the extent possible, i.e., without violating the inventory constraint and only using the available liquidity. Before C is halted at time τ^C (cf. (3.1)), this strategy can formally be described via the solution to the SDE system

$$
\begin{aligned}
\tilde{Q}'_{0-} &= q', \quad && \diamond\, d\tilde{Q}' = 1_{\{\tilde{Q}' \in [-\bar{q},\bar{q}),\ \lambda^{C'} > \underline{\lambda}\}} \diamond dC^+ - 1_{\{\tilde{Q}' \in (-\bar{q},\bar{q}],\ \lambda^{C'} > \underline{\lambda}\}} \diamond dC^-, \\
\lambda^{C'}_{0-} &= \lambda', \quad && \diamond\, d\lambda^{C'} = \diamond\Big(-|d\tilde{Q}'| + \big(0 \vec{+} (d\tilde{L}' - |d\tilde{M}'|)\vec{+} 0\big)\Big), \\
\tilde{L}'_{0-} &= 0, \quad && d\tilde{L}' = 1_{\{y \le f(\lambda^{C'}_-),\ \lambda^{C'}_- \ge \underline{\lambda}\}} \rho(e) N(dt,de,dy), \\
\tilde{M}'_{0-} &= 0, \quad && d\tilde{M}' = 1_{\{y \le g(\lambda^{C'}_-),\ \lambda^{C'}_- \ge \underline{\lambda}\}} \eta(e) N(dt,de,dy).
\end{aligned}
$$

Indeed, we have

$$\diamond\, dC' := \diamond\,(dC'^{\ell} \vec{+} 0 \vec{+} dC'^{r,c}) =: \diamond\, d\tilde{Q}' \qquad \text{on } [0,\tau^C) \cap [0,T].$$

We note that over $[0,\tau^C) \cap [0,T]$, $\lambda^{C'}$ can fall below $\underline{\lambda}$ only due to an exogenous order. If $\tau^C \le T$, the halt for C can be triggered for three reasons: First, if it is due to a left-jump trade ($\Delta^{\ell} C > 0$), C' liquidates its present position in this moment, possibly triggering a halt in the process, and stops operating afterwards. Second, if C's halt is triggered by an external order that has not halted C', the latter still copies, to the extent allowed by liquidity and position constraints, any left-jump trade preceding it, lets any external orders take effect, but after that closes shop by liquidating its position in a state-based trade, again possibly triggering a halt itself. Finally, third, if C's state-based trade triggered the halt, C' seizes operations after a final liquidating state-based trade, which again is allowed to trigger a halt.

With C' constructed in this way, notice that the dynamics of $\lambda^{C'}$ can be written as

$$\diamond\, d\lambda^{C'} = \diamond\,\big((-|dC'^{\ell}|) \vec{+} (d\tilde{L}' - |d\tilde{M}'|) \vec{+} (-|dC'^{r,c}|)\big),$$

in line with the dynamics (2.21) of λ^C. In particular, just like λ^C, also $\lambda^{C'}$ can drop below $\underline{\lambda}$, but with the exception of a possible final state-based trade by C' at time τ^C, this halt can only be due to external orders from $\tilde{L}'$ and $\tilde{M}'$. These two and also $\tilde{Q}'$ in turn do not move when $\lambda_{-}^{C'} < \underline{\lambda}$, i.e., right after C' has triggered this trading halt signal. The dynamics of $\tilde{Q}'$, $\tilde{L}'$ and $\tilde{M}'$ thus come to a halt as well. As a result, all these dynamics are in accordance with those of Sect. 3.1 in the sense that $\tilde{L}' = \tilde{L}^{C'}$, $\tilde{M}' = \tilde{M}^{C'}$, $\tilde{Q}' = \tilde{Q}^{C'}$.

Lemma B.2 *Assume that $\rho, \eta \in L^2(\nu)$ and $g(\infty) < \infty$. There is a constant c with the properties* (i)–(iii) *below which depends neither on $(\lambda, q),\ (\lambda', q') \in [\underline{\lambda}, \bar{\lambda}] \times [-\bar{q}, \bar{q}]$ nor on the choice of $C \in \tilde{\mathcal{C}}(\lambda, q)$, provided that*

$$\begin{aligned}&\textit{C only actively triggers a halt when reducing its position } |Q^C|,\\ &\textit{but not while expanding it.}\end{aligned} \tag{B.7}$$

(i) *Until it runs into a trading halt, the realisable wealth from strategy C' remains close to that of C in the sense that*

$$\begin{aligned}|\tilde{W}^C - \tilde{W}^{C'}| &\le c(|\lambda - \lambda'| + |q - q'| + V^{C,C'})\\ &\quad \times (1 + \lambda \vee \lambda' - \underline{\lambda} + \bar{L}^{+} + |Y|) \qquad \textit{on } [\![0, \tau^{C'} \wedge T]\!],\end{aligned} \tag{B.8}$$

where

$$V_{\cdot}^{C,C'} := V_{[0,\cdot]}(\tilde{L}^C - \tilde{L}^{C'}) + V_{[0,\cdot]}(\tilde{M}^C - \tilde{M}^{C'})$$

denotes the cumulative external order volume that exclusively affects the trading environment of just one of the strategies.

(ii) *In expectation, the exclusive order volume* $V^{C,C'}$ *is controlled in the sense that*

$$\mathbb{E}\big[V^{C,C'}_{T\wedge\tau^C\wedge\tau^{C'}}\big] \leq c(|\lambda-\lambda'|+|q-q'|) \tag{B.9}$$

and

$$\mathbb{E}[(V^{C,C'}_{T\wedge\tau^C\wedge\tau^{C'}})^2]^{1/2} \leq c(|\lambda-\lambda'|+|q-q'|)^{1/2}. \tag{B.10}$$

(iii) *The probability of an early halt for* C' *is*

$$\mathbb{P}[\tau^{C'}<\tau^C\wedge T] \leq c\Big(w_{\rho^-+|\eta|}\big(c(|\lambda-\lambda'|+|q-q'|)\big)+|\lambda-\lambda'|+|q-q'|\Big),$$

where $w_{\rho^-+|\eta|}$ *is the concave envelope of the modulus of continuity for* $\nu(\rho^-+|\eta|>\cdot)$ *on* $(0,\infty)$.

Remark B.3 Lemma B.2 establishes a precise control about how closely C' constructed above manages to track C. Property (B.7) is only violated by clearly suboptimal strategies which are of no interest for the value function; cf. also the proof of Corollary B.4.

Proof For ease of notation, let us introduce $\tau:=\tau^C$, $\tau':=\tau^{C'}$, $\tilde{\lambda}:=\tilde{\lambda}^C$, $\tilde{\lambda}':=\tilde{\lambda}^{C'}$, $\tilde{W}:=\tilde{W}^C_+$, $\tilde{W}':=\tilde{W}^{C'}_+$ as well as $\tilde{Q}:=\tilde{Q}^C$, $\tilde{L}:=\tilde{L}^C$, $\tilde{M}:=\tilde{M}^C$.

The key observation here is that on $[\![0,\tau'\wedge T]\!]$, the distance $|\tilde{Q}-\tilde{Q}'|+|\tilde{\lambda}-\tilde{\lambda}'|$ can only grow when an external order affects just one of the strategies in the sense that $V:=V^{C,C'}$ grows. Indeed, with $d\hat{Q}:=d\tilde{Q}-d\tilde{Q}'$ denoting the extra trades of C over C', and similarly $d\hat{L}$, $d\hat{M}$ the differences in order flow for C and C', we compute

$$\begin{aligned}
&\diamond\, d(|\tilde{Q}-\tilde{Q}'|+|\tilde{\lambda}-\tilde{\lambda}'|)\\
&=\operatorname{sgn}(\tilde{Q}-\tilde{Q}')\diamond d\hat{Q}+\operatorname{sgn}(\tilde{\lambda}-\tilde{\lambda}')\diamond\Big(-|d\hat{Q}|+\big(0\vec{+}(d\hat{L}-|d\hat{M}|)\vec{+}0\big)\Big)\\
&=\big(\operatorname{sgn}(\tilde{Q}-\tilde{Q}')-\operatorname{sgn}(\tilde{\lambda}-\tilde{\lambda}')\big)\diamond d\hat{Q}^+\\
&\quad+\big(-\operatorname{sgn}(\tilde{Q}-\tilde{Q}')-\operatorname{sgn}(\tilde{\lambda}-\tilde{\lambda}')\big)\diamond d\hat{Q}^-\\
&\quad+\operatorname{sgn}(\tilde{\lambda}-\tilde{\lambda}')\diamond\big(0\vec{+}(d\hat{L}-|d\hat{M}|)\vec{+}0\big).
\end{aligned}$$

Before time $\tau\wedge\tau'$, an extra buy order by C where C' cannot follow suit ("$d\hat{Q}^+>0$") can only happen when $\tilde{Q}'=\bar{q}$ or $\tilde{\lambda}'=\underline{\lambda}$. Therefore the $d\hat{Q}^+$-integrand in the preceding expression is negative in such a moment. Similarly, a sell order exclusive to C ("$d\hat{Q}^->0$") can only happen when $\tilde{Q}'=-\bar{q}$ or $\tilde{\lambda}'=\underline{\lambda}$, and then the above $d\hat{Q}^-$-integrand is negative. As a result, orders from C or C' can only decrease the distance $|\tilde{Q}-\tilde{Q}'|+|\tilde{\lambda}-\tilde{\lambda}'|$, and this distance increases by at most $dV=|d\hat{L}|+|d\hat{M}|$. It follows that

$$|\tilde{Q}-\tilde{Q}'|+|\tilde{\lambda}-\tilde{\lambda}'|\leq|q-q'|+|\lambda-\lambda'|+V \qquad \text{on } [\![0,\tau\wedge\tau'\wedge T]\!]. \tag{B.11}$$

Note that we can include $\tau\wedge\tau'\wedge T$ in (B.11) as by (B.7) and by construction of C', C and C' both unwind their position in this moment if $\tau\leq\tau'\wedge T$. If $\tau'<\tau\wedge T$, this can only happen due to external orders, and so (B.11) holds in this moment because its left-hand side remains unchanged while the right-hand side can only increase. Moreover, we can estimate the excess volume of trades by C over those of C'

$$\begin{aligned}\int_0^{\cdot} \diamond\, |d\hat{Q}| &= \int_0^{\cdot} 1_{\{\tilde{Q}'=\bar{q}\}} \diamond d\hat{Q}^+ + \int_0^{\cdot} 1_{\{\tilde{Q}'=-\bar{q}\}} \diamond d\hat{Q}^- \\ &\quad + \int_0^{\cdot} 1_{\{\lambda'=\underline{\lambda},\, \tilde{Q}'\in(-\bar{q},\bar{q})\}} \diamond |d\hat{Q}| \\ &\le 2(|\tilde{Q}-\tilde{Q}'| + |\tilde{\lambda}-\tilde{\lambda}'|) \\ &\le 2(|q-q'| + |\lambda-\lambda'| + V) \qquad \text{on } [\![0, \tau\wedge\tau'\wedge T]\!]. \end{aligned} \tag{B.12}$$

Here we use that when $\tilde{Q}' = \bar{q}$, the control C can increase the position $\tilde{Q}$ at most by $\bar{q} - \tilde{Q} = \tilde{Q}' - \tilde{Q}$ before reaching the upper position limit itself. An analogous consideration applies with $\tilde{Q}' = -\bar{q}$ for the amount by which C can decrease $\tilde{Q}$. Similarly, when $\tilde{\lambda}' = \underline{\lambda}$, C can trade at most a volume $\tilde{\lambda} - \underline{\lambda} = \tilde{\lambda} - \tilde{\lambda}'$ before itself exhausting the available liquidity.

Let us start with the proof of (i). In the following, we denote by c a generic constant which may change from line to line, but never depends on λ, λ', q, q', C, C'. Let c be a joint Lipschitz constant for $\mathtt{w}(T, 0, \cdot, 0, \cdot)$, $\mathtt{r}$ and for the functions

$$a(q,\lambda) := I(|q|,\lambda) - |q|\iota(\lambda) \qquad \text{and} \qquad b^{\pm}(q,\lambda) := I(|q|,\lambda) - 2q^{\pm}\iota(\lambda).$$

Representing both $\tilde{W}$ and $\tilde{W}'$ as in (B.3) allows us to write

$$\begin{aligned} &\tilde{W} - \tilde{W}' \\ &= \mathtt{w}(T,0,q,0,\lambda) - \mathtt{w}(T,0,q',0,\lambda') \\ &\quad + \int_0^{\cdot} a(\tilde{Q},\tilde{\lambda}) \diamond (0\vec{+}d\tilde{L}^C\vec{+}0) - \int_0^{\cdot} a(\tilde{Q}',\tilde{\lambda}') \diamond (0\vec{+}d\tilde{L}^{C'}\vec{+}0) \\ &\quad - \left(\int_0^{\cdot} b^+(\tilde{Q},\tilde{\lambda}) \diamond (0\vec{+}d\tilde{M}^{C,+}\vec{+}0) - \int_0^{\cdot} b^+(\tilde{Q}',\tilde{\lambda}') \diamond (0\vec{+}d\tilde{M}^{C',+}\vec{+}0) \right) \\ &\quad - \left(\int_0^{\cdot} b^-(\tilde{Q},\tilde{\lambda}) \diamond (0\vec{+}d\tilde{M}^{C,-}\vec{+}0) - \int_0^{\cdot} b^-(\tilde{Q}',\tilde{\lambda}') \diamond (0\vec{+}d\tilde{M}^{C',-}\vec{+}0) \right) \\ &\quad - \left(\int_0^{\cdot} 2b^+(\tilde{Q},\tilde{\lambda}) \diamond d\tilde{Q}^+ - \int_0^{\cdot} 2b^+(\tilde{Q}',\tilde{\lambda}') \diamond d\tilde{Q}'^+ \right) \\ &\quad - \left(\int_0^{\cdot} 2b^-(\tilde{Q},\tilde{\lambda}) \diamond d\tilde{Q}^- - \int_0^{\cdot} 2b^-(\tilde{Q}',\tilde{\lambda}') \diamond d\tilde{Q}'^- \right) \\ &\quad - \left(\zeta\left(\left(V_{[0,\cdot]}(C) + |\tilde{Q}|\right) - \left(V_{[0,\cdot]}(C') + |\tilde{Q}'|\right) \right) \right) \\ &\quad + \left(Y\left(\mathtt{r}(\tilde{Q},\tilde{\lambda}) - \mathtt{r}(\tilde{Q}',\tilde{\lambda}') \right) \right). \end{aligned}$$

Note that in contrast to (B.3), we opt here to express the difference in spread costs via the difference in position variation and the difference of terminal positions that need to be liquidated. The first amounts to $\int_0^{\cdot} \diamond\, |d\hat{Q}|$ and is thus controlled by (B.12); the second is controlled by (B.11). Both controls are of the desired form.

Since c is a Lipschitz constant for the function $\mathtt{w}(T, 0, \cdot, 0, \cdot)$, the first term contributes at most

$$|\mathtt{w}(T, 0, q, 0, \lambda) - \mathtt{w}(T, 0, q', 0, \lambda')| \leq c(|q - q'| + |\lambda - \lambda'|). \tag{B.13}$$

Also $\mathtt{r}$ is Lipschitz-continuous with this constant, and so the last term contributes at most $|Y|c(|\tilde{Q} - \tilde{Q}'| + |\tilde{\lambda} - \tilde{\lambda}'|)$, which by (B.11) is not more than

$$|Y|c(|q - q'| + |\lambda - \lambda'| + V) \qquad \text{on } [\![0, \tau \wedge \tau' \wedge T]\!]. \tag{B.14}$$

In fact, we can proceed similarly for the other terms. When a limit order or cancellation affects the environment of both C and C', the integrators for the a-terms coincide and the absolute value of the net contribution is found to be no more than

$$|a(\tilde{Q}, \tilde{\lambda}) - a(\tilde{Q}', \tilde{\lambda}')| \diamond (0\vec{+}|d\tilde{L}^C|\vec{+}0) \leq c(|\tilde{Q} - \tilde{Q}'| + |\tilde{\lambda} - \tilde{\lambda}'|) \diamond (0\vec{+}d\tilde{L}^{C,+}\vec{+}0).$$

Using (B.11) in the same way as we did for the Y-term, this gives a contribution

$$\begin{aligned}&|a(\tilde{Q}, \tilde{\lambda}) - a(\tilde{Q}', \tilde{\lambda}')| \diamond (0\vec{+}|d\tilde{L}^C|\vec{+}0)\\ &\leq c(|q - q'| + |\lambda - \lambda'| + V) \diamond (0\vec{+}|d\tilde{L}^C|\vec{+}0)\end{aligned}$$

for limit orders and cancellations $d\tilde{L}^C$ common to C and C'. We similarly find

$$\begin{aligned}&|b^{\pm}(\tilde{Q}, \tilde{\lambda}) - b^{\pm}(\tilde{Q}', \tilde{\lambda}')| \diamond (0\vec{+}d\tilde{M}^{C,\pm}\vec{+}0)\\ &\leq c(|q - q'| + |\lambda - \lambda'| + V) \diamond (0\vec{+}d\tilde{M}^{C,\pm}\vec{+}0)\end{aligned}$$

for common market buy and sell orders. Upon aggregation, we obtain a contribution from non-exclusive external orders which is no more than

$$c(|q-q'|+|\lambda-\lambda'|+V)\big(V_{[0,\cdot]}(2\tilde{M}^C)+V_{[0,\cdot]}(\tilde{L}^C)\big) \qquad \text{on } [\![0, \tau\wedge\tau'\wedge T]\!]. \tag{B.15}$$

When a market buy or a sell order affects only one of the environments but not the other, the resulting absolute change in realisable wealth due to the $b^{\pm}$-integrals is no more than its volume times $\max|b^{\pm}| \leq 2\bar{q}\iota(\underline{\lambda})$. Similarly, from the a-integrals, we collect no more than the exclusive volume times $\max|a| \leq \bar{q}\iota(\underline{\lambda})$. All of these contributions aggregate to no more than a multiple of

$$\bar{q}\iota(\underline{\lambda})V \qquad \text{on } [\![0, \tau \wedge \tau' \wedge T]\!]. \tag{B.16}$$

We can treat the $dQ^{\pm}$-integrals similarly as well: For common trades of C and C', we can net the integrands collecting a contribution of no more than

$$c(|q - q'| + |\lambda - \lambda'| + V)V_{[0,\cdot]}(\tilde{Q}) \qquad \text{on } [\![0, \tau \wedge \tau' \wedge T]\!]. \tag{B.17}$$

The contributions when only C trades while C' has to pause are bounded by $2\bar{q}\iota(\underline{\lambda})$ times the aggregate extra buy or sell volume $\int_0^{\cdot} \diamond\ |d\hat{Q}|$ that C does on top of C'. Due to (B.12), we thus pick up a contribution of at most

$$2\bar{q}\iota(\underline{\lambda})(|q-q'|+|\lambda-\lambda'|+V) \qquad \text{on } [\![0,\tau\wedge\tau'\wedge T]\!] \tag{B.18}$$

from these C-exclusive orders.

Adding the contribution estimates (B.13)–(B.18) to the spread costs controlled by (B.12) and (B.11), we find that

$$\begin{aligned}&|\tilde{W}-\tilde{W}'|\\&\le\ c(|q-q'|+|\lambda-\lambda'|+V)\big(1+|Y|+2V_{[0,\cdot]}(\tilde{M}^C)+V_{[0,\cdot]}(\tilde{L}^C)+2V_{[0,\cdot]}(\tilde{Q})\big)\\&\quad+2c\bar{q}\iota(\underline{\lambda})(|q-q'|+|\lambda-\lambda'|+V)+2\zeta(|q-q'|+|\lambda-\lambda'|+V)\end{aligned}$$

on $[\![0,\tau\wedge\tau'\wedge T]\!]$. Due to (B.2), the above total variation terms are jointly dominated by $c(\lambda\vee\lambda'-\underline{\lambda}+\tilde{L}^+)$. Finally, yet again exchanging c for a higher constant again denoted by c and still independent of the initial data and C and C', we arrive at (B.8) over $[\![0,\tau\wedge\tau'\wedge T]\!]$. Since in case $\tau\le\tau'\wedge T$, both C and C' are unwinding their position in this moment, their realisable portfolio values do not change (cf. also (2.26)) and thus remain as close as before. Since C' also stops trading after τ, (B.8) extends to $[\![0,\tau'\wedge T]\!]$ in any case.

For the proof of (ii), we express V in terms of the point process N as

$$\begin{aligned}&V_{t\wedge\tau\wedge\tau'}\\&=\int_{[\![0,t\wedge\tau\wedge\tau']\!]\times E\times[0,\infty)}\big(1_{\{g(\tilde{\lambda}_{s-})\wedge g(\tilde{\lambda}'_{s-})\le y<g(\tilde{\lambda}_{s-})\vee g(\tilde{\lambda}'_{s-})\}}|\rho(e)|\\&\qquad\qquad+1_{\{f(\tilde{\lambda}_{s-})\wedge f(\tilde{\lambda}'_{s-})\le y<f(\tilde{\lambda}_{s-})\vee f(\tilde{\lambda}'_{s-})\}}|\eta(e)|\big)N(ds,de,dy).\end{aligned} \tag{B.19}$$

Upon taking expectations, we can pass to the compensator $dt\otimes\nu(de)\otimes dy$ to obtain

$$\begin{aligned}\mathbb{E}[V_{t\wedge\tau\wedge\tau'}]=\mathbb{E}\bigg[\int_0^{t\wedge\tau\wedge\tau'}\bigg(&|g(\tilde{\lambda}_{s-})-g(\tilde{\lambda}'_{s-})|\int_E|\rho(e)|\nu(de)\\&+|f(\tilde{\lambda}_{s-})-f(\tilde{\lambda}'_{s-})|\int_E|\eta(e)|\nu(de)\bigg)ds\bigg].\end{aligned}$$

Taking c to be a Lipschitz constant for f and g and using (B.11), we get

$$\begin{aligned}\mathbb{E}[V_{t\wedge\tau\wedge\tau'}]&\le\mathbb{E}\bigg[\int_0^{t\wedge\tau\wedge\tau'}c(|\lambda-\lambda'|+|q-q'|+V_s)\\&\qquad\times\bigg(\int_E|\rho(e)|\nu(de)+\int_E|\eta(e)|\nu(de)\bigg)ds\bigg]\\&\le\int_E\big(|\rho(e)|+|\eta(e)|\big)\nu(de)\\&\qquad\times\bigg(c(|\lambda-\lambda'|+|q-q'|)t+c\int_0^t\mathbb{E}[V_{s\wedge\tau\wedge\tau'}]ds\bigg).\end{aligned}$$

Our estimate for $\mathbb{E}[V_{t\wedge\tau\wedge\tau'}]$ now follows (with a modified c taking into account the above ν-integral) from Gronwall's lemma.

To get the bound for $\mathbb{E}[V^2_{t\wedge\tau\wedge\tau'}]^{1/2}$, we consider again (B.19) and use Lemma C.1 to conclude that for some constant c depending only on $g(\infty)\vee f(\underline{\lambda})$,

$$\begin{aligned}
&\mathbb{E}[V^2_{T\wedge\tau\wedge\tau'}]^{1/2}\\
&\le c\bigg(\int_0^T\int_E\int_0^{g(\infty)\vee f(\underline{\lambda})}\mathbb{E}\Big[\big(1_{\{g(\tilde{\lambda}_{s-})\wedge g(\tilde{\lambda}'_{s-})\le y<g(\tilde{\lambda}_{s-})\vee g(\tilde{\lambda}'_{s-})\}}|\rho(e)|^2\\
&\qquad\qquad + 1_{\{f(\tilde{\lambda}_{s-})\wedge f(\tilde{\lambda}'_{s-})\le y<f(\tilde{\lambda}_{s-})\vee f(\tilde{\lambda}'_{s-})\}}|\eta(e)|^2\big)\\
&\qquad\qquad \times 1_{\{s\le T\wedge\tau\wedge\tau'\}}\Big]dy\,\nu(de)ds\bigg)^{1/2}\\
&= c\bigg(\int_0^T\mathbb{E}\Big[\Big(|g(\tilde{\lambda}_{s-})-g(\tilde{\lambda}'_{s-})|\int_E|\rho(e)|^2\nu(de)\\
&\qquad + |f(\tilde{\lambda}_{s-})-f(\tilde{\lambda}'_{s-})|\int_E|\eta(e)|^2\nu(de)\Big)1_{\{s\le T\wedge\tau\wedge\tau'\}}\Big]ds\bigg)^{1/2}.
\end{aligned}$$

From here, we can again proceed by a Lipschitz estimate and use (B.11) to arrive at

$$\begin{aligned}
\mathbb{E}[V^2_{t\wedge\tau\wedge\tau'}]^{1/2} &\le c\bigg(Tc(|\lambda-\lambda'|+|q-q'|+\mathbb{E}[V_{T\wedge\tau\wedge\tau'}])\\
&\qquad \times\int_E\big(|\rho(e)|^2+|\eta(e)|^2\big)\nu(de)\bigg)^{1/2}.
\end{aligned}$$

An application of (B.9) now yields (B.10) by passing to a higher constant c.

For the proof of (iii), let us denote by $\Gamma_s(z(e,y))$, $\Gamma'_s(z(e,y))$ the signal-based trades of C and C', respectively, from Lemma 2.4. Observe that $\tau'<\tau\wedge T$ can only happen if at time $\tau'=s$, the point process $N(ds,de,dy)$ puts a mark (e,y) which results in too high a market order or cancellation to be covered by the liquidity $R'_s(e,y):=\tilde{\lambda}'_{s-}-|\Gamma'_s(z(e,y))|-\underline{\lambda}$ remaining after a possible signal-based trade by C', while C is either not exposed to this order or its remaining liquidity $R_s(e,y):=\tilde{\lambda}_{s-}-|\Gamma_s(z(e,y))|-\underline{\lambda}$ suffices to cover that order without triggering a halt. Recalling that $\rho(e)\eta(e)\equiv 0$, this means in particular that a mark (e,y) has to be placed in one of the mark sets

$$\begin{aligned}
M^1_s &:= \{(e,y)\in E\times[0,g(\infty)\vee f(\underline{\lambda})] : R'_s(e,y)<\rho^-(e)+|\eta(e)|\le R_s(e,y)\},\\
M^2_s &:= \Big\{(e,y)\in E\times[0,g(\infty)\vee f(\underline{\lambda})] : y\in\big(g(r),g(r')\big]\cup\big(f(r),f(r')\big],\text{ where}\\
&\qquad\qquad r=R_s(e,y)+\underline{\lambda},\ r'=R'_s(e,y)+\underline{\lambda}\Big\}.
\end{aligned}$$

Notice that with w denoting a modulus of continuity for $\nu(\rho^-+|\eta|>\cdot)$ on $(0,\infty)$, we can estimate for any y-section $M^{1,y}_s$ of M^1_s its ν-measure

$$\nu(M^{1,y}_s)\le w\big(R_s(e,y)-R'_s(e,y)\big)\le w(\tilde{\lambda}_{s-}-\tilde{\lambda}'_{s-}),$$

where for the second estimate, we used that $|\Gamma'(z(e,y))| \leq |\Gamma(z(e,y)|$ (recalling that C' always trades at most as much as C). For the same reason, we can estimate the dy-measure of the e-sections $M_s^{2,e}$ of M_s^2 by

$$\begin{aligned} dy(M_s^{2,e}) &\leq 2 \max_{y\in[0,g(\infty)\vee f(\underline{\lambda})]} \Big(\big|g\big(r'(y)\big) - g\big(r(y)\big)\big| \vee \big|f\big(r'(y)\big) - f(r)\big|\Big) \\ &\leq c|\tilde{\lambda}_{s-} - \tilde{\lambda}'_{s-}|, \end{aligned}$$

where as before $r(y) = R_s(e,y) + \underline{\lambda}$, $r'(y) = R'_s(e,y) + \underline{\lambda}$. It follows that

$$\begin{aligned} &\mathbb{P}[\tau' < \tau \wedge T] \\ &\leq \mathbb{E}\Big[\int_{[\![0,\tau'\wedge T]\!]\times E\times[0,g(\infty)\vee f(\underline{\lambda})]} \big(1_{M_s^1}(e,y) + 1_{M_s^2}(e,y)\big) N(ds,de,dy)\Big] \\ &\leq \mathbb{E}\Big[\int_0^{\tau'\wedge T} \Big(\big(g(\infty)\vee f(\underline{\lambda})\big) w(\tilde{\lambda}_{s-} - \tilde{\lambda}'_{s-}) + \nu(E)c|\tilde{\lambda}_{s-} - \tilde{\lambda}'_{s-}|\big)\Big) ds\Big] \\ &\leq c\int_0^T \mathbb{E}\big[\big(w(\delta)+\delta\big)\big|_{\delta=|\lambda-\lambda'|+|q-q'|+V_{\tau'\wedge T}}\big] ds \\ &\leq cT\big(w_{\rho^-+|\eta|}(\delta)+\delta\big)\big|_{\delta=|\lambda-\lambda'|+|q-q'|+\mathbb{E}[V_{\tau'\wedge T}]}, \end{aligned}$$

where we used (B.11) in the second-to-last step and Jensen's inequality for the concave envelope $w_{\rho^-+|\eta|}$ of w in the last one. Assertion (iii) now follows from the estimate in (ii).

□

With Lemma B.2 at hand, we can now establish the main result of this section.

Corollary B.4 *Assume that $\nu(\rho^- + \eta > \cdot)$ is continuous on $(0,\infty)$ and suppose that $g(\infty) < \infty$ and $\eta, \rho \in L^2(\nu)$ with ρ^+ even satisfying the exponential integrability condition (3.9). Then $v_0(T,\cdot,\cdot) = v(T,\cdot,\cdot,0,0)$ is continuous on $[\underline{\lambda},\infty)\times[-\bar{q},\bar{q}]$.*

Proof Choose $\bar{\lambda} > \underline{\lambda}$. Let us start by proving the exponential moment estimate

$$\sup_{(\lambda,q)\in[\underline{\lambda},\bar{\lambda}]\times[-\bar{q},\bar{q}]} \sup_{C\in\tilde{\mathcal{C}}(\lambda,q)} \mathbb{E}[\exp(-cW^C_{T+})] < \infty \qquad \text{for any } c > 0. \tag{B.20}$$

For this, consider again the terminal wealth representation (B.3) and observe that it yields the lower bound

$$\begin{aligned}
\tilde{W}^C_{T+} \geq\ & \mathrm{w}(0,q,0,\lambda) + Y\mathrm{r}(\tilde{Q}^C_{T+},\tilde{\lambda}^C_{T+}) \\
& - \int_0^{T+} I(|\tilde{Q}^C|,\tilde{\lambda}^C) \diamond \big(0\vec{+}(d\tilde{L}^{C,-} + |d\tilde{M}^C|)\vec{+}0\big) \\
& - 2\int_0^{T+} \big(I(|\tilde{Q}^C|,\tilde{\lambda}^C) - |\tilde{Q}^C|\iota(\tilde{\lambda}^C) + \zeta\big)(1_{\{\tilde{Q}^C>0\}} \diamond d\tilde{Q}^{C,+} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad + 1_{\{\tilde{Q}^C<0\}} \diamond d\tilde{Q}^{C,-}) \\
\geq\ & \mathrm{w}(0,q,0,\lambda) + Y\mathrm{r}(\tilde{Q}^C_{T+},\tilde{\lambda}^C_{T+}) - \bar{q}\iota(\underline{\lambda})\big(\tilde{L}^{C,-}_T + V_{[0,T]}(\tilde{M}^C)\big) \\
& - 2\big(\bar{q}\iota(\underline{\lambda}) + \zeta\big)V_{[0,T]}(\tilde{Q}^C).
\end{aligned}$$

With (B.2) and using the independence of Y from $\mathcal{F}_T$ and its symmetry (see (3.5)), we thus get with the inventory bound from (2.13) for any $c > 0$ that

$$\begin{aligned}
\mathbb{E}[\exp(-c\tilde{W}^C_T)] \leq \exp\Big(& -c\big(\mathrm{w}(0,q,0,\lambda) - \bar{q}\iota(\underline{\lambda})(\lambda-\underline{\lambda})\big) + \mathcal{L}_Y(c\bar{q})\Big) \\
& \times \mathbb{E}\Big[\exp\Big(c\big(3\bar{q}\iota(\underline{\lambda}) + 2\zeta\big)\bar{L}^+_T\Big)\Big].
\end{aligned}$$

By the Lévy–Khintchine formula, the right-hand side is finite under condition (3.9). Since the above estimate is uniform in $C \in \tilde{\mathcal{C}}(\lambda,q)$ and $\mathrm{w}(0,\lambda,0,q)$ is bounded from below uniformly in $(\lambda,q) \in [\underline{\lambda},\bar{\lambda}] \times [-\bar{q},\bar{q}]$, this estimate yields (B.20).

Take now $(\lambda,q),(\lambda',q') \in [\underline{\lambda},\bar{\lambda}] \times [-\bar{q},\bar{q}]$. For any $\epsilon > 0$, we can take an ϵ-optimal $C \in \tilde{\mathcal{C}}(\lambda,q)$ and assume that it triggers a halt only when reducing its exposure $|Q^C|$, i.e., it satisfies (B.7). Indeed, triggering a halt while expanding the exposure means executing a loss-leading round-trip because of the forced unwinding of any remaining position after a halt has been triggered.

Now consider the associated $C' \in \tilde{\mathcal{C}}(\lambda',q')$ from Lemma B.2 for the estimate

$$\begin{aligned}
v_0(T,\lambda,q) - v_0(T,\lambda',q') &\leq \epsilon + \mathbb{E}[U(W^C_{T+})] - \mathbb{E}[U(\tilde{W}^{C'}_{T+})] \\
&\leq \epsilon + \mathbb{E}\big[U'(\tilde{W}^C_{T+})(\tilde{W}^C_{T+} - \tilde{W}^{C'}_{T+})1_{\{\tau^{C'}\geq\tau^C\wedge T\}}\big] \\
&\quad + \mathbb{E}\big[\big(U(\tilde{W}^C_{T+}) - U(\tilde{W}^{C'}_{T+})\big)1_{\{\tau^{C'}<\tau^C\wedge T\}}\big], \qquad \text{(B.21)}
\end{aligned}$$

where we used the concavity of the utility function U. Let us denote in the following by c again a generic constant that does not depend on the initial data and neither on C or C'. Note that on $\{\tau^C \leq \tau^{C'} < T\}$, both C and C' liquidate their positions at the same time and then stop operating. So $\tilde{W}^C_{T+} = \tilde{W}^C_{\tau^{C'}+} = \tilde{W}^C_{\tau^{C'}}$ and $\tilde{W}^{C'}_{T+} = \tilde{W}^{C'}_{\tau^{C'}+} = \tilde{W}^{C'}_{\tau^{C'}}$ on this set. Using this, applying then Lemma B.2 (i) followed by Hölder's inequality, we find that the first expectation in (B.21) is

$$\begin{aligned}
&\mathbb{E}\big[U'(\tilde{W}^C_{T+})(\tilde{W}^C_{T+} - \tilde{W}^{C'}_{T+})1_{\{\tau^{C'} \geq \tau^C \wedge T\}}\big] \\
&= \mathbb{E}\big[U'(\tilde{W}^C_{T+})(\tilde{W}^C_{\tau^{C'} \wedge T+} - \tilde{W}^{C'}_{\tau^{C'} \wedge T+})1_{\{\tau^{C'} \geq \tau^C \wedge T\}}\big] \\
&\leq \|U'(\tilde{W}^C_{T+})\|_{L^4(\mathbb{P})} \|1 + \bar{\lambda} - \underline{\lambda} + \bar{L}^+_T + |Y|\|_{L^4(\mathbb{P})} \\
&\quad \times \|c(|\lambda - \lambda'| + |q - q'| + V^{C,C'}_{\tau^C \wedge \tau^{C'} \wedge T})\|_{L^2(\mathbb{P})} \\
&\leq c\big(|\lambda - \lambda'| + |q - q'| + (|\lambda - \lambda'| + |q - q'|)^{1/2}\big),
\end{aligned}$$

where the final inequality is due to (B.20) and Lemma B.2 (ii). Clearly, there is $\delta(\epsilon) > 0$ such that for initial data from $[\underline{\lambda}, \bar{\lambda}] \times [-\bar{q}, \bar{q}]$ with $|\lambda - \lambda'| + |q - q'| =: \delta < \delta(\epsilon)$, the above expression is less than ϵ.

Again by (B.20), the utilities in the second expectation in (B.21) are from a uniformly integrable family of random variables. They are integrated over the set $\{\tau^{C'} < \tau^C \wedge T\}$ whose probability, due to Lemma B.2 (iii), is uniformly controlled by $c(w_{\rho^- + |\eta|}(c\delta) + \delta)$, where again $\delta = |\lambda - \lambda'| + |q - q'|$. So we can decrease $\delta(\epsilon) > 0$ even further and still in a way which does not dependent on (λ, q), (λ', q') or C, C' such that for all initial data satisfying $|\lambda - \lambda'| + |q - q'| = \delta < \delta(\epsilon)$, the second expectation in (B.21) is less than ϵ.

It follows that if $|\lambda - \lambda'| + |q - q'| < \delta(\epsilon)$, then

$$v_0(T, \lambda, q) - v_0(T, \lambda', q') \leq 3\epsilon.$$

Exchanging the roles of (λ, q) and (λ', q') in the above argument yields $\delta(\epsilon) > 0$ such that

$$|v_0(T, \lambda, q) - v_0(T, \lambda', q')| \leq 3\epsilon$$

for all $(\lambda, q), (\lambda', q') \in [\underline{\lambda}, \bar{\lambda}] \times [-\bar{q}, \bar{q}]$ with $|\lambda - \lambda'| + |q - q'| < \delta(\epsilon)$. This proves continuity of $v_0(T, \cdot, \cdot)$ on this domain. □

B.5 Continuity of the value function

Continuity of the value function on $[\underline{\lambda}, \infty) \times [-\bar{q}, \bar{q}]$ follows by recalling its multiplicative structure (3.8) and by combining Lemma B.1 with Corollary B.4.

Appendix C: A moment estimate for Poisson integrals

For the proof of continuity of the value function, we used the following moment estimate for the integral of a predictable mark-dependent process against a Poisson point process. We record it here for the sake of completeness as we could not find a result to this effect in the literature.

Lemma C.1 *On $(\Omega, \mathcal{F}, \mathbb{P})$, let $N = N(ds, de)$ be a marked Poisson point process with finite intensity measure $ds \otimes \nu(de)$ and marks in $(E, \mathcal{E})$. Suppose that $(\mathcal{F}_t)$ is a filtration to which N is adapted in the sense that for any $E' \in \mathcal{E}$, the random variable $N([0,t] \times E')$ is $\mathcal{F}_t$-measurable, $t \geq 0$. Then for $n = 0, 1, \ldots$, there is a constant c_n depending only on $T\nu(E)$ such that for any $(\mathcal{P} \otimes \mathcal{E})$-measurable α, we have*

$$\left\| \int_{[0,T]\times E} |\alpha_s(e)| \, N(ds, de) \right\|_{L^{2^n}(\mathbb{P})} \leq c_n \|\alpha\|_{L^{2^n}(\mathbb{P}\otimes ds|_{[0,T]}\otimes\nu)}. \tag{C.1}$$

Proof As the claim is clearly true for $n = 0$ with $c_0 := 1$, we can proceed by induction.

It suffices to consider $\alpha \in L^1(\mathbb{P} \otimes ds \otimes \nu(de))$. For such α, the compensated integral $\int_{[0,\cdot]\times E} |\alpha_s(e)|\bar{N}(ds, de)$ is a martingale. This allows us to apply the Burkholder–Davis–Gundy inequality to obtain a universal constant b_n such that

$$\begin{aligned}
\left\| \int_{[0,T]\times E} |\alpha_s(e)|\bar{N}(ds, de) \right\|_{L^{2^n}(\mathbb{P})} &\leq b_n \left\| \left[\int_{[0,\cdot]\times E} |\alpha_s(e)|\bar{N}(ds, de) \right]_T^{1/2} \right\|_{L^{2^n}(\mathbb{P})} \\
&= b_n \left\| \int_{[0,T]\times E} |\alpha_s(e)|^2 N(ds, de) \right\|_{L^{2^{n-1}}(\mathbb{P})}^{1/2} \\
&\leq b_n (c_{n-1} \| |\alpha|^2 \|_{L^{2^{n-1}}(\mathbb{P}\otimes ds|_{[0,T]}\otimes\nu)})^{1/2} \\
&= b_n \, c_{n-1}^{1/2} \|\alpha\|_{L^{2^n}(\mathbb{P}\otimes ds|_{[0,T]}\otimes\nu)},
\end{aligned}$$

where the last estimate follows by applying the claimed result for $n' := n - 1$ to $\alpha' := |\alpha|^2$. Moreover, the compensating integral satisfies

$$\begin{aligned}
&\left\| \int_{[0,T]\times E} |\alpha_s(e)| \, ds \otimes \nu(de) \right\|_{L^{2^n}(\mathbb{P})} \\
&= \left(\int_{([0,T]\times E)^{2^n}} \mathbb{E}\left[\prod_{i=1}^{2^n} |\alpha_{s_i}(e_i)| \right] \bigotimes_{i=1}^{2^n} \big(ds_i \otimes \nu(de_i)\big) \right)^{1/2^n} \\
&\leq \left(\int_{([0,T]\times E)^{2^n}} \prod_{i=1}^{2^n} \|\alpha_{s_i}(e)\|_{L^{2^n}(\mathbb{P})} \bigotimes_{i=1}^{2^n} \big(ds_i \otimes \nu(de_i)\big) \right)^{1/2^n} \\
&= \int_{[0,T]\times E} \|\alpha_s(e)\|_{L^{2^n}(\mathbb{P})} \big(ds \otimes \nu(de)\big) \\
&\leq \big(T\nu(E)\big)^{1-1/2^n} \left(\int_{[0,T]\times E} \mathbb{E}[|\alpha_s(e)|^{2^n}] \big(ds \otimes \nu(de)\big) \right)^{1/2^n} \\
&= \big(T\nu(E)\big)^{1-1/2^n} \|\alpha\|_{L^{2^n}(\mathbb{P}\otimes ds|_{[0,T]}\otimes\nu)},
\end{aligned}$$

the last estimate following from Hölder's inequality with $q = 2^n$, $1/p = 1 - 1/2^n$. By the triangle inequality, the previous two estimates give

$$\begin{aligned}\left\| \int_{[0,T]\times E} |\alpha_s(e)|\, N(ds, de) \right\|_{L^{2^n}(\mathbb{P})} \\ \leq \left(b_n c_{n-1}^{1/2} + \left(\nu(E)T\right)^{1-1/2^n} \right) \|\alpha\|_{L^{2^n}(\mathbb{P}\otimes ds|_{[0,T]}\otimes\nu)},\end{aligned}$$

and we obtain (C.1) for $c_n := b_n c_{n-1}^{1/2} + (\nu(E)T)^{1-1/2^n}$. □

Acknowledgements L. Körber and P. Bank are grateful to the Oxford-Man Institute of Quantitative Finance for their generous hospitality. We are also grateful to participants at the Mathematical Finance Seminar at Imperial College, the German Probability and Statistics Days in Essen, the SIAM Conference on Financial Mathematics and Engineering in Philadelphia and the Women in Mathematical Finance Conference at Rutgers University. We also thank Martin Schweizer for his careful reading of our manuscript and helping us to remove numerous typos, inconsistencies, and unclear formulations.

Funding information Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests The authors declare no competing interests.

References

1. Alfonsi, A., Blanc, P.: Dynamic optimal execution in a mixed-market-impact Hawkes price model. Finance Stoch. **20**, 183–218 (2016)
2. Almgren, R., Chriss, N.: Optimal execution of portfolio transactions. J. Risk **3**(2), 5–39 (2000)
3. Back, K.: Insider trading in continuous time. Rev. Financ. Stud. **5**, 387–409 (1992)
4. Bank, P., Besslich, D.: Modelling information flows by Meyer-σ-fields in the singular stochastic control problem of irreversible investment. Ann. Appl. Probab. **30**, 2923–2962 (2020)
5. Bank, P., Körber, L.: Merton's optimal investment problem with jump signals. SIAM J. Financ. Math. **13**, 1302–1325 (2022)
6. Belak, C., Muhle-Karbe, J., Ou, K.: Liquidation in target zone models. Mark. Microstruct. Liq. **04**, 1950010 (2018)
7. Cartea, Á., Pérez Arribas, I., Sánchez-Betancourt, L.: Double-execution strategies using path signatures. SIAM J. Financ. Math. **13**, 1379–1417 (2022)
8. Cartea, Á., Jaimungal, S.: Incorporating order-flow into optimal execution. Math. Financ. Econ. **10**, 339–364 (2016)
9. Cartea, Á., Jaimungal, S., Ricci, J.: Algorithmic trading, stochastic control, and mutually exciting processes. SIAM Rev. **60**, 673–703 (2018)
10. Cartea, Á., Jaimungal, S., Wang, Y.: Spoofing and price manipulation in order-driven markets. Appl. Math. Finance **27**, 67–98 (2020)

11. Cartea, Á., Sánchez-Betancourt, L.: Brokers and informed traders: dealing with toxic flow and extracting trading signals. SIAM J. Financ. Math. **16**, 243–270 (2025)
12. Cartea, Á., Wang, Y.: Market making with alpha signals. Int. J. Theor. Appl. Finance **23**, 2050016 (2020)
13. Casgrain, P., Jaimungal, S.: Trading algorithms with learning in latent alpha models. Math. Finance **29**, 735–772 (2019)
14. Cayé, T., Muhle-Karbe, J.: Liquidation with self-exciting price impact. Math. Financ. Econ. **10**, 15–28 (2016)
15. Chen, H., Petukhov, A., Wang, J., Xing, H.: The dark side of circuit breakers. J. Finance **79**, 1405–1455 (2024)
16. Chevyrev, I., Korepanov, A., Melbourne, I.: Superdiffusive limits beyond the Marcus regime for deterministic fast-slow systems. Commun. Am. Math. Soc. **4**, 746–786 (2024)
17. El Karoui, N.: Les aspects probabilistes du controle stochastique. In: Hennequin, P.L. (ed.) École d'Été de Probabilités de Saint-Flour IX-1979. Lecture Notes in Mathematics, vol. 876, pp. 73–238. Springer, Berlin (1981)
18. Fu, G., Horst, U., Xia, X.: Portfolio liquidation games with self-exciting order flow. Math. Finance **32**, 1020–1065 (2022)
19. Huberman, G., Stanzl, W.: Price manipulation and quasi-arbitrage. Econometrica **72**, 1247–1275 (2004)
20. Kyle, A.S.: Continuous auctions and insider trading. Econometrica **53**, 1315–1335 (1985)
21. Kyle, A.S.: Trading halts and price limits. Rev. Futures Mark. **7**, 426–434 (1988)
22. Lehalle, C.-A., Mounjid, O.: Limit order strategic placement with adverse selection risk and the role of latency. Mark. Microstruct. Liq. **3**, 1750009 (2017)
23. Lehalle, C.-A., Neuman, E.: Incorporating signals into optimal trading. Finance Stoch. **23**, 275–311 (2019)
24. Lenglart, E.: Tribus de Meyer et théorie des processus. In: Azéma, J., Yor, M. (eds.) Séminaire de Probabilités XIV. Lecture Notes in Mathematics, vol. 784, pp. 500–546. Springer, Berlin (1980)
25. Lewis, M.: Flash Boys: A Wall Street Revolt. Norton (2014)
26. Marcus, S.I.: Modeling and approximation of stochastic differential equations driven by semimartingales. Stochastics 4, 223–245 (1980/81)
27. Neuman, E., Voß, M.: Optimal signal-adaptive trading with temporary and transient price impact. SIAM J. Financ. Math. **13**, 551–575 (2022)
28. Obizhaeva, A.A., Wang, J.: Optimal trading strategy and supply/demand dynamics. J. Financ. Mark. **16**, 1–32 (2013)
29. Pástor, L'., Stambaugh, R.F.: Liquidity risk and expected stock returns. J. Polit. Econ. **111**, 642–685 (2003)