Market impact with multi-timescale liquidity

M. Benzaquen & J.-P. Bouchaud

Quantitative Finance, 2018
https://doi.org/10.1080/14697688.2018.1444279

Market impact with multi-timescale liquidity

M. BENZAQUEN*†‡ and J.-P. BOUCHAUD‡

†Ladhyx, UMR CNRS 7646, École Polytechnique, 91128 Palaiseau Cedex, France
‡Capital Fund Management, 23 rue de l'Université, 75007 Paris, France

(*Received 23 October 2017; accepted 16 February 2018; published online 8 June 2018*)

Finite-memory effects on the dynamics of the latent order book can be accounted for by allowing finite cancellation and decomposition rates within a continuous reaction-diffusion set-up

1. Introduction

Understanding the price formation mechanisms is undoubtably among the most exciting challenges of modern finance. *Market impact* refers to the way market participants' actions mechanically affect prices. Significant progress has been made in this direction during the past decades (Hasbrouck 2007, Bouchaud *et al*. 2008, Weber and Rosenow 2005, Bouchaud 2010). A notable breakthrough was the empirical discovery that the aggregate price impact of a meta-order§ is a concave function (approximately square-root) of its size Q (Grinold and Kahn 2000, Almgren *et al*. 2005, Tóth *et al*. 2011, Donier and Bonart 2015). In the recent past, so-called 'latent' order book models (Tóth *et al*. 2011, Mastromatteo *et al*. 2014a, 2014b, Donier *et al*. 2015) have proven to be a fruitful framework to theoretically address the question of market impact, among others.

As a precise mathematical incarnation of the latent order book idea, the zero-intelligence LLOB model of Donier *et al*. (2015) was successful at providing a theoretical underpinning to the square-root impact law. The LLOB model is based on a continuous mean field setting that leads to a set of reaction–diffusion equations for the dynamics of the latent bid and ask volume densities. In the infinite-memory limit (where the agents intentions, unless executed, stay in the latent book forever and there are no arrivals of new intentions), the latent order book becomes exactly linear and impact exactly square-root. Furthermore, this assumption leads to zero permanent impact of uninformed trades, and an inverse square-root decay of impact as a function of time. While the LLOB model is fully consistent mathematically, it suffers from at least two major difficulties when confronted with micro-data. First, a strict square-root law is only recovered in the limit where the execution rate m_0 of the meta-order is larger than the normal execution rate J of the market itself—whereas most meta-order impact data are in the opposite limit $m_0 \lesssim 0.1J$. Moreover, the regime $m_0 > J$ yields nearly deterministic impact trajectories

*Corresponding author. Email: michael.benzaquen@polytechnique.edu
§A 'meta-order' (or parent order) is a bundle of orders corresponding to a single trading decision. A meta-order is typically traded incrementally through a sequence of child orders.

that are clearly unrealistic (except for Bitcoin in the early days, see Donier and Bonart 2015). Second, the theoretical inverse square-root impact decay is too fast and leads to significant short time mean-reversion effects, not observed in real prices.

The aim of the present paper is to show that introducing different timescales for the renewal of liquidity allows one to cure both the above deficiencies. In view of the way financial markets operate, this step is very natural: agents are indeed expected to display a broad spectrum of timescales, from low-frequency institutional investors to High-Frequency Traders (HFT). We show that provided the execution rate m_0 is large compared to the low-frequency flow, but small compared to J, the impact of a meta-order crosses over from a linear behaviour at very small Q to a square-root law in a regime of Qs that can be made compatible with empirical data. We show that in the presence of a continuous, power-law distribution of memory times, the temporal decay of impact can be tuned to reconcile persistent order flow with diffusive price dynamics (often referred to as the *diffusivity puzzle*) (Bouchaud *et al.* 2004, 2008, Lillo and Farmer 2004). We argue that the permanent impact of uninformed trades is fixed by the slowest liquidity memory time, beyond which mean-reversion effects disappear. Interestingly, the *permanent* impact is found to be linear in the executed volume Q and independent of the trading rate, as dictated by no-arbitrage arguments.

Our paper is organized as follows. We first recall the LLOB model of Donier *et al.* (2015) in section 2. We then explore in section 3 the implications of finite cancellation and deposition rates (finite memory) in the reaction–diffusion equations, notably regarding permanent impact (section 4). We generalize the reaction–diffusion model to account for several deposition and cancellation rates. In particular, we analyse in section 5 the simplified case of a market with two sorts of agents: long memory agents with vanishing deposition and cancellation rates, and short memory high-frequency agents (somehow playing the role of market makers). Finally, we consider in section 6 the more realistic case of a continuous distribution of cancellation and deposition rates and show that such a framework provides an alternative way to solve the diffusivity puzzle (see Benzaquen and Bouchaud 2018) by adjusting the distribution of cancellation and deposition rates. Many details of the calculations are provided in the appendices.

2. Locally linear order book model

We here briefly recall the main ingredients of the locally linear order book (LLOB) model as presented by Donier *et al.* (2015). In the continuous 'hydrodynamic' limit, we define the latent volume densities of limit orders $\varphi_b(x,t)$ (bid side) and $\varphi_a(x,t)$ (ask side) in the reference frame of the fair price† at relative position x and time t. The latent volume densities obey the following set of partial differential equations:

$$\partial_t \varphi_b = D\partial_{xx}\varphi_b - \nu\varphi_b + \lambda\Theta(x_t - x) - R_{ab}(x) \quad (1a)$$

$$\partial_t \varphi_a = D\partial_{xx}\varphi_a - \nu\varphi_a + \lambda\Theta(x - x_t) - R_{ab}(x)\,, \quad (1b)$$

where the different contributions on the right-hand side respectively signify (from left to right): heterogeneous reassessments of agents intentions with diffusivity D (diffusion terms), cancellations with rate ν (death terms), arrivals of new intentions with intensity λ (deposition terms), and matching of buy/sell intentions (reaction terms). The relative transaction price x_t is conventionally defined through the equation $\varphi_b(x_t,t) = \varphi_a(x_t,t)$. The non-linearity arising from the reaction term in equations (1a) and (1b) can be abstracted away by defining $\phi(x,t) = \varphi_b(x,t) - \varphi_a(x,t)$, which solves:

$$\partial_t \phi = D\partial_{xx}\phi - \nu\phi + s(x,t)\,, \quad (2)$$

where the source term reads $s(x,t) = \lambda\,\mathrm{sign}(x_t - x)$ and the price x_t is defined as the solution of:

$$\phi(x_t,t) = 0\,. \quad (3)$$

Setting $\xi = x - x_t$, the stationary order book can easily be obtained as: $\phi^{st}(\xi) = -(\lambda/\nu)\,\mathrm{sign}(\xi)[1 - \exp(-|\xi|/\xi_c)]$ where $\xi_c = \sqrt{D\nu^{-1}}$ denotes the typical length scale below which the order book can be considered to be linear: $\phi^{st}(\xi) \approx -\mathcal{L}\xi$ (see figure 1). The slope $\mathcal{L} := \lambda/\sqrt{\nu D}$ defines the *liquidity* of the market, from which the total execution rate J can be computed since:

$$J := \partial_\xi \phi^{st}(\xi)\big|_{\xi=0} = D\mathcal{L}. \quad (4)$$

Donier *et al.* (2015) focused on the *infinite memory* limit, namely $\nu, \lambda \to 0$ while keeping $\mathcal{L} \sim \lambda\nu^{-1/2}$ constant, such that the latent order book becomes exactly linear since in that limit $\xi_c \to \infty$. This limit considerably simplifies the mathematical analysis, in particular concerning the impact of a meta-order. An important remark must however be introduced at this point: although the limit $\nu \to 0$ is taken in Donier *et al.* (2015), it is assumed that the latent order book is still able to reach its stationary state $\phi^{st}(\xi)$ before a meta-order is introduced. In other words, the limit $\nu \to 0$ is understood in a way such that the starting time of the meta-order is large compared to ν^{-1}.

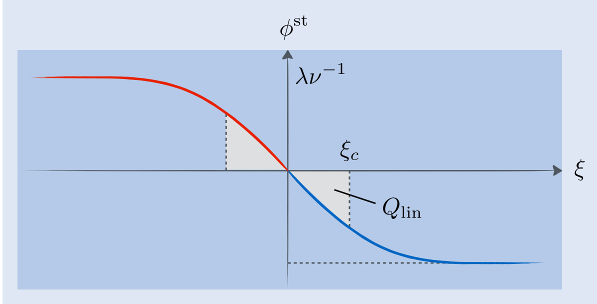

Figure 1. Stationary order book $\phi^{st}(\xi)$ as computed by Donier *et al.* (2015). The linear approximation holds up to $\xi_c = \sqrt{D\nu^{-1}}$ and the volume $Q_{lin.}$ of the grey triangles is of order $Q_{lin.} := \mathcal{L}\xi_c^2 = J\nu^{-1}$.

†The variable x denotes the reservation price relative to the 'fair' price $\hat{p}_t$ such that the true reservation price reads $p = \hat{p}_t + x$. We here assume that the fair price $\hat{p}_t$ encodes all informational aspects of prices and itself performs (on short time scales) an additive random walk with diffusivity coefficient D.

3. Price trajectories with finite cancellation and deposition rates

As mentioned in the introduction we here wish to explore the effects of non-vanishing cancellation and deposition rates, or said differently the behaviour of market impact for execution times comparable to or larger than ν^{-1}. The general solution of equation (2) is given by:

$$\phi(x,t) = (\mathcal{G}_\nu \star \phi_0)(x,t) + \int \mathrm{d}y \int_0^\infty \mathrm{d}\tau\, \mathcal{G}_\nu(x-y, t-\tau) s(y,\tau)\,, \quad (5)$$

where $\phi_0(x) = \phi(x,0)$ denotes the initial condition, and where $\mathcal{G}_\nu(x,t) = e^{-\nu t}\mathcal{G}(x,t)$ with $\mathcal{G}$ the diffusion kernel:

$$\mathcal{G}(x,t) = \Theta(t)\frac{e^{-\frac{x^2}{4Dt}}}{\sqrt{4\pi Dt}}\,. \quad (6)$$

Following Donier *et al.* (2015), we introduce a buy (sell) meta-order as an extra point-like source of buy (sell) particles with intensity rate m_t such that the source term in equation (2) becomes: $s(x,t) = m_t\delta(x-x_t)\cdot \mathbb{1}_{[0,T]} + \lambda\,\mathrm{sign}(x_t - x)$, where T denotes the time horizon of the execution. In all the following we shall focus on buy meta-orders—without loss of generality since within the present framework everything is perfectly symmetric. Performing the integral over space in equation (5) and setting $\phi_0(x) = \phi^{\mathrm{st}}(x)$ yields:

$$\phi(x,t) = \phi^{\mathrm{st}}(x)e^{-\nu t} + \int_0^{\min(t,T)} \mathrm{d}\tau\, m_\tau \mathcal{G}_\nu(x - x_\tau, t-\tau) - \lambda\int_0^t \mathrm{d}\tau\, \mathrm{erf}\left[\frac{x - x_\tau}{\sqrt{4D(t-\tau)}}\right] e^{-\nu(t-\tau)}\,. \quad (7)$$

The equation for price (3) is not analytically tractable in the general case, but different interesting limit cases can be investigated. In particular, focusing on the case of constant participation rates $m_t = m_0$, one may consider:

(i) Small participation rate $m_0 \ll J$ *vs* large participation rate $m_0 \gg J$.
(ii) Fast execution $\nu T \ll 1$ (the particules in the book are barely renewed during the meta-order execution) *vs* slow execution $\nu T \gg 1$ (the particles in the book are completely renewed, and the memory of the initial state has been lost).
(iii) Small meta-order volumes $Q := m_0 T \ll Q_{\mathrm{lin.}}$ (for which the linear approximation of the stationary book is appropriate, see figure 1) *vs* large volumes $Q \gg Q_{\mathrm{lin.}}$ (for which the linear approximation is no longer valid).

So in principle, one has to consider $2^3 = 8$ possible limit regimes. However, some regimes are mutually exclusive so that only six of them remain. A convenient way to summarize the results obtained for each of the limit cases mentioned above is to expand the price trajectory x_t up to first order in $\sqrt{\nu}$ as:†

$$x_t = \alpha\left[z_t^0 + \sqrt{\nu} z_t^1 + O(\nu)\right]\,, \quad (8)$$

where z_t^0 and z_t^1 denote, respectively, the zeroth-order and first-order contributions. Table 1 gathers the results for fast execution ($\nu T \ll 1$) and small meta-order volumes ($Q \ll Q_{\mathrm{lin.}}$). Note that the leading correction term z_t^1 is negative, i.e. the extra incoming flux of limit orders acts to lower the impact of the meta-order, see figure 2. The price trajectory for slow execution and/or large meta-order volumes, on the other hand, simply reads:

$$x_t = \frac{m_0\nu}{\lambda}t\,. \quad (9)$$

The corresponding calculations and explanations are given in appendix 1.

†Note that working at constant $\mathcal{L}$ implies $\lambda = O(\sqrt{\nu})$.

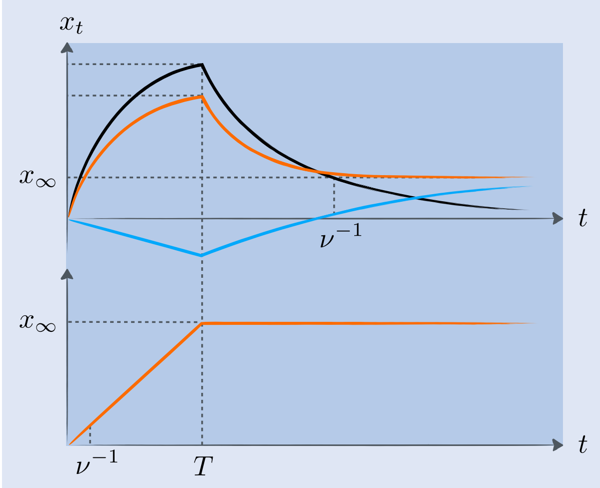

Figure 2. Top graph: Price trajectory during and after a buy meta-order execution for $\nu T \ll 1$. (Black curve) zeroth-order result from Donier *et al.* (2015). (Orange curve) first-order result. (Blue curve) first-order correction (see equation (8)). Bottom graph: Price trajectory for $\nu T \gg 1$. Note that the x-axis is not to scale since $\nu^{-1} \ll$ (resp. $\gg$) T.

4. Permanent impact as a finite-memory effect

As mentioned in the introduction, the impact relaxation following the execution is an equally important question. We here compute the impact decay after a meta-order execution. In the limit of small cancellation rates, we look for a scaling solution of the form $z_t^1 = TF(\nu t)$ (see equation (8)) where F is a dimensionless function. We consider the case where $\nu T \ll 1$ and $Q \ll Q_{\mathrm{lin.}}$. Long after the end of the execution of the meta-order, i.e. when $t \gg T$, equation (3) together with equations (7) and (8) becomes (to leading order):

$$0 = -\frac{\lambda\alpha T}{\sqrt{D}}F(\nu t)e^{-\nu t} - 2\lambda\alpha\int_0^t \mathrm{d}\tau\,\frac{z_t^0 - z_\tau^0}{\sqrt{4\pi D(t-\tau)}}e^{-\nu(t-\tau)} - 2\lambda\alpha T\sqrt{\nu}\int_0^t \mathrm{d}\tau\,\frac{F(\nu t) - F(\nu\tau)}{\sqrt{4\pi D(t-\tau)}}e^{-\nu(t-\tau)}\,. \quad (10)$$

Letting $u = \nu t$ and $z_t^0 = \beta/\sqrt{u}$ (see table 1) yields:

$$0 = \sqrt{\pi}e^{-u}F(u) + \beta\int_0^u \mathrm{d}v\,\frac{\sqrt{v}-\sqrt{u}}{\sqrt{uv(u-v)}}e^{v-u} + \int_0^u \mathrm{d}v\,\frac{F(u)-F(v)}{\sqrt{u-v}}e^{v-u}\,. \quad (11)$$

Table 1. Price trajectories for different impact regimes (see equation (8)). We set $\beta_0 := \frac{1}{2}[m_0/(2\pi J)]^{1/2}$. $F_\beta(u)$ is the solution of Eq. (11).

	α	z_t^0		z_t^1	
		$t \leq T$	$t > T$	$t \leq T$	$t > T$
$m_0 \ll J$	$\frac{m_0}{\mathcal{L}\sqrt{\pi D}}$	$\sqrt{t}$	$\sqrt{t} - \sqrt{t-T}$	$\left(\sqrt{\pi}/2 - 2/\sqrt{\pi}\right) t$	$T F_{\beta=1/2}(\nu t)$
$m_0 \gg J$	$\sqrt{\frac{2m_0}{\mathcal{L}}}$	$\sqrt{t}$	$\sqrt{t} - \sqrt{t-T}$ for $t \gtrsim T$ $\beta_0 T/\sqrt{t}$ for $t \gg T$	$-\frac{1}{3}\left(\frac{J}{2m_0}\right)^{1/2} t$	$T F_{\beta=\beta_0}(\nu t)$

Finally seeking F asymptotically of the form $F(u) = F_\infty + Bu^{-\gamma} + Cu^{-\delta}e^{-u}$ one can show that:

$$F(u) = F_\infty - \frac{\beta}{\sqrt{u}}\left[1 - e^{-u}\right] \qquad (u \gg 1)\,, \tag{12}$$

with the permanent component given by $F_\infty = \beta\sqrt{\pi}$, where β depends on the fast/slow nature of the execution (see table 1).

Injecting the solution for $F(u)$ in equation (8), and taking the limit of large times, one finds that the $t^{-1/2}$ decay of the zeroth-order term is exactly compensated by the $\beta u^{-1/2}$ term coming from $F(u)$, showing that the asymptotic value of the impact, given by $I_\infty = \alpha\sqrt{\nu}T F_\infty$, is reached exponentially fast as $\nu t \to \infty$ (see figure 2). This result can be interpreted as follows. At the end of execution (when the peak impact is reached), the impact starts decaying towards zero in a slow power-law fashion (see Donier *et al*. 2015) until approximately $t \sim \nu^{-1}$, beyond which all memory is lost (since the book has been globally renewed). Impact cannot decay anymore, since the previous reference price has been forgotten. Note that in the limit of large meta-order volumes and/or slow executions, all memory is already lost at the end of the execution and the permanent impact trivially matches the peak impact (see figure 2).

An important remark is in order here. Using table 1, one finds that $I_\infty = \frac{1}{2}\xi_c(Q/Q_{\text{lin.}})$ in both the small and large participation regime. In other words, we find that the permanent impact is *linear* in the executed volume Q, as dictated by no-arbitrage arguments (Huberman and Stanzl 2004, Gatheral 2010) and compatible with the classical Kyle framework (Kyle 1985). Nevertheless, the origin of this permanent impact is purely statistical here, and not necessarily related to 'true' information (which we have subsumed in the dynamics of the fair price $\hat{p}_t$). In other words, even random trades have a non-zero permanent impact as soon as the latent order book has a finite memory, precisely as in the zero-intelligence Santa-Fe model (Smith *et al*. 2003) where diffusive prices are generated from random trades.

5. Impact with fast and slow traders

5.1. Set up of the problem

As stated in the introduction, one major issue in the impact results of the LLOB model as presented by Donier *et al*. (2015) is the following. Empirically, the impact of meta-orders is only weakly dependent on the participation rate m_0/J (see e.g. Tóth *et al*. 2011). The corresponding *square-root law* is commonly written as:

$$I_Q := \langle x_T \rangle = Y\sigma\sqrt{\frac{Q}{V}}\,, \tag{13}$$

where σ is the daily volatility, V is the daily traded volume, and Y is a numerical constant of order unity. Note that I_Q only depends on the total volume of the meta-order $Q = m_0 T$, and not on m_0 (or equivalently on the time T).

As one can check from table 1, the independence of impact on m_0 only holds in the large participation rate limit ($m_0 \gg J$). However, most investors choose to operate in the opposite limit of small participation rates $m_0 \ll J$, and all the available data are indeed restricted to $m_0/J \lesssim 0.1$. In addition, the regime $m_0 \gg J$ yields deterministic square-root impact trajectories that would be easily detectable† and would lead to arbitrage opportunities in the absence of true information, as extensively discussed in Farmer *et al*. (2013), Gomes and Waelbroeck (2015), Bershova and Rakhlin (2013). Here, we offer a possible way out of this conundrum. The intuition is that the total market turnover J is actually dominated by high-frequency traders/market makers, whereas resistance to slow meta-orders can only be provided by slow participants on the other side of the book. More precisely, consider that only two sorts of agents co-exist in the market (see section 6 for a continuous range of frequencies):

(i) Slow agents with vanishing cancellation and deposition rates: $\nu_s T \to 0$, while keeping the corresponding liquidity $\mathcal{L}_s := \lambda_s/\sqrt{\nu_s D}$ finite; and
(ii) Fast agents with large cancellation and deposition rates, $\nu_f T \gg 1$, such that $\mathcal{L}_f := \lambda_f/\sqrt{\nu_f D} \gg \mathcal{L}_s$.

The system of partial differential equations to solve now reads:

$$\partial_t \phi_s = D\partial_{xx}\phi_s - \nu_s\phi_s + s_s(x,t) \tag{14a}$$

$$\partial_t \phi_f = D\partial_{xx}\phi_f - \nu_f\phi_f + s_f(x,t)\,, \tag{14b}$$

where $s_k(x,t) = \lambda_k\,\text{sign}(x_{kt} - x) + m_{kt}\delta(x - x_{kt})$, together with the conditions:

$$m_{st} + m_{ft} = m_0 \tag{15}$$

$$x_{st} = x_{ft} = x_t\,. \tag{16}$$

†Indeed, the price trajectory during a meta-order execution results from the combination of a square-root $x_t = \alpha\sqrt{t}$ (see table 1) and the wander about of the fair price $\hat{p}_t \sim \sqrt{Dt}$ (see Sect. 1). The deterministic square-root signal is detectable if it exceeds the noise level, that is $\alpha\sqrt{t} > \sqrt{Dt}$ which precisely corresponds to $m_0 > J$. In the regime $m_0 \ll J$ (equivalent to $\alpha\sqrt{t} \ll \sqrt{Dt}$) the signal is hidden in the overall noise, which explains why square-root trajectories are seldom observed in real price time series.

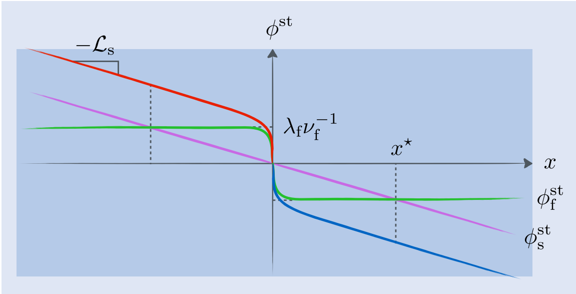

Figure 3. Stationary double-frequency order book $\phi^{st}(x) = \phi_s^{st}(x)$ (purple) + $\phi_f^{st}(x)$ (green) (see section 5).

Equation (15) means that the meta-order is executed against slow and fast agents, respectively, contributing to the rates m_{st} and m_{ft}. Equation (16) simply means that there is a unique transaction price, the same for slow and for fast agents. The total order book volume density is then given by $\phi = \phi_s + \phi_f$. In particular, in the limit of slow/fast agents discussed above the stationary order book is given by the sum of $\phi_s^{st}(x) \approx -\mathcal{L}_s x$ and $\phi_f^{st}(x) \approx -(\lambda_f/\nu_f)\mathrm{sign}(x)$ (see figure 3). The total transaction rate now reads

$$J = D\left|\partial_x\left[\phi_s^{st} + \phi_f^{st}\right]\right|_{x=0} = J_s + J_f, \tag{17}$$

where $J_f \gg J_s$ (which notably implies that $J \approx J_f$).

5.2. *From linear to square-root impact*

We now focus on the regime where the meta-order intensity is large compared to the average transaction rate of slow traders, but small compared to the total transaction rate of the market, to wit: $J_s \ll m_0 \ll J$. In this limit equations (14a) and (14b), together with the corresponding price setting equations $\phi_k(x_{kt}, t) \equiv 0$ yield (see appendix 2):

$$x_{st} = \left(\frac{2}{\mathcal{L}_s}\int_0^t \mathrm{d}\tau\, m_{s\tau}\right)^{1/2} \tag{18a}$$

$$x_{ft} = \frac{\nu_f}{\lambda_f}\int_0^t \mathrm{d}\tau\, m_{f\tau}\ . \tag{18b}$$

Differentiating equation (16) with respect to time together with equations (18) and using equation (15) yields:

$$m_{ft} = \frac{m_0}{\sqrt{1+\frac{t}{t^\star}}}, \quad \text{with} \quad t^\star := \frac{1}{2\nu_f}\frac{J_f^2}{J_s m_0}, \tag{19}$$

and $m_{st} = m_0 - m_{ft}$. Equation (19) indicates that most of the incoming meta-order is executed against the rapid agents for $t < t^\star$ but the slow agents then take over for $t > t^\star$ (see figure 4). The resulting price trajectory reads:

$$x_t = \frac{\lambda_f}{\mathcal{L}_s \nu_f}\left(\sqrt{1+\frac{t}{t^\star}} - 1\right), \tag{20}$$

which crosses over from a linear regime when $t \ll t^\star$ to a square-root regime for $t \gg t^\star$ (see figure 4). For a meta-order of volume Q executed during a time interval T, the corresponding impact is linear in Q when $T < t^\star$ and square-root (with I_Q independent of m_0) when $T > t^\star$. This last

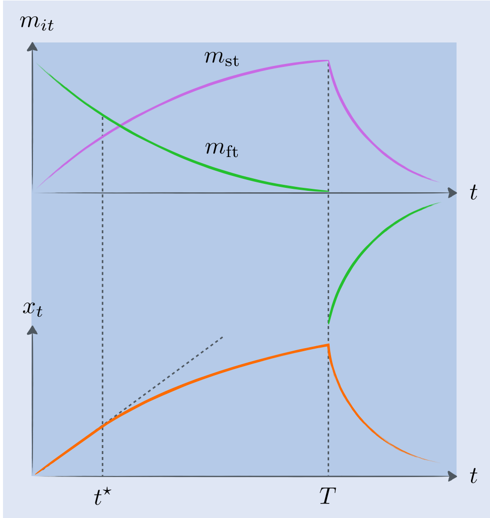

Figure 4. Execution rates m_{it} (top) and price trajectory (bottom) within the double-frequency order book model (see section 5).

regime takes place when $Q > m_0 t^\star$, which can be rewritten as:

$$\frac{Q}{V_d} > \frac{1}{\nu_f T_d}\frac{J}{J_s}, \tag{21}$$

where V_d is the total daily volume and T_d is one trading day. Numerically, with a HFT cancellation rate of—say—$\nu_f = 1\,\mathrm{s}^{-1}$ and $J_s = 0.1J$, one finds that the square-root law holds when the participation rate of the meta-order exceeds $3\,10^{-4}$, which is not unreasonable when compared with impact data. Interestingly, the cross-over between a linear impact for small Q and a square-root for larger Q is consistent with the data presented by Zarinelli *et al.* (2015) (note that the logarithmic impact curve proposed in Zarinelli *et al.* (2015) is indeed linear for small Q).

5.3. *Impact decay*

Regarding the decay impact for $t > T$, the problem to solve is that of equations (14a), (14b) and (16) only where equation (15) becomes:

$$m_{st} + m_{ft} = 0\ . \tag{22}$$

The solution behaves asymptotically ($t \gg T$) to zero as $x_t \sim t^{-1/2}$ (see appendix 2). Given the results of section 4 in the presence of finite-memory agents, the absence of permanent impact may seem counter-intuitive. In order to understand this feature of the double-frequency order book model in the limit $\nu_s T \to 0$, $\nu_f T \gg 1$, one can look at the stationary order book. As one moves away from the price, the ratio of slow over fast volume fractions (ϕ_s/ϕ_f) grows linearly to infinity. Hence, the shape of the latent order book for $|x| \gg x^\star$ matches that of the infinite memory single-agent model originally presented by Donier *et al.* (2015) (see figure 3). This explains the mechanical return of the price to its initial value before execution, encoded in the infinite memory latent order book when $\nu_s = 0$. However, the permanent impact for small but non-zero ν_s is of order $\sqrt{\nu_s}$, as obtained in section 4.

5.4. The linear regime

The regime of very small participation rates for which $m_0 \ll J_s, J_f$ is also of conceptual interest. In such a case equation (18a) must be replaced with:

$$x_{st} = \frac{1}{\mathcal{L}_s} \int_0^t \mathrm{d}\tau \, \frac{m_{s\tau}}{\sqrt{4\pi D(t-\tau)}} \,, \tag{23}$$

which together with equations (18b), (15) and (16) yields, in Laplace space (see appendix 2):

$$\widehat{m}_{1p} = \frac{1}{p} \frac{m_0}{1+\sqrt{pt^\dagger}} \,, \tag{24}$$

where $t^\dagger = (m_0/\pi J_s)t^\star$, with $t^\star$ defined in equation (19). For small times ($t \ll t^\dagger$) one obtains $m_{st} = 2m_0\sqrt{t/t^\dagger}$ while for larger times ($t^\dagger \ll t < T$), $m_{st} = m_0[1 - \sqrt{t^\dagger/(\pi t)}]$. Finally using again equations (18b), (15) and (16) yields $x_t = (\nu_f/\lambda_f)m_0 t$ for $t \ll t^\dagger$ and $x_t = (\nu_f/\lambda_f)m_0\sqrt{tt^\dagger/\pi}$ for $t^\dagger \ll t < T$, identical in terms of scaling to the price dynamics observed in the case $J_s \ll m_0 \ll J_f$ discussed above. The asymptotic impact decay is identical to the one obtained in that case as well.

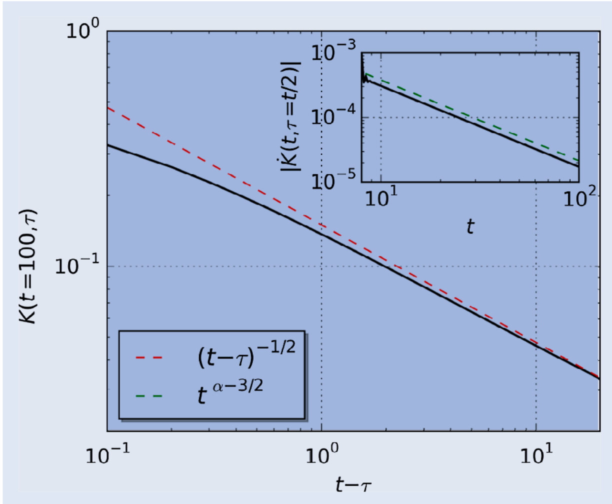

Figure 5. Numerical determination of the kernel $K(t,\tau) := M^{-1}(t,\tau)$, for $\alpha = 0.25$. One clearly sees that K decays as $(t-\tau)^{-1/2}$ at large lags. The inset shows that $K(t,t/2)$ behaves as $t^{\alpha-1/2}$, as expected.

6. Multi-frequency order book

The double-frequency framework presented in section 5 can be extended to the more realistic case of a continuous range of cancellation and deposition rates. Formally, one has to solve an infinite set of equations, labelled by the cancellation rate ν:†

$$\partial_t \phi_\nu = D\partial_{xx}\phi_\nu - \nu\phi_\nu + s_\nu(x,t) \,, \tag{25}$$

where $\phi_\nu(x,t)$ denotes the contribution of agents with typical frequency ν to the latent order book, and $s_\nu(x,t) = \lambda_\nu$ sign $(x_{\nu t} - x) + m_{\nu t}\delta(x - x_{\nu t})$, with $\lambda_\nu = \mathcal{L}_\nu\sqrt{\nu D}$. Equation (25) must then be completed with:

$$\int_0^\infty \mathrm{d}\nu \rho(\nu) m_{\nu t} = m_t \tag{26a}$$

$$x_{\nu t} = x_t \qquad \forall \nu \,, \tag{26b}$$

where $\rho(\nu)$ denotes the distribution of cancellation rates ν, and where we have allowed for an arbitrary order flow m_t. Solving exactly the above system of equations analytically is an ambitious task. In the following, we present a simplified analysis that allows us to obtain an approximate scaling solution of the problem for a power-law distribution of frequencies ν.

6.1. The propagator regime

We first assume, for simplicity, that the order flow J_ν is independent of frequency (see later for a more general case), and consider the case when $m_t \ll J, \forall t$. Although not trivially true, we assume (and check later on the solution) that this implies $m_{\nu t} \ll J \; \forall \nu$, such that we can assume linear response for all ν. Schematically, there are two regimes, depending on whether $t \gg \nu^{-1}$—in which case the corresponding density $\phi_\nu(x,t)$ has lost all its memory, or $t \ll \nu^{-1}$. In the former case the price trajectory follows equation (23), while in the latter case it is rather equation (18b) that rules the dynamics. One thus has:

$$\text{For } \nu t \ll 1 \quad x_t = \frac{1}{\mathcal{L}\sqrt{D}} \int_0^t \mathrm{d}\tau \, \frac{m_{\nu\tau}}{\sqrt{4\pi(t-\tau)}} \tag{27a}$$

$$\text{For } \nu t \gg 1 \quad x_t = \frac{\nu^{1/2}}{\mathcal{L}\sqrt{D}} \int_0^t \mathrm{d}\tau \, m_{\nu\tau} \,. \tag{27b}$$

Inverting equations (27b) and defining $\Psi(t) := 2/\sqrt{\pi t}$ yields (see appendix 2 and in particular equation (B7)):

$$\text{For } \nu t \ll 1 \quad m_{\nu t} = \mathcal{L}\sqrt{D} \int_0^t \mathrm{d}\tau \, \Psi(t-\tau)\dot{x}_\tau \tag{28a}$$

$$\text{For } \nu t \gg 1 \quad m_{\nu t} = \mathcal{L}\sqrt{D}\nu^{-1/2}\dot{x}_t \,. \tag{28b}$$

Our approximation is to assume that $m_{\nu t}$ in equation (26a) is effectively given by equation (28a) as soon as $\nu < 1/t$ and by equation (28b) when $\nu > 1/t$ such that equation (26a) becomes:

$$\int_0^{1/t} \mathrm{d}\nu\rho(\nu)\left[\int_0^t \mathrm{d}\tau\Psi(t-\tau)\dot{x}_\tau\right] + \int_{1/t}^\infty \mathrm{d}\nu\rho(\nu)\left[\nu^{-1/2}\dot{x}_t\right] = \frac{m_t}{\mathcal{L}\sqrt{D}} \,. \tag{29}$$

Equation (29) may be conveniently re-written as $\int_0^t \mathrm{d}\tau\left[G(t)\Psi(t-\tau) + H(t)t_c^{1/2}\delta(t-\tau)\right]\dot{x}_\tau = m_t/(\mathcal{L}\sqrt{D})$, with:

$$G(t) := \int_0^{1/t} \mathrm{d}\nu\rho(\nu) \,, \qquad H(t) := t_c^{-1/2} \int_{1/t}^\infty \mathrm{d}\nu\rho(\nu)\nu^{-1/2} \,. \tag{30}$$

Formally inverting the kernel $M(t,\tau) := \left[G(t)\Psi(t-\tau) + H(t)t_c^{1/2}\delta(t-\tau)\right]$ then yields the price dynamics $\dot{x}_t$ as a linear convolution of the past order flow $m_{\tau \le t}$. Note that when $m_t \to 0$, $\dot{x}_t$ is also small and hence, using equations (28), all $m_{\nu t}$ are

†In full generality, the diffusion coefficient should be allowed to depend on ν. This can be simply implemented by changing the distribution of frequencies as: $\rho(\nu) \to \rho(\nu)D_\nu/D$.

all small as well, justify our use of equations (27b) for all frequencies.

6.2. *Resolution of the 'diffusivity puzzle'*

Let us now compute the functions $G(t)$ and $H(t)$ for a specific power-law distribution $\rho(\nu)$ defined as:

$$\rho(\nu) = Z\nu^{\alpha-1}e^{-\nu t_c}\ , \tag{31}$$

where $\alpha > 0$, t_c is a high-frequency cut-off, and $Z = t_c^{\alpha}/\Gamma(\alpha)$.† For such a distribution, one obtains $G(t) = 1 - \Gamma(\alpha, t_c/t)/\Gamma(\alpha)$ and $H(t) = \Gamma(\alpha - 1/2, t_c/t)/\Gamma(\alpha)$. In the limit $t \ll t_c$, $G(t) \approx 1$ and $H(t) \approx 0$. In the limit $t \gg t_c$, $G(t) \approx (t/t_c)^{-\alpha}/[\alpha\Gamma(\alpha)]$, and the dominant term in the first-order expansion of $H(t)$ depends on whether $\alpha \lessgtr 1/2$. One has $H(t|_{\alpha<1/2}) \approx 2(t/t_c)^{1/2-\alpha}/[\Gamma(\alpha)(1-2\alpha)]$ and $H(t|_{\alpha>1/2}) \approx \Gamma(\alpha - 1/2)/\Gamma(\alpha)$. Focusing on the interesting case $\alpha < 1/2$, one finds (see figure 5) that inversion of the kernel $M(t,\tau)$ is dominated, at large times, by the first term $G(t)\Psi(t-\tau)$. Hence, one finds in that regime:†

$$x_t \approx \frac{\alpha\Gamma(\alpha)}{\mathcal{L}t_c^{\alpha}\sqrt{D}}\int_0^t \mathrm{d}\tau\,\frac{m_\tau \tau^{\alpha}}{\sqrt{4\pi(t-\tau)}}\ . \tag{32}$$

Let us now show that this equation can lead to a diffusive price even in the presence of a long-range correlated order flow. Assuming that $\langle m_t m_{t'}\rangle \sim |t-t'|^{-\gamma}$ with $0<\gamma<1$ (defining a long memory process, as found empirically (Bouchaud *et al.* 2004, Bouchaud *et al.* 2008), one finds from equation (32) that the mean square price is given by:

$$\langle x_t^2\rangle \propto \iint_0^t \mathrm{d}\tau\mathrm{d}\tau'\,\frac{\langle m_\tau m_{\tau'}\rangle(\tau\tau')^{\alpha}}{\sqrt{(t-\tau)(t-\tau')}}\ . \tag{33}$$

Changing variables through $\tau \to tu$ and $\tau' \to tv$ easily yields $\langle x_t^2\rangle \propto t^{1+2\alpha-\gamma}$. Note that the LLOB limit corresponds to a unique low-frequency ν for the latent liquidity. This limit can be formally recovered when $\alpha \to 0$. In this case, we recover the 'disease' of the LLOB model, namely a mean-reverting, subdiffusive price $\langle x_t^2\rangle \propto t^{1-\gamma}$ for all values of $\gamma > 0$. Intuitively, the latent liquidity in the LLOB case is too persistent and prevents the price from diffusing. Imposing price diffusion, i.e. $\langle x_t^2\rangle \propto t$ finally gives a consistency condition similar in spirit to the one obtained in Bouchaud *et al.* (2004):

$$\alpha = \frac{\gamma}{2} < \frac{1}{2}\ . \tag{34}$$

Equation (34) states that for persistent order flow to be compatible with diffusive price dynamics, the long-memory of order flow must be somehow buffered by a long-memory of the liquidity, which makes sense. The present resolution of the diffusivity puzzle—based on the memory of a multi-frequency self-renewing latent order book—is similar to, but different from that developed in Benzaquen and Bouchaud (2018). In

†Note that rigorously speaking, one should also introduce a low-frequency cut-off $\nu_{\rm LF}$ to ensure the existence of a stationary state of the order book in the absence of meta-order. Otherwise, $\langle\nu^{-1}\rangle = \infty$ when $\alpha \le 1$ and the system does not reach a stationary state (see the end of section 2 and Benzaquen and Bouchaud (2018) for a further discussion of this point).

†Taking into account the $H(t)$ contribution turns out not to change the following scaling argument.

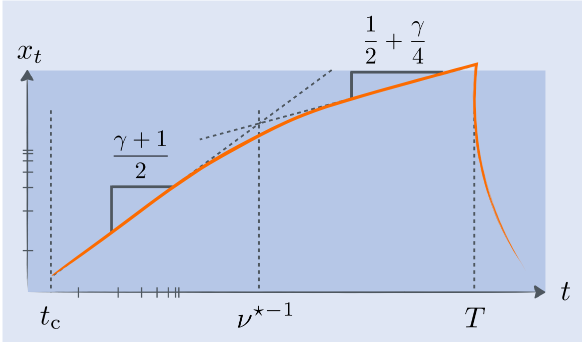

Figure 6. Price trajectory during a constant rate meta-order execution within the multi-frequency order book model. For $\gamma = 1/2$, the impact crosses over from a $t^{3/4}$ to a $t^{5/8}$ regime.

the latter study, we assumed the reassessment time of the latent orders to be fat-tailed, leading to a 'fractional' diffusion equation for $\phi(x,t)$.

6.3. *Meta-order impact*

We now relax the constraint that $\lambda_\nu \propto \sqrt{\nu}$ and define $J_\nu := J_{\rm hf}(\nu t_c)^{\zeta}$ with $\zeta > 0$, meaning that HFT is the dominant contribution to trading, since in this case:

$$J = \int_0^\infty \mathrm{d}\nu\rho(\nu)J_\nu = J_{\rm hf}\frac{\Gamma(\zeta+\alpha)}{\Gamma(\alpha)}. \tag{35}$$

(The case $\zeta < 0$ could be considered as well, but is probably less realistic).

We consider a meta-order with constant execution rate $m_0 \ll J_{\rm hf}$. Since J_ν decreases as the frequency decreases, there must exist a frequency $\nu^\star$ such that $m_0 = J_{\nu^\star}$, leading to $\nu^\star t_c = (m_0/J_{\rm hf})^{1/\zeta}$.‡ When $\nu \ll \nu^\star$, we end up in the non-linear, square-root regime where $m_0 \gg J_\nu$ and equation (18a) holds. Proceeding as in the previous section, we obtain the following approximation for the price trajectory:

$$G_\zeta(t)\left[\int_0^t \mathrm{d}\tau\,\Psi(t-\tau)\dot{x}_\tau \mathbb{1}_{\{t\le\nu^{\star-1}\}} + \frac{x_t\dot{x}_t}{2\sqrt{D}}\mathbb{1}_{\{t>\nu^{\star-1}\}}\right] + t_c^{1/2}H_\zeta(t)\dot{x}_t = \frac{m_0\sqrt{D}}{J_{\rm hf}}\ . \tag{36}$$

where, in the limit $t \gg t_c$ and $\alpha + \zeta < 1/2$:

$$G_\zeta(t) := \int_0^{1/t}\mathrm{d}\nu\rho(\nu)(\nu t_c)^{\zeta} \approx \left(\frac{t_c}{t}\right)^{\alpha+\zeta}\frac{1}{\Gamma(\alpha)(\alpha+s)} \tag{37a}$$

$$H_\zeta(t) := \int_{1/t}^\infty \mathrm{d}\nu\rho(\nu)(\nu t_c)^{\zeta-1/2} \approx \left(\frac{t_c}{t}\right)^{\alpha+\zeta-1/2} \times \frac{1}{\Gamma(\alpha)(1/2-\alpha-s)}\ . \tag{37b}$$

At short times $t \ll \nu^{\star-1}$, equation (36) boils down to equation (29) with $\alpha \to \alpha+\zeta$ and one correspondingly finds:

$$x_t \propto x_c\frac{m_0}{J_{\rm hf}}\left(\frac{t}{t_c}\right)^{\frac{1}{2}+\alpha+\zeta}\ , \tag{38}$$

‡Here again, we assume that $\nu^\star$ is much larger than the implicit low-frequency cut-off $\nu_{\rm LF}$.

where $x_c := \sqrt{Dt_c}$. For $t \gg \nu^{\star -1}$, the second term in equation (36) dominates over both the first and the third terms, leading to a generalized square-root law of the form:

$$x_t \propto x_c \sqrt{\frac{m_0}{J_{\rm hf}}} \left(\frac{t}{t_c}\right)^{\frac{1+\alpha+\zeta}{2}}, \quad (39)$$

Compatibility with price diffusion imposes now that $\alpha + \zeta = \gamma/2$, which finally leads to (see figure 6):

$$x_t \propto x_c \frac{m_0}{J_{\rm hf}} \left(\frac{t}{t_c}\right)^{\frac{1+\gamma}{2}}, \quad \text{when} \quad t_c \ll t \ll t_c \left(\frac{J_{\rm hf}}{m_0}\right)^{1/\zeta} \quad (40a)$$

$$x_t \propto x_c \sqrt{\frac{m_0}{J_{\rm hf}}} \left(\frac{t}{t_c}\right)^{\frac{2+\gamma}{4}}, \quad \text{when} \quad t \gg t_c \left(\frac{J_{\rm hf}}{m_0}\right)^{1/\zeta}. \quad (40b)$$

In the latter case, setting $\gamma = 1/2$ and $Q = m_0 T$, one finds an impact $I_Q := x_T$ behaving as§ $Q^{5/8}$ as soon as $Q > \upsilon(J_{\rm hf}/m_0)^{(1-\zeta)/\zeta}$, where we have introduced an elementary volume $\upsilon := J_{\rm hf} t_c$, which is the volume traded by HFT during their typical cancellation time. For small meta-orders such that $T \lesssim t_c$, impact is linear in Q.

7. Conclusion

In this work, we have extended the LLOB latent liquidity model (Donier *et al.* 2015) to account for the presence of agents with different memory timescales. This has allowed us to overcome several conceptual and empirical difficulties faced by the LLOB model. We have first shown that whenever the longest memory time is finite (rather than divergent in the LLOB model), a permanent component of impact appears, even in the absence of any 'informed' trades. This permanent impact is *linear* in the traded quantity and independent of the trading rate, as imposed by no-arbitrage arguments. We have then shown that the square-root impact law holds provided the meta-order participation rate is large compared to the trading rate of 'slow' actors, which can be small compared to the total trading rate of the market—itself dominated by high-frequency traders. In the original LLOB model where all actors are slow, a square-root impact law independent of the participation rate only holds when the participation rate is large compared to the total market rate, a regime not consistent with empirical data, as it would lead to nearly deterministic square-root impact trajectories.

Finally, the multi-scale latent liquidity model offers a new resolution of the diffusivity paradox, i.e. how an order flow with long-range memory can give rise to a purely diffusive price. We show that when the liquidity memory times are themselves fat-tailed, mean-reversion effects induced by a persistent order book can exactly offset trending effects induced by a persistent order flow, as in the propagator model (Bouchaud *et al.* 2004).

We therefore believe that the multi-timescale latent order book view of markets, encapsulated by equations (25) and (26b), is rich enough to capture a large part of the subtleties of the dynamics of markets. It suggests an alternative framework to build agent based models of markets that generate realistic price series, that complement and maybe simplify previous attempts (Tóth *et al.* 2011, Mastromatteo *et al.* 2014b). A remaining outstanding problem, however, is to reconcile the extended LLOB model proposed in this paper with some other well known 'stylized facts' of financial price series, namely power-law distributed price jumps and clustered volatility. We hope to report progress in that direction soon. Another, more mathematical endeavour is to give a rigorous meaning to the multi-timescale reaction model underlying equations (25) and (26b) and to the approximate solutions provided in this paper. It would be satisfying to extend the no-arbitrage result of Donier *et al.* (2015), valid for the LLOB model, to the present multi-timescale setting. Although more difficult to prove, we believe that our multi-timescale model is arbitrage free, in the sense that any round trip incurs positive costs on average.

§Note that $5/8 \approx 0.6$ is very close to the empirical impact results reported by Almgren *et al.* (2005) and Brokmann *et al.* (2015) in the case of equities, for which γ is usually close to 1/2.

Acknowledgements

We thank J. Bonart, A. Darmon, J. de Lataillade, J. Donier, Z. Eisler, A. Fosset, S. Gualdi, I. Mastromatteo, M. Rosenbaum and B. Tóth for extremely fruitful discussions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

M. Benzaquen http://orcid.org/0000-0002-9751-7625

References

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Direct estimation of equity market impact. *Risk*, 2005, **18**(7), 57–62.

Benzaquen, M. and Bouchaud, J.P., A fractional reaction-diffusion description of supply and demand. *Eur. Phys. J. B*, 2018, **91**, 23.

Bershova, N. and Rakhlin, D., The non-linear market impact of large trades: Evidence from buy-side order flow. *Quant. Finance*, 2013, **13**(11), 1759–1778.

Bouchaud, J.-P., Price impact. In *Encyclopedia of Quantitative Finance*, edited by R. Cont, 2010 (John Wiley & Sons: Hoboken, NJ).

Bouchaud, J.-P., Farmer, J.D. and Lillo, F., How markets slowly digest changes in supply and demand. In *Handbook of Financial Markets: Dynamics and Evolution*, edited by T. Hens and K.R. Schenk-Hoppe, pp. 57–160, 2008 (Elsevier: North-Holland).

Bouchaud, J.-P., Gefen, Y., Potters, M. and Wyart, M., Fluctuations and response in financial markets: The subtle nature of "random" price changes. *Quant. Finance*, 2004, **4**(2), 176–190.

Brokmann, X., Serie, E., Kockelkoren, J. and Bouchaud, J.-P., Slow decay of impact in equity markets. *Market Microstruct. Liquidity*, 2015, **1**(2), 1550007.

Donier, J. and Bonart, J., A million metaorder analysis of market impact on the bitcoin. *Market Microstruct. Liquidity*, 2015, **1**(2), 1550008.

Donier, J., Bonart, J.F., Mastromatteo, I. and Bouchaud, J.-P., A fully consistent, minimal model for non-linear market impact. *Quant. Finance*, 2015, **15**(7), 1109–1121.

Farmer, J.D., Gerig, A., Lillo, F. and Waelbroeck, H., How efficiency shapes market impact. *Quant. Finance*, 2013, **13**, 1743–1758.

Gatheral, J., No-dynamic-arbitrage and market impact. *Quant. Finance*, 2010, **10**(7), 749–759.
Gomes, C. and Waelbroeck, H., Is market impact a measure of the information value of trades? Market response to liquidity vs. informed metaorders. *Quant. Finance*, 2015, **15**(5), 773–793.
Grinold, R.C. and Kahn, R.N., *Active Portfolio Management*, 2000 (McGraw Hill: New York).
Hasbrouck, J., *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*, 2007 (Oxford University Press: Oxford).
Huberman, G. and Stanzl, W., Price manipulation and quasi-arbitrage. *Econometrica*, 2004, **72**(4), 1247–1275.
Kyle, A.S., Continuous auctions and insider trading. *Econometrica*, 1985, **53**, 1315–1335.
Lillo, F. and Farmer, J.D., The long memory of the efficient market. *Stud. Nonlinear Dyn. Econom.*, 2004, **8**(3), 1–33 .
Mastromatteo, I., Tóth, B. and Bouchaud, J.-P., Anomalous impact in reaction-diffusion financial models. *Phys. Rev. Lett.*, 2014a, **113**, 268701.
Mastromatteo, I., Tóth, B. and Bouchaud, J.-P., Agent-based models for latent liquidity and concave price impact. *Phys. Rev. E*, 2014b, **89**(4), 042805.
Smith, E., Farmer, J.D., Gillemot, L.s. and Krishnamurthy, S., Statistical theory of the continuous double auction. *Quant. Finance*, 2003, **3**(6), 481–514.
Tóth, B., Lemperiere, Y., Deremble, C., De Lataillade, J., Kockelkoren, J. and Bouchaud, J.P., Anomalous price impact and the critical nature of liquidity in financial markets. *Phys. Rev. X*, 2011, **1**(2), 021006.
Weber, P. and Rosenow, B., Order book approach to price impact. *Quant. Finance*, 2005, **5**(4), 357–364.
Zarinelli, E., Treccani, M., Farmer, J.D. and Lillo, F., Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstruct. Liquidity*, 2015, **1**(2), 1550004.

Appendix 1

We here provide the calculations that link equation (8) and table 1 during a meta-order execution ($t \leq T$); the impact decay computations ($t > T$) are given and discussed in section 4.

In the limit of slow execution of the meta-order, one has $(x_t - x_\tau)^2 \ll 4D(t-\tau)$ such that equation (7) together with equation (3) becomes:

$$0 = \phi^{\mathrm{st}}(x_t)e^{-\nu t} + \int_0^t \mathrm{d}\tau\, \frac{m_0}{\sqrt{4\pi D(t-\tau)}} e^{-\nu(t-\tau)} - 2\lambda \int_0^t \mathrm{d}\tau\, \frac{x_t - x_\tau}{\sqrt{4\pi D(t-\tau)}} e^{-\nu(t-\tau)} \,. \tag{A1}$$

Interestingly, slow and short execution is only compatible with small meta-order volume† (indeed, combining $m_0 \ll J$ and $\nu T \ll 1$ implies $m_0 T \ll J\nu^{-1}$). Thus for slow and short execution, using the linear approximation $\phi^{\mathrm{st}}(x_t) = -\mathcal{L}x_t$ and letting equation (8) into equation (A1) yields:

$$0 = -\mathcal{L}\alpha z_t^0 + m_0\sqrt{\frac{t}{\pi D}} \tag{A2a}$$

$$0 = -\mathcal{L}\sqrt{\nu} z_t^1 - 2\lambda \int_0^t \mathrm{d}\tau\, \frac{z_t^0 - z_\tau^0}{\sqrt{4\pi D(t-\tau)}} \,. \tag{A2b}$$

Equation (A2a) yields $\alpha = m_0/(\mathcal{L}\sqrt{\pi D})$ and $z_t^0 = \sqrt{t}$, and it follows from equation (A2b) that $z_t^1 = -kt$ where $k = \sqrt{4/\pi} - \sqrt{\pi/4}$.

In the limit of fast execution, one has $(x_t - x_\tau)^2 \gg 4D(t-\tau)$ such that the meta-order term can be approximated through the saddle point method. Letting $x_\tau \approx x_t - (t-\tau)\dot{x}_t$ into the price equation now yields:

$$0 = \phi^{\mathrm{st}}(x_t)e^{-\nu t} + \int_0^t \mathrm{d}\tau\, m_0 \frac{e^{-\frac{\dot{x}_t^2(t-\tau)}{4D}}}{\sqrt{4\pi D(t-\tau)}} e^{-\nu(t-\tau)} - \lambda \int_0^t \mathrm{d}\tau\, e^{-\nu(t-\tau)} \,. \tag{A3}$$

Letting $u = t - \tau$ and given $4D/\dot{x}_t^2 \ll t$ such that $\int_0^t \mathrm{d}u \approx \int_0^\infty \mathrm{d}u$, equation (A3) becomes:

$$0 = \phi^{\mathrm{st}}(x_t)e^{-\nu t} + \frac{m_0}{\sqrt{\dot{x}_t^2 + 4D\nu}} + \frac{\lambda}{\nu}\left(e^{-\nu t} - 1\right) . \tag{A4}$$

For short execution with small meta-order volume (we use $\phi^{\mathrm{st}}(x_t) = -\mathcal{L}x_t$), letting equation (8) into equation (A4) yields:

$$0 = -\mathcal{L}\alpha z_t^0 + \frac{m_0}{\alpha|\dot{z}_t^0|} \tag{A5a}$$

$$0 = -\mathcal{L}\alpha\sqrt{\nu} z_t^1 - \frac{\sqrt{\nu} m_0}{\alpha} \frac{\dot{z}_t^1}{(\dot{z}_t^0)^2} - \lambda t \,. \tag{A5b}$$

Equation (A5a) yields $\alpha = \sqrt{2m_0/\mathcal{L}}$ and $z_t^0 = \sqrt{t}$, and thus equation (A5b) becomes $\dot{z}_t^1 + z_t^1/(2t) = -\frac{1}{2}\sqrt{J/(2m_0)}$. It follows that $z_t^1 = -\frac{t}{3}\sqrt{J/(2m_0)}$. For a fast, short and large meta-order, x_t is expected to go well beyond the linear region of the order book such that in a hand-waving static approach (consistent with fast and short execution) one can match $m_0 t$ and the area of a rectangle of sides x_t and $\lambda\nu^{-1}$ (see figure 1). Letting $x_t = bt$ yields $b = m_0\nu/\lambda$. Note that this result can be recovered by letting $x_t = bt$ and $\phi^{\mathrm{st}}(x_t) = -\lambda\nu^{-1}$ into equation (A4). Indeed, at leading order one obtains:

$$0 = -\frac{\lambda}{\nu} + \frac{m_0}{|\dot{x}_t|} \,, \tag{A6}$$

from which the result trivially follows.

For long execution ($\nu T \gg 1$) the memory of the initial book is rapidly lost and one expects Markovian behaviour. Letting again $x_t = bt$ into the price equation and changing variables through $\tau = t(1-u)$ yields:

$$0 = m_0\sqrt{t}\int_0^1 \mathrm{d}u\, \frac{e^{-\frac{b^2 tu}{4D}}}{\sqrt{4\pi Du}} e^{-\nu tu} - \lambda \int_0^1 \mathrm{d}u\, e^{-\nu tu}\, \mathrm{erf}\sqrt{\frac{b^2 tu}{4D}} = \left(m_0 - \frac{\lambda b}{\nu}\right)\frac{1}{\sqrt{b^2 + 4D\nu}}\, \mathrm{erf}\sqrt{\left(\frac{b^2}{4D} + \nu\right)t} \,. \tag{A7}$$

Interestingly, equation (A7) yields $b = m_0\nu/\lambda$ (regardless of execution rate and meta-order size), which is exactly the result obtained above in the case of fast and short execution of a large meta-order but for different reasons.

Appendix 2

We here provide the calculations underlying the double-frequency order book model presented in section 5. In particular for the case $J_{\mathrm{s}} \ll m_0 \ll J_{\mathrm{f}}$, equations (18) are obtained as follows. In the limit of large trading intensities the saddle point methods (as detailed in appendix 1) can also be applied to the case of nonconstant execution rates (one lets $m_\tau \approx m_t$ about which the integrand is evaluated, see Donier *et al.* 2015), in particular one obtains (equivalent to equation (A5b)):

$$\mathcal{L}_{\mathrm{s}} x_{\mathrm{st}} |\dot{x}_{\mathrm{st}}| = m_{\mathrm{st}} \,, \tag{B1}$$

which yields equation (18a). For the rapid agents ($\nu_{\mathrm{f}} T \gg 1$) we must consider the case of long execution. In particular, an equation tantamount to equation (A7) can also be derived in the case of nonconstant execution rates. Proceeding in the same manner, one easily obtains:

$$0 = \left(m_{\mathrm{ft}} - \frac{\lambda_{\mathrm{f}}\dot{x}_{\mathrm{ft}}}{\nu_{\mathrm{f}}}\right)\frac{1}{\sqrt{\dot{x}_{\mathrm{ft}}^2 + 4D\nu_{\mathrm{f}}}}\, \mathrm{erf}\sqrt{\left(\frac{\dot{x}_{\mathrm{ft}}^2}{4D} + \nu_{\mathrm{f}}\right)t} \,, \tag{B2}$$

†Equivalently, rapid and long execution is only consistent with large meta-order volume (combining $m_0 \gg J$ and $\nu T \gg 1$ implies $m_0 T \gg J\nu^{-1}$).

which yields $\dot{x}_{\mathrm{f}t} = m_{\mathrm{f}t}\nu_{\mathrm{f}}/\lambda_{\mathrm{f}}$ and thus equation (18b). Then, as mentioned in section 5, the asymptotic impact decay is obtained from equations (14a), (14b) and (16) only where for $t > T$ we replace equation (15) with equation (22). Using equation (7) together with equation (3) in the limit $\nu_{\mathrm{s}}T \to 0$, and $\nu_{\mathrm{f}}T \gg 1$ together with (16) yields ($t > T$):

$$\mathcal{L}_{\mathrm{s}} x_t = \int_0^T + \int_T^t \mathrm{d}\tau \, \frac{m_{\mathrm{s}\tau}}{\sqrt{4\pi D(t-\tau)}} \tag{B3a}$$

$$0 = \int_0^T + \int_T^t \mathrm{d}\tau \, \frac{e^{-\nu_{\mathrm{f}}(t-\tau)}}{\sqrt{4\pi D(t-\tau)}} \big[m_{\mathrm{f}\tau} - 2\lambda_{\mathrm{f}}(x_t - x_\tau) \big] . \tag{B3b}$$

Asymptotically ($t \gg T$) the system of equations (B3b) becomes:

$$\mathcal{L}_{\mathrm{s}} x_t = \int_0^T \frac{m_{\mathrm{s}\tau}\mathrm{d}\tau}{\sqrt{4\pi D(t-\tau)}} + \int_T^t \frac{m_{\mathrm{s}\tau}\mathrm{d}\tau}{\sqrt{4\pi D(t-\tau)}} \tag{B4a}$$

$$0 = \int_0^t \mathrm{d}\tau \, \frac{e^{-\nu_{\mathrm{f}}(t-\tau)}}{\sqrt{4\pi D(t-\tau)}} \left[m_{\mathrm{f}\tau} - 2\lambda_{\mathrm{f}}(x_t - x_\tau) \right] . \tag{B4b}$$

We expect the asymptotic impact decay to be of the form $x_t = x_\infty + B/\sqrt{t}$. In addition equation (B4b) indicates that $m_{\mathrm{f}t} \sim \dot{x}_t$. We thus let $m_{\mathrm{s}t} = -m_{\mathrm{f}t} = C/t^{3/2}$. Injecting into equation (B4a) yields $x_\infty = 0$ (no permanent impact) and:

$$\frac{\mathcal{L}_{\mathrm{s}} B}{\sqrt{t}} = \frac{1}{\sqrt{t}} \left[\frac{m_0 f_T}{\sqrt{4\pi D}} + \frac{C}{\sqrt{\pi D T}} \right] , \tag{B5}$$

where $f_T = T$ if $t^\star \ll T$ and $f_T = T^2/(3t^\star)$ if $t^\star \gg T$. On the other hand, letting $u = t - \tau$ in equation (B4b) and using $x_t - x_s \approx (t-s)\dot{x}_t$ yields at leading order:

$$0 = \int_0^\infty \mathrm{d}u \, \frac{e^{-\nu_{\mathrm{f}} u}}{\sqrt{u}} \left[-\frac{C}{t^{3/2}} + \frac{\lambda_{\mathrm{f}} B u}{t^{3/2}} \right] = \sqrt{\frac{\pi}{\nu_{\mathrm{f}} t^3}} \left[-C + \frac{\lambda_{\mathrm{f}} B}{\nu_{\mathrm{f}}} \right] , \tag{B6}$$

which combined with equation (B5) easily leads to the values of B and C.

For the case $m_0 \ll J_{\mathrm{s}}, J_{\mathrm{f}}$, the calculations are slightly more subtle. Inverting equation (23) in Laplace space yields:

$$m_{st} = 2\mathcal{L}_{\mathrm{s}} \sqrt{D} \int_0^t \mathrm{d}\tau \, \frac{\dot{x}_{\mathrm{s}\tau}}{\sqrt{\pi(t-\tau)}} . \tag{B7}$$

One can easily check this result by re-injecting equation (B7) into equation (23). In turn, inverting equation (18b) is straightforward and yields $m_{\mathrm{f}t} = (\lambda_{\mathrm{f}}/\nu_{\mathrm{f}})\dot{x}_{\mathrm{f}t}$. Injecting $\dot{x}_{\mathrm{s}t} = \dot{x}_{\mathrm{f}t}$ into equation (B7) and using equation (15) yields:

$$m_{st} = \frac{1}{\sqrt{t^\dagger}} \int_0^t \mathrm{d}\tau \, \frac{m_0 - m_{\mathrm{s}\tau}}{\sqrt{\pi(t-\tau)}} , \tag{B8}$$

which can be written as:

$$\int_0^t \mathrm{d}\tau \, m_{\mathrm{s}\tau} \Phi(t-\tau) = 2m_0\sqrt{t} , \quad \text{with} \quad \Phi(t) := \delta(t)\sqrt{\pi t^\dagger} + \frac{\theta(t)}{\sqrt{t}} . \tag{B9}$$

Taking the Laplace transform of equation (B9) one obtains $\widehat{\Phi}(p)\widehat{m}_{sp} = m_0\sqrt{\pi}/p^{3/2}$ with $\widehat{\Phi}(p) = \sqrt{\pi t^\dagger} + \sqrt{\pi/p}$, which in turn yields equation (24).