Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: www.tandfonline.com/journals/rquf20

No-dynamic-arbitrage and market impact

Jim Gatheral

Quantitative Finance, Vol. 10, No. 7, August–September 2010, 749–759

No-dynamic-arbitrage and market impact

JIM GATHERAL*

Bank of America Merrill Lynch and the Courant Institute, New York University, New York, NY 10080, USA

(*Received 31 October 2008; in final form 25 September 2009*)

Starting from a no-dynamic-arbitrage principle that imposes that trading costs should be non-negative on average and a simple model for the evolution of market prices, we demonstrate a relationship between the shape of the market impact function describing the average response of the market price to traded quantity and the function that describes the decay of market impact. In particular, we show that the widely assumed exponential decay of market impact is compatible only with linear market impact. We derive various inequalities relating the typical shape of the observed market impact function to the decay of market impact, noting that, empirically, these inequalities are typically close to being equalities.

Keywords: Stochastic volatility; Volatility modelling; Volatility smile fitting; Volatility surfaces; Stochastic jumps; Options volatility; Options pricing

1. Introduction

Market impact modeling and estimation has long been central to the market microstructure literature and also of course of great interest to traders. Indeed today, any pre-trade analytic software worthy of the name generates a pre-trade estimate of the expected cost of a proposed trade as a function of the trade size and other parameters such as liquidity and volatility.

Starting from the observations that the autocorrelation of trade signs decays very slowly with time and that the variance of the price changes is empirically linear in time, and assuming a simple model for the evolution of market prices, Bouchaud *et al*. (2004) argue that market impact is temporary and that it decays as a power law. The exponent γ of the decay of market impact and the exponent α of the power law of decay of autocorrelation of order signs are related as $\gamma \approx (1-\alpha)/2$. The empirically observed linear growth of the variance of price changes with time can be viewed as a consequence of the principle that price changes should be unpredictable.

In unrelated work, starting from a principle of no-quasi-arbitrage, Huberman and Stanzl (2004) show that permanent market impact must be linear in the trade quantity and symmetric between buys and sells.

In this article, we impose a no-dynamic-arbitrage principle that merely states that the expected cost of trading should be non-negative so that price manipulation is not possible. Starting from this principle, we extend the above-mentioned results by linking the decay of market impact to the shape of the market impact function.

In section 2 we describe the price process, showing with specific examples that it generalizes price processes previously considered in the literature. In section 3 we state the principle of no-dynamic-arbitrage and explore the special cases of permanent impact and trading in and out of a position at the same rate, independent of specific assumptions on the shapes of the market impact and decay functions. In section 4 we study exponential decay of temporary market impact, eliminating this assumption on empirical grounds. In section 5 we study power law decay of market impact, deriving inequalities imposed by no-dynamic-arbitrage and studying compatibility with different forms of the market impact function. In section 6 we explore the implications of various stylized facts: the observation by Bouchaud *et al*. (2004) that market impact must be temporary and decay as a power law, the remarkable success of the well-known square-root market impact formula, and the tail behavior of the limit order book. In section 7 we note a surprising constraint on the exponent α of decay of the autocorrelation function of order signs and discuss some potential implications. Finally, in section 8 we summarize our findings and discuss their implications.

*Email: jim.gatheral@baml.com

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2010 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/14697680903373692

2. Model setup

In the following we assume that the stock price S_t at time t is given by

$$S_t = S_0 + \int_0^t f(\dot{x}_s) G(t-s) \mathrm{d}s + \int_0^t \sigma \, \mathrm{d}Z_s, \tag{1}$$

where $\dot{x}_s$ is our rate of trading in dollars at time $s < t$, $f(\dot{x}_s)$ represents the impact of trading at time s and $G(t-s)$ is a decay factor.

S_t follows an arithmetic random walk with a drift that depends on the accumulated impacts of previous trades.† The cumulative impact of others' trading (the *trading crowd* in the terminology of Huberman and Stanzl) is assumed to be implicit in S_0 and the noise term. Drift is ignored for two reasons: drift is a lower order effect and when estimating market impact in practice, we typically average buys and sells.

In the language of Gerig (2007) and Bouchaud *et al.* (2008), $\mathbb{E}[S_t]$ gives the expected market impact at time t of a hidden order with rate of trading $\dot{x}_s$; we know how much and when we intend to trade, but other market participants do not. We assume that neither our trading nor our estimate of market impact is conditioned on the state of the market.

We refer to $f(\cdot)$ as the *instantaneous market impact function* and to $G(\cdot)$ as the *decay kernel*. The continuous time process (1) can be viewed as a limit of the discrete time process

$$S_t = \sum_{i<t} f(\delta x_i) G(t-i) + \text{noise}, \tag{2}$$

where $\delta x_i = \dot{x}_i \delta t$ is the quantity traded in some small time interval δt characteristic of the stock, and by abuse of notation, $f(\cdot)$ is the market impact function. $\delta x_i > 0$ represents a purchase and $\delta x_i < 0$ represents a sale. δt could be thought of as $1/\nu$ where ν is the trade frequency. Increasing the rate of trading $\dot{x}_i$ is equivalent to increasing the quantity traded each δt. As discussed in more detail in section 2.1, the formulation (2) corresponds to the picture of Bouchaud *et al.* (2004), where the market impact associated with each execution is independent of the state of the market.

We will show in section 2.4 by making explicit substitutions for $f(\cdot)$ and $G(\cdot)$, that expression (1) may be regarded as a generalization of processes previously considered by Bouchaud *et al.* (2004), Almgren *et al.* (2005), and Obizhaeva and Wang (2005).

2.1. More on the price process (1)

The price process (1) makes reference only to the quantity traded (or more precisely the trading rate $\dot{x}_s$). In particular, the assumed price process makes no reference to the state of the order book or the history of order flow. This reflects the practical reality that trading costs are estimated (for example, by Torre 1997, Almgren *et al.* 2005 and Engle *et al.* 2008) by aggregating all executions of a certain type, such as all VWAP‡ executions, and bucketing by trade characteristics such as duration, without conditioning on the state of the order book or order flow history. The instantaneous market impact function $f(\cdot)$ and the decay kernel $G(\cdot)$ in (1) should therefore be thought of as unconditional averages over different market conditions and so relevant to strategies that do not take the state of the market into account. In particular, we do not consider the price impact associated with strategies, such as liquidity-seeking strategies, that are conditioned on the state of the order book, order flow history or other aspects of the market.

In the econophysics literature, two distinct pictures of market impact have emerged. In the picture of Farmer *et al.* (2006) (henceforward the LF picture), market impact is permanent but depends on the state of the market. In contrast, in the picture of Bouchaud *et al.* (2004) (henceforward the BGPW picture), market impact is state-independent, temporary and decays as a power law. In a recent working paper, Bouchaud *et al.* (2008) show that these two seemingly incompatible pictures of market impact are equivalent if, in the LF picture, the assumption is made that agents forecast order flow using a long-memory linear model (ARFIMA); the price evolves identically in each picture (LF or BGPW), depending only on order flow history. In section 2 we showed that the process (1) is a continuous-time limit of a discrete price process (2) consistent with the BGPW picture. It follows that the process (1) is also consistent with the LF picture under the ARFIMA assumption.

On the other hand, as explained carefully by Gerig (2007) in his thesis, in the LF picture, if market agents condition their trading on more than just a linear forecast of order flow, the functional form (1) may no longer correctly describe the average evolution of the stock price. Indeed, Gerig (2007) interprets hidden order data from the London Stock Exchange as rejecting (2) as a realistic model for the price impact of a hidden order.

Thus, although the simple price process (1) does generalize models for market impact previously considered in the literature, it may nevertheless be inconsistent with empirical observation.

2.2. Price impact and slippage

The cost of trading can be decomposed into two components. First, the impact of our trading on the market price (the mid-price, for example): We refer to this effect as *price impact*. Our *price impact* is described by the

†The price process (1) thus violates the conditions for no price manipulation stated by Jarrow (1992) and subsequent price manipulation literature.

‡VWAP stands for 'volume-weighted average price'. The corresponding strategy attempts to deviate as little as possible from the VWAP benchmark by trading evenly in volume (or 'business') time. In this article, time should always be thought of as business time.

price update function in the terminology of Huberman and Stanzl. However, whereas in the Huberman and Stanzl setup, price impact must be permanent, in our setup, price impact may decay over time. In particular, price impact may decay as a power law, as argued by Bouchaud *et al.* (2004).

The second component of the cost of trading corresponds to market frictions such as effective bid–ask spread that affect only our execution price: We refer to this component of trading cost as *slippage* (*temporary impact* in the terminology of Huberman and Stanzl). For small volume fractions, we could think of slippage as being proxied by VWAP slippage, the average difference in price between an actual VWAP execution and the market VWAP. In what follows, we will neglect slippage; the inequalities we derive will all be weakened in practice to the extent that slippage is significant.

2.3. *Cost of trading*

Denote the number of shares outstanding at time t by x_t. Then from (1), neglecting slippage, the expected cost $C[\Pi]$ associated with a given trading strategy $\Pi = \{x_t\}$ is given by

$$C[\Pi] = \mathbb{E}\left[\int_0^T \dot{x}_t(S_t - S_0)\mathrm{d}t\right] = \int_0^T \dot{x}_t \mathrm{d}t \int_0^t f(\dot{x}_s)G(t-s)\mathrm{d}s. \tag{3}$$

The $\mathrm{d}x_t = \dot{x}_t\,\mathrm{d}t$ shares traded at time t are traded at an expected price

$$S_t = S_0 + \int_0^t f(\dot{x}_s)G(t-s)\mathrm{d}s,$$

which reflects the residual cumulative impact of all our prior trading. Obviously, cost by this definition is equivalent to expected implementation shortfall.

2.4. *Special cases*

2.4.1. *Almgren* et al. In our notation, the temporary component of the model of Almgren *et al.* (2005) corresponds to setting $G(t-s) = \delta(t-s)$ and $f(v) = \eta\sigma v^{\beta}$ with $\beta = 0.6$. Here, σ is volatility and $v_t = \dot{x}_t/V$ is a dimensionless measure of the rate of trading, where V is the market volume per unit time (average daily volume say).

In this model, temporary market impact decays instantaneously. Our trading affects only the price of our own executions; other executions are not affected. The cost of trading becomes

$$C[\Pi] = \int_0^T \dot{x}_t \mathrm{d}t \int_0^t f(\dot{x}_s)G(t-s)\mathrm{d}s = \eta\sigma\int_0^T \dot{x}_t^{\,1+\beta}\mathrm{d}t.$$

2.4.2. *Obizhaeva and Wang*. In the setup of Obizhaeva and Wang (2005), we have $G(\tau) = \mathrm{e}^{-\rho\tau}$ and $f(v) \propto v$. In this model, market impact decays exponentially and instantaneous market impact is linear in the rate of trading. The cost of trading becomes

$$C[\Pi] = \int_0^T \dot{x}_t \mathrm{d}t \int_0^t f(\dot{x}_s)G(t-s)\mathrm{d}s \propto \int_0^T \dot{x}_t \mathrm{d}t \int_0^t \dot{x}_s \exp\{-\rho(t-s)\}\mathrm{d}s.$$

Alfonsi *et al.* (2010) also assume exponential decay of market impact, but they assume a nonlinear market impact function.

2.4.3. *Bouchaud* et al. In the setup of Bouchaud *et al.* (2004), we have $f(v) \propto \log(v)$ and

$$G(t-s) \propto \frac{l_0}{(l_0 + t - s)^{\gamma}},$$

with $\gamma \approx (1-\alpha)/2$, where α is the exponent of the power law of autocorrelation of trade signs. In this model, market impact decays as a power law and instantaneous market impact is concave in the rate of trading. The cost of trading becomes

$$C[\Pi] = \int_0^T \dot{x}_t \mathrm{d}t \int_0^t f(\dot{x}_s)G(t-s)\mathrm{d}s \propto \int_0^T \dot{x}_t \mathrm{d}t \int_0^t \frac{\log(\dot{x}_s)}{(l_0 + t - s)^{\gamma}}\mathrm{d}s.$$

3. The principle of no-dynamic-arbitrage

Huberman and Stanzl define a *round trip trade* as a sequence of trades whose sum is zero. In our notation, a trading strategy $\Pi = \{x_t\}$ is a *round-trip trade* if

$$\int_0^T \dot{x}_t \mathrm{d}t = 0.$$

By analogy with another of Huberman and Stanzl's definitions, we define a *price manipulation* to be a round-trip trade Π whose expected cost $C[\Pi]$ is negative.

> *The principle of no-dynamic-arbitrage states that price manipulation is not possible.*

Equivalently, the cost of trading is non-negative, on average.

More formally, for any strategy $\{x_t\}$ such that $\int_0^T \dot{x}_t \mathrm{d}t = 0$,

$$C[\Pi] = \int_0^T \dot{x}_t \mathrm{d}t \int_0^t f(\dot{x}_s)G(t-s)\mathrm{d}s \geq 0.$$

We see that the no-dynamic-arbitrage condition imposes a relationship between the shape of the market impact function $f(\cdot)$ and the decay kernel $G(\cdot)$. We say that a market impact function $f(\cdot)$ and a decay kernel $G(\cdot)$ are *consistent* if the combination precludes price manipulation.

Remark 1: The no-dynamic-arbitrage condition refers only to the expected profit associated with a trading strategy. Dynamic arbitrage is not scalable nor does the trader know at which prices he will trade, so dynamic arbitrage is not classical arbitrage; dynamic arbitrage is closer to what is known to the practitioner as statistical arbitrage. Nor does the no-dynamic-arbitrage condition refer to the Sharpe ratio or the gain–loss ratio associated with a trading strategy. Dynamic arbitrage is thus a weaker concept than the 'quasi-arbitrage' of Huberman and Stanzl (2004) or the 'approximate arbitrage' of Bernardo and Ledoit (2000).

3.1. Permanent impact

Suppose we trade into a position at the rate $+v$ and out at the same rate $-v$. If market impact is permanent, without loss of generality, $G(\cdot)=1$ and the cost of trading becomes

$$
\begin{aligned}
C[\Pi] &= v f(v)\left\{\int_0^{T/2} \mathrm{d}t \int_0^t \mathrm{d}s - \int_{T/2}^{T} \mathrm{d}t \int_0^{T/2} \mathrm{d}s\right\} \\
&\quad - v f(-v) \int_{T/2}^{T} \mathrm{d}t \int_{T/2}^{t} \mathrm{d}s \\
&= v\frac{T^2}{8}\{-f(-v) - f(v)\}.
\end{aligned}
$$

If $f(v) > -f(-v)$, we could manipulate the market price by buying then selling at the same rate and, conversely, if $f(v) < -\mathrm{f}(-v)$, we could manipulate the market price by selling then buying at the same rate. Thus, as originally shown by Huberman and Stanzl (2004), no-dynamic-arbitrage imposes that if market impact is permanent, $f(v) = -f(-v)$.

Motivated by this observation and our own empirical observation from algorithmic trade data† that there appears to be no substantial difference between $f(v)$ and $-f(-v)$, we henceforth assume that $f(v) = -f(-v)$.

3.2. A specific strategy

Consider a strategy‡ where shares are accumulated at the (positive) constant rate v_1 and then liquidated again at the (positive) constant rate v_2. According to equation (3), and assuming $f(v) = -f(-v)$, the cost of this strategy is given by $C_{11} + C_{22} - C_{12}$ with

$$
\begin{aligned}
C_{11} &= v_1 f(v_1) \int_0^{\theta T} \mathrm{d}t \int_0^t G(t-s)\mathrm{d}s, \\
C_{22} &= v_2 f(v_2) \int_{\theta T}^{T} \mathrm{d}t \int_{\theta T}^{t} G(t-s)\mathrm{d}s, \qquad (4)\\
C_{12} &= v_2 f(v_1) \int_{\theta T}^{T} \mathrm{d}t \int_0^{\theta T} G(t-s)\mathrm{d}s,
\end{aligned}
$$

where θ is such that $v_1\theta T - v_2(T - \theta T) = 0$, so

$$\theta = \frac{v_2}{v_1 + v_2}.$$

The principle of no-dynamic-arbitrage imposes that

$$C_{11} + C_{22} - C_{12} \geq 0.$$

Intuitively, the cross-term C_{12} represents the component of the cost of stock sales (purchases) associated with the price impact of prior stock purchases (sales). If the cross-term $C_{12} = 0$, there is no dynamic arbitrage and price manipulation is not possible. In particular, in the model of Almgren *et al.* (2005) where market impact decays instantaneously and the market price has no history of prior trading, there is no dynamic arbitrage.

3.3. Trading in and out at the same rate

Now, suppose only that $G(\cdot)$ is strictly decreasing (again with $f(v) = -f(-v)$). Then the cost of acquiring a position at the constant rate v then liquidating it again at the same rate is given by

$$
\begin{aligned}
C[\Pi] &= v f(v)\left\{\int_0^{T/2} \mathrm{d}t \int_0^t G(t-s)\mathrm{d}s + \int_{T/2}^{T} \mathrm{d}t \int_{T/2}^{t} G(t-s)\mathrm{d}s \right.\\
&\quad \left. - \int_{T/2}^{T} \mathrm{d}t \int_0^{T/2} G(t-s)\mathrm{d}s\right\} \\
&= v f(v)\left\{\int_0^{T/2} \mathrm{d}t \int_0^t [G(t-s) - G(t+T/2-s)]\mathrm{d}s \right.\\
&\quad \left. + \int_{T/2}^{T} \mathrm{d}t \int_{T/2}^{t} [G(t-s) - G(T-s)]\mathrm{d}s\right\} > 0,
\end{aligned}
$$

and there is no price manipulation.

We conclude that if price manipulation is possible with this specific strategy, it must involve trading in and out of a position at different rates.

4. Exponential decay of market impact

Suppose now that the decay kernel has the form

$$G(t-s) = e^{-\rho(t-s)},$$

and that we acquire a position at rate v_1, liquidating it again at the rate v_2. Then, explicit computation of all the integrals in (4) gives

$$
\begin{aligned}
C_{11} &= v_1 f(v_1)\frac{1}{\rho^2}\{e^{-\rho\theta T} - 1 + \rho\theta T\}, \\
C_{12} &= v_2 f(v_1)\frac{1}{\rho^2}\{1 + e^{-\rho T} - e^{-\rho\theta T} - e^{-\rho(1-\theta)T}\}, \qquad (5)\\
C_{22} &= v_2 f(v_2)\frac{1}{\rho^2}\{e^{-\rho(1-\theta)T} - 1 + \rho(1-\theta)T\}.
\end{aligned}
$$

Again, the no-dynamic-arbitrage principle forces a relationship between the instantaneous impact function

† The empirical microstructure literature taken as a whole is inconclusive on the issue of symmetry of market impact between buys and sells.
‡ The constant rate trading strategy can be thought of as a continuous-time approximation to the previously mentioned and widely used VWAP ('Volume Weighted Average Price') algorithm as discussed, for example, by Konishi (2002).

$f(\cdot)$ and the decay kernel $G(\cdot,\cdot)$:

$$C_{11} + C_{22} - C_{12} \geq 0.$$

After making the substitution $\theta = v_2/(v_1 + v_2)$, we obtain

$$\begin{aligned} & v_1 f(v_1)\left[e^{-v_2\rho T/(v_1+v_2)} - 1 + \frac{v_2 \rho T}{v_1 + v_2}\right] \\ &\quad + v_2 f(v_2)\left[e^{-v_1\rho T/(v_1+v_2)} - 1 + \frac{v_1 \rho T}{v_1 + v_2}\right] \\ &\quad - v_2 f(v_1)\left[1 + e^{-\rho T} - e^{-y1\rho T/(v_1+v_2)} - e^{-v_2\rho T/(v_1+v_2)}\right] \geq 0. \end{aligned} \tag{6}$$

We note that the first two terms are always positive, so price manipulation is only possible if the third term (C_{12}) dominates the others.

4.0.1. Example: $f(v) = \sqrt{v}$. Let $v_1 = 0.2$, $v_2 = 1$, $\rho = 1$, $T = 1$. Then the cost of liquidation is given by

$$C = C_{11} + C_{22} - C_{12} = -0.001705 < 0.$$

Since this expression depends on ρ and T only through the product ρT, we see that, for any choice of ρ, we can find a combination $\{v_1, v_2, T\}$ such that a round trip with no net purchase or sale of stock is profitable. We conclude that if market impact decays exponentially, no-dynamic-arbitrage excludes a square root instantaneous impact function.

More generally, expanding expression (6) in powers of ρT, no-dynamic-arbitrage imposes that

$$\frac{v_1 v_2 [v_1 f(v_2) - v_2 f(v_1)](\rho T)^2}{2(v_1 + v_2)^2} + O((\rho T)^3) \geq 0.$$

We see that price manipulation is always possible for small (ρT) unless $f(v) \propto v$. That is, for any value of the decay rate ρ, we may choose T sufficiently small to achieve price manipulation and so we may state† the following lemma.

Lemma 4.1: *If temporary market impact decays exponentially, price manipulation is possible unless the instantaneous market impact function $f(v)$ is directly proportional to v.*

Taking the limit $\rho \to 0^+$, we obtain the following corollary.

Corollary 4.2: *Nonlinear permanent market impact is inconsistent with the principle of no-dynamic-arbitrage.*

Again, as originally shown by Huberman and Stanzl.

4.1. Linear permanent market impact

If $f(v) = \eta v$ for some $\eta > 0$ and $G(t-s) = 1$, the cost of trading becomes

$$C[\Pi] = \eta \int_0^T \dot{x}_t \mathrm{d}t \int_0^t \dot{x}_s \mathrm{d}s = \frac{\eta}{2}(x_T - x_0)^2.$$

The trading cost per share is then given by

$$\frac{C[\Pi]}{|x_T - x_0|} = \eta |x_T - x_0|,$$

which is *independent* of the details of the trading strategy (depending only on the initial and final positions) and linear in the trade quantity. In particular, a linear permanent component of market impact can never give rise to price manipulation because it contributes nothing to the cost of a round-trip trade.

4.2. Excluding exponential decay of market impact

Empirically (see Almgren *et al.* 2005, for example), market impact is concave in v for small v. Also, market impact must be convex for very large v; imagine submitting a sell order whose size is much greater than the quantity available on the bid side of the order book.‡ It follows that no-dynamic-arbitrage together with any reasonable instantaneous market impact function $f(\cdot)$ excludes exponential decay of market impact as a realistic assumption.

5. Power-law decay of market impact

Having excluded exponential decay of market impact as a realistic assumption, suppose instead that the decay kernel has the form

$$G(t-s) = \frac{1}{(t-s)^\gamma}, \quad 0 < \gamma < 1.$$

Then, explicit computation of all the integrals in (4) gives

$$\begin{aligned} C_{11} &= v_1 f(v_1) \frac{T^{2-\gamma}}{(1-\gamma)(2-\gamma)} \theta^{2-\gamma}, \\ C_{22} &= v_2 f(v_2) \frac{T^{2-\gamma}}{(1-\gamma)(2-\gamma)} (1-\theta)^{2-\gamma}, \\ C_{12} &= v_2 f(v_1) \frac{T^{2-\gamma}}{(1-\gamma)(2-\gamma)} \{1 - \theta^{2-\gamma} - (1-\theta)^{2-\gamma}\}. \end{aligned} \tag{7}$$

According to the principle of no-dynamic-arbitrage, substituting $\theta = v_2/(v_1 + v_2)$, we must have

$$\begin{aligned} & f(v_1)\{v_1 {v_2}^{1-\gamma} - (v_1 + v_2)^{2-\gamma} + {v_1}^{2-\gamma} + {v_2}^{2-\gamma}\} \\ &\quad + f(v_2) {v_1}^{2-\gamma} \geq 0. \end{aligned} \tag{8}$$

If $\gamma = 0$, the no-dynamic-arbitrage condition (8) reduces to

$$f(v_2) v_1 - f(v_1) v_2 \geq 0,$$

so, again, we must have $f(v) \propto v$.

If $\gamma = 1$, equation (8) reduces to

$$f(v_1) + f(v_2) \geq 0.$$

† Friz *et al.* have shown that $f(v) \propto v$ is consistent with any convex non-increasing decay kernel $G(\cdot)$. In a forthcoming paper, Schied *et al.* provide a complete characterization of all such functions $G(\cdot)$.
‡ As has happened in the past, notably on the Tokyo Stock Exchange.

So long as $f(\cdot) \geq 0$, there is no constraint on $f(\cdot)$ when $\gamma = 1$. We see that in contrast to the case of exponential decay of market impact, power-law decay of market impact may be compatible with realistic shapes of the market impact function $f(\cdot)$.

5.1. The limit $v_1 \ll v_2$ and $0 < \gamma < 1$

In this limit, we accumulate stock much more slowly than we liquidate it. Let $v_1 = \epsilon v$ and $v_2 = v$ with $\epsilon \ll 1$. Then, in the limit $\epsilon \to 0$, with $0 < \gamma < 1$, equation (8) becomes

$$\begin{aligned} & f(\epsilon v)\{\epsilon - (1+\epsilon)^{2-\gamma} + \epsilon^{2-\gamma} + 1\} + f(v)\epsilon^{2-\gamma} \\ & \quad \sim -f(\epsilon v)(1-\gamma)\epsilon + f(v)\epsilon^{2-\gamma} \geq 0, \end{aligned}$$

so for ϵ sufficiently small we have

$$\frac{f(\epsilon v)}{f(v)} \leq \frac{\epsilon^{1-\gamma}}{1-\gamma}. \tag{9}$$

If the condition (9) is not satisfied, price manipulation is possible by accumulating stock slowly, maximally splitting the trade, then liquidating it rapidly.

5.2. Special cases

5.2.1. Power-law impact: $f(v) \propto v^\delta$. If $f(v) \sim v^\delta$ as found by, for example, Almgren *et al.* (2005), the no-dynamic-arbitrage condition (9) reduces to $\epsilon^{1-\gamma-\delta} \geq 1-\gamma$, so we must have $\gamma + \delta \geq 1$. We can state this result formally as follows.

Lemma 5.1: (*small v no-dynamic-arbitrage condition*) *If $G(\tau) = \tau^{-\gamma}$ and $f(v) \propto v^\delta$, dynamic-no-arbitrage imposes that*

$$\gamma + \delta \geq 1.$$

5.2.2. Log impact: $f(v) \propto \log(v/v_0)$. Bouchaud *et al.* (2004) find that their empirical results are well-described by $f(v) \sim \log v$. Of course, $f(\cdot)$ must be non-negative, so the trading rate v must always be greater than some minimum trading rate v_0. For example, one could think of one share every trade as being the minimum rate. In the case of a stock that trades 10 million shares a day, 10,000 times, and where the average trade size is 1000, we would obtain $v_0 = 0.10\%$. Noting that

$$\log v = \lim_{\delta \to 0} \frac{v^\delta - 1}{\delta},$$

we might guess that price manipulation is possible for all $\gamma < 1$. In fact, the precise condition on γ depends on the minimum trading rate v_0.

For example, substituting $v_0 = 0.001$, $v_1 = 0.15$, $v_2 = 1.0$ and $\gamma = 1/2$ into the arbitrage condition (8) with $f(v) = \log(v/v_0)$ gives the expected cost of the round-trip trade as -0.0028 which constitutes price manipulation (i.e. dynamic arbitrage).

For the general case, suppose that we trade into a position at some rate v with $v/v_0 > 1$ and $v \ll 1$. Then substituting into (8) gives

$$\begin{aligned} & \log(v/v_0)\{v - (1+v)^{2-\gamma} + v^{2-\gamma} + 1\} + \log(1/v_0)v^{2-\gamma} \\ & \quad = \log(v/v_0)\{-(1-\gamma)v + O(v^{2-\gamma})\} + \log(1/v_0)v^{2-\gamma}. \end{aligned}$$

So, for every $\gamma < 1$, provided v_0 is sufficiently small, we can find a small enough $v > v_0$ such that price manipulation is possible.

The choice of market impact function $f(v) \sim \log(v)$ is therefore inconsistent with power-law decay (with $\gamma < 1$) of market impact in the limit $v_0 \to 0$.

5.3. The limit $v_1 > v_2$ and $0 < \gamma < 1$

Suppose we accumulate stock at some very high rate v_1 and liquidate at some lower rate v_2. This is the well-known *pump and dump* strategy.† Setting $v = v_2/v_1 < 1$ and substituting into the dynamic-no-arbitrage condition (8) with power-law decay of market impact, we obtain

$$f(v_1)\{v^{1-\gamma} - (1+v)^{2-\gamma} + 1 + v^{2-\gamma}\} + f(v_2) \geq 0. \tag{10}$$

Since $f(v_2)$ is always positive, price manipulation is possible only if

$$h(v,\gamma) := v^{1-\gamma} - (1+v)^{2-\gamma} + 1 + v^{2-\gamma} < 0. \tag{11}$$

Expression (11) is shown in appendix A to be equivalent to the condition

$$\gamma < \gamma^* := 2 - \frac{\log 3}{\log 2} \approx 0.415.$$

So if $\gamma > \gamma^*$, price manipulation is not possible.

Is price manipulation possible if $\gamma < \gamma^*$? We first note that $h(v, \gamma)$ decreases as $v \to 1$, so price manipulation is maximized near $v = 1$. From section 3.3, we already know that there is no dynamic arbitrage when trading in and out at the same rate, as can be checked again easily in this case by substituting $v_2 = v_1$ into equation (10) to obtain

$$f(v_1)\{4 - 2^{2-\gamma}\} \geq 0, \quad \text{for all } \gamma \geq 0.$$

Observe that, in practice, we cannot exceed some maximum rate of trading $v_{\max}$ corresponding, for example, to continuously exhausting the available quantity in the order book. Without loss of generality, set $v_{\max} = 1$. Then, as $v_i \to 1$, we must have $f(v_i) \to \infty$. Specifically, in section 6.3 we will argue that

$$f(v_i) \sim \frac{1}{(1-v_i)^\nu}, \quad \text{as } v_i \to 1,$$

for some $\nu > 0$.

With $\epsilon \ll 1$, substituting $v_1 = 1 - \epsilon^2$ and $v_2 = 1 - \epsilon$ into equation (10) and in the limit $\epsilon \to 0$, we see that price manipulation is possible if

$$\frac{3 - 2^{2-\gamma}}{\epsilon^{2\nu}} + \frac{1}{\epsilon^\nu} < 0.$$

† See http://www.sec.gov/answers/pumpdump.htm for a definition.

For any $\gamma < \gamma^*$, we can choose ϵ sufficiently small to ensure that the first term dominates the second, resulting in price manipulation.

We deduce that, for a market impact function $f(\cdot)$ of the above form with any exponent $\nu > 0$, the no-arbitrage condition is as follows.

Lemma 5.2: (*large size no arbitrage condition*)

$$\gamma \geq \gamma^* = 2 - \frac{\log 3}{\log 2}.$$

6. Stylized facts

In this section we explore the implications of three stylized facts, or empirical regularities, that have been noted either explicitly or implicitly in the literature.

6.1. *Power-law decay of market impact*

As mentioned earlier, Bouchaud *et al.* (2004) have argued convincingly that if price impact is independent of the state of the market as assumed in our price process (1), it must be temporary† and decay as a power law. We proceed to outline their argument.

Suppose market impact is permanent and proportional to some function $f(n)$ of the trade size n. Then the change in price after N trades is given by

$$\Delta P = \sum_{i}^{N} \eta \epsilon_i f(n_i),$$

where ϵ_i and n_i denote the sign and size of the ith trade, respectively. If $\text{Cov}[\epsilon_i, \epsilon_j] = 0$ for $i \neq j$, the variance of the price change is given by

$$\text{Var}[\Delta P] = \eta^2 N \mathbb{E}[f(n_i)^2],$$

which grows linearly with N. Empirically, however, we find that autocorrelation of trade signs shows power-law decay with an exponent $\alpha < 1$ (corresponding to very slow decay). In this case, with $\text{Cov}[\epsilon_i, \epsilon_j] \propto |j - i|^{-\alpha}$, the cross-term in the computation of daily variance dominates and we obtain

$$\begin{aligned} \text{Var}[\Delta P] &\sim \sum_{i \neq j} \text{Cov}[\epsilon_i, \epsilon_j] \mathbb{E}[f(n_i)] \mathbb{E}[f(n_j)] \\ &\sim N^{2-\alpha}, \quad \text{as } N \to \infty. \end{aligned}$$

The variance of price changes grows superlinearly with N.

Empirically, we find that, to a very good approximation, $\text{Var}[\Delta P] \propto N$. If the variance of price changes grew superlinearly as $N^{2-\alpha}$, returns would be serially correlated and market efficiency would be glaringly violated: simple trend-following strategies would be consistently profitable. We conclude that market impact cannot be permanent.

If, on the other hand, market impact were temporary and decayed as $1/T^{\gamma}$, we would have

$$\begin{aligned} \text{Var}[\Delta P] &\sim \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \frac{\mathbb{E}[f(n_i)] \mathbb{E}[f(n_j)]}{(N-i)^{\gamma} (N-j)^{\gamma} |j-i|^{\alpha}} \\ &\sim N^{2-\alpha-2\gamma}, \quad \text{as } N \to \infty. \end{aligned}$$

In this case, the variance of price changes grows linearly with N only if

$$\gamma = \frac{1-\alpha}{2}. \tag{12}$$

For the French stock (France Télécom) considered by Bouchaud *et al.* (2004), the exponent $\alpha \approx 0.2$, so $\gamma \approx 0.4$.

6.2. *The square-root formula, γ and δ*

The *square-root formula* has the form

$$\text{Cost} = \text{Spread term} + c\sigma \sqrt{\frac{n}{V}},$$

where n is the number of shares to be traded, σ is the volatility of the stock (in daily units), and V is the average daily volume of the stock. With the terminology of section 2.2, we could think of the spread term as representing slippage and the second square-root term as representing price impact.

The square-root formula has been widely used in practice for many years to generate a pre-trade estimate of transactions cost. As noted in chapter 16 of Grinold and Kahn (1995), this formula is consistent with the trader rule-of-thumb that says that it costs roughly one day's volatility to trade one days' volume. Moreover, various studies of market impact costs, notably the study documented in chapter 7 of the *Barra Market Impact Model Handbook* (Torre 1997), have found the square-root formula to fit transactions cost data remarkably well.

Interestingly, the square-root formula implies that the cost of liquidating a stock is independent of the time taken: the formula refers neither to the duration of the trade nor to the trading strategy adopted. Fixing market volume and volatility, price impact depends only on trade size. At first, this claim seems to contradict our intuition that expected price impact should increase as the rate of trading v increases. Indeed, this intuition is empirically verified for very large trading rates that necessitate trading in sizes large relative to the quantity available in the order book. For reasonable trading rates, however (volume fractions of 1% to 25%, say), it does seem to be the case that price impact is roughly independent of trade duration, as can be checked, for example, by referring to tables 1 and 2 of Engle *et al.* (2008), where conditioning on trade size, cost seems to be only weakly dependent on trade duration.

† As mentioned earlier in section 4.1, even if there were a permanent component of market impact linear in trade quantity, it would contribute nothing to the cost of a round-trip trade and would leave all of our results unchanged.

According to our simple trade superposition model, from equation (7), the price impact associated with a VWAP execution with duration T is proportional to

$$v f(v) T^{2-\gamma}.$$

Noting that $v = n/(VT)$ (again with V denoting average daily volume), and putting $f(v) \propto v^{\delta}$, the impact cost per share is then proportional to

$$v^{1+\delta} T^{1-\gamma} = \left(\frac{n}{V}\right)^{\delta} T^{1-\gamma-\delta}.$$

If $\gamma + \delta = 1$, the cost per share is *independent* of T and, in particular, if $\gamma = \delta = 1/2$, the impact cost per share is proportional to $\sqrt{n/V}$, recovering the square-root formula.

We see that the square-root formula is consistent with both power-law decay of market impact and a power-law form of the market impact function $f(\cdot)$. Almgren *et al.* (2005) estimate $\delta \approx 0.6$ and from section 6.1, we have the estimate from Bouchaud *et al.* (2004) of $\gamma \approx 0.4$.

Recall the no-dynamic-arbitrage condition from section 5.1:

$$\gamma + \delta \geq 1.$$

Putting the two empirical estimates together, we have $\gamma + \delta \approx 0.4 + 0.6 = 1$!

6.3. *Very high trading rates*

Bouchaud *et al.* (2002) derive the following approximation to the average density $\rho(\hat{\Delta})$ of orders as a function of a re-scaled distance $\hat{\Delta}$ from the price level at which the order is placed to the current price:

$$\rho(\hat{\Delta}) = e^{-\hat{\Delta}} \int_0^{\hat{\Delta}} \mathrm{d}u \frac{\sinh(u)}{u^{1+\mu}} + \sinh(\hat{\Delta}) \int_{\hat{\Delta}}^{\infty} \mathrm{d}u \frac{e^{-u}}{u^{1+\mu}}, \quad (13)$$

where μ is the exponent in the empirical power-law distribution of new limit orders. With $\mu = 0.6$ as estimated by Bouchaud *et al.* (2002), we obtain the density plotted in figure 1. Computing the cumulative order density (book depth) as a function of the price level $\hat{\Delta}$ and switching the axes, we may compute the virtual impact function of Weber and Rosenow (2005), a function that measures the price impact conditional on trading a given quantity instantaneously using a single market order. With $\mu = 0.6$, we obtain the virtual impact function graphed in figure 2. The virtual impact function is convex for large size. Also, it is not possible to trade in a size greater than the quantity currently available in the order book so price impact increases without limit as $n \to n_{\max}$. Given the $\hat{\Delta}^{-1-\mu}$ shape of the tail of the order book, we have

$$n_{\max} - n(\hat{\Delta}) \sim \int_{\hat{\Delta}}^{\infty} \frac{\mathrm{d}u}{u^{1+\mu}} = \frac{\mu}{\hat{\Delta}^{\mu}},$$

where $n(\hat{\Delta})$ is the cumulative share quantity available up to level $\hat{\Delta}$ in the order book. Inverting this relationship, we see that instantaneous market impact has the tail behavior

$$\Delta P \sim \frac{1}{(n - n_{\max})^{1/\mu}}.$$

Then, following the discussion of section 2 where the trading rate v was argued to be a proxy for the size n of individual trades, we have

$$\Delta P \sim \frac{1}{(v - v_{\max})^{1/\mu}},$$

for sufficiently large v.

We observe that the form of this relationship is consistent with the assumptions of lemma 5.2, so we must have

$$\gamma \geq 2 - \frac{\log 3}{\log 2} \approx 0.415. \quad (14)$$

7. A surprising constraint on α

Equations (12) and (14) together impose the following constraint on the autocorrelation of trade-signs:

$$\alpha \leq 1 - 2\gamma^* \approx 0.17.$$

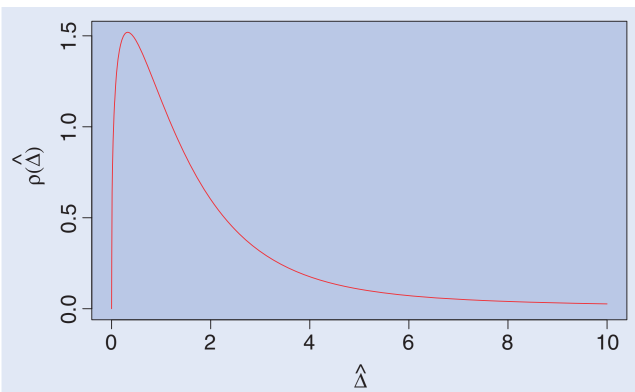

Figure 1. The red line is a plot of the order density $\rho(\hat{\Delta})$ with $\mu = 0.6$.

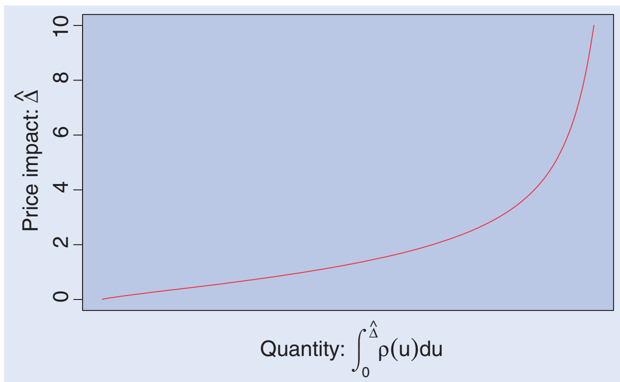

Figure 2. Switching the x and y axes in a plot of the cumulative order density gives the virtual impact function plotted. The red line corresponds to $\mu = 0.6$ as before.

Faster decay is ruled out by no-dynamic-arbitrage.

While this constraint is arguably compatible with the empirical results of Bouchaud *et al.* (2004) who find $\alpha \approx 0.2$ for France Télécom, it does not seem to be compatible with the later estimates in table 2 of Bouchaud *et al.* (2006) where α ranges from 0.125 to 0.69. Nor does it seem to be compatible with the empirical estimates of Lillo and Farmer (2004) (who estimate $\alpha \sim 0.6$) or the subsequent theoretical explanation in terms of order-splitting of Lillo *et al.* (2005) where α is related to the exponent of the power-law distribution of order sizes (again leading to $\alpha \sim 0.6$).

It could be that the high participation rate price manipulation strategy ('pump and dump') is not really practical (or legal), in which case the constraint $\gamma \geq \gamma^*$ would not apply. Or, there may actually be price manipulation in the market.

More interestingly, it could be that the form of our assumed price process (1), where, for example, we have assumed factorization of the dependence on volume and the dependence on time, is inconsistent with empirical observation.

8. Summary and discussion of results

Assuming a simple model for the price process that generalizes those considered by previous authors, we investigated the implications of no-dynamic-arbitrage for various combinations of the instantaneous market impact function $f(v)$ and the decay kernel $G(\tau)$. We showed that exponential decay of market impact $G(\tau) \sim \exp^{-\rho\tau}$ was consistent only with linear instantaneous market impact $f(v) \propto v$. Since market impact is well known to be concave in trade size (at least for reasonable sizes), we eliminated exponential decay of market impact as a reasonable assumption. Also, the combination $f(v) \propto \log(v/v_0)$ with $G(\tau) \sim \tau^{-\gamma}$ turned out to be problematic in the sense that price manipulation is always possible in the limit that the minimum trading rate $v_0 \to 0$.

We then showed that if the market impact function is of the form $f(v) \propto v^{\delta}$ and the decay kernel of the form $G(\tau) \sim \tau^{-\gamma}$, no-dynamic-arbitrage imposes that

$$\gamma + \delta \geq 1.$$

We noted that if the average cost of a (not-too-aggressive) VWAP execution is roughly independent of duration, as implied by the widely used square-root market impact formula, the exponent δ of the power law of market impact should satisfy

$$\delta + \gamma \approx 1.$$

By considering the tail of the limit-order book and a high trading rate price manipulation strategy, we deduced that

$$\gamma \geq \gamma^* := 2 - \frac{\log 3}{\log 2} \approx 0.415. \qquad (15)$$

Bouchaud *et al.* (2004) have shown that, for a state-independent market impact process like (1) with $G(\tau) \sim \tau^{-\gamma}$, the exponent γ of the power law of decay of market impact and the exponent α of the decay of autocorrelation of trade signs must balance to ensure diffusion (variance increasing linearly with time):

$$\gamma \approx (1 - \alpha)/2. \qquad (16)$$

In particular, if the autocorrelation of trade signs has long memory, we must have $\gamma \leq 1/2$.

Putting (16) and (15) together, we found the following constraint on the autocorrelation of trade signs:

$$\alpha \leq 1 - 2\gamma^* \approx 0.17.$$

Faster decay is ruled out by no-dynamic-arbitrage.

We noted some empirical estimates: $\gamma \approx 0.4$ (Bouchaud *et al.* 2004) and for not-too-aggressive trading strategies, $\delta \approx 0.6$ (Almgren *et al.* 2005).

Figure 3 shows a schematic representation of our results. In particular, we note that although values of δ and γ may be close to the boundary of allowable values for typical not-too-aggressive trading strategies, nothing precludes δ from moving away from this boundary when the trading strategy is aggressive and the rate of trading is very high. For example, an exponent $\mu = 0.6$ in the power law of limit order arrivals would give rise to $\delta = 1/0.6 \approx 1.67$.

We emphasize that none of the inequalities we derived are hard in practice. As noted earlier in section 2.2, all of the derived inequalities are weakened to the extent that market frictions such as slippage become significant. Moreover, as noted in section 3, our 'dynamic-arbitrage' is a weaker concept than the 'quasi-arbitrage' of Huberman and Stanzl (2004) or the 'approximate arbitrage' of Bernardo and Ledoit (2000). Imposing no quasi-arbitrage or no approximate arbitrage rather than no dynamic arbitrage may therefore weaken the inequalities we have derived under no-dynamic-arbitrage.

With these caveats and with the assumed price process (1), we have nevertheless found conditions under which there exist price manipulation strategies that work *on average*, without taking the state of the market into account; effectively submitting orders without either looking at the state of the order book or the history of order flow. *A forteriori*, under the same conditions, price manipulation must be possible if trading is conditioned on the state of the market.† No-dynamic arbitrage thus constrains modeling of the market impact of hidden orders, in particular eliminating exponential decay of market impact as a reasonable assumption.

Obviously, our results are all in the context of the simple price process (1): It could be, for example, that the decay of market impact is dependent on trade size so that the market impact of hidden orders is not separable into

† The converse is obviously not true: There could exist price manipulation strategies that are conditioned on the state of the market that are not profitable on average when not so conditioned.

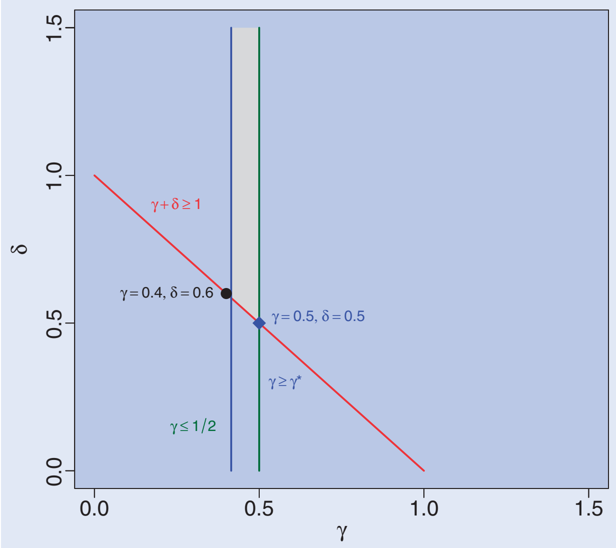

Figure 3. Combinations to the right of the red line satisfy $\gamma + \delta \geq 1$, to the right of the blue line $\gamma \geq \gamma^*$, to the left of the green line $\gamma \leq 1/2$ and in the shaded intersection, the allowable values of γ and δ consistent with the stylized facts of market impact. The black dot represents the empirical estimates $\gamma \approx 0.4$ and $\delta \approx 0.6$ and the blue diamond, the values $\gamma = 0.5$ and $\delta = 0.5$ consistent with the square-root formula.

volume- and time-dependent factors. Nor have we investigated every possible combination of price impact function $f(\cdot)$ and decay kernel $G(\cdot)$; understanding what combinations of functions are consistent with no-dynamic-arbitrage would be interesting in its own right. And even under the assumptions $f(v) \propto v^{\delta}$ and $G(\tau) \sim \tau^{-\gamma}$, we have only demonstrated that if the parameter inequalities we derived are violated, price manipulation is possible; we have not proved the converse. Specifically, price manipulation may be possible using alternative strategies (such as discrete trading strategies) even if the inequalities derived above are satisfied. If this were to be shown, the inequalities we derived would obviously be tightened further.

Finally, as discussed in section 2.1, the assumed price process (1) is consistent with the BGPW picture of market impact, which is itself equivalent to the LF picture under the ARFIMA assumption whereby market participants use a long-memory linear model to forecast order flow. The fact that empirically estimated values of α do not seem to satisfy $\alpha \leq 1 - 2\gamma^*$, thereby permitting price manipulation at high trading rates, may indicate that the price process (1) is too simplistic and, therefore, as suggested by Gerig (2007) and Bouchaud *et al.* (2008), that market participants condition their trading on more than just linear forecasts of order flow.

Acknowledgements

I wish to thank Peter Friz, Roger Lee and my Bank of America Merrill Lynch colleagues Joerg Osterrieder, Sergei Satchkov, Chris Smith and Bruno White for helpful comments and suggestions. I am also grateful to the anonymous referees for their valuable comments.

References

Alfonsi, A., Fruth, A. and Schied, A., Optimal execution strategies in limit order books with general shape functions. *Quant. Finance*, 2010, **10**, 143–157.
Almgren, R., Thum, C., Hauptmann, E. and Li, H., Equity market impact. *Risk*, 2005 July, 57–62.
Bernardo, A. and Ledoit, O., Gain, loss, and asset pricing. *J. Polit. Econ.*, 2000, **108**, 144–172.
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. Schenk-Hoppe, 2008 (Academic Press: New York).
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**, 176–190.
Bouchaud, J.-P., Kockelkoren, J. and Potters, M., Random walks, liquidity molasses and critical response in financial markets. *Quant. Finance*, 2006, **6**, 115–123.
Bouchaud, J.-P., Mézard, M. and Potters, M., Statistical properties of stock order books: empirical results and models. *Quant. Finance*, 2002, **2**, 251–256.
Engle, R.F., Ferstenberg, R. and Russell, J., Measuring and modeling execution cost and risk. Discussion paper, University of Chicago Graduate School of Business, 2008.
Farmer, J.D., Gerig, A., Lillo, F. and Mike, S., Market efficiency and the long-memory of supply and demand: is price impact variable and permanent or fixed and temporary? *Quant. Finance*, 2006, **6**, 107–112.
Gerig, A.N., A theory for market impact: How order flow affects stock price. PhD thesis, University of Illinois at Urbana-Champaign, 2007.
Grinold, R.C. and Kahn, R.N., *Active Portfolio Management*, 1995 (McGraw-Hill: New York).
Huberman, G. and Stanzl, W., Price manipulation and quasi-arbitrage. *Econometrica*, 2004, **72**, 1247–1275.
Jarrow, R.A., Market manipulation, bubbles, corners and short squeezes. *J. Financial Quant. Anal.*, 1992, **27**, 311–335.
Konishi, H., Optimal slice of a VWAP trade. *J. Financial Mkts*, 2002, **5**, 197–221.
Lillo, F. and Farmer, J.D., The long memory of the efficient market. *Stud. Nonlinear Dynam. Econometr.*, 2004, **8**, Article 1.
Lillo, F., Mike, S. and Farmer, J.D., Theory for long memory in supply and demand. *Phys. Rev. E*, 2005, **71**, 066122.
Obizhaeva, A. and Wang, J., Optimal trading strategy and supply/demand dynamics. SSRN eLibrary, 2005.
Torre, N., *BARRA Market Impact Model Handbook*, 1997.
Weber, P. and Rosenow, B., Order book approach to price impact. *Quant. Finance*, 2005, **5**, 357–364.

Appendix A: Proof of the large size no-arbitrage condition

We want to find the minimum value γ^* of the exponent γ in the power law of decay of market impact such that there is no dynamic arbitrage for $\gamma > \gamma^*$.

According to (10), there is arbitrage only if

$$h(v, \gamma) := v^{1-\gamma} - (1+v)^{2-\gamma} + 1 + v^{2-\gamma} < 0, \quad \text{for some } v \in (0, 1). \tag{A1}$$

Setting $v = 1$ in (A1), and solving for γ, we find that there is arbitrage only if

$$\gamma < 2 - \frac{\log 3}{\log 2} =: \gamma^*.$$

We note that

$$\begin{aligned}\partial_\gamma h(v,\gamma) &= (1+v)^{2-\gamma}\log(1+v) - v^{1-\gamma}(1+v)\log v \\ &\geq 0, \quad \forall v \in (0,1).\end{aligned}$$

So, if $h(v,\gamma^*)$ reaches its minimum at $v = 1$, the result is proved. To show this, with $\alpha := \log 3/\log 2$, define the function

$$\tilde{h}(v) := h(v,\gamma^*) - h(1,\gamma^*) = v^{\alpha-1} - (v+1)^\alpha + 1 + v^\alpha.$$

Then $\tilde{h}(1) = \tilde{h}(0) = 0$ and the second derivative of $\tilde{h}(\cdot)$ with respect to v is given by

$$\begin{aligned}\partial_{v,v}\tilde{h}(v) &= (\alpha-1)\left\{\frac{v\alpha+\alpha-2}{v^{3-\alpha}} - \frac{\alpha}{(1+v)^{2-\alpha}}\right\} \\ &\leq (\alpha-1)\left\{\frac{v\alpha+\alpha-2}{v^{3-\alpha}} - \frac{3}{4}\alpha\right\}.\end{aligned}$$

This latter expression reaches its maximum at

$$v^* := \frac{3}{\alpha} - 1 \approx 0.89.$$

Then

$$\begin{aligned}\partial_{v,v}\tilde{h}(v) &\leq (\alpha-1)\left\{\frac{v^*\alpha+\alpha-2}{v^{*3-\alpha}} - \frac{3}{4}\alpha\right\} \\ &= (\alpha-1)\left\{\frac{1}{v^{*3-\alpha}} - \frac{3}{4}\alpha\right\} \\ &< 0.\end{aligned}$$

Thus over the range $(0,1)$, $\tilde{h}(\cdot)$ is convex down and $\tilde{h}(1) = \tilde{h}(0) = 0$. Then $\tilde{h}(v) \geq 0$ for all $v \in (0,1)$, proving the result.