

Incorporating signals into optimal trading

Charles-Albert Lehalle^{1,2} · Eyal Neuman^{2,3}

Received: 26 April 2017 / Accepted: 2 November 2018 / Published online: 14 February 2019
© The Author(s) 2019

Abstract We incorporate a Markovian signal in the optimal trading framework which was initially proposed by Gatheral et al. (Math. Finance 22:445–474, 2012) and provide results on the existence and uniqueness of an optimal trading strategy. Moreover, we derive an explicit singular optimal strategy for the special case of an Ornstein–Uhlenbeck signal and an exponentially decaying transient market impact. The combination of a mean-reverting signal along with a market impact decay is of special interest, since they affect the short term price variations in opposite directions. Later, we show that in the asymptotic limit where the transient market impact becomes instantaneous, the optimal strategy becomes continuous. This result is compatible with the optimal trading framework which was proposed by Cartea and Jaimungal (Appl. Math. Finance 20:512–547, 2013). In order to support our models, we analyse nine months of tick-by-tick data on 13 European stocks from the NASDAQ OMX exchange. We show that order book imbalance is a predictor of the future price move and has some mean-reverting properties. From this data, we show that market participants, especially high-frequency traders, use this signal in their trading strategies.

Keywords Optimal portfolio liquidation · Market impact · Optimal stochastic control · Predictive signals

Mathematics Subject Classification (2010) 93E20 · 60H30 · 91G80

JEL Classification C02 · C61 · G11

✉ E. Neuman
e.neumann@imperial.ac.uk

C.-A. Lehalle
charles@lehalles.net

¹ Capital Fund Management, Paris, France

² CFM-Imperial College Institute, London, SW7 1NE, UK

³ Department of Mathematics, Imperial College London, London, SW7 1NE, UK

1 Introduction

The financial crisis of 2008/2009 raised concerns about the inventories kept by intermediaries. Regulators and policy makers took advantage of two main regulatory changes (Reg NMS in the US and MiFID in Europe) which were followed by the creation of worldwide trade repositories. They also enforced more transparency on the transactions and hence on market participants' positions, which pushed the trading processes towards electronic platforms [30, Introduction]. Simultaneously, consumers and producers of financial products asked for less complexity and more transparency.

This tremendous pressure on the business habits of the financial system shifted it from a customised and high margins industry, in which intermediaries could keep large (and potentially risky) inventories, to a mass market industry where logistics have a central role. As a result, investment banks nowadays unwind their risks as fast as possible. In the context of small margins and high velocity of position changes, trading costs are of paramount importance. A major factor of the trading costs is the market impact: the faster the trading rate, the more the buying or selling pressure will move the price in a detrimental way.

Academic efforts to reduce the transaction costs of large trades started with the seminal papers of Almgren and Chriss [6] and Bertsimas and Lo [11]. Both models deal with the trading process of one large market participant (for instance, an asset manager or a bank) who would like to buy or sell a large amount of shares or contracts during a specified duration. The cost minimisation problem turned out to be quite involved, due to multiple constraints on the trading strategies. On the one hand, the market impact (see [8] and references therein) demands to trade slowly, or at least at a pace which takes into account the available liquidity. On the other hand, traders have an incentive to trade rapidly, because they do not want to carry the risk of an adverse price move far away from their decision price.

The importance of optimal trading in the industry generated a lot of variations for the initial mean-variance minimisation of the trading costs (see [17, Chaps. 6 and 7], [25, Chap. 2.3] and [30, Chap. 3] for details). In this paper, we consider the mean-variance minimisation problem in the context of stochastic control (see e.g. [29, 12]). In this approach, some more realistic control variables which are related to order book dynamics and specific stochastic processes for the underlying price can be used (see [26, 32] for related work).

In this paper, we address the question of how to incorporate *signals*, which are predicting short term price moves, into optimal trading problems. Usually optimal execution problems focus on the tradeoff between market impact and market risk. However, in practice many traders and trading algorithms use short term price predictors. Most of such documented predictors relate to order book dynamics (see e.g. [18] and [16]). They can be divided into two categories: signals which are based on liquidity consuming flows [16], and signals that measure the imbalance of the current liquidity. In [31], an example of how to use liquidity imbalance signals within a very short trading tactic was studied. These two types of signals are closely related, since within short terms, price moves are driven by matching of liquidity supply and demand (i.e., current offers and consuming flows).

As mentioned earlier, one of the major influencers on transaction costs is the market impact. Empirical studies have shown that the influence of the market impact is *transient*, that is, it decays within a short time period after each trade (see [8] and references therein). In this paper, we focus on two frameworks which take into account different types of market impact:

- Gatheral, Schied and Slynko (GSS) framework [24], in which the market impact is transient and strategies have a fuel constraint, i.e., orders are finished before a given date T ;
- Cartea and Jaimungal (CJ) framework [13], where the market impact is instantaneous and the fuel constraint on the strategies is replaced by a smooth terminal penalisation.

Note that [24] is not the only framework with market impact decay. This kind of dynamics was originally introduced in [33] and reused in [2] as well as in some other papers. We decided to focus on these two frameworks since they are extensively used in the financial literature. The model and analysis which are developed in this paper could be applied also to other optimal trading frameworks.

The main theoretical result of this work deals with the addition of a Markovian signal into the optimal trading problem which was studied in [24]. We argue in Sect. 2.1 that this is modelled mathematically by adding a Markovian drift to the martingale price process. We formulate a cost functional which consists of the trading costs and the risk of holding inventory at each given time. Then we prove that there exists at most one optimal strategy that minimises this cost functional. The optimal strategy is formulated as a solution to an integral equation. We then derive explicitly the optimal strategy for the special case where the signal is an Ornstein–Uhlenbeck process. From the mathematical point of view, this is the first time that a non-martingale price process is incorporated into an optimal liquidation problem with a decaying market impact. Therefore the results of Theorems 2.3 and 2.4 extend [24, Proposition 2.9 and Theorem 2.11], respectively. Later we show that in the asymptotic regime where the transient market impact becomes instantaneous, the singular optimal strategies which were derived in the GSS framework become continuous. Moreover, we show that the asymptotics of the optimal strategy in the GSS framework coincide with the optimal strategy which is obtained in the CJ framework (see Remark 2.8 and Sect. 3). This benchmark between different trading frameworks provides researchers and practitioners a wider overview when they are facing realistic trading problems.

The use of predictive signals in optimal trading in the context described above is relatively new (see [16]). To the best of our knowledge, this is the first time that a Markovian signal and a transient market impact are confronted. The GSS framework already includes a transient market impact, without using signals. The CJ framework includes only a bounded Markovian signal and not a decaying market impact. Moreover, our results on optimal trading in the GSS framework incorporate a risk aversion term into the cost functional, which was not taken into account in the results of [24].

The main contribution of this work is in providing a new framework for optimal trading, which is an extension of the classical frameworks of [13] and [24], among others. The motivation to use this framework arises from market needs as our data analysis in Sect. 4 suggests. From a theoretical point of view, these models of trading

with signals provide some new mathematical challenges. We describe in short two of these challenges.

The optimal strategies that we derive in Theorem 2.4 and Corollary 2.7 (i.e., in the GSS framework) are deterministic, and they use only information on the signal at time 0. One of the challenging questions which remains open is how to optimise the trading costs over strategies which are adapted to the signal's filtration (see Remark 2.9).

An interesting phenomenon which arises from our results is that the optimal strategies may be not monotonic once we take into account trading signals (see Fig. 1). This implies that price manipulations triggered by trading strategies are possible. Another challenge is to establish conditions on the market impact kernel function and on the signal that prevent price manipulations (see Remark 2.10).

Another contribution of this paper is a statistical analysis of the imbalance signal and its use in actual trading, which we present in Sect. 4. In order to validate our assumptions and theoretical results, we use nine months of real data from Nordic European equity markets (the NASDAQ OMX exchange) to demonstrate the existence of a liquidity-driven signal. We focus the analysis on 13 stocks, accounting for more than 9 billions of transactions. We also show that practitioners are conditioning, at least partly, their trading rate on this signal. Up to 2014, this exchange provided with each transaction the identity of the buyer and the seller. This database was already used for some academic studies; hence the reader can refer to [38, Sect. 2] for more details. We added to these labelled trades a database of Capital Fund Management (CFM) that contains information on the state of the order book just before each transaction. Thanks to this hybrid database, we were able to compute the imbalance of the liquidity just before decisions are taken by participants (i.e., sending market orders which consume liquidity).

We divide most members of the NASDAQ OMX into four classes: global investment banks, institutional brokers, high-frequency market makers and high-frequency proprietary traders (the classification is detailed in the [Appendix](#)). Then we compute the average value of the imbalance just before each type of participant takes a decision (see Fig. 4). The conclusion is that some participants condition their trading rate on the liquidity imbalance. Moreover, we provide a few graphs that demonstrate a positive correlation between the state of the imbalance and the future price move. These graphs also provide evidence for the mean-reverting nature of the imbalance signal (see Figs. 5–7). In Fig. 9, we present the estimated trading speed of market participants as a function of the average value of the imbalance, within a medium time scale of 10 minutes. The exhibited relation between the trading rate and the signal in this graph is compatible with our theoretical findings.

This paper is structured as follows. In Sect. 2, we introduce a model with market impact decay, a Markovian signal and strategies with a fuel constraint (i.e., in the GSS framework). We provide general existence and uniqueness theorems, and then give an explicit solution for the case of an Ornstein–Uhlenbeck signal. The addition of a signal to the market impact decay is the central ingredient of this section. In Sect. 3, we compare our results from Sect. 2 to the corresponding results in the CJ framework. We show that the optimal strategy in the GSS framework coincides with the optimal strategy in the CJ framework in the asymptotic limit where the transient market impact becomes instantaneous and the signal is an Ornstein–Uhlenbeck pro-

cess. In Sect. 4, we provide empirical evidence for the predictability of the imbalance signal and its use by different types of market participants. We also perform a statistical analysis which supports our focus on an Ornstein–Uhlenbeck signal in the example which is given in Sect. 2. The last section is dedicated to the proofs of the main results.

2 Model setup and main results

2.1 Model setup and definition of the cost functional

In this section, we define a model which incorporates a Markovian signal into the GSS optimal trading framework. Definitions and results from [24] are used throughout this section.

We consider a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying the usual conditions, where \mathcal{F}_0 is trivial. Let $M = (M_t)_{t \geq 0}$ be a right-continuous martingale and $I = (I_t)_{t \geq 0}$ a homogeneous càdlàg Markov process, independent from M , and satisfying

$$\int_0^T E_\iota[|I_t|] dt < \infty \quad \text{for all } \iota \in \mathbb{R}, T > 0.$$

Here E_ι represents the expectation conditional on $I_0 = \iota$. In our model, I represents a signal that is observed by the trader.

We assume that the asset price process P , which is unaffected by trading transactions, is given by

$$dP_t = I_t dt + dM_t, \quad t \geq 0;$$

hence the signal interacts with the price through the drift term. This setting allows us to consider a large class of signals. The visible asset price, which is described later, also depends on the market impact that is created by the trader's transactions.

Let $[0, T]$ be a finite time interval and $x > 0$ the initial inventory of the trader. Let X_t be the amount of inventory held by the trader at time t . We say that X is an *admissible strategy* if it satisfies

- (i) $t \mapsto X_t$ is left-continuous and adapted;
- (ii) $t \mapsto X_t$ has \mathbb{P} -a.s. bounded total variation;
- (iii) $X_0 = x$ and $X_t = 0$ \mathbb{P} -a.s. for all $t > T$.

In what follows, we often consider integration, with respect to an admissible strategy X , as an integration with respect to the measure dX_t which is supported on $[0, T]$.

As in [24, 20, 19], we assume that the visible price $S = (S_t)_{t \geq 0}$ is affected by a transient market impact and is given by

$$S_t = P_t + \int_{[0,t)} G(t-s) dX_s, \quad t \geq 0, \tag{2.1}$$

where the *decay kernel* $G : (0, \infty) \rightarrow [0, \infty)$ is a measurable function such that

$$G(0) := \lim_{t \downarrow 0} G(t) \text{ exists.} \tag{2.2}$$

We interpret the integral in (2.1) as a Lebesgue–Stieltjes integral.

Next we derive the transaction costs which are associated with the execution of a strategy X . Note that if X_t is continuous in t , then the trading costs that arise by an infinitesimal order dX_t are $S_t dX_t$. When X_t has a jump of size ΔX_t at t , the price moves from S_t to $S_{t+} = S_t + G(0)\Delta X_t$, and the costs associated with the trade ΔX_t are given by (see [24, Sect. 2])

$$\frac{G(0)}{2}(\Delta X_t)^2 + S_t \Delta X_t.$$

It follows that the trading costs which arise from the strategy X are given by

$$\begin{aligned} \int_{[0,T]} S_t dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2 &= \int_{[0,T]} \int_0^t I_s ds dX_t \\ &\quad + \int_{[0,T]} \int_{[0,t)} G(t-s) dX_s dX_t \\ &\quad + \int_{[0,T]} M_t dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2, \end{aligned}$$

where the summation is performed over all the jumps of X in $[0, T]$. From [24, Lemma 2.3], we get a more convenient expression for the expected trading costs, namely

$$\begin{aligned} E \left[\int_{[0,T]} \int_0^t I_s ds dX_t + \int_{[0,T]} \int_{[0,t)} G(t-s) dX_s dX_t \right. \\ \left. + \int_{[0,T]} M_t dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2 \right] \\ = E \left[\int_{[0,T]} \int_0^t I_s ds dX_t + \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|) dX_s dX_t \right] - P_0 x. \end{aligned}$$

We are interested in adding a risk aversion term to our cost functional. A natural candidate is $\int_0^T X_t^2 dt$, which is considered as a measure for the risk associated with holding a position X_t at time t ; see [4, 22, 37] and the discussion in [35, Sect. 1.2]. Hence our cost functional which is the sum of the expected trading costs and the risk aversion term has the form, dropping the fixed term $-P_0 x$,

$$E \left[\int_{[0,T]} \int_0^t I_s ds dX_t + \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|) dX_s dX_t + \phi \int_0^T X_t^2 dt \right], \quad (2.3)$$

where $\phi \geq 0$ is a constant.

The main goal of this work is to minimise the cost functional (2.3) over the class of admissible strategies. Before we discuss our main results in this framework, we introduce the following class of kernels. We say that a continuous and bounded G is *strictly positive definite* if for every admissible strategy X , we have

$$\int_{[0,T]} \int_{[0,T]} G(|t-s|) dX_s dX_t > 0 \quad P\text{-a.s.} \quad (2.4)$$

We define \mathbb{G} to be the class of continuous, bounded and strictly positive definite functions $G : (0, \infty) \rightarrow [0, \infty)$.

Remark 2.1 Note that (2.2) is satisfied for every $G \in \mathbb{G}$. A characterisation of positive definite kernels (that is, when the inequality (2.4) is not strict but weak) is given in [24, Proposition 2.6].

Remark 2.2 An important subclass of \mathbb{G} is the class of bounded, nonincreasing convex functions $G : (0, \infty) \rightarrow [0, \infty)$ (see [3, Proposition 2]).

2.2 Results for a Markovian signal

In this section, we introduce our results on the existence and uniqueness of an optimal strategy when the signal is a càdlàg Markov process. As in [24, Sect. 2], we restrict our discussion to deterministic strategies. The minimisation of the cost functional over signal-adaptive random strategies is discussed in Remark 2.9.

We consider the class of strategies

$$\begin{aligned} \Xi(x) = \{X &: \text{deterministic admissible strategy with } X_0 = x \\ &\text{and support in } [0, T]\}. \end{aligned}$$

Note that for any X in $\Xi(x)$, the cost functional (2.3) has the form

$$\int_{[0, T]} \int_0^t E[I_s] ds dX_t + \frac{1}{2} \int_{[0, T]} \int_{[0, T]} G(|t - s|) dX_s dX_t + \phi \int_0^T X_t^2 dt. \quad (2.5)$$

In our first main result, we prove that there exists at most one strategy which minimises the cost functional (2.5).

Theorem 2.3 *Assume that $G \in \mathbb{G}$. Then there exists at most one minimiser to the cost functional (2.5) in the class $\Xi(x)$ of admissible strategies.*

In our next result, we give a characterisation for the minimiser of the cost functional (2.5).

Theorem 2.4 *$X^* \in \Xi(x)$ minimises the cost functional (2.5) over $\Xi(x)$ if and only if there exists a constant λ such that X^* solves*

$$\int_0^t E[I_s] ds + \int_{[0, T]} G(|t - s|) dX_s^* - 2\phi \int_0^t X_s^* ds = \lambda \quad \text{for all } 0 \leq t \leq T. \quad (2.6)$$

A few remarks are in order.

Remark 2.5 In the special case where the agent does not rely on a signal (i.e., $I = 0$) and there is zero risk aversion ($\phi = 0$), Theorems 2.3 and 2.4 coincide with [24, Proposition 2.9 and Theorem 2.11].

Remark 2.6 Dang [20] studied the case where the risk aversion term in (2.3) is nonzero, but again $I = 0$. In [20, Sect. 4.2], a necessary condition for the existence of an optimal strategy is given when the admissible strategies are deterministic and absolutely continuous. Our condition in (2.6) coincides with Dang's result in that case. Note, however, that the question whether the condition in [20] is also sufficient and the uniqueness of the optimal strategy remained open even in the special case $I = 0$.

2.3 Result for an Ornstein–Uhlenbeck signal

As mentioned in the introduction, special attention is given to the case where the signal I is an Ornstein–Uhlenbeck process,

$$\begin{aligned} dI_t &= -\gamma I_t dt + \sigma dW_t, & t \geq 0, \\ I_0 &= \iota, \end{aligned} \tag{2.7}$$

where W is a standard Brownian motion and $\gamma > 0$, $\sigma > 0$ are constants. In the following corollary, we derive an explicit formula for the optimal strategy in the case of zero risk aversion and when G has an exponential decay. This generalises the result of Obizhaeva and Wang [33] who solved this control problem when there is no signal.

Corollary 2.7 *Let I be defined as in (2.7). Assume that $\phi = 0$ and $G(t) = \kappa\rho e^{-\rho t}$, where $\kappa > 0$, $\rho > 0$ are constants. Then there exists a unique minimiser $X^* \in \Xi(x)$ to the cost functional (2.5), which is given by*

$$X_t^* = (1 - b_0(t))x + \frac{\iota}{2\kappa\rho^2\gamma} \left(\frac{\rho^2 - \gamma^2}{\gamma} b_1(t) - (\rho + \gamma)b_2(t) - (\rho + \gamma)b_3(t) \right), \tag{2.8}$$

where

$$\begin{aligned} b_0(t) &= \frac{\mathbb{1}_{\{t>0\}} + \mathbb{1}_{\{t>T\}} + \rho t}{2 + \rho T}, \\ b_1(t) &= 1 - e^{-\gamma t} - b_0(t)(1 - e^{-\gamma T}), \\ b_2(t) &= \mathbb{1}_{\{t>T\}} + \rho t - b_0(t)(1 + \rho T), \\ b_3(t) &= (b_0(t) - \mathbb{1}_{\{t>T\}})e^{-\gamma T}. \end{aligned}$$

Note that $1 - b_0(T+) = 0$ and $b_i(T+) = 0$ for $i = 1, 2, 3$; moreover, the optimal strategy is linear in both x and ι .

In Fig. 1, we present some examples of the optimal strategy with the parameters $\gamma = 0.9$, $\kappa = 0.1$, $T = 10$, $x = 10$. These particular values are compatible with the empirical parameters which are estimated at the end of Sect. 4.2. Arbitrary initial values (-0.5 , 0 and $+0.5$) are taken for the signal ι . The special case where $\iota = 0$ gives results similar to Obizhaeva and Wang [33]. The parameter ρ , which controls the market impact decay, cannot be estimated from the data that we have; hence we take two arbitrary but realistic values (1.0 and 2.5). We observe that for large values

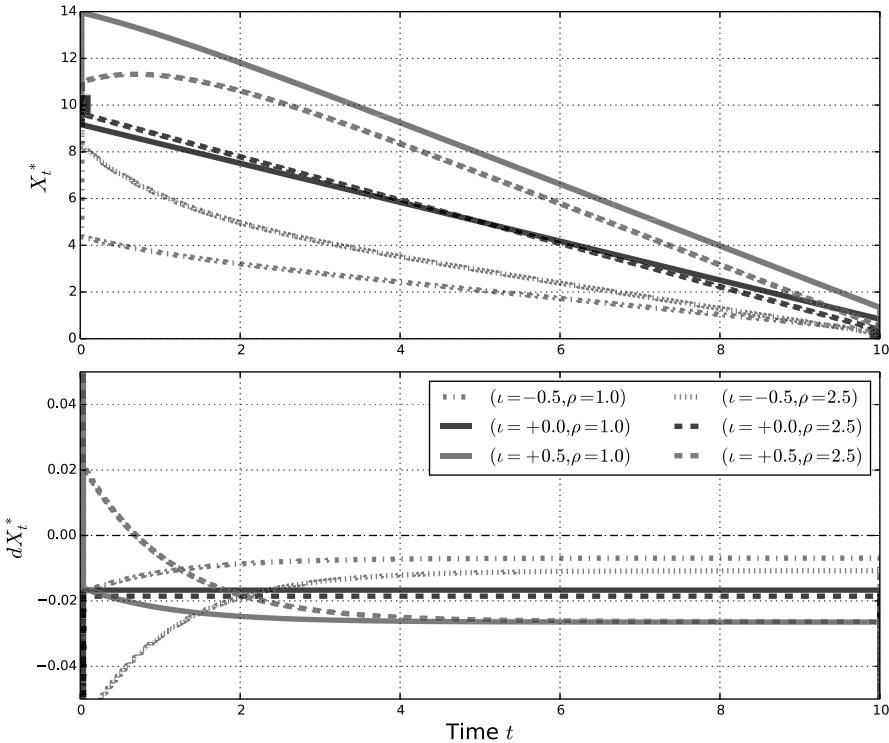


Fig. 1 Optimal trading strategies according to (2.8) for $\gamma = 0.9$, $\kappa = 0.1$, $T = 10$ and $x = 10$. We demonstrate different scenarios for selling 10 shares: without a signal, with a positive signal, and with a negative signal. We distinguish between a slow decay of the market impact (solid lines) and a fast decay (dashed lines). In the top graph, we show the remaining inventory; in the bottom graph, the trading speed (for $0 < t < 10$) is presented

of ρ , the initial jump in the optimal trading strategy is larger than the corresponding jump in the small ρ strategies, but the trading speed tends to have less variation. We particularly notice that when the initial signal is in the opposite direction to the trading ($\iota > 0$ for a sell order), the trading starts with purchases as expected, and afterwards the trading speed eventually becomes negative. On the other hand, when the initial signal is in the same direction as the trading, it is optimal to start selling immediately, and most of the inventory is sold before $T/2$.

In the following remarks, we discuss the result of Corollary 2.7.

Remark 2.8 Note that in the limit where $\rho \rightarrow \infty$, the market impact term in (2.5), $\frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|) dX_s dX_t$, formally corresponds to the costs arising from an instantaneous market impact, that is, $G(dt) = \kappa \delta_0(dt)$, where δ_0 is Dirac's delta measure. We briefly discuss the asymptotics of the optimal strategy $X^* = X^*(\rho)$ in (2.8) when $\rho \rightarrow \infty$. It is easy to verify that in the limit, the jumps of X^* vanish (see A and D in (5.11) for the explicit expression of the jumps), and the limiting optimal

strategy $X^*(\infty)$ is a smooth function which is given by

$$X_t^*(\infty) = X + \frac{\iota}{2\kappa\gamma^2} (1 - e^{-\gamma t}) - \frac{\iota}{2\kappa\gamma} t.$$

Motivated by these asymptotic results, we further explore in the next section absolutely continuous strategies which minimise the trading costs/risk aversion functional. We assume there that the market impact is instantaneous, that is, $G(dt) = \kappa\delta_0(dt)$, and drop the fuel constraint ($X_t = 0$ for $t > T$) from the admissible strategies. Then explicit formulas for the optimal strategy are derived when the risk aversion term is nonzero.

Remark 2.9 (An adaptive version of (2.8)) Equation (2.8) gives an optimal strategy for a trader with inventory $X_0 = x$ at $t = 0$, who is observing the initial value of the signal $\iota = I_0$ and wishes to minimise (2.5) for an exponentially decaying kernel and $\phi = 0$. The cost functional is therefore given by

$$\begin{aligned} U([0, T]) := & \int_{[0, T]} \int_0^\tau E[I_s | \mathcal{F}_0^W] ds dX_\tau \\ & + \frac{1}{2} \int_{[0, T]} \int_{[0, T]} \kappa\rho e^{-\rho|\tau-s|} dX_s dX_\tau, \end{aligned}$$

where $(\mathcal{F}_t^W)_{t \geq 0}$ is the natural filtration of W in (2.7). In this setting, once the trading has started, it is no longer possible to update the strategy by taking into account new information, i.e., new values of the signal. This can be compared to simpler frameworks like the one of Sect. 3, in which the optimal strategy is updated for any $0 \leq t \leq T$. We therefore add a short discussion on an adaptive framework for (2.8).

A natural way to update the optimal strategy at any time t is to define the process $(\tilde{X}_s)_{t \leq s \leq T}$ as the optimal strategy of the cost functional

$$\begin{aligned} \tilde{U}([t, T]) := & \int_{[t, T]} \int_t^\tau E[I_s | \mathcal{F}_t^W] ds dX_\tau \\ & + \frac{1}{2} \int_{[t, T]} \int_{[t, T]} \kappa\rho e^{-\rho|\tau-s|} dX_s dX_\tau. \end{aligned}$$

Note, however, that

$$U([0, T]) = U([0, t]) + \Delta_1 U(t, T) + \Delta_2 U(t, T),$$

where

$$\Delta_1 U(t, T) = \int_{(t, T)} \int_0^\tau E[I_s | \mathcal{F}_0^W] ds dX_\tau + \frac{1}{2} \int_{(t, T)} \int_{[0, t]} \kappa\rho e^{-\rho|\tau-s|} dX_s dX_\tau$$

and

$$\Delta_2 U(t, T) = \frac{1}{2} \int_{[0, T]} \int_{(t, T)} \kappa\rho e^{-\rho|\tau-s|} dX_s dX_\tau.$$

This implies that if \tilde{X} is used in place of X^* for some $\tau \in (t, T)$, the trader will have an (\mathcal{F}_t) -adapted control, but it will not necessarily be consistent with X^* which minimises $U([0, T])$. Therefore, in practice, one can choose between the following options:

- the optimal strategy X^* , limited to the information on the signal at $t = 0$;
- an approximate strategy \tilde{X} updated at each time $t \in (0, T)$, which takes into account the whole trajectory of (I_t) ;
- the optimal strategy which corresponds to a market impact without a decay (as shown in Sect. 3).

The question which of these strategies gives the best results remains open.

Note that in the cost functional U , the time-inconsistency is a result of the transient market impact term. In [36], time-inconsistent optimal liquidation problems were also studied. However, the inconsistency of the problems in [36] arises from the risk aversion term.

Remark 2.10 (Price manipulation) Market impact models admit transaction-triggered price manipulations if the expected costs of a sell (buy) strategy can be reduced by intermediate buy (sell) trades (see [3, Definition 1]). Theorem 2.20 in [24] implies that transaction-triggered price manipulations are impossible for the cost functional in (2.6), over the class of admissible strategies, in the case where $I \equiv 0$ and $\phi = 0$. However, Fig. 1 shows that adding signals to the same market impact model can create optimal strategies which are not monotonically decreasing, and therefore implies a possible price manipulation. It would be very interesting to investigate the conditions on the market impact kernel and the trading signals which ensure that there are no price manipulations. A study of the possible implications of these price manipulations for other market participants is also of major importance.

3 Optimal strategy for temporary market impact

In this section, we study an optimal trading problem that has some common features with the problem introduced in Sect. 2.1. We consider again a price process which incorporates a Markovian signal. The main change in this section is that the market impact in (2.1) is temporary, i.e., the kernel is given by $G(dt) = \kappa \delta_0(dt)$, where δ_0 is Dirac's delta measure and $\kappa > 0$ is a constant. Note that this type of kernel is not included in the class \mathbb{G} of kernels introduced in (2.1). The main goal of this section is to show how to incorporate trading signals in the CJ framework [13]. The results we obtain could be compared to the results of Sect. 2 (see Remark 2.8). Recall that we heuristically obtained the optimal strategy when the kernel $G = G_\rho$ ‘converges’ to Dirac's delta measure as $\rho \rightarrow \infty$.

We continue to assume that I is a càdlàg Markov process as in the beginning of Sect. 2, but we add the assumption that

$$E_\iota[|I_t|] \leq C(T)(1 + |\iota|) \quad \text{for all } \iota \in \mathbb{R}, 0 \leq t \leq T, \quad (3.1)$$

for some constant $C(T) > 0$. For the sake of simplicity, we assume that $M = \sigma^P W$ so that

$$dP_t = I_t dt + \sigma^P dW_t,$$

where $(W_t)_{t \geq 0}$ is a Brownian motion and $\sigma^P > 0$ is a constant.

In the following example, the fuel constraint on the admissible strategies is replaced with a terminal penalty function. This allows us to consider absolutely continuous strategies as in the framework of Cartea and Jaimungal (see e.g. [14, 15, 16]). We introduce some additional definitions and notations which are relevant to this setting.

Let \mathcal{V} denote the class of progressively measurable control processes $r = (r_t)_{t \geq 0}$ for which $\int_0^T |r_t| dt < \infty$ P -a.s. For any $x \geq 0$, we define

$$X_t^r = x - \int_0^t r_s ds. \quad (3.2)$$

Here X_t^r is the amount of inventory held by the trader at time t . We often suppress the dependence of X on r to ease the notation. The price process, which is affected by the linear instantaneous market impact, is given by

$$S_t = P_t - \kappa r_t, \quad t \geq 0,$$

where $\kappa > 0$. Note that S_t here corresponds to (2.1) when $G(dt) = \kappa \delta_0(dt)$. The investor's cash \mathcal{C}_t satisfies

$$d\mathcal{C}_t := S_t r_t dt = (P_t - \kappa r_t) r_t dt,$$

with $\mathcal{C}_0 = c$.

For the sake of consistency with earlier work of Cartea and Jaimungal in [14, 15, 16], we define the liquidation problem as a maximisation of the difference between the cash and the risk aversion. As mentioned earlier, the fuel constraint on the admissible strategies is replaced by a penalty function, which is given by $X_T(P_T - \varrho X_T)$, where $\varrho > 0$ is a constant. The cost functional is given by

$$V^r(t, \iota, c, x, p) = E_{t, \iota, c, x, p} \left[\mathcal{C}_T - \phi \int_t^T X_s^2 ds + X_T(P_T - \varrho X_T) \right], \quad (3.3)$$

where $\phi \geq 0$ is a constant and $E_{t, \iota, c, x, p}$ represents the expectation conditioned on $I_t = \iota$, $\mathcal{C}_t = c$, $X_t = x$, $P_t = p$. The value function is

$$V(t, \iota, c, x, p) = \sup_{r \in \mathcal{V}} V^r(t, \iota, c, x, p).$$

Note that this control problem could be easily transformed to a minimisation of the trading costs and risk aversion as in Sect. 2.

Let \mathcal{L}^I be the generator of the process I . Then the corresponding HJB equation is

$$\begin{aligned} 0 = \partial_t V + \iota \partial_p V + \frac{1}{2} (\sigma^P)^2 \partial_p^2 V + \mathcal{L}^I V - \phi x^2 \\ + \sup_r (r(p - \kappa r) \partial_c V - r \partial_x V), \end{aligned} \quad (3.4)$$

with the terminal condition

$$V(T, \iota, c, x, p) = c + x(p - \varrho x).$$

Let $E_{t,\iota}$ represent the expectation conditional on $I_t = \iota$. In the following result, we derive a solution to (3.4). The proof of Proposition 3.1 follows the same lines as the proof of [16, Proposition 1].

Proposition 3.1 *Assume that $\varrho \neq \sqrt{\kappa\phi}$. Then there exists a solution to (3.4), which is given by*

$$V(t, \iota, c, x, p) = c - xp + v_0(t, \iota) + xv_1(t, \iota) + x^2v_2(t), \quad (3.5)$$

where

$$\begin{aligned} v_2(t) &= \sqrt{\kappa\phi} \frac{1 + \zeta e^{2\beta(T-t)}}{1 - \zeta e^{2\beta(T-t)}}, \\ v_1(t, \iota) &= \int_t^T e^{\frac{1}{\kappa} \int_t^s v_2(u) du} E_{t,\iota}[I_s] ds, \\ v_0(t, \iota) &= \frac{1}{4\kappa} \int_t^T E_{t,\iota}[v_1^2(s, I_s)] ds, \end{aligned}$$

and the constants ζ and β are given by

$$\zeta = \frac{\varrho + \sqrt{\kappa\phi}}{\varrho - \sqrt{\kappa\phi}}, \quad \beta = \sqrt{\frac{\phi}{\kappa}}.$$

In the following result, we prove that the solution to (3.4) is indeed an optimal control to (3.3).

Proposition 3.2 *Assume that $\varrho \neq \sqrt{\kappa\phi}$. Then:*

(a) (3.5) *maximises the cost functional in (3.3). The optimal trading speed r^* is given by*

$$r_t^* = -\frac{1}{2\kappa} \left(2v_2(t)X_t + \int_t^T e^{\frac{1}{\kappa} \int_t^s v_2(u) du} E[I_s | I_t] ds \right), \quad 0 \leq t \leq T.$$

(b) *Assume further that I follows an Ornstein–Uhlenbeck process as in (2.7). Then the optimal trading speed r^* is given by*

$$r_t^* = -\frac{1}{2\kappa} \left(2v_2(t)X_t + I_t \int_t^T e^{-\gamma(s-t) + \frac{1}{\kappa} \int_t^s v_2(u) du} ds \right), \quad 0 \leq t \leq T.$$

The proofs of Propositions 3.1 and 3.2 are given in Sect. 5.2.

In the following remarks, we compare the results of Sects. 2 and 3.

Remark 3.3 If we set the risk aversion and penalty coefficients ϕ, ϱ in (3.3) to 0, then from the proof of Proposition 3.1, it follows that $v_2 \equiv 0$. Under the same assumptions on the signal as in Proposition 3.2 (b), the optimal strategy is given by

$$r_t^* = -\frac{I_t}{2\kappa\gamma}(1 - e^{-\gamma(T-t)}), \quad 0 \leq t \leq T, \quad (3.6)$$

which is consistent with $X_t^*(\infty)$ from Remark 2.8.

Remark 3.4 One can heuristically impose a ‘fuel constraint’ on the optimal strategy in Proposition 3.2 (b) by using the asymptotics of r_t^* when $\varrho \rightarrow \infty$. In this case, $\zeta \rightarrow 1$ and the limiting optimal speed which we denote by r_t^f is

$$r_t^f = -\frac{1}{2\kappa} \left(2\bar{v}_2(t)X_t + I_t \int_t^T e^{-\gamma(s-t)+\frac{1}{\kappa} \int_t^s \bar{v}_2(u) du} ds \right), \quad 0 \leq t \leq T,$$

where

$$\bar{v}_2(t) = \sqrt{\kappa\phi} \frac{1 + e^{2\beta(T-t)}}{1 - e^{2\beta(T-t)}}.$$

Remark 3.5 It is important to notice that (2.8) gives the optimal strategy on the time horizon $[0, T]$ in the GSS framework by using only information on the OU signal at $t = 0$. On the other hand, (3.6), which is the optimal trading speed r_t in the CJ framework, is using the information on the signal at time t . A crucial point here is that if one tries to solve repeatedly the control problem in the GSS framework on time intervals $[t, T]$ for any $t > 0$, by using I_t and S_t as an input, the optimal strategy will not necessarily minimise the cost functional (2.3) on $[0, T]$. The reason is that the control problem in (2.3) may be inconsistent. The market impact (and therefore the transaction costs) created on $[0, t]$ affects the cost functional on $[t, T]$ (see Remark 2.9 for more details). Note that this phenomenon does not occur in the instantaneous market impact case (i.e., in the CJ framework).

In Fig. 2, we simulate the optimal inventory X^* which corresponds to the optimal trading speed r^* from Proposition 3.2 (b). In the black solid line, we present the optimal inventory in the case where there is no signal. In this case, the optimal strategy is deterministic. The red region in Fig. 2 is a ‘heat map’ of 1000 realisations of the optimal inventory X^* . The parameters of the signal (2.7) are $\gamma = 0.1$, $\sigma = 0.1$ and $I_0 = 0$. We also set $T = 10$, $\kappa = 0.5$, $\phi = 0.1$, $X_0 = 10$ and $\varrho = 10$ in the cost functional (3.3). We observe that the random strategies are a perturbation of the classical deterministic optimal strategy. In Fig. 3, we present the value function (3.5) at $t = 0$ under the same assumptions as in Fig. 2, that is, assuming that I is an OU process and that the model parameters are similar. More precisely, we plot $V(0, \iota, c, x, p) - (c - xp)$; hence we omit constants which do not contribute to the behaviour of the model. We observe that the revenue which corresponds to the optimal sell strategy r^* is affected by the direction and value ι of the signal. The revenue of a sell strategy when the signal is positive, which indicates a potential price increase, is higher than with negative signal scenarios.

Fig. 2 Simulation of the optimal inventory X^* which corresponds to the trading speed r^* from Proposition 3.2 (b). In the black curve, we present the optimal inventory in the absence of a signal. The red region is a plot of 1000 trajectories of the optimal inventory X^* . The parameters of the model are $\gamma = 0.1$, $\sigma = 0.1$, $I_0 = 0$, $T = 10$, $\kappa = 0.5$, $\phi = 0.1$, $X_0 = 10$ and $\varrho = 10$

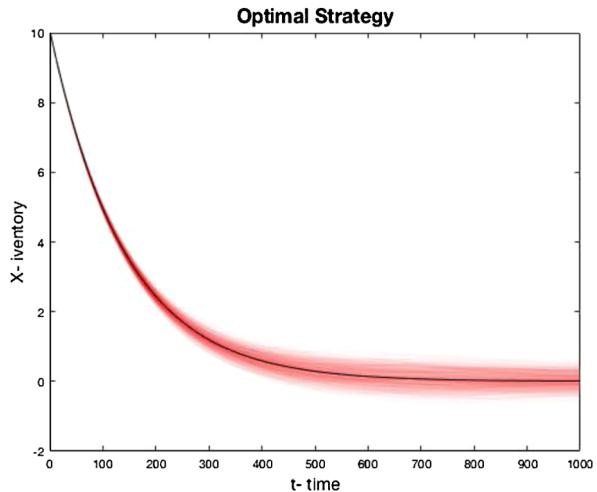
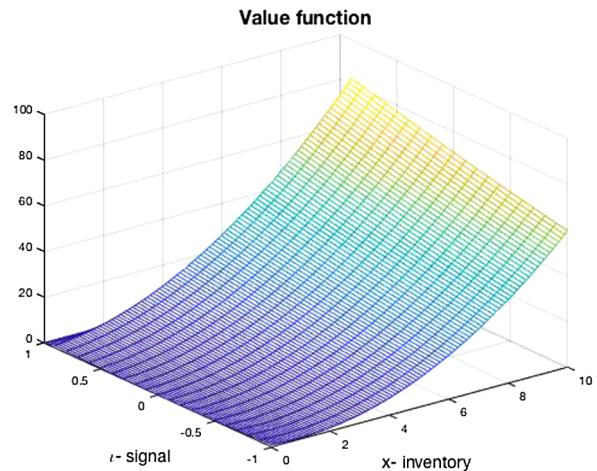


Fig. 3 Plot of the value function $V(0, t, c, x, p) - (c - xp)$ from (3.5) when the signal is an OU process. The parameters of the model are $\gamma = 0.1$, $\sigma = 0.1$, $T = 10$, $\kappa = 0.5$, $\phi = 0.1$, $X_0 = 10$ and $\varrho = 10$



4 Evidence for the use of signals in trading

In this section, we analyse financial data which is related to the limit order book imbalance. The data analysis in this section is directed to support the models which were introduced in Sects. 2 and 3. In Sect. 4.1, we describe our data base and provide empirical evidence for the use of the imbalance signal, which is a liquidity-driven signal. In Sect. 4.2, we study the statistical properties of the signal and motivate our model from Sect. 2.3 of an Ornstein–Uhlenbeck signal. Finally, in Sect. 4.3, we study the use of this signal during liquidation by different market participants. Note that in Sects. 2 and 3, we also discussed more general signals which are not necessarily liquidity-driven.

Before we start with the detailed analysis of the limit order book imbalance signal, we survey some related work on other processes which are known to affect asset prices and have mean-reverting properties. Each of these processes may serve as a

signal in the optimal trading framework of Sect. 2.3. We mention these specific examples as they demonstrate predictive signals which are affective at different time scales.

The order flow imbalance has been extensively studied in the literature (see e.g. [18] and references therein). The correlation between the current order flow and the future price move in 10 seconds intervals was studied by Cont et al. in [18]. The mean-reverting properties of the order flow were studied by Bechler and Ludkovski in [9] (see also [10]).

Pairs-trading refers to the case where two assets Q and P are in the same industry or have similar characteristics. In this case, one expects the returns of these two assets to track each other (see e.g. [7, Sect. 1]). Let S^Q and S^P be the price processes of the assets. Then the difference between the weighted returns of P and Q , $dJ_t := dS_t^P/S_t^P - \beta dS_t^Q/S_t^Q$, for a certain constant $\beta > 0$, can be approximated in many cases by a stationary mean-reverting process. Hence a trader who wants to liquidate a large amount of asset P , for example, may consider J_t as a trading signal. The typical mean-reversion time of such signals may vary between half a day to a month (see [7, Fig. 8]). More examples of trading signals which are used in optimal execution can be found in a presentation by Robert Almgren [5]. In Sect. 4.3, we show that the LOB imbalance signal affects the trading speed of high-frequency proprietary traders in the following 10 minutes time interval.

4.1 The database: NASDAQ OMX trades

The database which is used in this section is made of transactions on the NASDAQ OMX exchange. This exchange used to publish the identity of the buyer and seller of each transaction until 2014. To obtain order book data, we use recordings made by Capital Fund Management (CFM) on the same exchange, which were matched with NASDAQ OMX trades thanks to the timestamp, quantity and price of each trade. On a typical month, the accuracy of such matchings is more than 97% (see Table 2).

The NASDAQ OMX trades were already used for academic studies (see [38] and [31] for details). We study 13 stocks traded on NASDAQ OMX Stockholm from January 2013 to September 2013. The purpose of this section is not to conduct an extensive econometric study on this database; such work deserves a paper of its own. Our goal here is to show qualitative evidence for the existence of the order book imbalance signal and to study how market participants' decisions depend on its value. The 13 stocks which are used in this section have been selected for this research since high-frequency proprietary traders took part in at least 100'000 trades on each of them during the studied period. More details on the classification of the traders into different classes are given later in this section.

Table 1 shows descriptive statistics on the considered stocks in the database. Stocks are ranked by the average daily traded value (in units of 10^6 of the local currency, the Swedish krona), which can be considered as an indicator of liquidity. We also included in Table 1 the average price during the study period, since European exchanges apply dynamic tick size schedules: the lower the average stock price, the lower is the tick size (see [30, Chap. 1, Sect. 3]). The *minimum tick size* is the smallest tick size which was applied to the stock price during our study period. If the

Table 1 Statistics of the 13 studied stocks. Values and prices are in Swedish krona. The Garman and Klass (GK) volatility is estimated yearly. The table is sorted by the average daily traded value over 180 trading days

Company name (code)	Daily traded value (10^6)	Average price	Average bid–ask spread	Volatility (GK)	Minimum tick
Volvo AB (VOLVb.ST)	431.20	94.87	0.057	15.08%	0.05
Nordea Bank AB (NDA.ST)	384.48	76.09	0.053	15.02%	0.05
Telefonaktiebolaget LM					
Ericsson (ERICb.ST)	373.20	78.41	0.054	15.20%	0.05
Hennes & Mauritz AB (HMB.ST)	361.66	232.89	0.112	11.37%	0.10
Atlas Copco AB (ATCOa.ST)	329.94	175.19	0.110	16.13%	0.10
Swedbank AB (SWEDA.ST)	313.18	151.97	0.108	15.29%	0.10
Sandvik AB (SAND.ST)	296.09	90.88	0.067	17.01%	0.05
SKF AB (SKFB.ST)	255.99	161.11	0.112	16.47%	0.10
Skandinaviska Enskilda					
Banken AB (SEBa.ST)	221.23	66.85	0.053	15.56%	0.05
Nokia OYJ (NOKLST)	209.77	28.84	0.019	36.89%	0.01
Telia Co AB (TLSN.ST)	207.09	45.14	0.014	10.13%	0.01
ABB Ltd (ABB.ST)	179.51	144.35	0.108	11.89%	0.10
AstraZeneca PLC (AZN.ST)	168.06	318.57	0.127	12.09%	0.10

price changes are large enough, different tick sizes could have been applied during the study period; therefore we also added the yearly estimated Garman and Klass (GK) volatility to the table (see [23]). Last but not least, the *average bid–ask spread* has to be compared with the tick size: for all these stocks, the bid–ask spread lies between one and two ticks. All these stocks are therefore liquid and large-tick stocks.

The NASDAQ OMX database contains the identity of the buyer and the seller *from the viewpoint of the exchange*, that is, the members of the exchange who made the transactions. Asset managers, for example, are not direct members of the exchange. On the other hand, brokers, banks and some other specific market participants are members. We classify the market members into four types (for more details, see Appendix A.1, Tables 5–7):

- global investment banks (GIB);
- institutional brokers (IB);
- high-frequency market makers (HFMM);
- high-frequency proprietary traders (HFPT).

Table 2 gives some plain statistics about the number of trades on each stock of our database involving these types of participants. Keep in mind that the database covers 180 trading days. It can be read in the last line that on average, global investment banks are involved in 58% of the trades while high-frequency traders are involved in 32% of them; the remaining 10% involve institutional brokers. The percentage of identified participants is on average 78%, that is, 22% of the trades took place between two participants which we could not associate with any of our four classes (GIB, IB, HFMM, HFPT). Moreover, we had to filter around 2% of the trades (see

Table 2 Statistics on labelled trades involving each kind of market participant. *Trades count* is the sum of trades involving at least one labelled participant. *Pct. ident.* represents the percentage of trades involving at least one participant out of the four types that we focus on. *Pct. LOB matched* is the percentage of trades for which we found a matching quote in our LOB database. The averages in the bottom line are calculated over all identified trades

Code	Global banks	HFMM	Instit. brokers	HFPT	Trades count	Pct. ident.	pct. LOB matched
VOLVb.ST	56.9%	17.1%	10.6%	15.3%	927,467	76.7%	97.4%
NDA.ST	60.7%	10.6%	9.9%	18.7%	694,509	76.8%	97.4%
ERICb.ST	57.8%	17.6%	7.7%	16.9%	811,931	81.0%	97.2%
HMb.ST	58.5%	16.0%	8.9%	16.6%	716,644	76.8%	97.8%
ATCOa.ST	58.2%	13.7%	10.5%	17.6%	677,981	79.1%	98.0%
SWEDa.ST	61.2%	12.2%	9.5%	17.2%	600,655	74.6%	97.7%
SAND.ST	61.0%	15.2%	10.4%	13.4%	701,961	77.4%	96.9%
SKFb.ST	60.9%	13.8%	10.4%	14.9%	587,088	77.1%	97.0%
SEBa.ST	61.5%	12.1%	8.8%	17.7%	515,743	75.8%	97.8%
NOKLST	54.5%	8.1%	8.9%	28.5%	710,173	79.6%	99.2%
TLSN.ST	61.2%	10.0%	10.6%	18.2%	548,602	68.9%	97.8%
ABB.ST	50.1%	15.6%	5.2%	29.2%	359,067	86.2%	98.1%
AZN.ST	51.4%	12.8%	9.0%	26.8%	411,118	89.6%	98.8%
Average	58.3%	13.6%	9.4%	18.7%	–	77.7%	–

last column) because in some cases we could not match limit order book records with the observed transactions.

We expect institutional brokers to execute orders for clients without taking additional risks (i.e., act as ‘pure agency brokers’). Such brokers often have medium-size clients and local asset managers. They do not spend a lot of resources such as technology or quantitative analysts to study the microstructure, and they do not react fast to microscopic events.

Global investment banks can take risks at least on a fraction of their order flow. Most of them already had proprietary trading desks and high-frequency trading activities in 2013 (i.e., during the recording of the data). They usually have large international clients and have the capability to react to changes in the state of the order book.

High-frequency market makers are providing liquidity on both sides of the order book. They have a very good knowledge on market microstructure. As market makers, we expect them to focus on adverse selection and not to keep large inventories. On the other hand, high-frequency proprietary traders take their own risks in order to earn money, while taking profit of their knowledge of the order book dynamics.

The data in Table 3 is compatible with our prior knowledge on the different classes of traders:

- HFMM trade far more with limit orders (73%), than with market orders;
- IB use more market orders than limit orders;
- on average, HFPT and GIB have balanced order flows.

Table 3 Descriptive statistics of market participants on an ‘average stock’. All the trades are normalised as if all orders were buy orders. The imbalance is positive when its sign is in the direction of the trade

Participant class	Trade type	Average imbalance	Average number	Pct.
Global banks	limit	-0.41	103,418	48.2%
	market	0.56	111,082	51.8%
HFMM	limit	-0.31	30,747	73.0%
	market	0.62	11,818	27.0%
HFPT	limit	-0.37	28,763	47.2%
	market	0.63	31,858	52.8%
Instit. brokers	limit	-0.56	9,984	33.6%
	market	0.33	19,505	66.4%

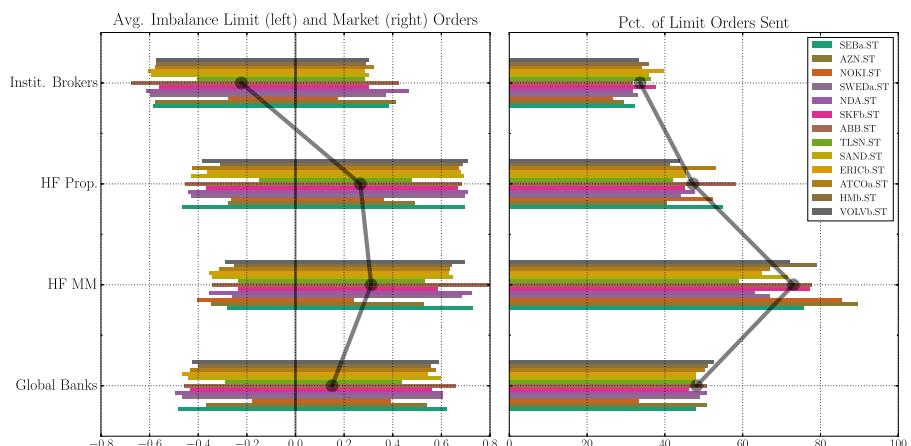


Fig. 4 Use of limit and market orders vs. the state of the imbalance before a trade, for each type of market participant. (Left panel) Average imbalance just before a limit order (left part, negative), and average imbalance just before a market order (right part, positive). The dark line with the large dots represents the average over all trades for all stocks. (Right panel) Percentage of trades with limit orders out of all orders. The dark line is the average over all stocks

Moreover, high-frequency participants (HFMM and HFPT) both use market orders to consume liquidity on the *weak side of the book* (i.e., buying when the imbalance is on average 0.60 and selling when it is on average -0.60), and provide liquidity when the imbalance is less intense than -0.5. The later observation is compatible with HF participants who contribute to stabilise the price with their limit orders.

These numbers are only averages; in Fig. 4, we give their dispersion across our 13 stocks. It can be seen in Fig. 4 that the asymmetry between HFPT and IB is observed for all stocks (see left panel). Moreover, the left panel suggests that high-frequency participants use market orders and limit orders when the imbalance is in their favour.

4.2 The imbalance signal

The order book imbalance has been identified as one of the main drivers of liquidity dynamics. It plays an important role in order book models, and more specifically it drives the rate of insertions and cancellations of limit orders near the mid price (see [1, Chap. 10] and [28]). As an illustration of the theoretical results of this paper, we document here the *imbalance signal* and its use by different types of participants. This signal is computed by using the quantity Q_B of the best bid and Q_A of the best ask of the order book,

$$\text{Imb}(\tau) = \frac{Q_B(\tau) - Q_A(\tau)}{Q_B(\tau) + Q_A(\tau)},$$

just before the occurrence of a transaction at time $\tau+$. Note that our 13 stocks are considered as “*large-tick stocks*”—except for Sandvik AB (SAND.ST) and Telia Co AB (TLSN.ST)—for which the average bid–ask spread is greater than 1.4 times the tick size. This means that the liquidity at the best bid and ask gives a substan-

Fig. 5 Predictive power of the imbalance: the average price move for the next 10 trades (y-axis) as a function of the current imbalance (x-axis)

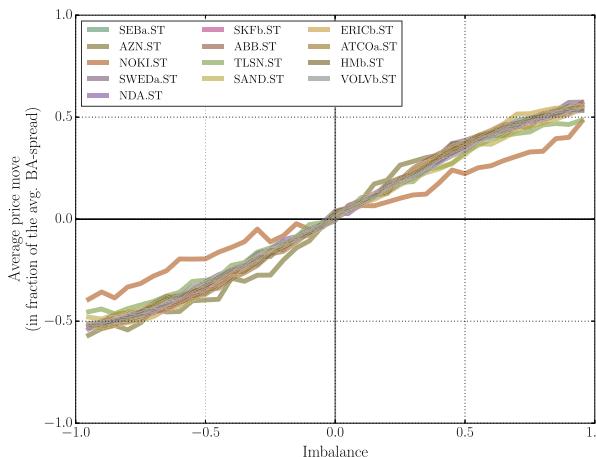


Fig. 6 Predictive power of the imbalance for the AstraZeneca stock: the average price move for the next 10 trades (y-axis) as a function of the current imbalance (x-axis), with confidence levels of upper and lower 5%

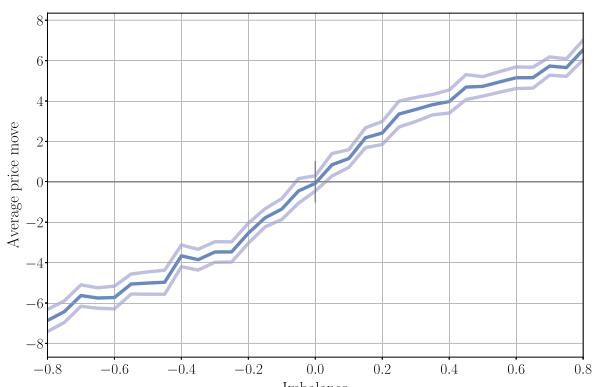


Table 4 Results of linear regressions involving the imbalance. The first column is the result of a regression of the price move after 10 trades given the imbalance immediately before the first of these trades. This can also be shown in the slope of Fig. 5. The p -value is very close to zero for all stocks, meaning they are highly significant. The R^2 varies between 1% (Nokia) to 16% (Volvo AB and Nordea Bank AB). Other columns are the results of the regression of future imbalance (respectively after 3, 5, 7, 10 and 100 trades) with respect to the imbalance immediately before the first of these trades, given that the imbalance is between -0.5 and 0.5 . This regression corresponds to the slopes at the center of Fig. 7. All p -values are significant at more than 99.99%

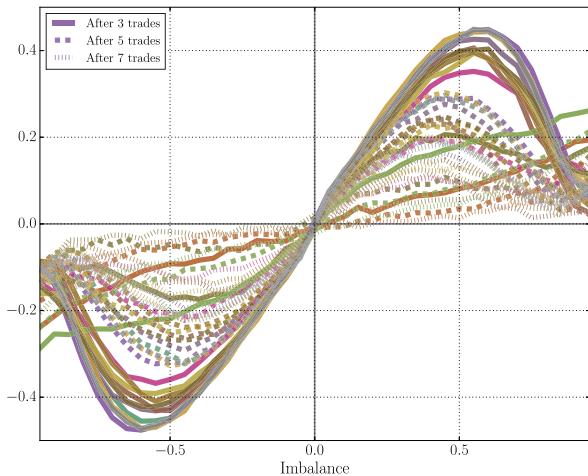
	d price	R^2	Imb. 3 t	Imb. 5 t	Imb. 7 t	Imb. 10 t	Imb. 100 t
VOLVb.ST	0.58	0.16	0.91	0.72	0.49	0.26	0.03
NDA.ST	0.58	0.16	0.90	0.71	0.51	0.30	0.04
ERICb.ST	0.62	0.15	0.93	0.74	0.53	0.30	0.03
HMb.ST	0.59	0.08	0.84	0.62	0.41	0.21	0.02
ATCOa.ST	0.60	0.13	0.85	0.58	0.34	0.13	0.02
SWEDA.ST	0.62	0.14	0.87	0.67	0.45	0.23	0.02
SAND.ST	0.56	0.15	0.81	0.57	0.37	0.20	0.03
SKFb.ST	0.59	0.13	0.76	0.49	0.28	0.13	0.01
SEBa.ST	0.61	0.15	0.91	0.73	0.51	0.28	0.03
NOKI.ST	0.41	0.01	0.18	0.08	0.05	0.03	0.00
TLSN.ST	0.54	0.04	0.43	0.22	0.13	0.08	0.02
ABB.ST	0.59	0.11	0.86	0.61	0.33	0.11	0.03
AZN.ST	0.64	0.04	0.47	0.20	0.09	0.05	0.02

tial information on the price pressure (see [27] for details about the role of the tick size in liquidity formation). For smaller-tick stocks, several price levels need to be aggregated in order to obtain the same level of prediction for future price moves.

In order to demonstrate the predictive power of the imbalance, we consider the average mid-price move after 10 trades as a function of the current imbalance (see Figs. 5 and 6). Table 4 gives data which is associated to these curves. The column ‘*d price*’ shows the price change renormalised by the average bid–ask spread on each stock after 10 trades. This price move is on average close to 0.6 times the imbalance just before the first of these trades.

Mean-reversion of the imbalance Figure 7 shows the average value of the imbalance after $\Delta T = 3, 5$ and 7 trades as a function of its current value. The colours of the curves represent the same stocks as in Fig. 5. The decreasing slopes around $\text{Imb}(t) = 0$ are underlined by the columns 4–7 of Table 4. This demonstrates the mean-reverting property of the imbalance. We do not comment too much on the decreasing slopes for large imbalance values. We just mention that a strong imbalance may imply a future price change, which in turn can create a depletion of the ‘weak side of the order book’ (in the sense of [21]). This phenomenon may cause an inversion of the imbalance, since the queue in the second-best price level of the order book, which is now ‘promoted’ to be the first level, could be large. See [28] for details about queues dynamics in order books.

Fig. 7 Mean-reversion of the imbalance: the average value of the imbalance after 3 (solid lines), 5 (dashed lines) and 7 (dotted lines) trades (y -axis) as a function of the current imbalance (x -axis). The colours of the lines represent the same stocks as in Fig. 5



To approximately fit Ornstein–Uhlenbeck (OU) dynamics to the imbalance data, we use ‘trade time’ instead of ‘calendar time’ (i.e., seconds) in order to compensate on different frequencies of trading for each of our 13 stocks (see columns dt in Table 9 in the Appendix). This yields a discrete version of an OU process,

$$I_{n+\Delta n} - I_n = -\gamma I_n \Delta n + \sigma \sqrt{\Delta n} \xi_{n+\Delta n},$$

where Δn is the number of future trades, γ the mean-reversion parameter and σ the standard deviation of the innovation $\xi_{n+\Delta n}$. The linear regressions on the last columns of Table 4 are following the model

$$I_{n+\Delta n} = a_{\Delta n} I_n + \tilde{\sigma}_{\Delta n} \epsilon_{n+\Delta n}.$$

This leads to the estimators of γ and σ as

$$\hat{\gamma} \simeq \frac{1 - a_{\Delta n}}{\Delta n}, \quad \hat{\sigma} \simeq \frac{\tilde{\sigma}_{\Delta n}}{\sqrt{\Delta n}}.$$

Figure 8 shows the frequencies of values of $\hat{\gamma}$ for the 13 stocks over the scales $\Delta n = 3, 5, 7, 100$. Table 8 gives the associated values of these $\hat{\gamma}$. In Table 9, different estimates of $\hat{\sigma}$ are given.

Some numerical values of the model parameters At a time scale of 35 seconds or 7 trades, γ should be taken close to 0.92 and σ close to 0.22. We also provide an estimator for the instantaneous market impact κ using the empirical average of the mid-price move¹ after a trade times the sign of the trade. Table 9 in the Appendix shows that the average value of κ divided by the average bid-ask spread is close to 0.1.

¹The mid price is the middle of the best bid and best ask prices.

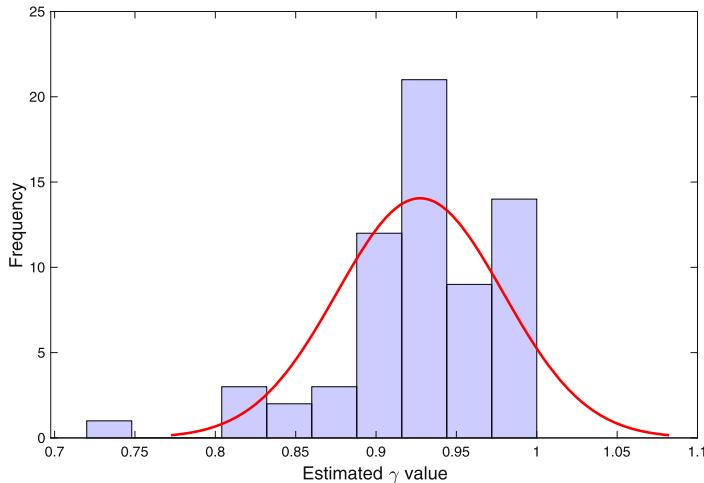


Fig. 8 Histogram for the estimators of γ

We summarise the main findings of this section:

- (i) the imbalance can be considered as a liquidity-driven short-term signal;
- (ii) this signal has mean-reverting properties;
- (ii) market participants, especially high-frequency traders, take the imbalance into account while trading (see Table 3 and Fig. 4).

4.3 Use of signals by market participants

As previously mentioned, we expect HF proprietary traders, HF market makers and global investment banks to pay more attention to order book dynamics than institutional brokers. However, as market makers, HFMM are expected to earn money by buying and selling when the mid price does not change much (relying on the *bid–ask bounce*). On the other hand, HFPT are typically alternating between intensive buy and sell phases which are based on price moves.

Our expectations are met in Table 3, where the average imbalance just before a trade is shown for each type of market participant. All the trades in this table are normalised as if all orders were buy orders. The imbalance is positive when its sign is in the direction of the trade, and negative if it is in an opposite direction.

We notice the following behaviour:

- When the transaction is obtained via a market order, the market participant had the opportunity to observe the imbalance before consuming liquidity.
- When the transaction is obtained via a limit order, fast participants have the opportunity to cancel their orders to prevent an execution and potential adverse selection.

Table 3 underlines that HF participants and GIB make ‘better choices’ on trading according to the market imbalance. Institutional brokers seems to be the less ‘imbalance aware’ when they decide to trade. This could be explained either by the fact that they invest less in microstructure research, quantitative modelling and automated

trading, or because they have less freedom to be opportunistic. Since they act as pure agency brokers, they do not have the choice to retain clients' orders, and this could prevent them from waiting for the best imbalance to trade.

Strategic behaviour Once we suspect that some participants take into account the imbalance in their trading decisions, we can look for a relation between the trading rate and the corresponding imbalance for each type of participant. This is motivated by the optimal trading frameworks of the previous sections, where we used the trading rate as a control.

In order to learn more about the relation between the imbalance signal and the trading speed, we compute the *imbalance-conditioned trading rates* R_+ and R_- for each type of market participant, during all consecutive intervals of 10 minutes from January 2013 to September 2013 (within the trading hours, i.e., from 9:00 to 17:30). Note that in the following analysis, the signal, time and trading quantities are discrete.

Definition 4.1 The *imbalance-conditioned trading rates* of market participants of type \mathcal{P} during the time interval \mathcal{T} are given by

$$R_{\pm}(\mathcal{T}, \mathcal{P} | \iota) = \frac{1}{N(\mathcal{T}, \mathcal{P}, \iota)} \sum_{t \in \mathcal{T}} \bar{\delta}_{\varepsilon(t)\text{sign}(\text{Imb}(t))}(\pm\iota) A_t \bar{\delta}_{\mathcal{P}}(t) \bar{\delta}_{|\text{Imb}(t)|}(\iota),$$

where

- $\varepsilon(t)$ is the sign of the trade at time t ;
- $\bar{\delta}_{\varepsilon(t)\text{sign}(\text{Imb}(t))}(\pm 1)$ is 1 if at time t the imbalance sign times the sign of the trade is equal to ± 1 , and 0 otherwise;
- A_t is the traded amount of the trade at time t ;
- $\bar{\delta}_{\mathcal{P}}(t)$ is 1 if the trade at time t involved a participant of type \mathcal{P} , and 0 otherwise;
- $\bar{\delta}_{|\text{Imb}(t)|}(\iota)$ is 1 if the absolute value of the imbalance at time t equals ι , and 0 otherwise;
- $N(\mathcal{T}, \mathcal{P}, \iota)$ is the number of trades involving participant \mathcal{P} in \mathcal{T} when the imbalance equals ι .

Qualitatively, R_{\pm} have the following interpretation:

- $R_+(\mathcal{T}, \mathcal{P} | \iota)$ is an estimate of the amount traded *in the direction of the imbalance* when the absolute value of the imbalance is ι , by participants of type \mathcal{P} during the time interval \mathcal{T} ;
- $R_-(\mathcal{T}, \mathcal{P} | \iota)$ is an estimate of the amount traded *in the opposite direction of the imbalance* when the absolute value of the imbalance is ι , by participants of type \mathcal{P} during the time interval \mathcal{T} .

In order to get the imbalance-conditioned trading rates, we renormalise R_{\pm} by

$$A(\mathcal{T} | \iota) = \sum_{\mathcal{P}} (R_+(\mathcal{T}, \mathcal{P} | \iota) + R_-(\mathcal{T}, \mathcal{P} | \iota)).$$

Note that $A(\mathcal{T} | \iota)$ is the traded amount during the interval \mathcal{T} given that the imbalance is ι . Then $R_+(\mathcal{T}, \mathcal{P} | \iota)$ divided by $A(\mathcal{T} | \iota)$ is an estimate of *the probability that a*

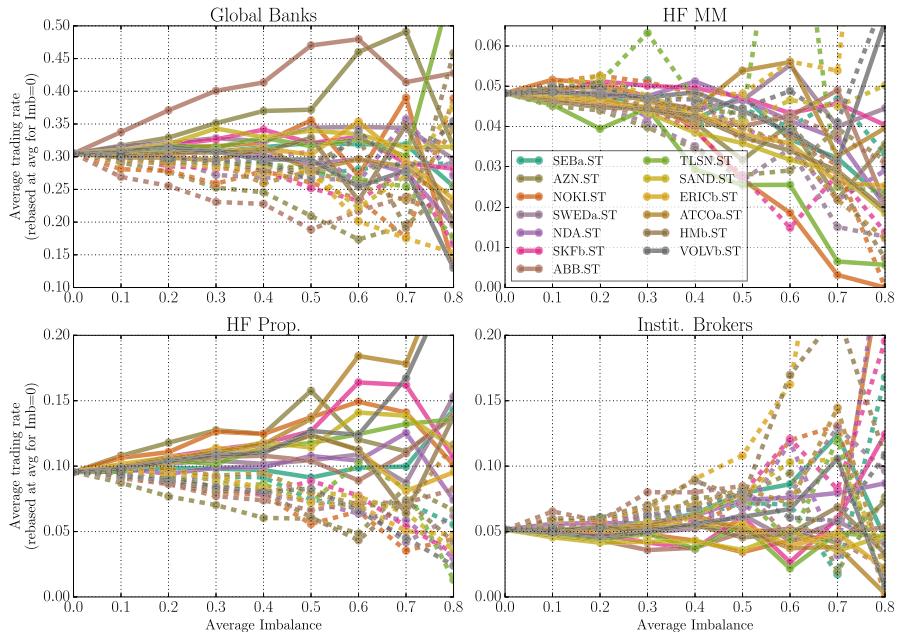


Fig. 9 Renormalised average trading rate in the direction of the imbalance \hat{r}_+ (solid line) and in the opposite direction \hat{r}_- (dotted line), during 10 consecutive minutes, for each type of participant

stock is traded by a participant of type \mathcal{P} during interval \mathcal{T} in the direction of the imbalance, given that the imbalance is ι . Then $R_-(\mathcal{T}, \mathcal{P} | \iota)$ divided by $A(\mathcal{T} | \iota)$ is an estimate of the probability that a stock is traded by a participant of type \mathcal{P} during interval \mathcal{T} in the opposite direction of the imbalance, given that the imbalance is ι .

Let $N_{\mathcal{T}}$ be the number of ten minutes intervals in our data base. We define

$$r_+(\mathcal{P} | \iota) = \frac{1}{N_{\mathcal{T}}} \sum_{\mathcal{T}} \frac{R_+(\mathcal{T}, \mathcal{P} | \iota)}{A(\mathcal{T} | \iota)}, \quad r_-(\mathcal{P} | \iota) = \frac{1}{N_{\mathcal{T}}} \sum_{\mathcal{T}} \frac{R_-(\mathcal{T}, \mathcal{P} | \iota)}{A(\mathcal{T} | \iota)},$$

which are unbiased estimators for the probability that a participant of type \mathcal{P} trades in the direction (respectively, opposite direction) of the imbalance, given that the absolute value of the imbalance is ι . To be able to put all the stocks on the same graph, we draw

$$\hat{r}_{\pm}^k(\mathcal{P} | \iota) = r_{\pm}^k(\mathcal{P} | \iota) / \bar{r}_{\pm}(\mathcal{P} | 0)$$

in Fig. 9. Here $\bar{r}_{\pm}(\mathcal{P} | 0)$ is the average of $r_{\pm}^k(\mathcal{P} | 0)$ over all stocks k . Figure 9 shows the variations of \hat{r}_+ (the relative speed of trading in the direction of the imbalance, in solid lines) and \hat{r}_- (the relative speed of trading in the opposite direction to the imbalance, in dashed lines) with respect to the imbalance ι before the trade, for each type of market participant and for each stock. From this graph, we observe the following:

- For high-frequency market makers, the higher the imbalance in the order book, the less they trade. This effect does not seem to be related to the direction of their trades. It corresponds to an expected behaviour from market makers.
- For high-frequency proprietary traders, the higher the imbalance, the more they trade in a similar direction, and the less they trade in the opposite direction.
- Institutional brokers do not seem to be influenced by the imbalance. Additional data analysis shows that they trade more with limit orders when the imbalance is intense; this may drive the price to move in the opposite direction.
- The behaviour of global banks seems to be influenced by the imbalance for part of the stocks in our sample.

Towards a theory of strategic use of signals The analysis in this section suggests that some market participants are using liquidity-driven signals in their trading strategies. The liquidity imbalance, computed from the best bid and ask prices of the order book for medium- and large-tick stocks, appears to be a good candidate. Moreover, its dynamics exhibit mean-reverting properties.

The theory developed in Sects. 2 and 3 can be regarded as a tentative framework to model the behaviour of the following participants. Global investment banks who execute large orders seem to be a typical example for participants who adopt the type of strategies that we model. High-frequency proprietary traders who are combining slow signals (which may be considered as execution of large orders) along with fast signals could also use our framework. We could moreover hope that thanks to the availability of such frameworks, institutional brokers could optimise their trading and increase the profits for more final investors.

5 Proofs

5.1 Proofs of Theorems 2.3, 2.4 and Corollary 2.7

The proofs of Theorems 2.3 and 2.4 use ideas from the proofs of [24, Proposition 2.9 and Theorem 2.11].

Proof of Theorem 2.3 Let $x \geq 0$. For any $X \in \Xi(x)$, define

$$C(X) := C_1(X) + C_2(X) + K(X), \quad (5.1)$$

where

$$\begin{aligned} C_1(X) &= \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|) dX_s dX_t, \\ C_2(X) &= \phi \int_0^T X_s^2 ds, \\ K(X) &= \int_{[0,T]} \int_0^t E_t[I_s] ds dX_t. \end{aligned}$$

Note that $C(X)$ is the cost functional in (2.5).

Since G is strictly positive definite, we have for any $X \in \Xi(x)$ that

$$C_1(X) > 0. \quad (5.2)$$

Also $C_2(\cdot)$ is quadratic in X , and therefore we have

$$C_2(X) \geq 0. \quad (5.3)$$

Let $X, Y \in \Xi(x)$. We define the cross-functionals

$$\begin{aligned} C_1(X, Y) &= \frac{1}{2} \int_{[0, T]} \int_{[0, T]} G(|t - s|) dX_s dY_t, \\ C_2(X, Y) &= \phi \int_0^T X_s Y_s ds. \end{aligned}$$

Note that

$$C_i(X, Y) = C_i(Y, X) \quad \text{for } i = 1, 2$$

and

$$C_i(X - Y) = C_i(X) + C_i(Y) - 2C_i(X, Y) \quad \text{for } i = 1, 2. \quad (5.4)$$

From (5.2), it follows that $C_1(X - Y) > 0$, and together with (5.4), we get

$$C_1\left(\frac{1}{2}X + \frac{1}{2}Y\right) = \frac{1}{4}C_1(X) + \frac{1}{4}C_1(Y) + \frac{1}{2}C_1(X, Y) < \frac{1}{2}C_1(X) + \frac{1}{2}C_1(Y).$$

Repeating the same steps, using (5.3) instead of (5.2), we get

$$C_2\left(\frac{1}{2}X + \frac{1}{2}Y\right) \leq \frac{1}{2}C_2(X) + \frac{1}{2}C_2(Y).$$

Since $K(X)$ is linear in X , we have

$$K\left(\frac{1}{2}X + \frac{1}{2}Y\right) = \frac{1}{2}K(X) + \frac{1}{2}K(Y).$$

From (5.1), it follows that

$$C\left(\frac{1}{2}X + \frac{1}{2}Y\right) < \frac{1}{2}C(X) + \frac{1}{2}C(Y).$$

Let $\alpha \in (0, 1)$. The claim that

$$C(\alpha X + (1 - \alpha)Y) < \alpha C(X) + (1 - \alpha)C(Y)$$

follows from the continuity of $C(\alpha X)$ in α by a standard extension argument. As $C(\cdot)$ is strictly convex, we get that there exists at most one minimiser to $C(X)$ in $\Xi(x)$. \square

Proof of Theorem 2.4 First we prove that the condition (2.6) is necessary for optimality. Let $0 \leq t < t_0 \leq T$ and consider the round trip

$$dY_s = \delta_{t_0}(ds) - \delta_t(ds).$$

For all $\alpha \in \mathbb{R}$, we have

$$C_i(X^* + \alpha Y) = C_i(X^*) + \alpha^2 C_i(Y) + 2\alpha C_i(X^*, Y), \quad i = 1, 2, \quad (5.5)$$

and

$$K(X^* + \alpha Y) = K(X^*) + \alpha K(Y). \quad (5.6)$$

Let $Z := X^* + \alpha Y$, and recall that $C(Z) = C_1(Z) + C_2(Z) + K(Z)$. Using (5.5) and (5.6), we can differentiate $C(Z)$ with respect to α and get

$$\frac{\partial C(Z)}{\partial \alpha} = K(Y) + \sum_{i=1,2} (2\alpha C_i(Y) + 2C_i(X^*, Y)).$$

From optimality, we have $C(X^*) \leq C(Z)$, and therefore we expect that

$$\left. \frac{\partial C(Z)}{\partial \alpha} \right|_{\alpha=0} = K(Y) + 2 \sum_{i=1,2} C_i(X^*, Y) = 0. \quad (5.7)$$

Note that

$$\begin{aligned} C_1(X^*, Y) &= \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|r-s|) dX_s^* dY_r \\ &= \frac{1}{2} \int_{[0,T]} G(|t_0-s|) dX_s^* - \frac{1}{2} \int_{[0,T]} G(|t-s|) dX_s^*, \\ C_2(X^*, Y) &= \phi \int_0^T X_s^* Y_s ds = -\phi \int_t^{t_0} X_s^* ds, \\ K(Y) &= \int_{[0,T]} \int_0^r E_\ell[I_s] ds dY_r = \int_t^{t_0} E_\ell[I_s] ds. \end{aligned}$$

We get that (5.7) is equivalent to

$$\begin{aligned} &\int_{[0,T]} G(|t_0-s|) dX_s^* - 2\phi \int_0^{t_0} X_s^* ds + \int_0^{t_0} E_\ell[I_s] ds \\ &= \int_{[0,T]} G(|t-s|) dX_s^* - 2\phi \int_0^t X_s^* ds + \int_0^t E_\ell[I_s] ds. \end{aligned}$$

Since t and t_0 were chosen arbitrarily, this implies (2.6).

Assume now that there exists $X^* \in \Xi(x)$ satisfying (2.6). We show that X^* minimises $C(\cdot)$. Let X be any other strategy in $\Xi(x)$. Define $Z = X - X^*$. Then

from (2.6), we have

$$\begin{aligned}
 C_1(X^*, Z) &= \frac{1}{2} \int_{[0,T]} \int_{[0,T]} G(|t-s|) dX_s^* dZ_t \\
 &= \frac{1}{2} \int_{[0,T]} \left(\lambda + 2\phi \int_0^t X_s^* ds - \int_0^t E_t[I_s] ds \right) dZ_t \\
 &= \frac{\lambda}{2} \left(X([0, \infty)) - X^*([0, \infty)) \right) + \phi \int_{[0,T]} \int_0^t X_s^* ds dZ_t \\
 &\quad - \frac{1}{2} \int_{[0,T]} \int_0^t E_t[I_s] ds dZ_t \\
 &= \phi \int_{[0,T]} \int_0^t X_s^* ds dZ_t - \frac{1}{2} K(Z),
 \end{aligned} \tag{5.8}$$

where we have used the fact that $X([0, \infty)) = X^*([0, \infty)) = x$ in the last equality. From (5.5) and (5.8), we have

$$\begin{aligned}
 C_1(X) &= C_1(Z + X^*) \\
 &= C_1(Z) + C_1(X^*) + 2C_1(X^*, Z) \\
 &= C_1(Z) + C_1(X^*) - K(Z) + 2\phi \int_{[0,T]} \int_0^t X_s^* ds dZ_t, \\
 C_2(X) &= C_2(Z + X^*) \\
 &= C_2(Z) + C_2(X^*) + 2C_2(X^*, Z) \\
 &= C_2(Z) + C_2(X^*) + 2\phi \int_0^T X_s^* Z_s ds.
 \end{aligned}$$

From the linearity of $K(\cdot)$, we get

$$K(X) = K(Z) + K(X^*).$$

It follows that

$$\begin{aligned}
 C(X) &= \sum_{i=1,2} C_i(X) + K(X) \\
 &= C_1(X^*) + C_2(X^*) + K(X^*) + C_1(Z) + C_2(Z) \\
 &\quad + 2\phi \int_{[0,T]} \int_0^t X_s^* ds dZ_t + 2\phi \int_0^T X_s^* Z_s ds \\
 &= C(X^*) + C_1(Z) + C_2(Z) \\
 &\quad + 2\phi \int_{[0,T]} \int_0^t X_s^* ds dZ_t + 2\phi \int_0^T X_s^* Z_s ds.
 \end{aligned}$$

Recall that $Z_0 = 0$ and $Z_t = 0$ for every $t > T$; hence integration by parts gives

$$0 = \int_{[0,T]} \int_0^t X_s^* ds dZ_t + \int_0^T X_t^* Z_t dt,$$

and since $C_i(Z) \geq 0$ for $i = 1, 2$, we get

$$C(X) \geq C(X^*). \quad \square$$

Proof of Corollary 2.7 From (2.7), it follows that $E_t[I_t] = \iota e^{-\gamma t}$. Since $\phi = 0$, (2.6) reduces to

$$\frac{\iota}{\gamma} (1 - e^{-\gamma t}) + \kappa \rho \int_{[0,T]} e^{-\rho|t-s|} dX_s^* = \lambda. \quad (5.9)$$

Moreover, we have the fuel constraint

$$\int_{[0,T]} dX_t^* = -x. \quad (5.10)$$

Motivated by the example in Obizhaeva and Wang [33], we guess a solution of the form

$$dX_t^* = A\delta_0(dt) + (Be^{-\gamma t} + C)dt + D\delta_T(dt),$$

where δ_x is Dirac's delta measure at x and A, B, C, D are some constants.

Note that

$$\begin{aligned} \kappa \rho \int_0^t e^{-\gamma s} e^{-\rho(t-s)} ds &= \frac{\kappa \rho}{\rho - \gamma} (e^{-\gamma t} - e^{-\rho t}), \\ \kappa \rho \int_t^T e^{-\gamma s} e^{-\rho(s-t)} ds &= \frac{\kappa \rho}{\rho + \gamma} (e^{-\gamma t} - e^{-\gamma T - \rho(T-t)}), \end{aligned}$$

and therefore

$$\begin{aligned} \kappa \rho \int_{[0,T]} e^{-\rho|t-s|} dX_s^* &= \kappa \rho e^{-\rho t} A + B \frac{\kappa \rho}{\rho - \gamma} (e^{-\gamma t} - e^{-\rho t}) \\ &\quad + B \frac{\kappa \rho}{\rho + \gamma} (e^{-\gamma t} - e^{-\gamma T - \rho(T-t)}) \\ &\quad + C \kappa (1 - e^{-\rho t}) + C \kappa (1 - e^{-\rho(T-t)}) + D \kappa \rho e^{-\rho(T-t)}. \end{aligned}$$

From (5.9), it follows that

$$\lambda = 2\kappa C + \frac{\iota}{\gamma},$$

and together with (5.10), we get the linear system

$$\begin{aligned}
-\frac{\iota}{\gamma}e^{-\gamma t} + B \frac{\kappa\rho}{\rho - \gamma} e^{-\gamma t} + B \frac{\kappa\rho}{\rho + \gamma} e^{-\gamma t} &= 0, \\
A\kappa\rho e^{-\rho t} - B \frac{\kappa\rho}{\rho - \gamma} e^{-\rho t} - C\kappa e^{-\rho t} &= 0, \\
-B \frac{\kappa\rho}{\rho + \gamma} e^{-\gamma T - \rho(T-t)} - C\kappa e^{-\rho(T-t)} + D\kappa\rho e^{-\rho(T-t)} &= 0, \\
A + \frac{B}{\gamma}(1 - e^{-\gamma T}) + CT + D &= -x.
\end{aligned}$$

From the first equation, we get

$$B = \iota \frac{\rho^2 - \gamma^2}{2\kappa\rho^2\gamma},$$

and then

$$\begin{aligned}
A &= \frac{1}{2 + T\rho} \\
&\times \left(\frac{\iota}{2\kappa\rho^2\gamma} \left((\rho + \gamma)(1 + T\rho - \frac{\rho - \gamma}{\gamma}(1 - e^{-\gamma T})) - (\rho - \gamma)e^{-\gamma T} \right) - x \right), \\
C &= \rho A - \iota \frac{\rho + \gamma}{2\kappa\rho\gamma}, \\
D &= A - \frac{\iota}{2\kappa\rho^2\gamma} ((\rho + \gamma) - (\rho - \gamma)e^{-\gamma T}). \tag{5.11}
\end{aligned}$$

The optimal strategy is therefore

$$X_t^* = x + \mathbb{1}_{\{t>0\}} A + Ct + \frac{B}{\gamma}(1 - e^{-\gamma t}) + \mathbb{1}_{\{t>T\}} D,$$

which is equivalent to (2.8). \square

5.2 Proofs of Propositions 3.1 and 3.2

Proof of Proposition 3.1 The proof follows the same lines as the proof of [16, Proposition 1]. Plugging in the ansatz $V(t, \iota, c, x, p) := c + xp + v(t, x, \iota)$, we get

$$0 = \partial_t v + \mathcal{L}^I v + \iota x - \phi x^2 + \sup_r (-r^2\kappa - r\partial_x v).$$

Optimising over r , it follows that

$$r^* = -\frac{\partial_x v}{2\kappa},$$

and we get the PDE

$$\partial_t v + \mathcal{L}^I v + \frac{1}{4\kappa}(\partial_x v)^2 + \iota x - \phi x^2 = 0, \tag{5.12}$$

where $v(T, x, \iota) = -\varrho x^2$. As in [16, Eq. (A.2)], we have linear and quadratic x -terms in (5.12) along with a quadratic terminal condition. Hence we make for the solution the ansatz

$$v(t, x, \iota) = v_0(t, \iota) + x v_1(t, \iota) + x^2 v_2(t, \iota).$$

By comparing terms with similar powers of x , we get the system of PDEs

$$\partial_t v_0 + \mathcal{L}^\iota v_0 + \frac{1}{4\kappa} v_1^2 = 0, \quad (5.13)$$

$$\partial_t v_1 + \mathcal{L}^\iota v_1 + \frac{1}{\kappa} v_2 v_1 + \iota = 0, \quad (5.14)$$

$$\partial_t v_2 + \mathcal{L}^\iota v_2 + \frac{1}{\kappa} v_2^2 - \phi = 0, \quad (5.15)$$

with the terminal conditions

$$v_0(T, \iota) = 0, \quad v_1(T, \iota) = 0, \quad v_2(T, \iota) = -\varrho.$$

We first find a solution to (5.15). Note that since the terminal condition is independent of ι , we might be able to find a ι -independent solution, that is, $v_2(t) := v_2(t, \iota)$, which satisfies

$$\partial_t v_2 + \frac{1}{\kappa} v_2^2 - \phi = 0.$$

This is a Riccati equation which has the solution (see the proof of [16, Proposition 1])

$$v_2(t) = \sqrt{\kappa\phi} \frac{1 + \zeta e^{2\beta(T-t)}}{1 - \zeta e^{2\beta(T-t)}},$$

where

$$\zeta = \frac{\varrho + \sqrt{\kappa\phi}}{\varrho - \sqrt{\kappa\phi}}, \quad \beta = \sqrt{\frac{\phi}{\kappa}}.$$

Let $E_{t,\iota}$ represent the expectation conditional on $I_t = \iota$. Using v_2 , we can find a Feynman–Kac representation for the solution of (5.14) as

$$v_1(t, \iota) = E_{t,\iota} \left[\int_t^T e^{\frac{1}{\kappa} \int_t^s v_2(u) du} I_s ds \right] = \int_t^T e^{\frac{1}{\kappa} \int_t^s v_2(u) du} E_{t,\iota}[I_s] ds.$$

Again by the Feynman–Kac formula, we derive a solution to (5.13) as

$$v_0(t, \iota) = E_{t,\iota} \left[\frac{1}{4\kappa} \int_t^T v_1^2(s, I_s) ds \right] = \frac{1}{4\kappa} \int_t^T E_{t,\iota}[v_1^2(s, I_s)] ds. \quad \square$$

Proof of Proposition 3.2 (a) Note that V is a classical solution to (3.4). By standard arguments (see e.g. [34, Theorem 3.5.2]), in order to prove that V in (3.5) is the value function of (3.3), it is enough to show that r^* is admissible and that

$$|V(t, \iota, c, x, p)| \leq C(1 + \iota^2 + c^2 + x^2 + p^2) \quad \text{for all } t \geq 0, \iota, c, x, p \in \mathbb{R}. \quad (5.16)$$

Clearly, $\sup_{t \in [0, T]} |v_2(t)| < \infty$. From our conditions on I , we have

$$E_t[|I_t|] \leq C(1 + |\iota|) \quad \text{for all } \iota \in \mathbb{R}, 0 \leq t \leq T,$$

and therefore

$$\begin{aligned} |xv_1(t, \iota)| &\leq C|x|(1 + |\iota|) \leq C(1 + \iota^2 + x^2) \quad \text{for all } t \geq 0, \iota, x \in \mathbb{R}, \\ |v_0(t, \iota)| &\leq C(1 + \iota^2) \quad \text{for all } t \geq 0, \iota \in \mathbb{R}. \end{aligned}$$

This implies (5.16). To prove that r^* is admissible, it is enough to show that we have $\int_0^T |r_t^*| dt < \infty$. Since v_2 is bounded, we notice that

$$\begin{aligned} |r_t^*| &\leq \frac{1}{2\kappa} \left(2|v_2(t)||X_t| + \int_t^T e^{\frac{1}{\kappa} \int_t^s |v_2(u)| du} E_{t,\iota}[|I_s|] ds \right) \\ &\leq C_1 |X_t| + C_2 T(1 + |\iota|) \\ &\leq (C_2 + C_1)(x + T(1 + |\iota|)) + C_1 \int_0^t |r_s| ds, \end{aligned}$$

where we used (3.2) in the last inequality. From the Gronwall inequality, we have

$$|r_t^*| \leq (C_2 + C_1)(x + T(1 + |\iota|))e^{C_1 T};$$

hence r^* is admissible.

(b) Note that an Ornstein–Uhlenbeck process satisfies (3.1). Hence the proof follows immediately from (a). \square

Acknowledgements We are very grateful to anonymous referees and to the editors for their careful reading of the manuscript, and for a number of useful comments and suggestions that significantly improved this paper. We also thank Mikko Pakkanen whose useful comments greatly improved the manuscript.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Tables and complementary statistics

A.1 Composition of market participants groups

Table 5 Composition of the group of HFT and the classification of ‘high-frequency market makers’ and ‘high-frequency proprietary traders’

High-frequency traders			
Name	NASDAQ-OMX member code(s)	Market maker	Prop. trader
All Options International B.V.	AOI		
Hardcastle Trading AG	HCT		
IMC Trading B.V.	IMC, IMA	Yes	
KCG Europe Limited	KEM, GEL	Yes	
MMX Trading B.V.	MMX		
Nyenburgh Holding B.V.	NYE		
Optiver VOF	OPV	Yes	
Spire Europe Limited	SRE, SREA, SREB	Yes	
SSW-Trading GmbH	IAT		
WEBB Traders B.V	WEB		
Wolverine Trading UK Ltd	WLV		

Table 6 Composition of the group of global investment banks

Global investment banks	
Name	NASDAQ-OMX member code(s)
Barclays Capital Securities Limited Plc	BRC
Citigroup Global Markets Limited	SAB
Commerzbank AG	CBK
Deutsche Bank AG	DBL
HSBC Bank Plc	HBC
Merrill Lynch International	MLI
Nomura International Plc	NIP

Table 7 Composition of the group of institutional brokers

Institutional brokers	
Name	NASDAQ-OMX member code(s)
ABG Sundal Collier ASA	ABC
Citadel Securities (Europe) Limited	CDG
Erik Penser Bankaktiebolag	EPB
Jefferies International Limited	JEF
Neonet Securities AB	NEO
Remium Nordic AB	REM
Timber Hill Europe AG	TMB

A.2 Complementary statistics

Table 8 Estimates of the speed of mean-reversion γ , using different time scales

$\hat{\gamma}$ estimates	3 trades	5 trades	7 trades	10 trades	100 trades
VOLVb	0.97	0.94	0.93	0.93	0.99
NDA	0.97	0.94	0.93	0.93	0.99
ERICb	0.98	0.95	0.93	0.93	0.99
HMb	0.95	0.92	0.92	0.92	0.99
ATCOa	0.95	0.92	0.91	0.91	0.99
SWEDA	0.96	0.93	0.92	0.92	0.99
SAND	0.94	0.91	0.91	0.92	0.99
SKFb	0.92	0.90	0.90	0.91	0.99
SEBa	0.97	0.95	0.93	0.93	0.99
NOKI	0.73	0.82	0.86	0.90	0.99
TLSN	0.81	0.84	0.88	0.91	0.99
ABB	0.95	0.92	0.90	0.91	0.99
AZN	0.82	0.84	0.87	0.90	0.99

Table 9 Estimate of κ divided by the average bid-ask spread, average time between two trades, and level of noise in the estimated dynamics of the imbalance signal. Each column of $\hat{\sigma}$ is estimated for a different time scale (i.e., number of trades). The decay in the estimates of σ shows that innovations of the imbalance are sub-diffusive

$\hat{\kappa}$ over spread	\hat{dt} (s)	$\hat{\sigma}$ 3 trades	$\hat{\sigma}$ 5 trades	$\hat{\sigma}$ 7 trades	$\hat{\sigma}$ 10 trades	$\hat{\sigma}$ 100 trades
VOLVb	0.088	5.30	0.25	0.23	0.22	0.19
NDA	0.098	7.20	0.26	0.24	0.22	0.20
ERICb	0.092	6.60	0.25	0.23	0.22	0.19
HMb	0.095	6.68	0.27	0.25	0.22	0.19
ATCOa	0.109	7.77	0.27	0.25	0.23	0.20
SWEDA	0.105	7.73	0.27	0.25	0.23	0.20
SAND	0.101	7.24	0.28	0.25	0.22	0.19
SKFb	0.108	8.53	0.28	0.25	0.23	0.20
SEBa	0.099	9.13	0.26	0.24	0.22	0.19
NOKI	0.172	10.24	0.33	0.26	0.22	0.19
TLSN	0.134	7.74	0.31	0.26	0.22	0.19
ABB	0.113	15.13	0.28	0.26	0.24	0.20
AZN	0.163	15.51	0.32	0.26	0.23	0.19

References

1. Abergel, F., Anane, M., Chakraborti, A., Jedidi, A., Toke, I.M.: Limit Order Books (Physics of Society: Econophysics and Sociophysics). Cambridge University Press, Cambridge (2016)
2. Alfonsi, A., Fruth, A., Schied, A.: Optimal execution strategies in limit order books with general shape functions. Quant. Finance **10**, 143–157 (2010)

3. Alfonsi, A., Schied, A., Slynko, A.: Order book resilience, price manipulation, and the positive portfolio problem. *SIAM J. Financ. Math.* **3**, 511–533 (2012)
4. Almgren, R.: Optimal trading with stochastic liquidity and volatility. *SIAM J. Financ. Math.* **3**, 163–181 (2012)
5. Almgren, R.: In: Real Time Trading Signals. Presentation from $K \times 25$, the International kdb + User Conference, NYC (2018). <https://kx.com/media/2018/05/Almgren-Kx25-May2018.pdf>
6. Almgren, R., Chriss, N.: Optimal execution of portfolio transactions. *J. Risk* **3**(2), 5–39 (2000)
7. Avellaneda, M., Lee, J.H.: Statistical arbitrage in the US equities market. *Quant. Finance* **10**, 761–782 (2010)
8. Bacry, E., Luga, A., Lasnier, M., Lehalle, C.A.: Market impacts and the life cycle of investors orders. *Mark. Microstruct. Liq.* **1**(2), 1550009 (2015)
9. Bechler, K., Ludkovski, M.: Optimal execution with dynamic order flow imbalance. *SIAM J. Financ. Math.* **6**, 1123–1151 (2015)
10. Bechler, K., Ludkovski, M.: Order flows and limit order book resiliency on the meso-scale. *Mark. Microstruct. Liq.* **3**(4), 1850006 (2017)
11. Bertsimas, D., Lo, A.W.: Optimal control of execution costs. *J. Financ. Mark.* **1**, 1–50 (1998)
12. Bouchard, B., Dang, N.M., Lehalle, C.A.: Optimal control of trading algorithms: a general impulse control approach. *SIAM J. Financ. Math.* **2**, 404–438 (2011)
13. Cartea, Á., Jaimungal, S.: Modelling asset prices for algorithmic and high-frequency trading. *Appl. Math. Finance* **20**, 512–547 (2013)
14. Cartea, Á., Jaimungal, S.: Optimal execution with limit and market orders. *Quant. Finance* **15**, 1279–1291 (2015)
15. Cartea, Á., Jaimungal, S.: Risk metrics and fine tuning of high-frequency trading strategies. *Math. Finance* **25**, 576–611 (2015)
16. Cartea, Á., Jaimungal, S.: Incorporating order-flow into optimal execution. *Math. Financ. Econ.* **10**, 339–364 (2016)
17. Cartea, Á., Jaimungal, S., Penalva, J.: Algorithmic and High-Frequency Trading (Mathematics, Finance and Risk). Cambridge University Press, Cambridge (2015)
18. Cont, R., Kukanov, A., Stoikov, S.: The price impact of order book events. *J. Financ. Econom.* **12**, 47–88 (2014)
19. Curato, G., Gatheral, J., Lillo, F.: Optimal execution with non-linear transient market impact. *Quant. Finance* **17**, 41–54 (2017)
20. Dang, N.M.: Optimal execution with transient impact. Working paper (2014). Available online at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2183685
21. Fodra, P., Pham, H.: Semi Markov model for market microstructure. *Appl. Math. Finance* **22**, 261–295 (2015)
22. Forsyth, P., Kennedy, J., Tse, T.S., Windcliff, H.: Optimal trade execution: a mean-quadratic-variation approach. *J. Econ. Dyn. Control* **36**, 1971–1991 (2012)
23. Garman, M.B., Klass, M.J.: On the estimation of security price volatility from historical data. *J. Bus.* **53**, 67–78 (1980)
24. Gatheral, J., Schied, A., Slynko, A.: Transient linear price impact and Fredholm integral equations. *Math. Finance* **22**, 445–474 (2012)
25. Guéant, O.: The Financial Mathematics of Market Liquidity: From Optimal Execution to Market Making. Chapman & Hall/CRC Press, London/Boca Raton (2016)
26. Guéant, O., Lehalle, C.A., Fernandez-Tapia, J.: Optimal portfolio liquidation with limit orders. *SIAM J. Financ. Math.* **13**, 740–764 (2012)
27. Huang, W., Lehalle, C.A., Rosenbaum, M.: How to predict the consequences of a tick value change? Evidence from the Tokyo Stock Exchange pilot program. *Mark. Microstruct. Liq.* **02**(03n04), 1750001 (2016)
28. Huang, W., Lehalle, C.A., Rosenbaum, M.: Simulating and analyzing order book data: The queue-reactive model. *J. Am. Stat. Assoc.* **110**, 107–122 (2015)
29. Kharroubi, I., Pham, H.: Optimal portfolio liquidation with execution cost and risk. *SIAM J. Financ. Math.* **1**, 897–931 (2010)
30. Lehalle, C.A., Laruelle, S., Burgot, R., Pelin, S., Lasnier, M.: Market Microstructure in Practice. World Scientific, Singapore (2013)
31. Lehalle, C.A., Mounjid, O.: Limit order strategic placement with adverse selection risk and the role of latency. *Mark. Microstruct. Liq.* **3**(1), 1750009 (2017)
32. Neuman, E., Schied, A.: Optimal portfolio liquidation in target zone models and catalytic superprocesses. *Finance Stoch.* **20**, 495–509 (2016)

33. Obizhaeva, A.A., Wang, J.: Optimal trading strategy and supply/demand dynamics. *J. Financ. Mark.* **16**, 1–32 (2013)
34. Pham, H.: Continuous-Time Stochastic Control and Optimization with Financial Applications. Springer, Berlin (2009)
35. Schied, A.: A control problem with fuel constraint and Dawson–Watanabe superprocesses. *Ann. Appl. Probab.* **23**, 2472–2499 (2013)
36. Schöneborn, T.: Optimal trade execution for time-inconsistent mean–variance criteria and risk functions. *SIAM J. Financ. Math.* **6**, 1044–1067 (2015)
37. Tse, S.T., Forsyth, P.A., Kennedy, J.S., Windcliff, H.: Comparison between the mean-variance optimal and the mean-quadratic-variation optimal trading strategies. *Appl. Math. Finance* **20**, 415–449 (2013)
38. van Kervel, V., Menkveld, A.: High-frequency trading around large institutional orders. *J. Finance* (2018, in press) <https://doi.org/10.2139/ssrn.2619686>