ANZIAM J. 61(2019), 161–194
doi:10.1017/S1446181119000038

PRICE IMPACT OF LARGE ORDERS USING HAWKES PROCESSES

L. R. AMARAL✉[1] and A. PAPANICOLAOU[1]

(Received 19 February, 2018; accepted 30 January, 2019; first published online 6 May 2019)

Abstract

We introduce a model for the execution of large market orders in limit order books, and use a linear combination of self-exciting Hawkes processes to model asset-price dynamics, with the addition of a *price-impact function* that is concave in the order size. A criterion for a general price-impact function is introduced, which is used to show how specification of a concave impact function affects order execution. Using our model, we examine the immediate and permanent impacts of large orders, analyse the potential for price manipulation, and show the effectiveness of the time-weighted average price strategy. Our model shows that price depends on the balance between the intensities of the Hawkes process, which can be interpreted as a dependence on order-flow imbalance.

2010 *Mathematics subject classification*: primary 91B26; secondary 60G55, 91B70.

Keywords and phrases: price-impact function, limit order books, execution of large orders, Hawkes processes.

1. Introduction

The phrase "price impact" refers to the changes in an order book which are caused by an incoming order to buy or sell. When a market *buy* order is sent to an exchange, a matching engine finds limit *sell* orders at the best prices available and assigns them to fill the order. After the order has been executed, those limit *sell* orders are removed from the order book, and depending on the size of the market order, there may be a slight overall rise in the best bid and the best ask prices. The same sequence of events happens for market *sell* orders. Most trading activities have no price impact, but large market orders will have an impact, with the general principle being that *buy* orders push the price up and *sell* orders push the price down.

Impactful market orders may arise when a broker has a client request to buy or sell a very large amount of stock. In this case, a trader must decide how to best divide this request into market *buy*/*sell* orders for execution. These market *buy*/*sell* orders will

[1]Department of Finance and Risk Engineering, NYU Tandon School of Engineering, 6 MetroTech Center, Brooklyn, NY 11201, USA; e-mail: lra286@nyu.edu, ap1345@nyu.edu.

also be large relative to other market orders and the amount of liquidity available at the best bid and best ask prices. The trader understands beforehand that the first orders to be executed will push the order book prices in the wrong direction for his/her later orders, and with this in mind he/she will look for an optimal strategy for how and when to execute such orders.

To gain an understanding of what we mean by "large orders", consider a situation where a client needs 1 million shares to be executed by the end of the day. Below are some options for the trader:

- send one market order of 1 million shares;
- send one market order of 500 000 shares, wait for 3 minute, then send the remaining shares;
- send 1000 market orders of 1000 shares waiting 1 minute between the orders.

The first two options do not seem to be the best strategy, because the order book probably does not have the depth (that is, at a given time there are not enough posted limit orders to handle trade sizes of magnitude 1 million or 500 000 shares). The depth of the order book will take some time to replenish after a large order is executed, which means that the third option is probably the only viable choice. However, this option will require approximately 2 days to be fully completed, which increases the market risk of the operation. Moreover, the client gives a restriction that the order must be executed by the end of the day. Therefore, the best strategy will be for the trader to estimate the impact on the order book for orders of various sizes as well as the order book's replenishment rate, and then execute an optimal multi-trade strategy.

Optimal execution with price impact has been considered in many papers to date [1, 2, 16, 19]. In this paper, we denote by ψ the impact of a large market order, and we consider the immediate impact of a trade on the mid-price as

$$S(t) = S(t^-) + \psi(t^-, q),$$

where t is time, $S(t^-) = \lim_{\Delta t \searrow 0} S(t - \Delta t)$ is the pre-impacted mid price, $S(t)$ is the impacted mid-price, $q \in \mathbb{R}$ is the number of shares being bought or sold, and $\psi(t^-, \cdot) = \lim_{\Delta t \searrow 0} \psi(t - \Delta t, \cdot)$ is the impact function immediately prior to a trade at time t. One of the first impact functions to be proposed was the linear function [2, 24]

$$\psi(q) \propto q.$$

However, empirical evidence [12, 33, 35] suggests a volume dependence that is sub-linear, and described by a power law

$$\psi(q) \propto \operatorname{sgn}(q) \times |q|^{1/b}, \quad b > 1,$$

with an intraday temporal component that reduces impact with the passing of the minutes in the trading day. Also, Bouchaud [12] argued that the exponent of price impact is a function of the chosen partition time, typically ranging from 0.1 to 1.0. Using a data set from the Citigroup US equity trading desks, Almgren et al. [3]

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

estimated the impact to have $b = 5/3$. There are many other studies which provide evidence to support the idea that immediate price impact is a concave function; see [16, 18, 29, 33, 35] and the dimensional analysis approach of Pohl et al. [30].

Some important features to consider are the empirically observed dependence on order-flow imbalance (OFI; see [11, 15, 16]), and also the possibility that a proposed model might be flawed if it allows for so-called *quasi-arbitrage*, that is, the model is constructed in such a way that a trader can execute a sequence of orders that will terminate with a zero net inventory, and are expected to return a positive profit (see [19, 22]). Both of these ideas are taken into consideration as we develop the model in this paper.

1.1. Literature review Hawkes processes are shown by Bacry et al. [7] to reproduce volatility clustering effects and the so-called Epps effect in the correlation structure between two or more assets. Moreover, the differences of Hawkes processes have a natural appeal, since they represent price movements as jumps on a discrete grid and with random arrival times. Elsewhere in the literature, Hawkes processes with impact from market orders are considered by Alfonsi and Blanc [1] and Cartea et al. [14], and more general Hawkes applications to finance are included in the papers [6, 8–10]. In [26] a trader holds off execution until enough liquidity accumulates at the best bid/ask, which is different from this paper but related because it shows that traders can benefit by waiting. With regard to the permanence of price impact, Donier et al. [18] considered a model where impact is not expected to have complete decay, which is similar to the result we have in this paper. We also discuss the addition of a mean reversion term to eliminate permanent impact similar to the model of Alfonsi and Blanc [1]. There is also the issue of tick size in the order book, which is addressed by Smith et al. [33] where it is explained how the gaps in an order book change the shape of the impact function. To complete the review of the literature, we cite high-frequency market-making papers, such as [5, 13] where the market maker must estimate the impact of trades as part of his/her calculations in finding the optimal placement for his/her own limit of orders.

1.2. Results and structure of this paper In this paper we consider a standard Hawkes-process model and propose a criterion for adding price-impact functions. We investigate large-order impact, when market prices are modelled as a linear combination of self-exciting Hawkes processes. In particular, we model the mid-point price (that is, the half-way mark between the best bid and best ask prices) as the difference of two mutually exciting scalar-valued Hawkes processes, upon which we calculate the immediate and permanent impacts of an exogenously inserted market order.

We give some general properties that an impact function should have. More specifically, we propose a logarithmic impact function taking as input an order's volume and the Hawkes intensity processes; we also consider the far more tractable linear impact function. Based on the Hawkes model and our impact functions, we study the behaviour of prices after a large order has been executed. We give

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

expressions for temporary impact, permanent impact, and we give a measure of how much time is required for the effects of an order to decay. To argue for the soundness of our model, we examine the possibility of quasi-arbitrage and show that for two consecutive orders it is not possible for traders to obtain extreme profits from impacts caused by these orders. Finally, we present analysis to show improvement in the average price when a large order is executed in blocks using a time-weighted average price strategy.

The rest of this paper proceeds as follows. Section 2 introduces Hawkes processes and the model that we use throughout the paper. Section 3 introduces our proposed price-impact functions and examines the temporary and permanent impact of executed orders; Section 3.4.1 shows how to eliminate permanent impact including a rate of mean reversion in the mid-price process. Section 4 provides a discussion on the possibilities of quasi-arbitrage and the potential for price manipulation. Section 5 explores price impact for time-weighted average price strategies. Section 6 concludes.

2. Hawkes processes

A Hawkes process takes jumps with an intensity that is a function of time and history of past jumps [20]. Recently, Hawkes processes have been used in financial models for high-frequency trading (see, for example, [1, 7, 17, 25, 34]). We develop our model using the probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ [31], upon which we define a Hawkes process.

Definition 2.1 (Hawkes process). A Hawkes process is a pair $(N(t), \lambda(t))$ with $t \geq 0$ being time, that is a self-exciting jump process where $N(t)$ is a counting process and $\lambda(t)$ is the intensity of arrivals, and which has the dynamics

$$\lambda(t) = \mu + \int_{-\infty}^{t} \phi(t-s)\, dN(s),$$

where $\mu > 0$ is the minimum intensity level, function $\phi(t) \geq 0$ is a kernel to control the rate of memory decay, and $N(t)$ has increments that are conditionally distributed as

$$dN(t) \sim \text{Poisson}(\lambda(t)\, dt), \quad \text{with } N(0) = 0.$$

In this paper we use the kernel function

$$\phi(t-s) = \alpha e^{-\beta(t-s)},$$

where parameter $\beta > 0$ is the rate at which $\lambda(t)$ reverts towards μ, and parameter $\alpha > 0$ amplifies the effect on $\lambda(t)$ caused by a jump in $N(t)$. The Hawkes process is then written in terms of an initial condition

$$\lambda(t) = e^{-\beta t}\lambda(0) + (1 - e^{-\beta t})\mu + \alpha \int_0^t e^{-\beta(t-s)}\, dN(s).$$

It should be pointed out that choosing an exponential for ϕ is convenient because it makes (N, λ) an $\mathcal{F}_t$-adapted Markov process. Other kernels (for example, ϕ with polynomial decay) are more consistent with statistical analysis of market data, but may be non-Markov [6].

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

2.1. Stability of a Hawkes process The parameters α and β need to have some restriction in order for the Hawkes process to be used in modelling long-term behaviour. A necessary and sufficient condition for stability of a one-factor Hawkes process with exponential kernel function is

$$\alpha/\beta < 1 \tag{2.1}$$

(see [7]). To understand the reason for the condition $\alpha/\beta < 1$, consider the following: for general ϕ, take the limit of the expected value of $\lambda(t)$ for a one-factor Hawkes process

$$\begin{aligned}\mathbb{E}[\lambda(t)] &= \mu + \mathbb{E}\bigg[\int_{-\infty}^{t} \phi(t-s)\,dN(s)\bigg] \\ &= \mu + \int_{-\infty}^{t} \phi(t-s)\mathbb{E}[\lambda(s)]\,ds \\ &= \mu + \mathbb{E}[\lambda(t)]\int_{-\infty}^{t} \phi(t-s)\frac{\mathbb{E}[\lambda(s)]}{\mathbb{E}[\lambda(t)]}\,ds \\ &\to \frac{\mu}{1-\int_0^\infty \phi(v)\,dv},\end{aligned}$$

as $t \to \infty$, so that the expectation of the point process converges to

$$\mathbb{E}[N(t)] = \frac{\mu t}{1-\int_0^\infty \phi(v)\,dv}.$$

In the case of $\phi(t-s) = \alpha e^{-\beta(t-s)}$, we have $\mathbb{E}[N(t)] = \mu t/(1-\alpha/\beta)$, indicating long-term stability if $\alpha/\beta < 1$.

2.2. Asset-price model and two-factor Hawkes process Our model for an asset price is a two-factor Hawkes process. Specifically, we look at the difference between two Hawkes processes that are mutually exciting.

Definition 2.2 (Asset-price model from a two-factor Hawkes process). A two-factor mutually exciting Hawkes process with exponential kernel function is $(N_1(t), \lambda_1(t))$ and $(N_2(t), \lambda_2(t))$, for $t \geq 0$, where $N_1(t)$ and $N_2(t)$ are counting processes whose respective intensities are

$$\begin{aligned}\lambda_1(t) &= \mu + \alpha\int_{-\infty}^{t} e^{-\beta(t-s)}\,dN_2(s), \\ \lambda_2(t) &= \mu + \alpha\int_{-\infty}^{t} e^{-\beta(t-s)}\,dN_1(s),\end{aligned}$$

where $N_1(t)$ and $N_2(t)$ have increments conditionally distributed as

$$\begin{aligned}dN_1(t) &\sim \text{Poisson}(\lambda_1(t)\,dt), \\ dN_2(t) &\sim \text{Poisson}(\lambda_2(t)\,dt)\end{aligned}$$

with $N_1(0) = N_2(0) = 0$. This process is mutually exciting, because the counting processes will excite the intensities of one another.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

We denote by $S(t)$ the price of an asset, and, given the two-factor Hawkes process from Definition 2.2, we write $S(t)$ as a function of $N_1(t)$ and $N_2(t)$,

$$S(t) = S(0^-) + \delta \times (N_2(t) - N_1(t)), \tag{2.2}$$

where $\delta > 0$ is the tick size (for example, $\delta = 1/100$ or 1 cent). When a jump in the process $N_1(t)$ occurs, the value of $S(t)$ decreases and the intensity $\lambda_2(t)$ increases. Similarly, when a jump in $N_2(t)$ occurs, the value of $S(t)$ increases and the intensity $\lambda_1(t)$ increases. It is important to point out that the mutually exciting component in the model $S(t)$ creates some mean reversion in the short term, that is, an up tick in price causes an increase in the probability of a down tick. Mean reversion is a widely observed characteristic of high-frequency asset dynamics referred to as *retracement*; which is defined as the temporary reversal of an asset's expected trend. Retracement applies to market impact, because the newly impacted price is not expected to continue as if it were not impacted, but instead, it will regress back towards its pre-impacted price.

An important result for mutually exciting Hawkes processes and the price model proposed in equation (2.2) is that its limit in distribution is Brownian motion [23]. A proof of this limit is given by Bacry et al. [8], where the distribution of $(S(tT))_{t\leq 1}$ with $S(0-) = 0$ tends towards Brownian motion for $\delta = 1/\sqrt{T}$ and $T \to \infty$. The limit to Brownian motion is a nice characteristic, because it captures the following well-known phenomenon of financial data: asset prices exhibit mean reversion when observed at high frequency (that is, a frequency where prices look like Hawkes processes), but show less mean reversion when observed with coarser time frames (that is, a frequency where prices look like Brownian motion).

3. Price-impact functions

We now add the feature of price impact to our model. For a trader who is looking to place an order at time t, $\mathcal{F}_{t-}$ contains all information available at the instance their order is placed. *They cannot anticipate future movements of the market, and so they do not know $\mathcal{F}_t$ until time t has arrived.*

Let $q \in \mathbb{R}$ denote the size of a market order with

$$\begin{aligned} q > 0 \quad & \text{a } \textit{buy} \text{ order,} \\ q < 0 \quad & \text{a } \textit{sell} \text{ order.} \end{aligned}$$

A given order will have an immediate impact on the price, which we denote by $\psi(t, q)$. We consider a large order to be an exogenous impact on the Hawkes processes, and so the impact of a trade will augment the σ-algebra: a time-t market order of size q will have an impact of $\psi(t^-, q)$, which is an exogenous event that changes the Hawkes processes, and so the distribution of future price movements is conditioned on the joint σ-algebra

$$\mathcal{F}_{t-} \vee \{\psi(t^-, q)\}.$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

Formally, the price-impact function $\psi(t^-, q)$ is measurable with respect to $\mathcal{F}_{t^-}$ such that

$$\psi(t^-, q) = \text{impact on mid-price from a market order of size } q \text{ placed at time } t^-.$$

The function $\psi(t^-, \cdot)$ can be stochastic; in Section 3.3.1 we will introduce an impact function that depends on $\lambda_2(t^-) - \lambda_1(t^-)$. Since the trader does not have information to make anticipative trades, it follows that he/she cannot time a trade to coincide with an endogenous jump (that is, a jump in either N_1 or N_2 is endogenous). Hence, the probability that a market order is placed at the same time as an endogenous jump is zero, and we have

$$\mathcal{F}_t = \mathcal{F}_{t-} \vee \{\psi(t^-, q)\} \quad \text{almost surely.} \tag{3.1}$$

Remark 3.1. We have defined the σ-algebra $\mathcal{F}_t$ to be right continuous with a left-hand limit. The trading analysed in this paper will always manage the price history in this way. However, we should make the reader aware that the so-called *reactionary* strategies are permissible, that is, a trader can wait for the jump arrival in N_1 or N_2, and then immediately make a trade. If the exogenous impacts are of a reactionary form, then the σ-algebra is not right continuous, because

$$\mathcal{F}_{t^+} = \mathcal{F}_t \vee \{\psi(t, q)\}.$$

A reactionary strategy for price manipulation was discussed by Alfonsi and Blanc [1].

Definition 3.2 (Impacted mid-price). The instantaneous impact $\psi(t^-, q)$ (possibly stochastic) caused by a nonanticipative trade of size q executed in the market at time t^-, makes the immediate price impact

$$S(t) = S(t^-) + \psi(t^-, q), \tag{3.2}$$

where $S(t^-) = \lim_{\Delta t \searrow 0} S(t - \Delta t)$ and $\psi(t^-, \cdot) = \lim_{\Delta t \searrow 0} \psi(t - \Delta t, \cdot)$.

Remark 3.3 (Role of nonanticipative trades in Definition 3.2). Note that the assumption of trades being nonanticipative means that there is zero probability of the Hawkes process jumping at the trade time (as summarized by equation (3.1)), and so equation (3.2) holds almost surely.

Suppose at time t^- there is a market *buy* order with impact $\psi(t^-, q)$. Then

$$S(t) = S(0^-) + \delta(N_2(t) - N_1(t)) = S(0^-) + \delta(N_2(t^-) - N_1(t^-)) + \psi(t^-, q).$$

Assuming that *buy* orders have $\psi(t^-, q) > 0$, it follows that

$$N_2(t) = N_2(t^-) + \psi(t^-, q)/\delta,$$

and hence for $T > t$,

$$\lambda_1(T) = \mu + (\lambda_1(t^-) - \mu)e^{-\beta(T-t)} + \alpha \int_t^T e^{-\beta(T-s)}\, dN_2(s) + \frac{\alpha}{\delta}\psi(t^-, q)e^{-\beta(T-t)}.$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

This equation holds almost surely because of equation (3.1) and the nonanticipativeness of trades. Letting T tend down towards t, we see that this type of exogenous impact causes a jump in the intensity,

$$\lambda_1(t) = \lambda_1(t^-) + \frac{\alpha}{\delta}\psi(t^-, q), \tag{3.3}$$

at execution time t for order size $q > 0$; a similar effect occurs for $q < 0$ and $\lambda_2(t)$,

$$\lambda_2(t) = \lambda_2(t^-) - \frac{\alpha}{\delta}\psi(t^-, q), \tag{3.4}$$

at execution time t.

Remark 3.4. It should be pointed out that the exogenous impact relations of equation (3.3) and (3.4) cause N_1 and N_2 no longer to be counting processes. Instead, these impact relations mean that we are using a type of hybrid Hawkes-based model. The long-term stability condition in (2.1) does not ensure that the hybrid process has stability. Indeed, a series of destabilizing exogenous impacts could be designed. However, any sequence of finitely many impacts of finite size cannot destabilize the process; this is the only type of price-impact series that will be considered in this paper.

Given the state of the order book, the impact will depend on other factors such as order-book depth and OFI. However, it is generally accepted that ψ is concave for $q > 0$ and convex for $q < 0$. Figure 1 provides a visual explanation of the effects on the order book when a *buy* market order is sent at market price, and also the reason it becomes increasingly difficult to impact the price.

3.1. Order-flow imbalance The order-flow imbalance is a measure of supply and demand imbalance [16]. OFI can be seen in Figure 1, where in the moments after the market order has been placed there are fewer limit *sell* orders in the queue for the best ask price. According to our model, if there is a jump in $\lambda_1(t)$ the intensity of the process $N_2(t)$ increases, making it more likely that the asset price goes up rather than down. Similarly, there is also an imbalance in the order flow as the process $\lambda_2(t)$ jumps but in the opposite direction. Therefore, we interpret the difference $\lambda_2(t) - \lambda_1(t) \neq 0$ as an imbalance in the order flow.

Definition 3.5 (Order-flow imbalance). OFI is defined as a measure of the excess in buy or sell orders for a trading security during a period of time.

To test the connection between OFI and the Hawkes intensities, one could run a regression to find the factors

$$\text{OFI}_t = \beta_0 + \beta_1(\lambda_2(t) - \lambda_1(t)) + \text{“noise”}.$$

In practice, the OFI is something that can be computed from the data but the λs would need to be filtered (see [34] for filtering of Hawkes processes), and then a regression run on the filtered λs. Further discussion on the effects of filtering appears in Section 5.6.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

FIGURE 1. The order-book dynamic before and after the impact from a *buy* market order of 15 000 shares. If the order size had been slightly larger, there would have been nonzero price impact. However, notice how the order-book depth is such that limit *sell*-order queues are bigger for prices, away from the best ask price. This increase in depth is what creates the price impact's concavity.

3.2. Net effect of signed orders The order book is composed of limit *buy* and limit *sell* orders at different prices. Previously, in Figure 1 we illustrated the order-book dynamic when a market order is executed. It is natural to infer that two market *buy* orders or two market *sell* orders of quantity q_1 and q_2 sent in sequence with arbitrarily small waiting time Δt will have the same order-book impact of a single order with quantity $q_1 + q_2$ when both orders impact the same side of the order book; this concept is illustrated in Figure 2. On the other hand, when successive orders have different signs their net impact will not have the same impact as a single net order; for instance, the impact of sending one *buy* (*sell*) order of size q followed Δt time units later by a *sell* (*buy*) order of size q is not necessarily zero. Figure 2 illustrates an example of two market orders, one *buy* of size 12 000 and one *sell* of size 12 000, that are sent in short succession. Notice that the outstanding shares available at the best bid are less than the *sell* order size and at the best ask greater than the *buy* order size, so that the first order has no price impact, while the second order causes the mid-price to decrease. If the trader had recognized initially the net quantity of orders, then he/she could have placed an order of size zero, which obviously would have had zero impact.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

FIGURE 2. The impact of two market orders, one *buy* and one *sell*, both of size 12 000, sent at the same time. The net impact is not zero.

From this concept of net effect on the order book, we develop an important characteristic that should be included when constructing a price-impact function. Namely, for two orders with the same sign and sizes q_1 and q_2, their impact is the same as one order of size $q_1 + q_2$, when the time interval between orders is $\Delta t \to 0$. In general, for $q_1, q_2, q_3, \ldots, q_n$ being market orders with the same sign, and $t = t_1 < t_2 < t_3 < \cdots < t_n$ being the placement time for respective orders, the effect of the trades should be

$$\lim_{\Delta t \to 0} \sum_{i=1}^{n} \psi(t_i, q_i) = \psi\Big(t, \sum_{i=1}^{n} q_i\Big),$$

where $\Delta t = \max_i(t_i - t_{i-1})$.

3.3. Immediate impact of one large order arrival This subsection explores the immediate impact of a large order. The impact ψ is exogenously inserted into the asset-price model of equation (2.2), and then the impact on price evaluated. Based on our discussion thus far on q dependence, the relevance of OFI from Section 3.1 and the behaviour of net effects from Section 3.2, we list some desired characteristics for a price-impact function.

CONDITION 3.6 (Criterion for price-impact function). An impact function ψ satisfies the following criteria.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

(1) $\psi(t,0)=0$, $\psi(t,q)>0$ for $q>0$, and $\psi(t,q)<0$ for $q<0$.
(2) $|\psi|$ is strictly concave increasing in q for $q>0$ and strictly concave decreasing in q for $q<0$.
(3) $|\psi|$ is an increasing function of $\lambda_2(t)-\lambda_1(t)$ when q is positive and a decreasing function of $\lambda_2(t)-\lambda_1(t)$ when q is negative. The difference of the intensities measures how likely it is that the price will go up or down.
(4) For q_1 and q_2 of the same sign,

$$\lim_{\Delta t\to 0}\{\psi(t^-,q_1)+\psi(t+\Delta t,q_2)\}=\psi(t^-,q_1+q_2),$$

where Δt is the time interval between the order of size q_1 and q_2; for example, the expected impact of two consecutive market *buy*/*sell* orders must converge to the expected impact of a single *buy*/*sell* order that is in the quantity of the summed order sizes.

Bouchaud [12] suggested that dependence of ψ on volume is sub-linear, and it is described by a power law with parameter that is a function of the chosen partition interval Δt. However, Bouchaud argued that the power observed was typically close to $1/2$. Other studies have found a similar value of the parameter (see [35]); Almgren et al. [3] argued that impact should be around a $3/5$ power-law function.

3.3.1 *Impact-function examples.* We now present some examples that fit the criteria of Condition 3.6. In particular, we present a logarithmic function.

Example 3.7 (Linear impact function). The linear impact function ψ of a large order of size q is described as

$$\psi(t,q)=c(t)q, \tag{3.5}$$

where $c(t)>0$ is a continuous function of time that measures order-book liquidity at time t. This linear impact is similar to the model proposed by Kyle [24], except for the fact that $c(t)$ is a function of time; intraday dynamics of $c(t)$ are discussed by Cont et al. [16]. The linear impact function does not satisfy criteria (2) and (3) of Condition 3.6.

Example 3.8 (Logarithm-based impact function). A logarithmic impact function ψ for an order of size q is

$$\psi(t,q)=\operatorname{sgn}(q)b\ln\left(1+c(t)|q|\exp\left(\operatorname{sgn}(q)\frac{\lambda_2(t)-\lambda_1(t)}{b\alpha/\delta}\right)\right), \tag{3.6}$$

where $c(t)$ is a continuous function, $b>0$ is a shape parameter to provide the concavity, and $\operatorname{sgn}(\cdot)$ is the sign of the input quantity. This logarithmic impact function satisfies all points in Condition 3.6. Using the exogenous impact properties in equations (3.3) and (3.4), criterion (4) in Condition 3.6 for the logarithmic impact function is shown

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

to be

$$\begin{aligned}
&\lim_{\Delta t\to 0}\{\psi(t^-,q_1)+\psi(t+\Delta t,q_2)\}\\
&\quad=\operatorname{sgn}(q_1)b\ln(1+c(t)|q_1|\exp(\operatorname{sgn}(q_1)\{\lambda_2(t^-)-\lambda_1(t^-)\}/(b\alpha/\delta)))\\
&\qquad+\operatorname{sgn}(q_2)b\ln\left(1+c(t)|q_2|\exp\left(\operatorname{sgn}(q_2)\frac{-(\alpha/\delta)\psi(t^-,q_1)+\{\lambda_2(t^-)-\lambda_1(t^-)\}}{b\alpha/\delta}\right)\right)\\
&\quad=\operatorname{sgn}(q_1)b\ln\left(1+c(t)|q_1|\exp\left(\operatorname{sgn}(q_1)\frac{\lambda_2(t^-)-\lambda_1(t^-)}{b\alpha/\delta}\right)\right)\\
&\qquad+\operatorname{sgn}(q_2)b\ln\left(1+\frac{c(t)|q_2|\exp(\operatorname{sgn}(q_2)\{\lambda_2(t^-)-\lambda_1(t^-)\}/(b\alpha/\delta))}{(1+c(t)|q_1|\exp(\operatorname{sgn}(q_1)\{\lambda_2(t^-)-\lambda_1(t^-)\}/(b\alpha/\delta)))^{\operatorname{sgn}(q_1q_2)}}\right)\\
&\quad=\operatorname{sgn}(q_1)b\ln(1+c(t)|q_1|\exp(\operatorname{sgn}(q_1)\{\lambda_2(t^-)-\lambda_1(t^-)\}/(b\alpha/\delta))\\
&\qquad+c(t)|q_2|\exp(\operatorname{sgn}(q_2)\{\lambda_2(t^-)-\lambda_1(t^-)\}/(b\alpha/\delta)))\\
&\quad=\psi(t^-,q_1+q_2).
\end{aligned}$$

Note that this calculation requires orders q_1 and q_2 to have $\operatorname{sgn}(q_1)=\operatorname{sgn}(q_2)$.

Remark 3.9. At this point, two comments should be addressed. First, the impact functions presented above do not take value in the same tick grid as the asset price does in equation (2.2). Second, all orders have an impact on the mid-price but it has been shown that most orders, particularly orders of small quantity, do not have immediate impact on the mid-price (see [33]). Both issues could be addressed by truncating the impact functions so that the impacted price takes the nearest value on the tick grid. Specifically, any order for which the impact is less than one tick will result in zero immediate price impact in the mid-price. However, this change could lead us to less tractable mathematical solutions, and so this issue is forgone for the rest of this paper.

3.3.2 *Intraday dynamics of $c(t)$.* In Examples 3.7 and 3.8, the function $c(t)$ scales the order size q and reflects order-book liquidity at time t. It is widely known that liquidity changes during the trading day. Order-book liquidity is usually greater at the end of the day than in the middle of the day. It is then natural to say that the expected impact of an order at the end of the day is lower than in the middle of the day. This is a reason to have a coefficient c that depends on time. Almgren et al. [3] considered liquidity changes during a single day. In particular, they modelled liquidity using the volatility and an average daily-volume parameter. Another interesting study was by Cont et al. [16] where they discussed how the opening minutes of the market have depth that is two times lower than the average, and hence the impact of orders decreases throughout the day.

In addition to the function $c(t)$ being time dependent, one could let the Hawkes process parameters μ, α and β have some intraday dynamics, as well as the δ and b in the impact function.

3.4. Permanent impact of one large order Suppose that a large market order of size $q>0$ is placed in the order book at time t. The function $\psi(t,q)$ is the amount by which this order will *immediately* impact the price. Much of this immediate impact is

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

only temporary, and for the model given by equation (2.2) along with the effects of ψ as shown in equations (3.3) and (3.4), there is expected to be some permanent impact. In this section, we focus on *buy* orders ($q > 0$), but the same permanent-impact rules also apply to *sell* orders.

Given a trade at time t with impact $\psi(t^-, q)$, the permanent impact is defined as follows.

Definition 3.10 (Permanent impact function). The expected permanent impact function of an order with impact $\psi(t^-, q)$ at time t^- is

$$P(q) = \lim_{T\to\infty} (\mathbb{E}[S(T)|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}] - \mathbb{E}[S(T)|\mathcal{F}_{t^-}]).$$

Indeed, for the model in this paper and the type of price impact under consideration, there is a nonzero permanent impact.

Proposition 3.11 (Permanent impact of an order $\psi(t^-, q)$ at time t^-). *The expected permanent impact of an order $\psi(t^-, q)$ at time t^- is*

$$P(q) = \psi(t^-, q)\frac{\beta}{\alpha + \beta}. \tag{3.7}$$

Proof. See Appendix A.1. □

Recall the process stability condition, $\beta/\alpha > 1$. Therefore, the permanent impact is bounded by $|\psi|/2$, that is,

$$|P(q)| > \frac{|\psi(t^-, q)|}{2}.$$

Figure 3 displays the Monte Carlo average for the impact of a large order, from which the level of permanent impact can be seen.

3.4.1 *Zero permanent impact via reversion to a fundamental price.* The permanent impact of equation (3.7) may or may not be an appropriate feature to be allowed in a trading model. If the model is used for intraday trading with a finite time, then the presence of permanent impact is less of an issue. However, if long time-scales are considered then it will not be reasonable to allow permanent impact. For instance, if limit orders are likely to replenish the order book in a relatively short amount of time, then permanent impact could lead to profit-making strategies, which would be arbitrage in the sense that will be defined shortly in Section 4.2.

The models of Alfonsi and Blanc [1] and Obizhaeva and Wang [28] have the added feature of mean reversion to a *fundamental price*, which in turn reduces permanent impact to zero. For example, letting $F(t)$ denote the fundamental price, the mid-price will have the differential

$$dS(t) = \rho(F(t) - S(t))\,dt + \delta(dN_2(t) - dN_1(t)),$$

where $\rho > 0$ is the rate of mean reversion to F, and it assumes that trades affect the Poisson intensities in the same manner given by equations (3.3) and (3.4). This model

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

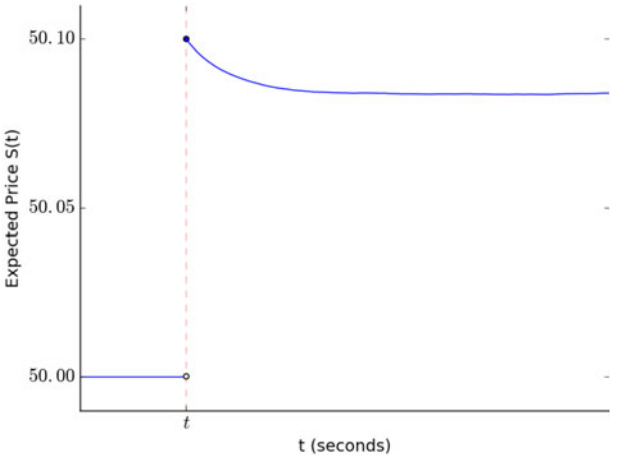

FIGURE 3. The expected value of $S(t)$ after being impacted at time $t = 0$ by a large order with $\psi(0, q) = 10$, $\alpha = 0.2, \beta = 1.0, S(0) = 50, \delta = 0.01$.

has no permanent impact because impacts and initial conditions are forgotten at an exponential rate,

$$\mathbb{E}[S(t)] = S(0)e^{-\rho t} + \rho \int_0^t e^{-\rho(t-u)}\mathbb{E}F(u)\,du + \frac{\delta(\lambda_2(0) - \lambda_1(0))e^{-\rho t}(1 - e^{-(\alpha+\beta-\rho)t})}{\alpha + \beta - \rho},$$

where there is a need for a model for $F(t)$. For example, $F(t)$ could be a process driven by a Brownian motion. In [1] they show the potential for price manipulation in this model; we address price manipulation in Section 4.

4. Expected profit and price manipulation

4.1. Average execution price As done by Rogers and Singh [32], there is a representation of a trade size with the integral of an order-book density. To get to an integral representation, we start by considering the number $m(q)$ equal to the number of queues that a large order q consumes in the limit order book. The order q is written as a sum

$$q = \sum_{j=0}^{m(q)} \tilde{q}_j(t^-),$$

where $\tilde{q}_j(t)$ is the number of limit orders consumed at time t from the queue at j ticks from the best bid/ask. The order is executed as follows: from the total size q, the first $\tilde{q}_0$ shares will be executed in the first level of the book, $\tilde{q}_1$ in the second level and so on, until the order is completely executed. The values of $\tilde{q}_0, \tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_m$ are estimated by inverting the impact function ψ, as shown in Figure 4.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

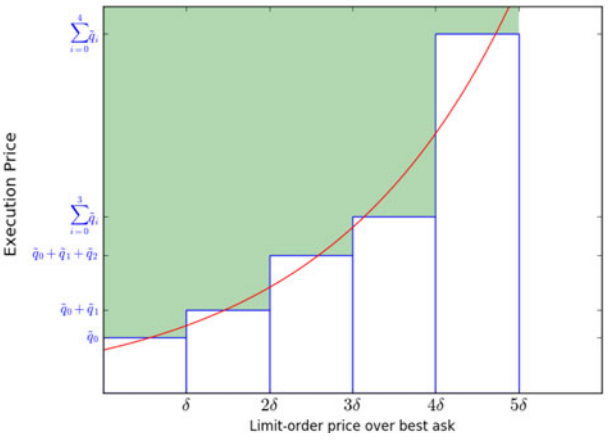

FIGURE 4. The cost of a *buy* order of size $q = \sum_{i=1}^{m(q)} \tilde{q}_i$. The horizontal axis marks the order queues for each tick size, but the height of each queue is the cumulative number of limit orders available at that price or lower. The convex line is the inverse of the impact function $\psi^{-1}(t^-, \cdot)$, and the shaded area is the total price paid, equal to $\int_0^q \psi(t^-, v)\, dv$. The average price is the shaded area divided by q.

Let parameter $\theta > 0$ denote the spread between the best bid and best ask prices. Given the $\tilde{q}_j$, the average execution price is

$$\text{Average executed price}(t^-, q) = \frac{1}{q} \sum_{j=0}^{m(q)} \tilde{q}_j(t^-) \left\{ S(t^-) + \text{sgn}(q) \left(j\delta + \frac{\theta}{2} \right) \right\}.$$

The first term in this expression is the mid-price, the second term is the impact cost of execution, and the third term is half the bid–ask spread to make the execution price begin at the best bid/ask. Based on the inverse relation with the impact function (as seen in Figure 4), a natural way to rewrite the above equation is by taking a continuum limit and using the function ψ as follows:

$$\text{Average executed price}(t^-, q) = S(t^-) + \frac{1}{q} \int_0^q \psi(t^-, v)\, dv + \text{sgn}(q) \frac{\theta}{2}.$$

It is also possible to add a term k that represents broker/exchange fees in order to get the net executed price of an order of volume q at time t, that is,

$$\text{Net average executed price}(t^-, q) = S(t^-) + \frac{1}{q} \int_0^q \psi(t^-, v)\, dv + \text{sgn}(q) \left(k + \frac{\theta}{2} \right).$$

Using the net executed profit, it is possible to extract the profit function. For the case in which the final position is zero, it is not necessary to mark to market, and so the profit function is given by

$$\text{Profit}(\Pi) = \underbrace{-\sum_{i=1}^{n} S(t_i^-) q_i}_{\text{gross profit}} \quad \underbrace{-\sum_{i=1}^{n} \int_0^{q_i} \psi(t_i^-, v)\, dv}_{\text{impact cost}} \quad \underbrace{-\left(k + \frac{\theta}{2} \right) \sum_{i=1}^{n} |q_i|}_{\text{fixed cost}}.$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

The first term of the above equation is the gross profit of the trading strategy. The second represents the impact cost for each order that composes the strategy, which could be made more elaborate and depend on the shape of the order book. The third term is fixed costs related to broker/exchange fees and the bid–ask spread.

4.2. Price-manipulating strategy In this subsection we investigate market conditions under which the price-impact function may allow *price manipulation*. First, we define some basic concepts, similar to those explained in the papers [1, 19].

Definition 4.1 (Trading strategy). A trading strategy is a sequence of orders $(q_1, q_2, q_3, \ldots, q_n)$ to be sent to the market in a finite interval $[0, T]$.

Definition 4.2 (Round-trip trading strategy). A round-trip trading strategy is a trading strategy that is pre-determined to have a net position of zero at terminal time T, that is,

$$\sum_{i=1}^{n} q_i = 0.$$

The definition of price-manipulating strategy was given by Huberman and Stanzl [22].

Definition 4.3 (Price-manipulating strategy). A price-manipulating strategy is a round-trip trading strategy in which the expected profit is greater than zero, that is,

$$\mathbb{E}[\text{Profit}(\Pi)] > 0, \quad \sum_{i=1}^{n} q_i = 0.$$

In the rest of this subsection some simple strategies are checked to see whether or not market conditions of the proposed model allow for price manipulation.

4.2.1 *Strategies with only two orders and linear impact function.* Consider a case in which a *buy* order of size q is sent to the market at time zero and is closed after the passage of t units of time. In this case, based on the profit function described in the previous section, the expected profit is

$$\mathbb{E}[\text{Profit}(\Pi)] = q\left(\psi_1(t^-, q) + \frac{[\alpha\psi_1(t^-, q) - \delta(\lambda_2(t^-) + \lambda_1(t^-))](e^{-\Delta t(\alpha+\beta)} - 1)}{\alpha + \beta}\right) - \int_0^q \psi_1(t^-, v)\, dv + \int_{-q}^{0} \mathbb{E}\psi_2(t + \Delta t, v)\, dv - q(2k + \theta),$$

where ψ_1 is the impact of the first order of size q and ψ_2 is the impact of the following order of size $-q$. For the linear impact function defined in equation (3.5), the expected profit is

$$\mathbb{E}[\text{Profit}(\Pi)] = q\left(c(t)q + \frac{[c(t)q\alpha - \delta(\lambda_2(t^-) + \lambda_1(t^-))](e^{-\Delta t(\alpha+\beta)} - 1)}{\alpha + \beta}\right) - \frac{c(t) + c(t + \Delta t)}{2} q^2 - q(2k + \theta).$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

This quantity is monotone in Δt, and so as $\Delta t \to 0$ (because $c(t)$ is a continuous function) we have price manipulation if

$$\mathbb{E}[\text{Profit}(\Pi)] = -q(2k+\theta) < 0,$$

or when $\Delta t \to \infty$, assuming that $\lim_{\Delta t\to\infty} c(t+\Delta t) = c(t)$, we might have

$$\mathbb{E}[\text{Profit}(\Pi)] \approx q\left(c(t)q - \frac{c(t)q\alpha - \delta(\lambda_2(t^-)+\lambda_1(t^-))}{\alpha+\beta}\right) - c(t)q^2 - q(2k+\theta) > 0,$$

and therefore, the trading strategy is profitable if

$$\lambda_2(t^-) - \lambda_1(t^-) > \frac{1}{\delta}((2k+\theta)(\alpha+\beta) + c(t)\alpha q). \tag{4.1}$$

Again, the lowest boundary is achieved when $q \to 0$, so the price manipulation is present in this case:

$$\lambda_2(t^-) - \lambda_1(t^-) > \frac{1}{\delta}(2k+\theta)(\alpha+\beta).$$

Similarly, it is possible first to send a *sell* order and then to close the position after t. To consider both cases, manipulation is present when

$$|\lambda_2(t^-) - \lambda_1(t^-)| > \frac{1}{\delta}(2k+\theta)(\alpha+\beta).$$

Note that the price manipulation strategy here is not only restricted by the OFI $\lambda_2 - \lambda_1$, but also by the order size given in (4.1).

4.2.2 *Strategies with only two orders and logarithmic impact function.* This is similar to the previous strategy , except that we consider the logarithmic impact function from Example 3.8 satisfying all points of Condition 3.6.

Proposition 4.4. *Consider a concave impact function such that, asymptotically,*

$$\psi(t,q) \sim \log(1+q), \quad q \gg 1,$$

for all $t \geq 0$. Then there is a negative expected profit for a large round-trip trading strategy consisting of two orders.

Proof. The expected profit is

$$\begin{aligned}
\mathbb{E}[\text{Profit}(\Pi)] &= \mathbb{E}\Bigg[q\Big\{\psi(t^-,q) + \frac{(\delta(\lambda_2(t^-)-\lambda_1(t^-)) - \alpha\psi(t^-,q))(1-e^{-(T-t)(\alpha+\beta)})}{\alpha+\beta}\Big\} \\
&\quad - \int_0^q \psi(t^-,v)\,dv + \int_{-q}^0 \psi(T^-,v)\,dv\Bigg] - q(2k+\theta) \\
&= q\ln(q+1)\underbrace{\Big\{-1 - \frac{\alpha(1-e^{-(T-t)(\alpha+\beta)})}{\alpha+\beta} + \frac{2}{\ln(q+1)} - \frac{2}{q}\Big\}}_{<0} \\
&\quad + q\Big\{\frac{\delta(\lambda_2(t^-)-\lambda_1(t^-))(1-e^{-(T-t)(\alpha+\beta)})}{\alpha+\beta} - (2k+\theta)\Big\} \\
&< 0, \quad \text{for } q \gg 0 \text{ and } \lambda_2(t^-) = \lambda_1(t^-).
\end{aligned}$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

The same asymptotic method could be applied to the case in which a *sell* order is sent and then the position is closed at T if $\lambda_2(t^-) = \lambda_1(t^-)$. □

5. Time-weighted average price

For the years leading up to 2016, it was estimated that high-frequency trading accounted for roughly 55% of trading volume in US equity markets and about 40% in European equity markets (see more on these statistics in [27]). From November 2012 to October 2014, automatic trading systems accounted for 79.9% of foreign-exchange futures trading volume and 62.3% of interest-rate futures trading volume (see [21]). The increasing number of trades made by algorithms has created a demand for execution strategies to reduce the market impact of large orders. Algorithms such as time-weighted average price (TWAP) and the volume-weighted average price (VWAP) (see [2, 4]) are some of the strategies widely used by brokers, hedge funds and banks to optimize their execution. Such strategies share the idea that it is better to slice a large order into several smaller orders as a means to reduce impact. The idea of a TWAP strategy is to slice the order into equally sized market orders to be executed successively at equally spaced times. In this section we study the benefits of using a TWAP strategy.

5.1. Optimal waiting time after a large order is executed

Suppose that a large *buy* order was executed at time t and a new large *buy* order needs to be executed soon after. If the time between these two orders is too short, the market will not mean-revert and the execution price for the second order will be too high. If the trader waits too long to make the second trade, then there is a good chance the market will move in an undesirable way and the trader will suffer a loss. Therefore, the trader should look for an optimal time interval to make this trade. Given a trade at time t with impact $\psi(t^-, q)$, the temporary impact is defined as follows.

Definition 5.1 (Temporary impact function). The expected temporary impact function of an order with impact $\psi(t^-, q)$ at time t is

$$T(q, \Delta t) = \mathbb{E}[S(t + \Delta t)|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}] - \mathbb{E}[S(t + \Delta t)|\mathcal{F}_{t^-}] - P(q),$$

where $P(q)$ is the permanent impact.

The problem of choosing optimal Δt can be formulated as the minimum time to wait for the expected impacted price to be within a constant value ξ of its long-term limit, that is,

$$\Delta t^* = \underset{\Delta t>0}{\operatorname{argmin}} |T(q, \Delta t)| \leq \xi \quad \text{where } \xi > 0. \tag{5.1}$$

Proposition 5.2 (Optimal waiting time). *If the price $S(t)$ follows a two-factor Hawkes process, the optimal waiting time after an order of size q to ensure that the expected price is no more than a constant ξ from the long-term limit is*

$$\Delta t^* = \left(-\frac{1}{\alpha + \beta} \ln \left(\frac{\xi(1 + \beta/\alpha)}{|\psi(t^-, q)|} \right) \right)^+.$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

Proof. See Appendix A.2. □

A trader who implements a TWAP strategy will split a large order into several equal-sized smaller orders, and then send them in successive time increments of $\Delta t > 0$. Hence, it is clear that the optimal Δt^* from Proposition 5.2 will be useful for TWAP.

5.2. Temporary impact of a sequence of large orders (TWAP strategy) The execution of a market order will have an immediate impact of $\psi(t^-, q)$, but the market will have some reaction to this impact. In particular, it has been empirically observed that markets exhibit some reversion towards the pre-impact price, or *retracement* as we henceforth will refer to it. For example, a large *buy* order will have impact $\psi(t^-, q)$, but in the time period $[t, t + \Delta t]$ the price will exhibit some retracement downward towards the pre-impact price. Figure 5 illustrates this idea.

Definition 5.3 (Retrace function). For a TWAP strategy with order size q, we define the retracement function

$$R(t, q; \Delta t) = \mathbb{E}[S(t + \Delta t) - S(t) | \mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}],$$

that is, the amount of expected price retrace Δt time units after the order is executed at time t.

Repeating the steps taken in the proof of Proposition 3.11 yields

$$\begin{aligned}
&\mathbb{E}[S(t + \Delta t) | \mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}] \\
&\quad = S(t^-) + \psi(t^-, q) + \frac{(\delta(\lambda_2(t^-) - \lambda_1(t^-)) - \alpha\psi(t^-, q))(1 - e^{-\Delta t(\alpha+\beta)})}{\alpha + \beta} \\
&\quad = S(t) + \underbrace{\frac{\delta(\lambda_2(t^-) - \lambda_1(t^-))(1 - e^{-\Delta t(\alpha+\beta)})}{\alpha + \beta} - \frac{\alpha\psi(t, q)(1 - e^{-\Delta t(\alpha+\beta)})}{\alpha + \beta}}_{\text{retrace}},
\end{aligned}$$

where $S(t) = S(t^-) + \psi(t^-, q)$ almost surely, and the expected retracement

$$R(t, q; \Delta t) = \frac{(\delta(\lambda_2(t^-) - \lambda_1(t^-)) - \alpha\psi(t^-, q))(1 - e^{-\Delta t(\alpha+\beta)})}{\alpha + \beta}. \tag{5.2}$$

The retrace function helps us to identify the components of an impact $\psi(t^-, q)$, namely, it helps to determine which part of the initial impact is permanent and which part is temporary.

5.3. TWAP effectiveness In this section, we investigate the effectiveness of the TWAP strategy for our proposed model. We find that, generally speaking, it is better to implement TWAP than to execute a single large order, but sometimes it may be better to finish the strategy early if a condition on $\lambda_2 - \lambda_1$ indicates poor performance otherwise. The latter idea is what we refer to as *quasi-TWAP*, which we will explain shortly.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

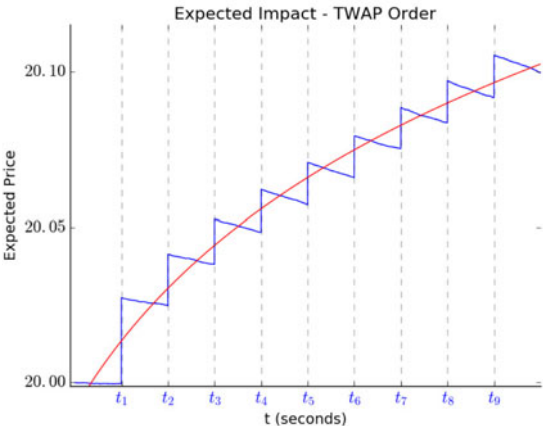

FIGURE 5. The immediate impact and subsequent retracement for a TWAP strategy. The TWAP strategy is to divide a large order into several equal-sized orders and then execute them successively at equally spaced time increments. In this case, *buy* orders have a positive impact on the price, and the retracement is downward towards the pre-impact price.

5.3.1 *Expected average execution price for TWAP under linear impact function.* In this section we show that the TWAP strategy is "better" than a single order in the sense that the expected average price per share for a TWAP strategy is better than the price per share for a single large order. Focusing on the *buy* order case, the average executed price (Avg Exec Price) of a single order of size q is

$$\text{Avg Exec Price 1 Order}(t^-, q) = S(t^-) + \frac{1}{q}\int_0^q \psi(t^-, v)\,dv + \text{sgn}(q)\frac{\theta}{2},$$

and the expected average execution price for the TWAP strategy is

$$\begin{aligned}&\text{Avg Exec Price } n \text{ Orders}(t^-, q)\\ &= \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\Big[S((t+i\Delta t)^-)|\mathcal{F}_{t^-} \vee \bigvee_{j=0}^{i}\{\psi((t+j\Delta t)^-, q/n)\}\Big]\\ &\quad + \frac{1}{q}\sum_{i=0}^{n-1}\int_0^{q/n}\mathbb{E}\Big[\psi(t+i\Delta t, v)|\mathcal{F}_{t^-} \vee \bigvee_{j=0}^{i-1}\{\psi((t+j\Delta t)^-, q/n)\}\Big]\,dv + \text{sgn}(q)\frac{\theta}{2}.\end{aligned}$$

Based on expected average execution price, a trader faced with linear price impact will have the following criterion to help him/her in choosing to either (a) place a single large market order or (b) implement TWAP.

PROPOSITION 5.4 (TWAP effectiveness under the linear impact function). *Consider the linear impact function $\psi(t, q) = c(t)q$, where $c(t)$ is continuous and independent with $\mathbb{E}[c(t+s)|\mathcal{F}_t] = c(t)$. Then the n-order TWAP strategy with buy orders of size $q/n > 0$*

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

scheduled at times $t_i = t + i\Delta t$, for $i = 0, 1, 2, \ldots, n-1$, has better average price than a single order of size q if and only if

$$\lambda_2(t^-) - \lambda_1(t^-) \le \frac{\alpha c(t) q}{\delta n} \times \left[\frac{(n-1)(1 - e^{-(\alpha+\beta)\Delta t})/2 - e^{-(\alpha+\beta)\Delta t} + e^{-(\alpha+\beta)\Delta t}(1 - e^{-(\alpha+\beta)n\Delta t})/n(1 - e^{-(\alpha+\beta)\Delta t})}{1 - e^{-(\alpha+\beta)\Delta t} - (1 - e^{-(\alpha+\beta)n\Delta t})/n} \right]. \tag{5.3}$$

For $\lambda_2(t^-) - \lambda_1(t^-) = 0$ and $n > 1$, it is clear that TWAP is better. For a TWAP of sell orders of size $q/n < 0$, the condition has the reverse inequality.

Proof. See Appendix A.3. □

The proof of Proposition 5.4 derives the following:

$$\begin{aligned} &\text{Avg Exec Price } n \text{ Orders}(t^-, q) \\ &\quad = \text{Exec Price 1 Order}(t^-, q) \\ &\qquad + \delta \frac{\lambda_2(t^-) - \lambda_1(t^-)}{n(\alpha+\beta)} \left(n - \frac{1 - e^{-(\alpha+\beta)n\Delta t}}{1 - e^{-(\alpha+\beta)\Delta t}} \right) \\ &\qquad - \frac{\alpha c(t) q}{n^2(\alpha+\beta)} \left(\frac{n(n-1)}{2} - \frac{n e^{-(\alpha+\beta)\Delta t}}{1 - e^{-(\alpha+\beta)\Delta t}} + \frac{1 - e^{-(\alpha+\beta)n\Delta t}}{(1 - e^{-(\alpha+\beta)\Delta t})^2} e^{-(\alpha+\beta)\Delta t} \right). \end{aligned}$$

For $\lambda_2(0^-) - \lambda_1(0^-) = 0$, the right-hand side of this expression is minimized as $\Delta t \to \infty$ and $n \to \infty$, and so there is a best possible expected average TWAP execution price for *buy* orders,

$$\text{Avg Exec Price } n \text{ Orders}(t^-, q) \ge \text{Exec Price 1 Order}(t^-, q) - \frac{\alpha c(t) q}{2(\alpha+\beta)} \quad \text{for } \lambda_2(0^-) - \lambda_1(0^-) = 0. \tag{5.4}$$

The bound in (5.4) will be demonstrated through simulations in Section 5.5.

5.4. A quasi-TWAP strategy The large single-order strategy has one advantage over the TWAP: it does not have price uncertainty. This price uncertainty is sometimes referred to as *market risk*, because the TWAP's average execution price may be worse than a single order if there is a significant price movement. Therefore, we propose a quasi-TWAP strategy, which is essentially a TWAP with an early-exit trigger. The quasi-TWAP always takes less or equal time to execute than a TWAP, and therefore it has less market risk. The following algorithm gives a step-by-step decision process for the quasi-TWAP.

For the linear impact function, the quasi-TWAP early-exit trigger is given in the following proposition.

Proposition 5.5 (quasi-TWAP: linear impact function). *We denote the quantity*

$$\tau(n, \Delta t) = \frac{(n-1)(1 - e^{-(\alpha+\beta)\Delta t})/2 - e^{-(\alpha+\beta)\Delta t} + e^{-(\alpha+\beta)\Delta t}(1 - e^{-(\alpha+\beta)n\Delta t})/n(1 - e^{-(\alpha+\beta)\Delta t})}{1 - e^{-(\alpha+\beta)\Delta t} - (1 - e^{-(\alpha+\beta)n\Delta t})/n}.$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

Algorithm quasi-TWAP algorithm (*buy* order)

for $i = 0$ to $n - 1$ **do**
  if Avg Exec Price $n - i$ Orders$(t_i^-, q(n-i)/n) <$ Exec Price 1 Order$(t_i^-, q(n-i)/n)$ **then**
    Send a *buy* Order of size q/n;
    Wait Δt;
  else
    Send a *buy* Order of size $q(n-i)/n$; **break**;
  end if
end for

Based on inequality (5.3)*, the early-exit trigger for the quasi-TWAP with linear price-impact function is*

$$\lambda_2(t^-) - \lambda_1(t^-) \geq \frac{\alpha c(t) q}{\delta n} \tau(n-i, \Delta t). \tag{5.5}$$

The trader should continue with TWAP if

$$\lambda_2(t^-) - \lambda_1(t^-) < \frac{\alpha c(t) q}{\delta n} \tau(n-i, \Delta t).$$

TWAP analysis for the logarithmic impact function in (3.6) is much harder to do analytically, as formulae like (5.3) and (5.4) are not easily obtained due to nonlinearity. However, we can apply the linear quasi-TWAP exit trigger from (5.5), and through simulation we can see improvement.

5.5. Numerical simulations We use numerical simulations to check the effectiveness of the TWAP and quasi-TWAP strategies when compared to simpler strategies. The metrics used to compare the performance of the strategies are the expected average executed price and the expected execution time, when the executed prices between the strategies are similar. The following three strategies are used in our analysis.

- *One order strategy*: send one *buy*(*sell*) order of size q at time $t = 0$.
- TWAP *strategy*: send n *buy*(*sell*) orders of size q/n at time $t = 0, \Delta t, 2\Delta t, \ldots, (n-1)\Delta t$.
- *Quasi*-TWAP *strategy*: send *buy*(*sell*) orders of size q/n following the rules defined in Section 5.4.

We ignore the broker and exchange fees (k) and the bid–ask spread (θ) since they have the same effect on the execution price for both strategies. Moreover, we assume that at time $t = 0$ there is no OFI , that is, $\lambda_2(0) = \lambda_1(0)$. Lastly, we assume the linear impact function has $c(t)$ that is constant during the execution of the entire TWAP order, that is, $c(t)$ is constant from $t = 0$ to $t = (n-1)\Delta t$. An example of a price-path simulation with the exogenous effect of a TWAP order, assuming a linear impact function, is shown in Figure 6.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

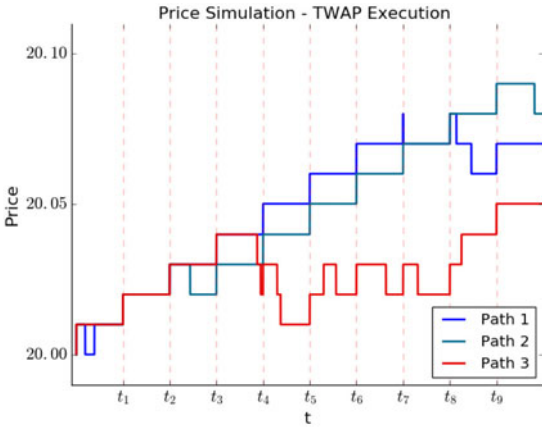

FIGURE 6. Three price paths simulation using the parameters $\mu = 0.008$, $\alpha = 0.004$, $\beta = 0.008$, $S(0) = 20.0$, $c(t) = 0.04/50\,000$, $\delta = 0.01$ with the exogenous effect of 10 orders, each of size 10 000, sent at times $t = 0, t_1, t_2, \ldots, t_9$.

SIMULATION 5.6 (Linear impact function). In the first simulation, we assume a linear price impact given in equation (3.5). Tables 1 and 3 show the expected average executed price and expected execution time of the strategies for different parameters α and β. The simulations in Table 1 show for all α, β and linear price impact that the TWAP strategy has a lower expected average price for the *buy* case than a one-order strategy. Table 3 shows that the quasi-TWAP strategy has a similar expected average executed price than the TWAP but less market exposure (measured by the expected execution time) due to the rule of early termination.

SIMULATION 5.7 (Logarithm-based impact function). In the second simulation, we assume a logarithm price impact given by (3.6). Now, the price impact is stochastic and a function of OFI, and therefore it is not as obvious as in the linear case that the TWAP and quasi-TWAP strategies will have a better expected average execution price than the alternative of sending only one large market order. The results of the simulations are similar to the results in the linear case. Table 2 shows the expected average executed price for the TWAP strategy for different parameters α and β. For most α and β in the simulation, the TWAP strategy has a lower expected average price for the *buy* case than a one-order strategy. The parameterization for which a one-order strategy is almost the same as TWAP or quasi-TWAP, is when α is very small, that is, the parameterizations where the Hawkes intensities are least impacted by trading. Analogously to the conclusions in the linear case, Table 4 shows that the quasi-TWAP strategy has an expected average price similar to the TWAP, but it has a lower expected execution time.

5.6. TWAP with linear impact and latent Hawkes intensity Finally, we consider the issue of latency in λ_1 and λ_2, and, using simulation, we test the effectiveness of

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

TABLE 1. The expected executed price for the TWAP strategy and one-order strategy using a linear impact function for 50 000 simulations, $\lambda_1(0) = \lambda_2(0) = 0.15$, $\mu = 0.1$, $q = 100\,000$, $n = 10$, $S(0) = 20.0$, $c(t) = 0.04/50\,000$, side = *buy*, $\delta = 0.01$.

Parameter α	Parameter β	One order execution	TWAP $\Delta t = 5$ s	TWAP $\Delta t = 15$ s	TWAP $\Delta t = 30$ s	Best TWAP
0.001	0.005	20.04	20.0394	20.0385	20.0372	20.0333
0.001	0.01	20.04	20.0394	20.0387	20.0382	20.0364
0.001	0.02	20.04	20.0396	20.039	20.0389	20.0381
0.001	0.05	20.04	20.0397	20.0393	20.0393	20.0392
0.001	0.1	20.04	20.0397	20.0395	20.0395	20.0396
0.001	0.2	20.04	20.0399	20.0399	20.0399	20.0398
0.001	0.5	20.04	20.04	20.04	20.04	20.0399
0.005	0.01	20.04	20.0373	20.0341	20.0318	20.0267
0.005	0.02	20.04	20.0376	20.0353	20.0336	20.032
0.005	0.05	20.04	20.0383	20.0371	20.0371	20.0364
0.005	0.1	20.04	20.0386	20.0383	20.0385	20.0381
0.005	0.2	20.04	20.0392	20.0392	20.0391	20.039
0.005	0.5	20.04	20.0396	20.0396	20.0395	20.0395
0.01	0.02	20.04	20.0354	20.0315	20.03	20.0267
0.01	0.05	20.04	20.0363	20.0349	20.0344	20.0333
0.01	0.1	20.04	20.0376	20.0372	20.0368	20.0364
0.01	0.2	20.04	20.0386	20.0385	20.0387	20.0381
0.01	0.5	20.04	20.0393	20.0393	20.0395	20.0392
0.02	0.05	20.04	20.0336	20.0309	20.0304	20.0286
0.02	0.1	20.04	20.0354	20.0344	20.0342	20.0333
0.02	0.2	20.04	20.037	20.0369	20.0368	20.0364
0.02	0.5	20.04	20.0387	20.0385	20.0383	20.0383
0.05	0.1	20.04	20.0303	20.0286	20.0283	20.0267
0.05	0.2	20.04	20.0335	20.0329	20.0326	20.032
0.05	0.5	20.04	20.0367	20.0367	20.0367	20.0364
0.1	0.2	20.04	20.0289	20.0281	20.028	20.0267
0.1	0.5	20.04	20.034	20.0338	20.0342	20.0333
0.2	0.5	20.04	20.0302	20.0297	20.0299	20.0286

TWAP and quasi-TWAP under this added uncertainty. Section 3.1 suggests that the λs would need to be filtered in practice, and so this section considers a model where $N_1(t)$ and $N_2(t)$ are observed but λ_1 and λ_2 are unobserved with additive noise.

The model we consider is the one studied by Vacarescu [34]. It is similar in its mutually exciting behaviour, but with a Brownian motion continuously adding movement to the intensity,

$$\begin{pmatrix}\lambda_1(t)\\ \lambda_2(t)\end{pmatrix} = \mu + \int_{-\infty}^{t} e^{-\beta(t-s)}\left(\alpha\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}dN_1(s)\\ dN_2(s)\end{pmatrix} + \sigma\begin{pmatrix}dW_1(s)\\ dW_2(s)\end{pmatrix}\right),$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

TABLE 2. The expected executed price for the TWAP strategy and one-order strategy using a logarithm impact function for 50 000 simulations, $\lambda_1(0) = \lambda_2(0) = 0.15$, $\mu = 0.1$, $q = 100\,000$, $n = 10$, $S(0) = 20.0$, $c(t) = 0.04/50\,000$, side = *buy*, $\delta = 0.01$, $b = 1$.

Parameter α	Parameter β	One order execution	TWAP $\Delta t = 5$ s	TWAP $\Delta t = 15$ s	TWAP $\Delta t = 30$ s
0.001	0.005	20.039	20.0384	20.0377	20.0367
0.001	0.01	20.039	20.0386	20.0378	20.0376
0.001	0.02	20.039	20.0388	20.0383	20.0379
0.001	0.05	20.039	20.0389	20.0391	20.0395
0.001	0.1	20.039	20.0392	20.0393	20.0394
0.001	0.2	20.039	20.0395	20.0397	20.0397
0.001	0.5	20.039	20.0398	20.0397	20.0397
0.005	0.01	20.039	20.0365	20.0336	20.0314
0.005	0.02	20.039	20.0369	20.035	20.0339
0.005	0.05	20.039	20.0376	20.0366	20.0368
0.005	0.1	20.039	20.0383	20.0384	20.0383
0.005	0.2	20.039	20.0389	20.0389	20.0388
0.005	0.5	20.039	20.0395	20.0396	20.0394
0.01	0.02	20.039	20.0347	20.0315	20.0298
0.01	0.05	20.039	20.036	20.0348	20.0344
0.01	0.1	20.039	20.0372	20.0366	20.0367
0.01	0.2	20.039	20.0381	20.0378	20.0385
0.01	0.5	20.039	20.0391	20.0393	20.0396
0.02	0.05	20.039	20.0331	20.0307	20.0303
0.02	0.1	20.039	20.0352	20.0342	20.0341
0.02	0.2	20.039	20.0369	20.0367	20.0365
0.02	0.5	20.039	20.0385	20.0386	20.0381
0.05	0.1	20.039	20.03	20.0285	20.0282
0.05	0.2	20.039	20.0335	20.0328	20.0329
0.05	0.5	20.039	20.0367	20.0367	20.0364
0.1	0.2	20.039	20.0287	20.028	20.0283
0.1	0.5	20.039	20.034	20.0341	20.0337
0.2	0.5	20.039	20.0299	20.0296	20.0296

where σ is a scalar noise parameter, W_1 and W_2 are independent Brownian motions, and each of the increments $dN_1(t)$ and $dN_2(t)$ is Poisson with intensity equal to the positive part of λ,

$$dN_1(t) \sim \text{Poisson}(\lambda_1^+(t)\,dt) \quad \text{and} \quad dN_2(t) \sim \text{Poisson}(\lambda_2^+(t)\,dt)$$

with $N_1(0) = N_2(0) = 0$ and $\lambda^+ = \max(\lambda, 0)$. Letting $\mathcal{G}_t$ denote the σ-algebra generated by the observable processes $\{S(0^-), (N_1(s), N_2(s))_{s\le t}\}$, the expected average execution price for the TWAP needs to be conditional on $\mathcal{G}_t$ instead of the larger, fully informed

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

TABLE 3. The expected executed price and expected execution time for the quasi-TWAP strategy using a linear impact function for 50 000 simulations, $\lambda_1(0) = \lambda_2(0) = 0.15$, $\mu = 0.1$, $q = 100\,000$, $n = 10$, $S(0) = 20.0$, $c(t) = 0.04/50\,000$, side = *buy*, $\delta = 0.01$.

Parameter α	Parameter β	Quasi-TWAP $\Delta t = 5$ s		Quasi-TWAP $\Delta t = 15$ s		Quasi-TWAP $\Delta t = 30$ s	
0.001	0.005	20.0394	43.85 s	20.0384	113.65 s	20.0372	194.48 s
0.001	0.01	20.0395	43.88 s	20.0387	114.86 s	20.0377	199.83 s
0.001	0.02	20.0395	43.94 s	20.0389	117.84 s	20.0385	208.24 s
0.001	0.05	20.0395	44.26 s	20.0394	122.96 s	20.0391	232.02 s
0.001	0.1	20.0398	44.47 s	20.0397	127.35 s	20.0398	250.0 s
0.001	0.2	20.04	44.49 s	20.0399	131.08 s	20.0398	261.43 s
0.001	0.5	20.04	44.68 s	20.0398	133.78 s	20.0401	267.59 s
0.005	0.01	20.0372	43.65 s	20.0337	111.47 s	20.0307	188.75 s
0.005	0.02	20.0376	43.74 s	20.0351	115.03 s	20.0337	200.95 s
0.005	0.05	20.0381	44.12 s	20.0372	121.66 s	20.0368	230.36 s
0.005	0.1	20.0387	44.37 s	20.0384	127.0 s	20.0383	249.73 s
0.005	0.2	20.0393	44.46 s	20.039	130.99 s	20.039	261.44 s
0.005	0.5	20.0396	44.67 s	20.0394	133.77 s	20.0397	267.52 s
0.01	0.02	20.0353	43.42 s	20.0313	111.17 s	20.0287	190.52 s
0.01	0.05	20.0364	43.89 s	20.0347	119.95 s	20.0341	227.2 s
0.01	0.1	20.0375	44.23 s	20.0369	126.52 s	20.0366	249.2 s
0.01	0.2	20.0385	44.41 s	20.0383	130.93 s	20.0383	261.44 s
0.01	0.5	20.0394	44.67 s	20.0391	133.76 s	20.0393	267.52 s
0.02	0.05	20.0334	43.32 s	20.0307	115.75 s	20.0297	219.42 s
0.02	0.1	20.0353	43.89 s	20.0343	125.59 s	20.0339	248.38 s
0.02	0.2	20.0371	44.32 s	20.0368	130.89 s	20.0366	261.28 s
0.02	0.5	20.0386	44.66 s	20.0385	133.76 s	20.0386	267.54 s
0.05	0.1	20.03	42.51 s	20.0282	120.48 s	20.0279	239.68 s
0.05	0.2	20.0334	43.92 s	20.033	130.32 s	20.0328	260.56 s
0.05	0.5	20.0368	44.63 s	20.0367	133.74 s	20.0367	267.43 s
0.1	0.2	20.0287	42.88 s	20.0281	128.25 s	20.0281	257.1 s
0.1	0.5	20.0341	44.57 s	20.0341	133.68 s	20.034	267.39 s
0.2	0.5	20.0299	44.36 s	20.0297	133.37 s	20.03	266.8 s

σ-algebra $\mathcal{F}_t$,

Avg Exec Price n Orders(t^-, q)

$$= \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\Big[S((t+i\Delta t)^-)|\mathcal{G}_{t^-} \vee \bigvee_{j=0}^{i}\{\psi((t+j\Delta t)^-, q/n)\}\Big]$$

$$+ \frac{1}{q}\sum_{i=0}^{n-1}\int_0^{q/n}\mathbb{E}\Big[\psi(t+i\Delta t, v)|\mathcal{G}_{t^-} \vee \bigvee_{j=0}^{i-1}\{\psi((t+j\Delta t)^-, q/n)\}\Big]dv + \operatorname{sgn}(q)\frac{\theta}{2}.$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

TABLE 4. The expected executed price and expected execution time for the quasi-TWAP strategy using a logarithm impact function for 50 000 simulations, $\lambda_1(0) = \lambda_2(0) = 0.15$, $\mu = 0.1$, $q = 100\,000$, $n = 10$, $S(0) = 20.0$, $c(t) = 0.04/50\,000$, side = *buy*, $\delta = 0.01$, $b = 1$.

Parameter α	Parameter β	Quasi-TWAP $\Delta t = 5$ s		Quasi-TWAP $\Delta t = 15$ s		Quasi-TWAP $\Delta t = 30$ s	
0.001	0.005	20.0384	43.83 s	20.0376	113.47 s	20.0364	193.69 s
0.001	0.01	20.0386	43.85 s	20.038	114.71 s	20.0374	199.95 s
0.001	0.02	20.0387	43.9 s	20.0384	117.78 s	20.0383	208.14 s
0.001	0.05	20.039	44.26 s	20.0392	122.76 s	20.0394	231.76 s
0.001	0.1	20.0393	44.44 s	20.0395	127.22 s	20.0395	249.83 s
0.001	0.2	20.0397	44.5 s	20.0397	131.01 s	20.0399	261.42 s
0.001	0.5	20.0397	44.67 s	20.0398	133.79 s	20.0396	267.54 s
0.005	0.01	20.0365	43.65 s	20.0333	111.62 s	20.0305	188.54 s
0.005	0.02	20.0368	43.72 s	20.0346	114.99 s	20.0334	200.69 s
0.005	0.05	20.0376	44.1 s	20.0369	121.68 s	20.0366	229.91 s
0.005	0.1	20.0383	44.36 s	20.0381	127.05 s	20.0379	249.54 s
0.005	0.2	20.039	44.46 s	20.0388	131.02 s	20.0391	261.52 s
0.005	0.5	20.0395	44.67 s	20.0395	133.73 s	20.0395	267.54 s
0.01	0.02	20.0346	43.39 s	20.0309	111.32 s	20.0285	190.24 s
0.01	0.05	20.0359	43.86 s	20.0345	119.79 s	20.034	227.39 s
0.01	0.1	20.037	44.22 s	20.0367	126.56 s	20.0367	249.39 s
0.01	0.2	20.0382	44.4 s	20.0383	131.01 s	20.0381	261.39 s
0.01	0.5	20.0391	44.67 s	20.0392	133.78 s	20.0392	267.55 s
0.02	0.05	20.033	43.32 s	20.0305	115.65 s	20.0295	218.79 s
0.02	0.1	20.035	43.89 s	20.034	125.54 s	20.0339	248.09 s
0.02	0.2	20.0369	44.3 s	20.0366	130.87 s	20.0366	261.37 s
0.02	0.5	20.0384	44.66 s	20.0382	133.75 s	20.0383	267.56 s
0.05	0.1	20.0297	42.5 s	20.0282	120.52 s	20.0279	239.73 s
0.05	0.2	20.0331	43.9 s	20.0328	130.34 s	20.0327	260.62 s
0.05	0.5	20.0365	44.63 s	20.0365	133.74 s	20.0366	267.49 s
0.1	0.2	20.0287	42.88 s	20.028	128.23 s	20.0281	257.27 s
0.1	0.5	20.034	44.57 s	20.0339	133.65 s	20.034	267.37 s
0.2	0.5	20.0297	44.36 s	20.0296	133.36 s	20.0297	266.72 s

For the linear impact function, the effects of filtering are minimal, as the TWAP effective criterion set forth in (5.3) is simply taken to be the filtered version

$$\mathbb{E}[\lambda_2(t^-) - \lambda_1(t^-)|\mathcal{G}_{t^-}] \leq \frac{\alpha c(t) q}{\delta n} \tau(n, \Delta t), \tag{5.6}$$

where $\tau(n, \Delta t)$ is the same as in Proposition 5.5. Following Vacarescu's work [34], the filtered intensities are calculated with a linear equation in between TWAP execution

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

TABLE 5. The expected executed price and expected execution time for the quasi-TWAP strategy using a linear impact function and the filter for 50 000 simulations, $\lambda_1(0) = \lambda_2(0) = 0.15$, $\mu = 0.1$, $q = 100\,000$, $n = 10$, $S(0) = 20.0$, $c(t) = 0.04/50\,000$, side = *buy*, $\delta = 0.01$, $b = 1$.

Parameter α	Parameter β	Parameter σ	Quasi-TWAP $\Delta t = 5$ s		Quasi-TWAP $\Delta t = 15$ s		Quasi-TWAP $\Delta t = 30$ s	
0.01	0.05	0.001	20.0366	43.89 s	20.0348	119.96 s	20.0341	227.3 s
0.01	0.05	0.005	20.0366	43.86 s	20.0348	119.5 s	20.0342	225.84 s
0.01	0.05	0.01	20.0366	43.75 s	20.035	118.19 s	20.0346	222.09 s
0.01	0.05	0.05	20.0383	41.72 s	20.0398	101.02 s	20.0412	175.45 s
0.01	0.1	0.001	20.0375	44.22 s	20.0369	126.63 s	20.0368	249.55 s
0.01	0.1	0.005	20.0375	44.21 s	20.0369	126.57 s	20.0368	249.05 s
0.01	0.1	0.01	20.0375	44.18 s	20.0369	126.16 s	20.0368	248.58 s
0.01	0.1	0.05	20.0384	43.0 s	20.0385	117.47 s	20.0387	227.8 s
0.01	0.5	0.001	20.0394	44.7 s	20.0395	133.91 s	20.0393	267.75 s
0.01	0.5	0.005	20.0393	44.7 s	20.0394	133.88 s	20.0393	267.72 s
0.01	0.5	0.01	20.0394	44.69 s	20.0391	133.86 s	20.0396	267.77 s
0.01	0.5	0.05	20.0393	44.66 s	20.0393	133.72 s	20.0393	267.43 s
0.05	0.1	0.001	20.0301	42.58 s	20.0282	120.75 s	20.0279	240.3 s
0.05	0.1	0.005	20.03	42.52 s	20.0283	120.51 s	20.0278	240.04 s
0.05	0.1	0.01	20.0301	42.53 s	20.0284	120.57 s	20.0279	239.86 s
0.05	0.1	0.05	20.0312	41.55 s	20.0302	114.34 s	20.0301	224.33 s
0.05	0.5	0.001	20.0368	44.66 s	20.0366	133.87 s	20.0368	267.71 s
0.05	0.5	0.005	20.0368	44.66 s	20.037	133.87 s	20.0368	267.76 s
0.05	0.5	0.01	20.0368	44.66 s	20.0368	133.85 s	20.0368	267.72 s
0.05	0.5	0.05	20.0369	44.62 s	20.0369	133.72 s	20.0371	267.39 s
0.1	0.5	0.001	20.0343	44.6 s	20.0341	133.8 s	20.0341	267.64 s
0.1	0.5	0.005	20.0342	44.6 s	20.0341	133.84 s	20.0343	267.62 s
0.1	0.5	0.01	20.0342	44.61 s	20.0341	133.82 s	20.0342	267.59 s
0.1	0.5	0.05	20.0342	44.59 s	20.0343	133.66 s	20.0341	267.4 s

times,

$$d\mathbb{E}[\lambda_2(t) - \lambda_1(t)|\mathcal{G}_t] = -\beta\mathbb{E}[\lambda_2(t^-) - \lambda_1(t^-)|\mathcal{G}_{t^-}] + \alpha(dN_1(t) - dN_2(t)), \tag{5.7}$$

and for linear impact function $\psi(t, q) = c(t)q$, at the time of a TWAP trade the filter is adjusted by the amount that the trade will affect intensities, namely,

$$\mathbb{E}[\lambda_2(t) - \lambda_1(t)|\mathcal{G}_t] = \mathbb{E}[\lambda_2(t^-) - \lambda_1(t^-)|\mathcal{G}_{t^-}] + \frac{\alpha}{\delta}c(t)\frac{q}{n},$$

at an execution time. Table 5 shows the simulation results for the filtered quasi-TWAP for linear impact function using the criterion set forth in equation (5.6).

We do not go further and address latency for the log impact function. The reason is because $\lambda_2(t) - \lambda_1(t)$ is observable at the times of TWAP trades for $\psi(t, q)$ given by equation (3.6). In this case trading would reveal the latent state, thereby resetting the filter's variance to zero, which does not seem quite realistic. Instead, it would perhaps be better to have the impact function take $\lambda_2(t) - \lambda_1(t)+$ "noise", for example, so that the process remains latent. In addition, the linear filter of (5.7) is not enough

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

for the nonlinear impact functions, but rather the entire posterior distribution function of $\lambda_2(t) - \lambda_1(t)$ is needed to evaluate the posterior expected price impact's contribution to the posterior. Hence, log impact and general concave ψ can be done but will require nonlinear filtering.

6. Conclusion

Using the model proposed by Bacry et al. [7] for the tick-by-tick asset price, this paper has analysed the price dynamic after the order book receives one or more large orders. Both linear and logarithmic impact functions are proposed; the logarithmic function is consistent with other studies which found execution prices to have sub-linear dependence on trading volume [3, 12, 35]. In the logarithmic case, the impact function has a dependence on the order-flow imbalance, which is a property that is also consistent with the literature. We show that for some specific strategies, price manipulation is present in specific market conditions determined by the parameters $\lambda_2(t) - \lambda_1(t)$, and that this arbitrage opportunity will vanish if traders exploit it with orders of growing size.

In Section 5 we study the particular case of the TWAP strategy. Our calculations show that this strategy can be effective for execution of large orders, and further evidence is provided in our numerical simulations. Using both logarithmic and linear impact, we notice that the TWAP strategy often has a better expected execution price than the strategy of sending only one market order at $t = 0$. In addition, we have proposed the quasi-TWAP strategy as a slight improvement to ordinary TWAP, with the improvement attributed to the strategy's consideration of order flow.

Interesting future work would include analysis of VWAP strategies, and calculation of optimal scheduled execution using the dynamic impact function proposed in this paper. The optimization under these conditions could lead to more robust solutions, since the dynamic impact function has the property of changing accordingly the incoming order flow.

Appendix A. Proofs

A.1. Proof of Proposition 3.11 The proposed model in equation (2.2) is driven by the following pair of stochastic differential equations:

$$d\lambda_1(t) = \beta(\mu - \lambda_1(t))\,dt + \alpha\,dN_2(t),$$
$$d\lambda_2(t) = \beta(\mu - \lambda_2(t))\,dt + \alpha\,dN_1(t).$$

The expected value of these two equations is given by

$$d\mathbb{E}\lambda_1(t) = \beta(\mu - \mathbb{E}\lambda_1(t))\,dt + \alpha\mathbb{E}\lambda_2(t)\,dt,$$
$$d\mathbb{E}\lambda_2(t) = \beta(\mu - \mathbb{E}\lambda_2(t))\,dt + \alpha\mathbb{E}\lambda_1(t)\,dt.$$

Taking the difference between the equations yields

$$\frac{d}{dt}\mathbb{E}[\lambda_2(t) - \lambda_1(t)] = -(\alpha + \beta)\mathbb{E}[\lambda_2(t) - \lambda_1(t)],$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

and the solution of the above differential equation is given by

$$\mathbb{E}[\lambda_2(T) - \lambda_1(T)|\mathcal{F}_{t^-}] = [\lambda_2(t^-) - \lambda_1(t^-)]e^{-(T-t)(\alpha+\beta)}.$$

Using this solution along with the impact of $\psi(t^-, q)$ at time t and using the impact relationships shown in equations (3.3) and (3.4), we have

$$\mathbb{E}[\lambda_2(T) - \lambda_1(T)|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}] = \left(\lambda_2(t^-) - \lambda_1(t^-) - \frac{\alpha}{\delta}\psi(t^-, q)\right)e^{-(T-t)(\alpha+\beta)},$$

where $T > t$. So, the expected price after t is

$$\begin{aligned}
&\mathbb{E}[S(T)|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}] \\
&\quad = S(t^-) + \psi(t^-, q) + \delta\mathbb{E}\left[\int_t^T (dN_2(u) - dN_1(u))\middle|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}\right] \\
&\quad = S(t^-) + \psi(t^-, q) + \delta\mathbb{E}\left[\int_t^T (\lambda_2(u) - \lambda_1(u))\,du\middle|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}\right] \\
&\quad = S(t^-) + \psi(t^-, q) + (\delta(\lambda_2(t^-) - \lambda_1(t^-)) - \alpha\psi(t^-, q))\int_t^T e^{-(u-t)(\alpha+\beta)}\,du \\
&\quad = S(t^-) + \psi(t^-, q) + \frac{(\delta(\lambda_2(t^-) - \lambda_1(t^-)) - \alpha\psi(t^-, q))(1 - e^{-(T-t)(\alpha+\beta)})}{\alpha+\beta}.
\end{aligned}$$

Thus, the permanent impact $P(q) = \psi(t^-, q)\beta/(\alpha+\beta)$, which completes the proof. □

A.2. Proof of Proposition 5.2 The proof of Proposition 3.11 showed the permanent impact

$$P(q) = \psi(t^-, q)\left(\frac{\beta}{\alpha+\beta}\right) \tag{A.1}$$

and the expected price after an order of size q as

$$\mathbb{E}[S(t+\Delta t)|\mathcal{F}_{t^-} \vee \{\psi(t^-, q)\}] - \mathbb{E}[S(t+\Delta t)|\mathcal{F}_{t^-}] = \psi(t^-, q) + \frac{\alpha\psi(t^-, q)}{\alpha+\beta}(e^{-\Delta t(\alpha+\beta)} - 1). \tag{A.2}$$

Subtracting equations (A.2) and (A.1) from each other, we get the objective in (5.1), which we seek to optimize over Δt. Evaluating the objective at any $\Delta t > 0$ yields

$$\Delta t \geq -\frac{1}{\alpha+\beta}\ln\left(\frac{\xi(1+\beta/\alpha)}{|\psi(t^-, q)|}\right),$$

which for threshold $\xi < |\psi(t^-, q)|/(1+\beta/\alpha)$ is maximized at

$$\Delta t^* = -\frac{1}{\alpha+\beta}\ln\left(\frac{\xi(1+\beta/\alpha)}{|\psi(t^-, q)|}\right).$$

It is possible that the choice for $\xi > |\psi(t^-, q)|/(1+\beta/\alpha)$ (that is, if the order size is not so large), in which case price impact does not pose a reason for waiting to trade.

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

A.3. Proof of Proposition 5.4 For a TWAP strategy of n orders at times $t_i = i\Delta t$ and order size $q/n > 0$, with linear impact function $\psi(q) = cq$, let us use the retracement function R in equation (5.2) to prove the statement of the proposition. For any $i > 0$, using the retracement in equation (5.2) and then proceeding inductively, the expectation of future retracements given the future n-order TWAP impacts is given by

$$\begin{aligned}
&\mathbb{E}\Big[S(t_{i+1}^-) - S(t_i^-)\Big|\mathcal{F}_{0^-} \vee \bigvee_{j=0}^{n-1}\{\psi(t_j^-, q/n)\}\Big] \\
&\quad = \frac{cq}{n} + \delta\Big(\frac{1-e^{-(\alpha+\beta)\Delta t}}{\alpha+\beta}\Big)\mathbb{E}\Big[\lambda_2(t_i^-) - \lambda_1(t_i^-) - \frac{acq}{\delta n}\Big|\mathcal{F}_{0^-} \vee \bigvee_{j=0}^{n-1}\{\psi(t_j^-, q/n)\}\Big] \\
&\quad = \frac{cq}{n} + \delta\Big(\frac{1-e^{-(\alpha+\beta)\Delta t}}{\alpha+\beta}\Big)\mathbb{E}\Big[\Big(\lambda_2(t_{i-1}^-) - \lambda_1(t_{i-1}^-) - \frac{acq}{\delta n}\Big)e^{-(\alpha+\beta)\Delta t} \\
&\qquad\quad - \frac{acq}{\delta n}\Big|\mathcal{F}_{0^-} \vee \bigvee_{j=0}^{n-1}\{\psi(t_j^-, q/n)\}\Big] \\
&\quad = \quad \vdots \\
&\quad = \frac{cq}{n} + \delta\Big(\frac{1-e^{-(\alpha+\beta)\Delta t}}{\alpha+\beta}\Big)\Big\{(\lambda_2(0^-) - \lambda_1(0^-))e^{-(\alpha+\beta)i\Delta t} - \frac{acq}{\delta n}\sum_{j=0}^{i} e^{-(\alpha+\beta)j\Delta t}\Big\}.
\end{aligned}$$

Summing the expected retracements yields expected total retracement up to time t_i^-,

$$\begin{aligned}
&\mathbb{E}\Big[S(t_i^-) - S(0^-)\Big|\mathcal{F}_{0^-} \vee \bigvee_{j=0}^{n-1}\{\psi(t_j^-, q/n)\}\Big] \\
&\quad = \sum_{j=0}^{i-1}\mathbb{E}\Big[S(t_{j+1}^-) - S(t_j^-)\Big|\mathcal{F}_{0^-} \vee \bigvee_{\ell=0}^{n-1}\{\psi(t_\ell^-, q/n)\}\Big] \\
&\quad = i\frac{cq}{n} + \delta\Big(\frac{1-e^{-(\alpha+\beta)\Delta t}}{\alpha+\beta}\Big)\sum_{j=0}^{i-1}\Big\{(\lambda_2(0^-) - \lambda_1(0^-))e^{-(\alpha+\beta)j\Delta t} - \frac{acq}{\delta n}\sum_{\ell=0}^{j} e^{-(\alpha+\beta)\ell\Delta t}\Big\},
\end{aligned}$$

and using geometric series, the summations can be simplified as

$$\begin{aligned}
&\mathbb{E}\Big[S(t_i^-) - S(0^-)\Big|\mathcal{F}_{0^-} \vee \bigvee_{j=0}^{n-1}\{\psi(t_j^-, q/n)\}\Big] \\
&\quad = i\frac{cq}{n} + \frac{\delta}{\alpha+\beta}\Big\{(\lambda_2(0^-) - \lambda_1(0^-))(1 - e^{-(\alpha+\beta)i\Delta t}) - \frac{acq}{\delta n}\sum_{j=0}^{i-1}(1 - e^{-(\alpha+\beta)(j+1)\Delta t})\Big\}.
\end{aligned}$$

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

Next, the average execution price of the n-order TWAP can be calculated as

$$\begin{aligned}
&\text{Avg Exec Price } n \text{ Orders}(t^-, q)\\
&= \frac{c(t)q}{2n} + \operatorname{sgn}(q)\frac{\theta}{2} + S(0^-) + \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\Big[S(t_i^-) - S(0^-)\Big|\mathcal{F}_{0^-} \vee \bigvee_{j=0}^{n-1}\{\psi(t_j^-, q/n)\}\Big]\\
&= \frac{c(t)q}{2n} + \operatorname{sgn}(q)\frac{\theta}{2} + S(0^-) + \frac{cq}{n^2}\sum_{i=0}^{n-1} i\\
&\quad + \frac{\delta}{n(\alpha+\beta)}\sum_{i=0}^{n-1}\Big\{(\lambda_2(0^-) - \lambda_1(0^-))(1 - e^{-(\alpha+\beta)i\Delta t})\\
&\quad - \frac{acq}{\delta n}\sum_{j=0}^{i-1}(1 - e^{-(\alpha+\beta)(j+1)\Delta t})\Big\}\\
&= \text{Exec Price 1 Order}(t^-, q)\\
&\quad + \delta\frac{\lambda_2(0^-) - \lambda_1(0^-)}{n(\alpha+\beta)}\sum_{i=0}^{n-1}(1 - e^{-(\alpha+\beta)i\Delta t}) - \frac{acq}{n^2(\alpha+\beta)}\sum_{i=0}^{n-1}\sum_{j=0}^{i-1}(1 - e^{-(\alpha+\beta)(j+1)\Delta t}).
\end{aligned}$$

The double sum can be simplified to a single summation,

$$\begin{aligned}
&\text{Avg Exec Price } n \text{ Orders}(t^-, q) = \text{Exec Price 1 Order}(t^-, q)\\
&+ \delta\frac{\lambda_2(0^-) - \lambda_1(0^-)}{n(\alpha+\beta)}\sum_{i=0}^{n-1}(1 - e^{-(\alpha+\beta)i\Delta t}) - \frac{acq}{n^2(\alpha+\beta)}\sum_{i=0}^{n-1}\Big(i - e^{-(\alpha+\beta)\Delta t}\frac{1 - e^{-(\alpha+\beta)i\Delta t}}{1 - e^{-(\alpha+\beta)\Delta t}}\Big),
\end{aligned}$$

and then geometric series applied once again to have summations completely removed:

$$\begin{aligned}
&\text{Avg Exec Price } n \text{ Orders}(t^-, q)\\
&\quad = \text{Exec Price 1 Order}(t^-, q)\\
&\qquad + \delta\frac{\lambda_2(0^-) - \lambda_1(0^-)}{n(\alpha+\beta)}\Big(n - \frac{1 - e^{-(\alpha+\beta)n\Delta t}}{1 - e^{-(\alpha+\beta)\Delta t}}\Big)\\
&\qquad - \frac{acq}{n^2(\alpha+\beta)}\Big(\frac{n(n-1)}{2} - \frac{ne^{-(\alpha+\beta)\Delta t}}{1 - e^{-(\alpha+\beta)\Delta t}} + \frac{1 - e^{-(\alpha+\beta)n\Delta t}}{(1 - e^{-(\alpha+\beta)\Delta t})^2}e^{-(\alpha+\beta)\Delta t}\Big).
\end{aligned}$$

From this it follows that the TWAP strategy has better expected average execution price than a single order if and only if equation (5.3) holds. In particular,

$$\text{Avg Exec Price } n \text{ Orders}(t^-, q) \le \text{Exec Price 1 Order}(t^-, q),$$

if the TWAP condition holds, that is,

$$\begin{aligned}
&\lambda_2(0^-) - \lambda_1(0^-) \le \frac{\alpha cq}{n\delta}\\
&\quad \times \Big[\frac{(n-1)(1 - e^{-(\alpha+\beta)\Delta t})/2 - e^{-(\alpha+\beta)\Delta t} + e^{-(\alpha+\beta)\Delta t}(1 - e^{-(\alpha+\beta)n\Delta t})/n(1 - e^{-(\alpha+\beta)\Delta t})}{1 - e^{-(\alpha+\beta)\Delta t} - (1 - e^{-(\alpha+\beta)n\Delta t})/n}\Big].
\end{aligned}$$

This completes the proof of Proposition 5.4. □

https://doi.org/10.1017/S1446181119000038 Published online by Cambridge University Press

References

[1] A. Alfonsi and P. Blanc, "Dynamic optimal execution in a mixed-market-impact Hawkes price model", *Finance Stoch.* **20** (2016) 183–218; doi:10.1007/s00780-015-0282-y.

[2] R. Almgren and N. Chriss, "Optimal execution of portfolio transactions", *J. Risk* **3** (2001) 5–39; doi:10.21314/JOR.2001.041.

[3] R. Almgren, C. Thum, E. Hauptmann and H. Li, "Direct estimation of equity market impact", *Risk* **18** (2005) 58–62; doi:10.1.1.146.1241.

[4] M. Avellaneda, *Algorithmic and high-frequency trading: an overview* (Quant Congress, USA, 2011); (Retrieved May 2013); https://www.math.nyu.edu/faculty/avellane/QuantCongressUSA2011AlgoTradingLAST.pdf.

[5] M. Avellaneda and S. Stoikov, "High-frequency trading in a limit order book", *Quant. Finance* **8** (2008) 217–224; doi:10.1080/14697680701381228.

[6] E. Bacry, K. Dayri and J. F. Muzy, "Non-parametric kernel estimation for symmetric Hawkes processes application to high frequency financial data", *Eur. Phys. J. B* **85** (2012) 1–12; doi:10.1140/epjb/e2012-21005-8.

[7] E. Bacry, S. Delattre, M. Hoffmann and J. F. Muzy, "Modelling microstructure noise with mutually exciting point processes", *Quant. Finance* **13** (2013) 65–77; doi:10.1080/14697688.2011.647054.

[8] E. Bacry, S. Delattre, M. Hoffmann and J. F. Muzy, "Some limit theorems for Hawkes processes and application to financial statistics", *Stochastic Process. Appl.* **123** (2013) 2475–2499; doi:10.1016/j.spa.2013.04.007.

[9] E. Bacry, I. Mastromatteo and J. F. Muzy, "Hawkes processes in finance", *Market Microstructure and Liquidity* **1** (2015) ID:1550005; doi:10.1142/S2382626615500057.

[10] E. Bacry and J. F. Muzy, "Hawkes model for price and trades high-frequency dynamics", *Quant. Finance* **14** (2014) 1147–1166; doi:10.1080/14697688.2014.897000.

[11] K. Bechler and M. Ludkovski, "Optimal execution with dynamic order flow imbalance", *SIAM J. Financial Math.* **6** (2015) 1123–1151; doi:10.1137/140992254.

[12] J. P. Bouchaud, "Price impact", in: *Encyclopedia of quantitative finance* (John Wiley and Sons, 2010); doi:10.1002/9780470061602.eqf18006.

[13] R. Carmona and K. Webster, "High frequency market making", Preprint, 2012, arXiv:1210.578.

[14] A. Cartea, S. Jaimungal and J. Ricci, "Buy low, sell high: a high frequency trading perspective", *SIAM J. Financial Math.* **5** (2014) 415–444; doi:10.1137/130911196.

[15] R. Cont, "Statistical modeling of high-frequency financial data", *IEEE Signal Process. Mag.* **28** (2011) 16–25; doi:10.1109/MSP.2011.941548.

[16] R. Cont, A. Kukanov and S. Stoikov, "The price impact of order book events", *J. Financial Econ.* **12** (2014) 47–88; doi:10.2139/ssrn.1712822.

[17] J. Da Fonseca and R. Zaatour, "Hawkes process: fast calibration, application to trade clustering, and diffusive limit", *J. Futures Markets* **34** (2014) 548–579; doi:10.1002/fut.21644.

[18] J. Donier, J. Bonart, I. Mastromatteo and J. P. Bouchaud, "A fully consistent, minimal model for non-linear market impact", *Quant. Finance* **15** (2015) 1109–1121; doi:10.1080/14697688.2015.1040056.

[19] J. Gatheral, "No-dynamic-arbitrage and market impact", *Quant. Finance* **10** (2010) 749–759; doi:10.1080/14697680903373692.

[20] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes", *Biometrika* **58** (1971) 83–90; doi:10.2307/2334319.

[21] R. Haynes and J. Roberts, Automated trading in futures markets, CFTC White Paper (2015); https://www.cftc.gov/sites/default/files/idc/groups/public/@economicanalysis/documents/file/oce_automatedtrading.pdf.

[22] G. Huberman and W. Stanzl, "Price manipulation and quasi-arbitrage", *Econometrica* **72** (2004) 1247–1275; doi:10.1111/j.1468-0262.2004.00531.x.

[23] I. Karatzas and S. Shreve, *Brownian motion and stochastic, calculus*, 2nd edn (Springer, New York, 1998).

[24] A. Kyle, "Continuous auctions and insider trading", *Econometrica* **53** (1985) 1315–1335; doi:10.2307/1913210.
[25] P. Laub, T. Taimre and P. Pollett, "Hawkes processes", Preprint, 2015, arXiv:1507.02822.
[26] I. Mastromatteo, "Apparent impact: the hidden cost of one-shot trades", *J. Stat. Mech. Theory Exp.* **6** (2015) ID: P06022; doi:10.1088/1742-5468/2015/06/P06022.
[27] R. S. Miller and G. Shorter, "High frequency trading: overview of recent developments", *Congressional Res. Serv.* (2016) 1–15; https://digital.library.unt.edu/ark:/67531/metadc847719/.
[28] A. Obizhaeva and J. Wang, "Optimal trading strategy and supply/demand dynamics", *J. Finance Markets* **16** (2013) 1–32; doi:10.1016/j.finmar.2012.09.001.
[29] V. Plerou, P. Gopikrishnan, X. Gabaix and H. E. Stanley, "Quantifying stock-price response to demand fluctuations", *Phys. Rev. E* **66** (2002) ID: 027104; doi:10.1103/PhysRevE.66.027104.
[30] M. Pohl, A. Ristig, W. Schachermayer and L. Tangpi, "The amazing power of dimensional analysis: quantifying market impact", *Market Microstructure and Liquidity* **3** (2017) ID: 1850004; doi:10.1142/S2382626618500041.
[31] A. Shriyaev, *Probability*, 2nd edn (Springer, New York, 1996).
[32] L. C. G. Rogers and S. Singh, "The cost of illiquidity and its effects on hedging", *Math. Finance* **20** (2010) 597–615; doi:10.1111/j.1467-9965.2010.00413.x.
[33] E. Smith, J. D. Farmer, L. Gillemot and S. Krishnamurthy, "Statistical theory of the continuous double auction", *Quant. Finance* **3** (2003) 481–514; doi:10.1088/1469-7688/3/6/307.
[34] A. Vacarescu, "Filtering and parameter estimation for partially observed generalized Hawkes processes", Ph.D. Thesis, Stanford University, 2011; https://purl.stanford.edu/tc922qd0500.
[35] P. Weber and B. Rosenow, "Order book approach to price impact", *Quant. Finance* **5** (2005) 357–364; doi:10.1080/14697680500244411.