How efficiency shapes market impact

J. Doyne Farmer[ab], Austin Gerig[c], Fabrizio Lillo[bde] & Henri Waelbroeck[f]

[a] Institute for New Economic Thinking at the Oxford Martin School and Mathematical Institute, University of Oxford, Eagle House, Walton Well Road, Oxford, OX2 6ED, UK.

[b] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.

[c] CABDyN Complexity Centre, Saïd Business School, University of Oxford, Oxford, UK.

[d] Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, Pisa 56126, Italy.

[e] Dipartimento di Fisica e Chimica, viale delle Scienze I-90128, Palermo, Italy.

[f] Portware LLC, 233 Broadway, 24th Floor, New York, NY 10279, USA.
Published online: 01 Nov 2013.

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

How efficiency shapes market impact

J. DOYNE FARMER*†‡, AUSTIN GERIG§, FABRIZIO LILLO‡¶‖
and HENRI WAELBROECK††

†Institute for New Economic Thinking at the Oxford Martin School and Mathematical Institute, University of Oxford,
Eagle House, Walton Well Road, Oxford, OX2 6ED, UK
‡Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
§CABDyN Complexity Centre, Saïd Business School, University of Oxford, Oxford, UK
¶Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, Pisa 56126, Italy
‖Dipartimento di Fisica e Chimica, viale delle Scienze I-90128, Palermo, Italy
††Portware LLC, 233 Broadway, 24th Floor, New York, NY 10279, USA

We develop a theory for the market impact of large trading orders, which we call *metaorders* because they are typically split into small pieces and executed incrementally. Market impact is empirically observed to be a concave function of metaorder size, i.e. the impact per share of large metaorders is smaller than that of small metaorders. We formulate a stylized model of an algorithmic execution service and derive a fair pricing condition, which says that the average transaction price of the metaorder is equal to the price after trading is completed. We show that at equilibrium the distribution of trading volume adjusts to reflect information, and dictates the shape of the impact function. The resulting theory makes empirically testable predictions for the functional form of both the temporal and permanent components of market impact. Based on the commonly observed asymptotic distribution for the volume of large trades, it says that market impact should increase asymptotically roughly as the square root of metaorder size, with average permanent impact relaxing to about two-thirds of peak impact.

1. Introduction

Market impact is the expected price change conditioned on initiating a trade of a given size and a given sign. Understanding market impact is important for several reasons. One motivation is practical: to know whether a trade will be profitable, it is essential to be able to estimate transaction costs, and in order to optimize a trading strategy to minimize such costs, it is necessary to understand the functional form of market impact.‡‡ Another motivation is ecological: impact exerts selection pressure against a fund becoming too large, and therefore is potentially important in determining the size distribution of funds.§§ Finally, an important motivation is theoretical: market impact reflects the shape of excess demand, the understanding of which has been a central problem in economics since the time of Alfred Marshall.

In this paper, we present a theory for the market impact of large trading orders that are split into pieces and executed incrementally. We call these *metaorders*.¶¶ The true size of metaorders is typically not public information, a fact that plays a central role in our theory. The strategic reasons for incremental execution of metaorders were originally analysed by Kyle (1985), who developed a model for an insider trader with monopolistic information about future prices. Kyle showed that the optimal strategy for such a trader is to break her metaorder into pieces and execute it incrementally at a uniform rate, gradually incorporating her information into the price. In Kyle's theory, the price increases linearly with time as the trading takes place, and all else being equal, the total impact is a linear function of size. The prediction of linearity is reinforced by Huberman and Stanzl (2004) who show that, providing

---

*Corresponding author. Email: Doyne.Farmer@inet.ox.ac.uk
‡‡See Bertismas and Lo (1998), Almgren and Chriss (1999), Almgren and Chriss (2000), Almgren (2003), Almgren *et al.* (2005), and Obizhaeva and Wang (2005).
§§See for example Berk and Green (2004). Schwartzkopf and Farmer (2010) have shown that managers of large mutual funds offset increases in market impact by lowering fees, slowing down trading and diversifying assets.

¶¶Other names used in the literature are large trades, packages or hidden orders.

liquidity is constant, to prevent arbitrage permanent impact must be linear.

Real data contradict these predictions: metaorders do not show linear impact. Empirical studies consistently find concave impact, i.e. incremental impact per share decreases with size.[†] It is in principle possible to reconcile the Kyle model with concave dependence on size by making the additional hypothesis that larger metaorders contain less information per share than smaller ones, for example because more informed traders issue smaller metaorders.[‡] A drawback of this hypothesis is that it is neither parsimonious nor easily testable, and as we will argue here, under the assumptions of our model it violates market efficiency.

Huberman and Stanzl are careful to specify that linearity only applies when liquidity is constant. In fact, liquidity fluctuates by orders of magnitude and has a large effect on price fluctuations.[§] Empirical studies find that order flow is extremely persistent, in the sense that the autocorrelation of order signs is positive and decays very slowly. No arbitrage arguments imply either fluctuating asymmetric liquidity as postulated by Lillo and Farmer (2004), or no permanent impact, as discussed by Bouchaud *et al.* (2004).[¶]

[†]An early study of metaorders, which links together the individual trades coming from a given client, is Chan and Lakonishok (1993, 1995). Later studies by Torre (1997), Almgren *et al.* (2005), Engle *et al.* (2008), Moro *et al.* (2009), and Toth *et al.* (2011a) find concave temporary impacts roughly consistent with a square root functional form. The functional form of permanent impact is harder to measure and more controversial. These studies should be distinguished from the large number of studies of the market impact of individual trades or the sum of trades in a given period of time, that do not attempt to link together the individual trades coming from a given client. See Hasbrouck (1991), Hausman *et al.* (1992), Keim and Madhavan (1996), Torre (1997), Kempf and Korn (1999), Plerou *et al.* (2002), Evans and Lyons (2002), Lillo *et al.* (2003), Potters and Bouchaud (2003), Gabaix *et al.* (2003, 2006), Chordia and Subrahmanyam (2004), Farmer *et al.* (2005), Weber and Rosenow (2006), and Hopman (2007).

[‡]This results in concave dependence on size but preserves linear dependence as a function of time. Our model in contrast predicts that both size and time follow the same concave functional form. The empirical results strongly support concave dependence on size, whereas the dependence on time is an open question.

[§]Farmer *et al.* (2004) show that fluctuations in instantaneous liquidity can span as much as three orders of magnitude for the same equity in the course of a year; Gillemot *et al.* (2006) show that liquidity fluctuations dominate volume fluctuations in driving clustered volatility. Some of the theoretical consequences of time varying liquidity have been investigated by Challet (2007) and Gatheral (2010).

[¶]Asymmetric liquidity means that the price response to a buy order differs from the price response to a sell order of the same size. The persistence of order flow implies predictability of order signs. If impact has a non-zero permanent component, then if the next order is likely to be a buy, unpredictability of returns implies that the price response to a buy order must be smaller than the price response to a sell order. See also Bouchaud *et al.* (2006), Farmer *et al.* (2006), Wyart *et al.* (2006), and Bouchaud *et al.* (2009). For a precursor of the theory developed here see the PhD thesis of Gerig (2007). For an early attempt to derive a theory yielding a square root market impact see Zhang (1999). These theories are for the impact of individual transactions, and the answers they give for the time and size dependence of impact are quite different than those we derive here for metaorders that are typically composed of many individual transactions.

The central goal of our model is to understand how order splitting affects market impact.[‖] Whereas Kyle assumed a single, monopolistic informed trader, our informed traders are competitive. They receive a common information signal and must independently choose the size of the order to submit. The orders are submitted to an algorithmic execution service that bundles them together as one large metaorder and executes them incrementally. We show that this leads to a Nash equilibrium in which order sizes are equilibrated with information. This equilibrium satisfies the condition that the final price after a metaorder is executed equals its average transaction price. We call this condition *fair pricing*, to emphasize the fact that under this assumption trading a metaorder is a breakeven deal—neither party makes a profit as a result of trading.[††] Our equilibrium is less general than Kyle's in that it assumes uniform execution, but it is more general in that it allows an arbitrary information distribution. This is key because, as we show, there is an equilibrium between information and metaorder size, and the metaorder size distribution can be observed directly from empirical data.

For a given information distribution, our theory predicts the metaorder size distribution and the average impact as a function of time during execution. Metaorder size and average impact are both observable, and in some data-sets it is even possible to know whether or not trades are informed (for example, Gomes and Waelbroeck (2013)). The fair pricing condition allows us to make several strong predictions based on a simple set of hypotheses. For a given metaorder size distribution, it predicts the average impact as a function of time both during and after execution. We thus predict the relationship between the functional form of two observable quantities with no a priori relationship, making our theory falsifiable in a strong sense. It is worth noticing that in a recent paper Bershova and Rakhlin (2013a) use proprietary data to perform an empirical analysis of a set of large institutional orders executed at Alliance Bernstein's buy-side trading desk and validate the predictions of our theory (more details will be given in the next sections).

The falsifiability of our model is in contrast to theories that make assumptions about the functional form of utility and/or behavioural or institutional assumptions about the informativeness of trades, which typically leave room for interpretation and require auxiliary assumptions to make empirical tests. For example, Gabaix *et al.* (2003, 2006) have also argued that the distribution of trading volume plays a central role in determining impact, and have derived a formula for impact that is concave under some circumstances. However, in contrast to our model, their prediction for market impact depends sensitively on the functional form for risk aversion.[‡‡] Our theory, in contrast, is based entirely on market efficiency and does not depend on the functional form of utility.

[‖]The persistence of order flow in the London Stock Exchange has been shown to be overwhelmingly due to order splitting rather than herding by Toth *et al.* (2011b).

[††]Neither party makes a profit in a one-way trade based on current prices, but for a round trip the aggressive party nonetheless incurs losses equal to the one way permanent impact.

[‡‡]Gabaix *et al.* argue that if risk aversion is proportional to σ^δ, where σ is the standard deviation of profits, the impact will increase with the size N of the metaorder as $N^{\delta/2}$. Thus, the impact is concave if $\delta < 2$, linear if $\delta = 2$ (i.e. if risk is proportional to variance), and convex otherwise.

An interesting alternative to our model, proposed by Toth *et al.* (2011a), links the shape of the impact to the diffusivity of prices, which gives a linear shape of the latent order book in the vicinity of the spread. The shape of the impact function depends on model parameters and on the execution speed; in the limit of high participation rates, their result is consistent with a square root function for market impact as a function of size.

Our work here is related to several papers that study market design. Viswanathan and Wang (2002), Glosten (2003) and Back and Baruch (2007) derive and compare the equilibrium transaction prices of orders submitted to markets with uniform vs. discriminatory pricing. Depending on the set-up of the model, these prices can be different so that investors will prefer one pricing structure to the other and can potentially be 'cream-skimmed' by a competing exchange.† The fair pricing condition we introduce here forces the average transaction price of a metaorder (which transacts at discriminatory prices) to be equal to the price that would be set under uniform pricing. Fair pricing, therefore, means investors have no preference between the two pricing structures. On the surface, this result is similar to the equivalence of uniform and discriminatory pricing in Back and Baruch (2007). However, in their paper, this equivalence results because orders are always allowed to be split, whereas ours is a true equivalence between the pricing of a split vs. unsplit order.

The organization of the paper is the following. In section 2, we give a description of the model and discuss its interpretation. In section 3, we develop the consequences of the martingale condition and show how this leads to zero overall profits and asymmetric price responses when order flow is persistent. In section 4, we show that any Nash equilibrium must satisfy the fair pricing condition. In section 5, we derive in general terms what this implies about market impact. In section 6, we introduce specific functional forms for the metaorder size distribution and explicitly compute the impact for these cases. Finally, in section 7, we discuss the empirical implications of the model, discuss how existing empirical literature supports our theory, and make some concluding remarks.

2. Model description

We study a stylized model of an algorithmic trading service combining and executing orders of long-term traders. This can be thought of as a broker-dealer receiving multiple orders on the same security, or as an institutional trading desk of a large asset manager combining the orders from multiple portfolio managers into one large metaorder. Our goal is to model the price impact of a metaorder during a trading period in which it may or may not be present. We set up the model in a stylized manner as a game in which trading takes place across multiple periods. Before the game starts, long-term traders receive an information signal; at the end of the trading period final prices, which reflect the signal, are revealed. While this is somewhat artificial in comparison to a real market (which has no such thing as a 'final' price), the framework is simple enough to allow us to find a solution, and the basic conclusions should

apply more broadly. The structure of the model is in many respects similar to the classic framework of Kyle (1985), but with several important differences. In our framework, the informed trader does not have a monopoly—they are competitive profit maximizers. Also, the total number of periods is not known, so the market makers must consider at each step the two possibilities that the game will either stop or continue on the next round. A point-by-point comparison with Kyle's model is made in the conclusions.

2.1. Framework

The trading of a single asset is organized in a sequence of auctions, which can be regarded as a game. The auctions take place at times $t = 1, \ldots, T$, where $1 \leq T \leq M$. There are three kinds of agents, long-term traders, market makers and a day trader, and an algorithmic trading firm that mechanically bundles orders together. We use different symbols to indicate different prices. Specifically, $\tilde{S}_t$ is the transaction price at time t, while $\tilde{X}_N$ is the final price after the metaorder of size N has finished. Moreover, as explained in section 2.3, we will average the prices over the different realizations of the day trader's signal. We indicate with S_t and X_N the averaged transaction and final price, respectively. The sequence of events can be summarized as follows (see also table 1 and figure 1).

$t = 0$: The $\mathcal{K}$ long-term traders receive a common information signal α, formulate orders of size $n_k(\alpha)$ and submit them to the algorithmic trading firm that bundles them together into a *metaorder* of size $\sum_k n_k$. The algorithmic firm divides the metaorder into equal-sized lots (assumed to be of unit size for convenience)‡ to be executed incrementally once the game starts. There may or may not be information: With probability μ the signal α is drawn from a distribution $P(\alpha)$, which has nonzero support over a continuous interval $-\alpha_{\max} \leq \alpha < \alpha_{\max}$, where $0 < \alpha_{\max} \leq \infty$. With probability $(1 - \mu)$ there is no information, i.e. $\alpha = 0$, which also implies that there will be no metaorder, i.e. $n_k = 0$ for all k. We assume that the orders of the long-term traders are bounded, i.e. $0 \leq n_k(\alpha) \leq M_k$, where $M_k = M/\mathcal{K}$ is a large positive integer.§ For simplicity, we discuss here the case where $\alpha \geq 0$ but the results apply equally well with obvious modifications when $\alpha \leq 0$.

$t = 1, \ldots, T$: At each time t, a representative *day trader* submits a market order of size $F(\eta_t)$, where η_t is a random signal, which can be represented as a zero mean IID noise process with an arbitrary distribution $\hat{P}(\eta_t)$, and F is an increasing function whose functional form is not important. If $\alpha \neq 0$, the algorithmic trading firm also

---

†For example, see Bernhardt *et al.* (2002).

‡Note that because we measure orders in lots, n_k can be a non-integer fractional size of one lot.

§The imposition of a maximum trade size M_k is a technical detail to avoid mathematical problems that occur in the limit $M \to \infty$. This is explained in the Appendix. For the typical situations we have in mind M, $\mathcal{K}$, and $M/\mathcal{K}$ are all large numbers.

Table 1. *Agents in the model*. The long-term investor is the key agent, who must choose an order size n_k at the beginning of the game based on information α. The market makers are competitive profit optimizers who set prices (but don't know α). The day trader and the algorithmic trading firm are mechanical agents; the day trader provides a background of noisy order flow and the algorithmic trading firm bundles up the long-term trades and submits them in lots of equal size.

Agent	Action	Period	Information
Long-term investor	Submits order $n_k(\alpha)$ to ATF	0	α
Algorithmic trading firm (ATF)	Submits lot from metaorder	t	$N = \sum_{k=1}^{\mathcal{K}} n_k(\alpha)$
Day trader	Submits order $F(\eta_t)$	t	η_t
Market maker (MM)	Price quote; MM with best quote executes transactions at price $\tilde{S}_t$	t	Combined order $s + F(\eta_t)$

submits a market order of size one. The market makers observe the net order flow $s + F(\eta_t)$, where $s = sign(\alpha)$. The market makers' quotes are formulated independently. All of the market orders are fully filled by the market maker(s) offering the best prices at the best quote $\tilde{S}_t$ (the meaning of tilde in the notation is explained in section 2.3) . If the metaorder is present, i.e. if $\alpha \neq 0$, then $T = N$ and the last auction occurs when the metaorder is fully executed, i.e. when

$$t = N = \sum_k n_k,$$

whereas if $\alpha = 0$, then $T = M$ and the last auction occurs at time† $t = M$. Given that η_t is IID, the typical transaction price sequence $\{\tilde{S}_t\} = \{\tilde{S}_1, \ldots, \tilde{S}_N\}$ will look like a random walk; if there is a metaorder present this induces a time-dependent drift on top of the random walk.

$t = T + 1$: The end of the game is announced along with the final price $\tilde{X}_N$ (again, the meaning of tilde in the notation is explained in section 2.3). Since the long-term traders' information signals are independent of those of the day trader, information is additive and the final price is

$$\tilde{X}_N \equiv X_0 + \alpha + \sum_{t=1}^{N} \eta_t, \qquad (1)$$

where X_0 is the initial price. Note that if no metaorder exists, i.e. $N = 0$, then $\tilde{X}_0 = X_0 + \sum_{t=1}^{M} \eta_t$.

We are not concerned here with the question of whether or not the algorithmic execution service's strategy of splitting the order into equal pieces is optimal. Our goal is instead to take this as given and to derive the implications for price impact.

2.2. Further comments about the agents

We now make some more comments about each of the agents. We assign a sharp division of labor—each agent plays an idealized role. In a more realistic setting, for example, market
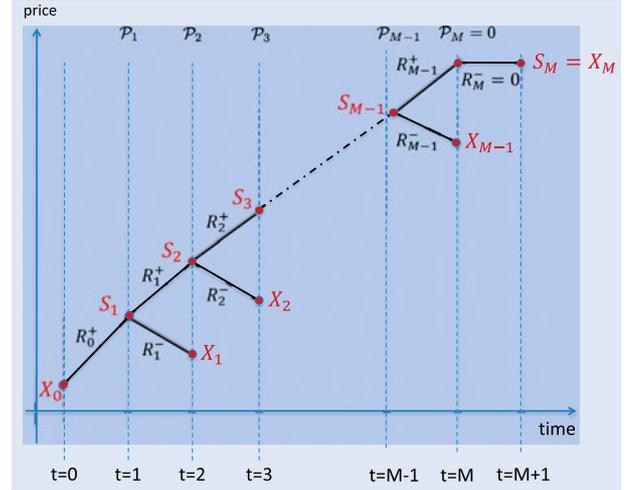

Figure 1.   The tree of possible price paths for a buy metaorder for different sizes N. For this figure, we assume the metaorder is present and show only expected price paths, averaged over the day trader's information (which is why the notation does not include tildes). The price is initially X_0; after the first lot is executed it is $S_1 = X_0 + R_0^+$. If $N = 1$ it is finished and the price reverts to $X_1 = S_1 - R_1^-$, but if $N > 1$ another lot is executed and it rises to $S_2 = S_1 + R_1^+$. This proceeds similarly until the execution of the metaorder is completed. At any given point, the probability that the metaorder has size $N > t$, i.e. that the order continues, is $\mathcal{P}_t$. If had we followed a typical price path under circumstances when the day trader's noisy information signal is large, rather than the expected price paths shown here, the sequence of prices would be a random walk with a time-varying drift caused by the metaorder's impact.

makers may also be somewhat informed about directional signals. We do not believe this alters the basic conclusions.

Long-term traders. Only the long-term traders receive the information signal α. For mathematical convenience but without loss of generality, we assume that the number of long-term traders is equal to $\mathcal{K}$, where $\mathcal{K}$ is a large number. The long-term traders are rational: they know that others have received the same signal, but each long-term trader decides $n_k(\alpha)$ independently. The decision as to what order size to submit based on the information α is the key decision in the game; this is done only once, in period zero.‡

---

†We choose the ending time when $\alpha = 0$ to be M because if there were an upper bound $M' < M$ the market maker would know a metaorder was present whenever $t > M'$.

‡It is interesting to note that in a recent paper using proprietary data, Bershova and Rakhlin (2013b) show that at any given time long-term investors mostly trade on one side of the market, i.e. at any given time they are either buyers or sellers, which gives support to our assumption of common information.

Day traders provide a noisy background of uncertain order flow that makes it impossible for the market makers to know with certainty whether there is a metaorder present. The day traders are treated as a single representative agent who receives a private information signal η_t at each period t of the game, which is independent of α. The day trader's decision is mechanical; at each period he submits a market order (either to buy or sell) of size $F(\eta_t)$. The day trader is trading on real information, but nonetheless plays a role in our model similar to that of noise traders in many other models. The key point is that the day traders do not engage in order splitting, i.e. they trade on the information they receive in given time step only in that time step. There is no restriction on the size of η_t, and in particular η_t can be large and of the opposite sign of α, so that the combined order flow $s + F(\eta_t)$ can be of either sign, even if the steady flow from the long-term traders imparts a bias.

Market makers. We are not assuming any special institutional privileges, such as those of the specialists in the NYSE; our market makers are simply competitive liquidity providers. At each time step, each market maker observes the combined market order and submits a quote without knowing the quotes of the other market makers. The combined order is fully executed by the market maker(s) offering the best price. The market makers are able to take past order flow and prices into account in setting their quotes.

We assume the market makers know the initial price X_0, the information distributions $P(\alpha)$ and $\hat{P}(\eta_t)$, the probability μ that a metaorder is present, and the function F relating the day trader's information to their order size. During the course of the game, they can update their prior μ to make a time dependent estimate μ_t'. However, they do not know the information signals α or η_t, and thus they do not know how much of the order flow to ascribe to the long-term trader vs. the day trader. Thus they do not know with certainty whether a metaorder is present, and if it is, they do not know its size.

2.3. Averaging

We are interested in computing statistical averages. $\tilde{S}_t$ denotes a specific realization of transaction prices, whereas S_t denotes an average price over the day trader's signals η_i, $\forall i \leq t$. Likewise, $\tilde{X}_N$ denotes a specific realization of the final price whereas X_N denotes an average final price over the day trader's signals η_i, $\forall i$. We use the notation $\hat{E}$ to represent an average over η,

$$\hat{E}[x(\eta)] = \int x(\eta)\hat{P}(\eta)d\eta. \qquad (2)$$

Therefore, $S_t(\alpha) = \hat{E}[\tilde{S}_t|\alpha]$ and $X(\alpha) = \hat{E}[\tilde{X}|\alpha] = X_0 + \alpha$. The goal of the paper is to compute the *average immediate impact* $\mathcal{I}_t \equiv S_t - X_0$ and the *average permanent impact* $I_N \equiv X_N - X_0$ of the metaorder. The corresponding incremental average impacts are $R_t^+ = S_{t+1} - S_t$ and $R_t^- = S_t - X_t$. As we will show, at the equilibrium the average impacts do not depend on μ or $\hat{P}(\eta)$. However, they do depend on the equilibrium distribution of metaorder lengths p_N, which is a key quantity that we compute.

In addition to taking expectations over the day trader's random signal η, which we denote by $\hat{E}$, we must compute expectations over metaorder sizes. The crux of our argument

hinges around the market makers' ignorance of α; when $\alpha \neq 0$ this translates into uncertainty about metaorder size. We will use the notation E_t to represent an average over all metaorders of size $N \geq t$, and as described above, $\hat{E}$ for averages over η_t. For a generic function f_N the average over metaorder sizes is

$$E_t[f_N] = \frac{\sum_{N=t}^M p_N f_N}{\sum_{N=t}^M p_N} = \frac{\sum_{i=0}^{M-t} p_{t+i} f_{t+i}}{\sum_{i=0}^{M-t} p_{t+i}}, \qquad (3)$$

where in the last term we made the substitution $N = t + i$.

Our main result is to derive the equilibrium relationship between the metaorder size distribution p_N and the average immediate and permanent impacts $\mathcal{I}_t$ and I_N.

2.4. Uncertainty of metaorder size and persistence of order flow

Assuming that a metaorder is present, the likelihood that it will persist depends on the distribution p_N and the number of executions t that it has already experienced. Let $\mathcal{P}_t$ be the probability that the metaorder will continue given that it is still active at timestep t, i.e.

$$\mathcal{P}_t = \frac{\sum_{i=t+1}^M p_i}{\sum_{i=t}^M p_i}. \qquad (4)$$

This makes precise how order splitting can make order flow positively autocorrelated. In particular, if p_N has tails heavier than an exponential $\mathcal{P}_t$ will increase with time and induce persistence in order flow.

2.5. Assumption of known starting and stopping times

The assumption that the starting and stopping times are known to the participants is similar to the one made in the Kyle model (1985). For long metaorders and sufficiently high participation rates, the starting time can be inferred from the imbalance in order flow.† Assuming an average participation rate of 20% (i.e. four times as much volume for day traders as from long-term traders) and two standard deviations to reject the null hypothesis of balanced order flow, this means that a metaorder can be detected after about 20 steps. Large metaorders are frequently executed in many small increments.‡ If there are 1000 increments the error for inferring the starting time is 2%; similarly 10 000 increments corresponds to an error of 0.2%. Thus for large metaorders it is not unrealistic to imagine that market makers can infer their presence and estimate their starting times and stopping times. Note, however, that we allow

---

†The starting time can be inferred by treating order signs as a binomial random process. The imbalance required to reject the null hypothesis of unbiased order flow is $q\sqrt{t}$, where q is the desired number of standard deviations of statistical significance. The accumulated imbalance after t steps is zt, where z is the participation rate. Equating these gives $t = (q/z)^2$. Hiding an order of size N requires $z < q/\sqrt{N}$. Thus, larger metaorders need to be executed more slowly to avoid detection. Since the time needed to complete execution is inversely proportional z, for a large metaorder this can become prohibitive—it is impossible to escape detection.

‡Metaorders can be extremely large. For example, the New York Times recently reported that Warren Buffet took 8 months to buy a 5.5% share of IBM.

for the possibility that there is no metaorder present at all, and that the probability that the order is not present can be arbitrarily large. Nonetheless, when N is large the market maker can infer the presence of the order, and the relevant starting and stopping times, with considerable accuracy.

In a more realistic setting, there may be many metaorders active at the same time. Here we assume that we focus attention on a single metaorder, i.e. we are interested in the outcome of an event study in which one examines price changes conditional on the presence of a single metaorder. As long as arrival of metaorders are uncorrelated this should be valid.

3. Martingale condition

In this section, we introduce a martingale condition and discuss its implications for liquidity and overall profitability.

3.1. Derivation of martingale condition

Market makers must set prices given only past and present order flow information, as well as past prices, without knowing whether the order flow originated from a metaorder or from day traders. Their decision function for setting prices is of the form

$$\tilde{S}_t = f(s + F(\eta_t), s + F(\eta_{t-1}), \dots, s + F(\eta_1)).$$

As we show below, we are able to finesse the difficult problem of computing this decision function by imposing a martingale condition and averaging, which is sufficient for the main goal of this paper of deriving the equilibrium between the impacts and p_N.

We assume that transaction prices are a martingale, so that the current transaction price $\tilde{S}_t$ is equal to the expected future price. Define an indicator variable m where $m = 1$ if the metaorder is present and $m = 0$ if it is absent. Recall that μ'_t is the market maker's best estimate of the probability that the metaorder is present. For the price in the next period it is, necessary to average over three possibilities:†

(1) With probability μ'_t the metaorder is present and with probability $\mathcal{P}_t$ trading continues. In this case, the expected price in the next period is $\hat{E}[\tilde{S}_{t+1}|\tilde{S}_t, m = 1] = \tilde{S}_t + R_t^+$.

(2) With probability μ'_t the metaorder is present and with probability $1 - \mathcal{P}_t$ time t is the last trading period. In this case, the expected price in the next period is $\hat{E}[\tilde{X}_{t+1}|\tilde{S}_t, m = 1] = \tilde{S}_t - R_t^-$.

(3) With probability $1 - \mu'_t$ the metaorder is not present. Since the day trader's information η_t is zero mean and IID, the expected average transaction price on the next time step conditioned on the current transaction price must satisfy $\hat{E}_t[\tilde{S}_{t+1}|\tilde{S}_t, m = 0] = \tilde{S}_t$, i.e. the average transaction price is unchanged.

Thus, the martingale condition can be written

$$\tilde{S}_t = \mu'_t \left(\mathcal{P}_t E[\tilde{S}_{t+1}|\tilde{S}_t, m = 1] + (1 - \mathcal{P}_t)E[\tilde{X}_t|\tilde{S}_t, m = 1] \right)$$
$$+ (1 - \mu'_t)E[\tilde{S}_{t+1}|\tilde{S}_t, m = 0].$$

---

†In these equations, the average is taken over every noise term η_{t+1}. However, because we condition on $\tilde{S}_t$, the operater $\hat{E}$ does not remove the tilde for this variable.

Substituting for the average next price conditioned on the current price for cases 1–3 gives:

$$\tilde{S}_t = (1 - \mu'_t)\tilde{S}_t + \mu'_t \left(\mathcal{P}_t(\tilde{S}_t + R_t^+) + (1 - \mathcal{P}_t)(\tilde{S}_t - R_t^-) \right). \quad (5)$$

The current transaction price $\tilde{S}_t$ and μ' both cancel and this reduces to

$$\mathcal{P}_t R_t^+ - (1 - \mathcal{P}_t)R_t^- = 0. \quad (6)$$

Equation (6) holds for $t = 1, 2, \dots, M - 1$. If the metaorder has maximal length at the end of the Mth interval by definition $\mathcal{P}_M = 0$, which implies that $S_M = X_M$.

Let us pause for a moment to digest this result. We started with a martingale condition for realized prices, including fluctuations caused by day traders, and then reduced it to a martingale condition for the average impact due to the presence of a metaorder. The reduced martingale no longer depends on μ'_t, $\hat{p}_N$, or the day trader's information. The fact that we assume a martingale for realized prices implies that arbitrage of the impact is impossible.

The ability to average away the day traders is a consequence of our assumption that α and η_t are independent, which implies additivity of information. This separates the problem of the metaorder's impact from that of the day trader's impact—the metaorder's impact effectively rides on top of the day trader's impact. As we will see, the virtue of this approach is that it allows us to infer quite a lot without needing to solve for the market makers' optimal quote setting function f.

3.2. Asymmetric price response

Equation (6) can be trivially rewritten in the form

$$\frac{R_t^+}{R_t^-} = \frac{1 - \mathcal{P}_t}{\mathcal{P}_t}, \quad (7)$$

where $R_t^+ = S_{t+1} - S_t$ and $R_t^- = S_t - X_t$. Thus the martingale condition fixes the ratio of the price responses R_t^- and R_t^+, but does not fix their scale. If $\mathcal{P}_t$ is large, corresponding to a metaorder that is likely to continue, then R_t^+/R_t^- small. This means that the price response if the order continues is much less it is than if it stops.

To complete the calculation, we need another condition to set the scale of the price responses R_t^- and R_t^+, which may change as t varies.‡ Such a condition is introduced in section 4. Even without such a condition, one can already see intuitively that 'all else equal', for a heavy-tailed metaorder distribution p_N, the impact will be concave. (Recall that heavy tails in p_N imply that $\mathcal{P}_t$ increases with t).

3.3. Zero overall profits

The martingale condition implies that the market makers break even overall, i.e. that their total profits summing over

---

‡Various authors have used alternative conditions. For example, Gerig (2007) and Gerig *et al.* (2011) use a symmetry condition, which can be derived from assumptions of linearity. The fair pricing condition that we derive here has the advantage that it can be justified based on equilibrium arguments.

metaorders of all sizes is zero. This is stated more precisely in proposition one.

PROPOSITION 1 *The average prices for each transaction are* $\{S_t\}$*, with* $t = 1, \ldots, N$*. At the final step the valuation is* $N X_N$*. The martingale condition implies zero overall profits, i.e.*

$$\Pi = E_1[N\pi_N] \equiv \sum_{N=1}^{M} p_N N \pi_N = 0, \qquad (8)$$

where

$$\pi_N \equiv \frac{1}{N} \sum_{t=1}^{N} S_t - X_N \qquad (9)$$

is the profit per lot transacted. The proof of proposition 1 is given in Appendix A. The phrase 'overall profits' emphasizes that the martingale condition only implies zero profits when averaged over metaorders of all sizes. It allows for the possibility that the market makers may make profits on metaorders in a given size range, as long as they take corresponding losses in other size ranges.

Surprisingly, Proposition 1 is not necessarily true when M is infinite. The basic problem is similar to the St. Petersburg paradox: as the metaorder size becomes infinite it is possible to have infinitely rare but infinitely large losses. The conditions under which this holds are more complicated, as discussed in Appendix A.

4. Fair pricing

The martingale condition derived in the previous section sets only the ratio R_t^+/R_t^- at each step t (see equation (7)). We therefore need another condition to derive the values of R_t^+ and of R_t^- and thus to obtain the expression for the immediate and the permanent impact. The condition we derive here is the *fair pricing condition*, which states that for any N

$$\pi_N = \frac{1}{N} \sum_{t=1}^{N} S_t - X_N = 0. \qquad (10)$$

Under fair pricing the average execution price is equal to the final price. We call this fair pricing for the obvious reason that both parties would naturally regard this as 'fair'. Fair pricing implies that the market makers break even on metaorders of any size, as opposed to the martingale condition, which only implies they break even when averaging over metaorders of all sizes.

We now derive the fair pricing condition by showing that it is a Nash equilibrium.[†] We prove the following:

PROPOSITION 2 *If the immediate impact* $\mathcal{I}_t = S_t - X_0$ *has a second derivative bounded below zero,[‡] in the limit where the number of informed traders* $\mathcal{K} \to \infty$*, any Nash equilibrium must satisfy the fair pricing condition* $\pi_N = 0$ *for* $1 < N < M$.

[†]Subsequent to our work here, Donier (2012) postulates a very similar condition based on perfect competition between market makers. Recently, Bershova and Rakhlin (2013a) showed that the fair pricing condition and its consequences on market impact are rather well empirically verified in a set of real metaorders.
[‡]Note that this is sufficient for concavity.

On average market makers profit from orders of length one and take (equal and opposite) losses from orders of length M.

This result is driven by competition between informed traders. All informed traders receive the same information signal α. The strategy of informed trader k is the choice of the order size $n_k(\alpha)$. The orders are then bundled together to determine the combined metaorder size $N = \sum_{k=1}^{\mathcal{K}} n_k(\alpha)$. The decision of each informed trader is made without knowing the decisions of others.

The derivation has two steps: first we examine the case $\pi_N \neq 0$ for $1 < N < M$, and show that if others hold their strategies constant, providing the impact is concave and $\mathcal{K}$ is sufficiently large, traders can increase profits by changing strategy. Secondly, we show that if $\pi_N = 0$ there is no incentive to change strategy. Then we return to examine the cases $N = 1$ and $N = M$, which must be treated separately. The derivation is given in the Appendix.

In contrast to the martingale condition, which only implies that immediate profits are zero when averaged over size, fair pricing means that they are identically zero for every size. It implies that no one pays any costs or makes any profits simply by trading in any particular size range.

Since informed traders must formulate order sizes knowing only α, the distribution of information $p(\alpha)$ implies the distribution of metaorder size p_N. We will use p_N as a proxy for $p(\alpha)$, which is the key fact allowing us to state our results in terms of the observable quantity p_N rather than $p(\alpha)$, which is much more difficult to observe.

5. General expressions for impact

In this section, we assume that the martingale condition holds for all N and the fair pricing condition holds for $1 < N < M$. This allows us to derive both the immediate impact $\mathcal{I}_N$ and the permanent impact I_N for any given metaorder size distribution p_N. We later argue that for realistic situations the metaorder size distribution gives rise to a concave impact function, consistent with the Nash equilibrium.

The martingale condition (equation (6)) and the fair pricing condition (equation (10)) define a system of linear equations for S_t and X_t at each value of t, which we can alternatively express in terms of the price differences $R_t^+ = S_{t+1} - S_t$ and $R_t^- = S_t - X_t$, where $t = 1, \ldots, M$. The martingale condition holds for $t = 1, \ldots, M$ and the fair pricing condition holds for $t = 2, \ldots, M - 1$. There are thus $2M - 2$ homogeneous linear equations with $2M - 1$ unknowns.[§] Because the number of unknowns is one greater than the number of conditions there is necessarily an undetermined constant, which we choose to be R_1^+.

PROPOSITION 3 *The system of martingale conditions* (equation (6)) *and fair pricing conditions* (equation 10) *has solution*

$$R_t^+ = \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^{M} p_i} \frac{1 - p_1}{\sum_{i=t}^{M} p_i} R_1^+ \quad t = 2, 3, \ldots, M-1 \quad (11)$$

$$R_t^- = \frac{\mathcal{P}_t}{1 - \mathcal{P}_t} R_t^+ \quad t = 1, 2, \ldots, M-1 \quad (12)$$

[§] The price S_{M+1} does not exist, so R_M^+ is not needed. This reduces the number of unknowns by one.

The proof is given in Appendix A.

An important property of the solution is the equivalence of the impact $\mathcal{I}_t$ as a function of either time or size. This is in contrast to the prediction of an 'extended' Kyle model under the assumption that traders of different sizes are differently informed, which yields linear impact as a function of time while varying non-linearly with size.

5.1. General solution for immediate impact

Summing equation (11) implies that for $N > 2$ the immediate impact is

$$\mathcal{I}_t = S_t - X_0 = R_0^+ + R_1^+ \left(1 + \sum_{k=2}^{t-1} \frac{1}{k} \frac{p_k}{\sum_{i=k+1}^{M} p_i} \frac{1 - p_1}{\sum_{i=k}^{M} p_i} \right),$$ (13)

For $t = 1$ the immediate impact is $\mathcal{I}_1 = S_1 - X_0 = R_0^+$ and for $t = 2$ it is $\mathcal{I}_2 = S_2 - X_0 = R_0^+ + R_1^+$. (The meaning of the undetermined constants R_1^+ and R_0^+ is discussed in a moment).

5.2. General solution for permanent impact

The permanent impact $X_N - X_0$ is easily obtained. Making some simple algebraic manipulations

$$X_N = X_N - S_N + S_N = S_N - R_N^- = S_N - \frac{\mathcal{P}_N}{1 - \mathcal{P}_N} R_N^+.$$

By combining equations (13) and (4) we get

$$I_N = X_N - X_0 = R_0^+ + R_1^+ \left(1 + \sum_{k=2}^{N-1} \frac{1}{k} \frac{p_k}{\sum_{i=k+1}^{M} p_i} \frac{1 - p_1}{\sum_{i=k}^{M} p_i} \right. $$
$$\left. - \frac{1 - p_1}{N \sum_{i=0}^{M-N} p_{N+i}} \right). \quad (14)$$

5.3. Setting the scale

We have expressed both the permanent and immediate impact purely in terms of p_N and the undetermined constants R_0^+ and R_1^+. The undetermined constants can in principle be fixed based on the information at the equilibrium. At the equilibrium, information signals in the range $\alpha \in [0, \alpha_1]$ will be assigned to metaorders of size one, with an average size $\bar{\alpha}_1$, signals in the range $\alpha \in (\alpha_1, \alpha_2]$ will be assigned to metaorders of size two, with an average size $\bar{\alpha}_2$, and so on. The scale of the impact is set by the relations

$$R_0^+ = I_1 = \bar{\alpha}_1 = \int_0^{\alpha_1} \alpha p(\alpha) d\alpha$$

$$R_0^+ + R_1^+ \frac{\mathcal{P}_1}{1 - \mathcal{P}_1} = I_2 = \bar{\alpha}_2 = \int_{\alpha_1}^{\alpha_2} \alpha p(\alpha) d\alpha$$

We have used the words 'in principle' because, unlike metaorder sizes, information is not easily observed.

Barring the ability to independently measure information, the constants R_0^+ and R_1^+ remain undetermined parameters. $R_1^+ > 0$ plays the important role of setting the scale of the impact. The constant R_0^+, in contrast, is unimportant—it is simply the impact of the first trade, before the metaorder has been detected.

6. Dependence on the metaorder size distribution

We have so far left the metaorder size distribution p_N unspecified. In this section, we compute the impact for the Pareto distribution,

$$p_N \sim \frac{1}{N^{\beta+1}}.$$ (15)

which we argue is well supported, at least as an approximation for large N, by empirical data.[†] In Appendix B, we consider the stretched exponential distribution, which is not supported by data, but provides a useful point of comparison. We also consider a lognormal distribution in Appendix B.

6.1. Empirical evidence supporting the Pareto distribution

There is now considerable accumulated evidence that in the large size limit in most major equity markets the metaorder size V is distributed as $P(V > v) \sim v^{-\beta}$, with $\beta \approx 1.5$.

- *Trade size.* In many different equity markets for large trades, the volume V has been observed by several groups to be distributed as a power law (Gopikrishnan *et al.* 2000, Gabaix *et al.* 2006).[‡] This relationship becomes sharper if only block trades are considered (Lillo *et al.* 2005).

- *Long memory in order flow.* The signs of order flow in many equity markets are observed to have long memory.[§] This means that the transaction sign autocorrelation function $C(\tau)$ decays in time as $C(\tau) \sim \tau^{-\gamma}$, where $0 < \gamma < 1$. Under a simple theory of order splitting the exponent $\beta = \gamma + 1$, Lillo *et al.* (2005) and Bouchaud *et al.* (2009). Since empirically it is $\gamma \simeq 0.5$ this implies $\beta \simeq 1.5$.

- *Reconstruction of large metaorders from brokerage data.* Vaglica *et al.* (2008) reconstructed metaorders Spanish stock exchange using data with brokerage codes and found that N is distributed as a power law[¶] for large N with $\beta \approx 1.7$.

[†]The notation $f(x) \sim g(x)$ means that there exists a constant $C \neq 0$ such that in the limit $x \to \infty$, $f(x)/g(x) \to C$. We use it to indicate that this relationship is only valid in the limit of large metaorder size N.

[‡]The value of β is somewhat controversial, however: Eisler and Kertecz (2006) and Racz *et al.* (2009) have argued that the correct value of $\beta > 2$.

[§]Long memory has been observed in the Paris Stock Market by Bouchaud *et al.* (2004), in the London and New York Stock Markets by Lillo and Farmer (2004), and in the Spanish Stock Market by Vaglica (2008). We use the term long memory in its more general sense to mean any process whose autocorrelation function is non-integrable (Beran 1994). This can include processes with structure breaks, such as that studied by Ding *et al.* (1993).

[¶]An interesting point that is relevant for the theory developed here is that the power law behaviour of metaorder size comes from the heterogeneity of market participants. Vaglica *et al.* (2008) showed that metaorder size distribution for individual brokerages is roughly a lognormal distribution, and that a power law only emerges when all the brokerages are combined. Here we have assumed that the long-term traders are homogeneous, and that p_N is determined by information; this suggests that other factors are at play determining p_N, and in particular that the heterogeneous size of investors may be an important factor.

● *Direct evidence from proprietary data.* In a recent paper Bershova and Rakhlin (2013a) use proprietary data to perform an empirical analysis of a set of large institutional metaorders executed at AllianceBernstein's buy-side trading desk in the US equity market. As mentioned above, their data are consistent with an asymptotically Pareto distribution with an estimated tail exponent $\beta = 1.56$.

There is thus good evidence that metaorders have a power law distribution, though more study is of course needed.

6.2. Market impact for Pareto metaorder size

In this section, we derive the functional form of the impact for Pareto metaorder size. We do this by using equation (13) in the limit as $M \to \infty$, and return in section 6.3 and in Appendix B to discuss how this is modified when M is finite.

While we only care about the asymptotic form for large N, for convenience we assume an exact Pareto distribution for all N, i.e.

$$p_N = \frac{1}{\zeta(\beta)} \frac{1}{N^{\beta+1}}, \quad N \geq 1 \tag{16}$$

where the normalization constant $\zeta(\beta)$ is the Riemann zeta function. For the Pareto distribution the probability $\mathcal{P}_t$ that an order of size t will continue is

$$\mathcal{P}_t = \frac{\zeta(1+\beta, t+1)}{\zeta(1+\beta, t)} \simeq \left(\frac{t}{t+1}\right)^{\beta} \sim 1 - \frac{\beta}{t}. \tag{17}$$

where $\zeta(s, a)$ is the generalized Riemann zeta function (also called the Hurwitz zeta function). The approximations are valid in the large t limit.

6.2.1. Immediate impact.
The immediate impact can be easily calculated from equation (13). For Pareto distributed metaorder sizes, using equations (12) and (16), R_t^+ is

$$R_t^+ = \left(1 + \frac{1}{t^{2+\beta}} \frac{\zeta(1+\beta) - 1}{\zeta(1+\beta, t)\zeta(1+\beta, t+1)}\right) R_1^+ \sim \frac{1}{t^{2-\beta}}. \tag{18}$$

Thus the immediate impact $\mathcal{I}_t = S_t - X_0$ behaves asymptotically for large t as

$$\mathcal{I}_t \sim \begin{cases} t^{\beta-1} & \text{for } \beta \neq 1 \\ \log(t+1) & \text{for } \beta = 1 \end{cases} \tag{19}$$

The exponent β has a dramatic effect on the shape of the impact. For Lorenzian distributed metaorder size ($\beta = 1$), the impact is logarithmic, for $\beta = 1.5$ it increases as a square root, for $\beta = 2$ it is linear, and for $\beta > 2$ it is superlinear. Thus as we vary β the impact goes from concave to convex, with $\beta = 2$ as the borderline case.† Figure 2 illustrates the reversion process for $\beta = 1.5$ and shows how the shape of the impact varies with β.

---

†The reason $\beta = 2$ is special is that for $\beta < 2$ the second moment of the Pareto distribution is undefined. Under the theory of Lillo *et al.* (2005), long-memory requires $\beta < 2$.
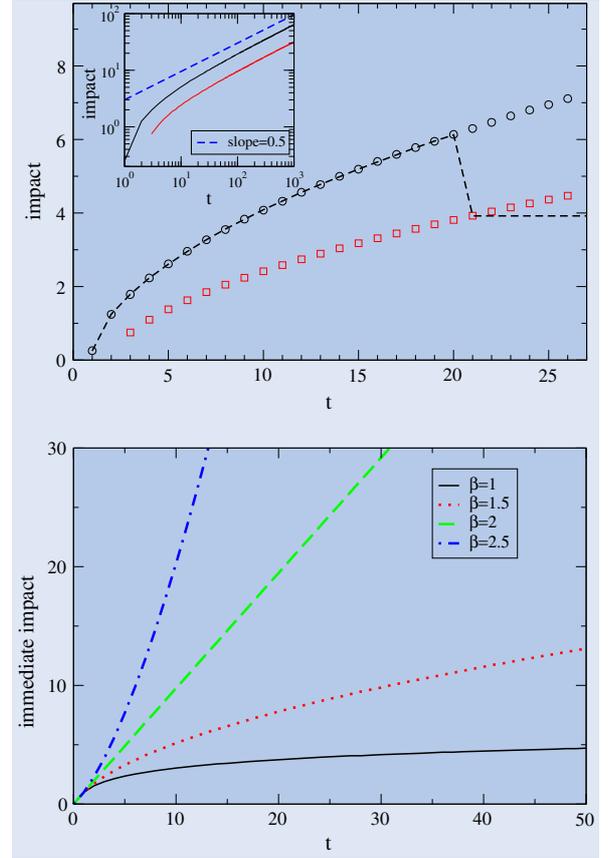

Figure 2. An illustration of predicted market impact for Pareto distributed metaorder size. *Top panel*: Immediate impact $\mathcal{I}_t$ (black circles) and permanent impact I_t (red squares) for $\beta = 1.5$. The dashed line is the price profile of a metaorder of size $N = 20$, demonstrating how the price reverts from immediate to permanent impact when metaorder execution is completed. The inset shows a similar plot in double logarithmic scale for a wider range of sizes (from $N = 1$ to $N = 1000$). The blue dashed line is a comparison to the asymptotic square root scaling. *Bottom panel*: Expected immediate impact $\mathcal{I}_t$ as a function of time t for tail exponents $\beta = 1, 1.5, 2$ and 2.5, illustrating how the impact goes from concave to convex as β increases.
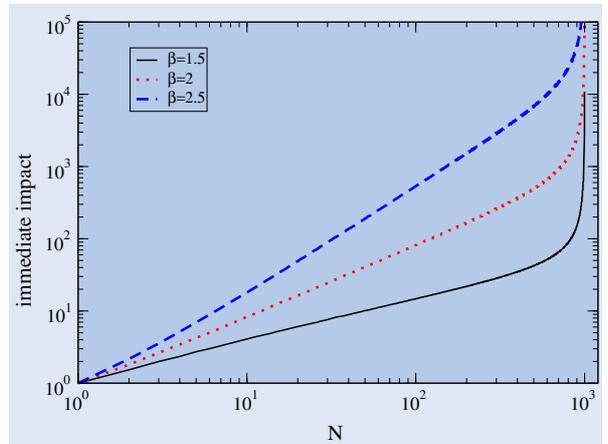

Figure 3. Immediate impact when the metaorder size has finite support, i.e when N has a maximum value M. Plot is in log log scale with $M = 1000$ and $\beta = 1.5, 2$, and 2.5.

6.2.2. Permanent impact.
The permanent impact under the Pareto assumption is easily computed using equation (14). A

direct calculation shows that†

$$I_N \sim \frac{1}{N} \int^N x^{\beta-1} dx = \frac{1}{\beta} N^{\beta-1}. \qquad (20)$$

Equations (19) and (20) imply that the ratio of the permanent to the immediate impact is

$$\frac{I_N}{\mathcal{I}_N} = \frac{1}{\beta}. \qquad (21)$$

For example, if $\beta = 1.5$ the model predicts that on average the permanent impact is equal to two-thirds of the maximum immediate impact, i.e. following the completion of a metaorder the price should revert by one-third from its peak value.

6.3. Effect of maximum order size

In the previous section, we have assumed that $N \ll M$, so that we can treat the problem as if M were infinite. In real markets, the maximum order size is probably quite large, a significant fraction of the market capitalization of the asset. Thus we doubt that the finite support of N has much practical importance, except perhaps for extremely large metaorders.

From a conceptual point of view, however, having an upper bound on metaorder size creates some interesting effects. As we have already mentioned, Proposition 1 fails to hold when $M = \infty$, so this must be handled with some care. In Appendix C, we illustrate how the results change when $N \approx M$. What we observe is that when $N < M/2$ the impact is roughly unchanged from its behaviour in the limit $M \to \infty$, but when $N \gtrsim M/2$ the impact becomes highly convex. This is caused by the fact that the market maker knows the metaorder must end when $t = M$. Since by definition $\mathcal{P}_M = 0$, the martingale condition requires that $X_M = S_M$, i.e. there is no reversion when the metaorder is completed. This propagates backward and when $N \approx M$ it significantly alters the impact, as seen in figure 3.

Nonetheless, from a practical point of view we do not think this is an important issue, which is why we have relegated the details of the discussion to Appendix C.

6.4. Relation between volume and information

One of the main contributions of this paper is to derive a relationship between information and metaorder size. We now make this more explicit and show that, under the assumption that returns reflect information, it results in the expected relationship between returns and volume.

In the continuous limit for large N the distributions for information and metaorder size are related by conservation of probability as

$$p(\alpha) = p_N \frac{dN}{d\alpha}. \qquad (22)$$

So for example, if the empirical metaorder size is p_N is asymptotically Pareto distributed, as argued in the next section, $p_N \sim N^{-(\beta+1)}$ and $\alpha = N^{\beta-1}$ as shown below, we have

†This is the same impact vs. size derived by Gabaix *et al.* (2006). Their derivation is based on quite different reasoning, and requires mean-variance utility with a linear (rather than the usual quadratic) risk aversion term. They predict a different permanent impact.

$p(\alpha) = \alpha^{(1-2\beta)/(\beta-1)}$. Based on the empirically observed value $\beta \approx 1.5$, this gives $p(\alpha) = \alpha^{-4}$, which means that the cumulative scales as $P(\alpha > x) \sim x^{-3}$. This is what is typically observed for price returns in American stock markets (Plerou *et al.* 1999). This is consistent with (Gabaix *et al.* 2003), with the exception that our analysis is specifically for metaorders.

7. Discussion

7.1. Comparison to the Kyle model

The three types of agents in our model are similar to Kyle's; his informed trader is replaced by our long-term traders, and his noise traders are replaced by our day traders. In both cases, we assume a final liquidation. There are also several key differences. In our model:

- Our long-term traders do not have a monopoly, but rather have common information and compete in setting the size of their orders. Their orders are bundled together and executed as a package by an algorithmic execution service. These two facts are essential to show that the fair pricing condition is a Nash equilibrium.
- The number of periods N for execution is uncertain, and depends on the information of the long-term traders. This is important because the martingale condition is based on the market makers' uncertainty about when the metaorder will terminate.
- The distribution $p(\alpha)$ is arbitrary (whereas Kyle assumed a normal distribution). Our key result is that information is almost fully reflected in metaorder size, i.e. that p_N and $p(\alpha)$ are closely related. This means our results apply to any empirical size distribution p_N.
- Our day traders respond to ongoing information signals that are permanent in the sense that they affect the final price, in contrast to Kyle's noise traders, who are completely uninformed. This means that the information in the final liquidation price is incrementally revealed.
- The most important difference is that our model has a different purpose. Kyle assumed an information signal and solved for the optimal strategy to exploit it. We assume metaorder execution may or may not be going on in the background, and solve for its impact on prices.

7.2. Empirical implications and tests of the model

The theory presented here makes several predictions with clear empirical implications. In this section, we summarize what these are and outline a few of the problems that are likely to be encountered in empirical testing.

(1) The fair pricing condition, equation (10), is directly testable, although it requires a somewhat arbitrary choice about when enough time has elapsed since the metaorder has completed for reversion to occur. (One wants to minimize this time because of the diffusive nature of prices, but one wants to allow enough time to make sure that reversion is complete).

(2) The asymmetric price response predicted by equation (7) is testable. However, this only tests the martingale

condition, which is the less controversial part of our model.

(3) The equivalence of impact as a function of time and size is directly testable. Under our theory, for $N > t$ the immediate impact from the first t steps is the same, regardless of N. This is in contrast to the Kyle model which predicts linear impact as a function of time, but can explain concavity in size only by postulating variable informativeness of trades vs. metaorder size.

(4) The prediction of immediate and permanent impact based on p_N is directly testable through equations (13) and (14).

(5) If the metaorder distribution is a power law (Pareto distribution), then for large N the immediate impact scales as $\mathcal{I}_t \sim t^{\beta-1}$ and the ratio of the permanent to the immediate impact of the last transaction is $I_N/\mathcal{I}_N = 1/\beta$. See section 6.2.

Prediction (2) has been tested and confirmed by Lillo and Farmer (2004), Farmer *et al.* (2006) and Gerig (2007). Preliminary results seem to support, or at least not contradict, prediction (5). The only studies of which we are aware that attempted to fit functional form to the impact of metaorders are by Torre (1997), Almgren *et al.* (2005), Moro *et al.* (2009), Toth *et al.* (2011a), and Bershova and Rakhlin (2013a). They find immediate impact roughly consistent with a square root functional form. Moro *et al.* also tested the ratio of permanent to immediate impact and found 0.51 for the Spanish stock market and 0.69 for the London stock market, with large error bars. Very recently (subsequent to the appearance of our paper in preprint form), Bershova and Rakhlin (2013a) and Gomes and Waelbroeck (2013) each investigated separate proprietary sets of large metaorders on US equity markets. They independently verify the fair pricing condition. Moreover Bershova and Rakhlin (2013a) show that the permanent impact is approximated by a square root function of execution time and that its ratio to the immediate impact is close to 2/3. These last studies are the strongest piece of direct evidence for our theory.

7.3. Information revelation

Though the market makers in our model are uncertain whether or not a metaorder is present, if it is present, they know when its execution begins and ends. The ability to detect metaorders from imbalances in order flow using brokerage codes has been demonstrated (see Vaglica (2008), Toth *et al.* (2010)). A recent study of metaorders based on brokerage code information found average participation rates of 17% for the Spanish stock market (BME) and 34% for the London Stock Market,† for metaorders whose average size was just under 100 in both markets, making such metaorders difficult to hide. The detection problem introduces uncertainties in starting and stopping times that may affect shape of the price impact.

---

†Participation rate is defined as the fraction of trades that a given agent participates in.

7.4. Final thoughts

The traditional view in finance is that market impact is just a reflection of information. This point of view often goes a step further and postulates that the functional form of impact is determined by behavioural and institutional factors, such as how informed the agents are who trade with a given volume. This hypothesis is difficult to test because it is inherently complicated and information is difficult to measure independently of impact. Within the framework developed here, such anomalies would violate the fair pricing condition.

In this paper, we embrace the view that impact reflects information, but we show how at equilibrium the trading volume reflects the underlying information and makes it possible to compute the impact. The metaorder size distribution determines the shape of the impact but does not set its scale. Metaorder size has the important advantage of being a measurable quantity, and thus predictions based on it are much more testable than those based directly on information.

The fair pricing condition that we have derived here may well hold on its own, even without informational efficiency. This could be true for purely behavioural reasons: the fair pricing condition holds because it can be measured reliably, and both parties view it as fair. Thus while the main results here are consistent with rationality, they do not necessarily depend on it.

We provide an example solution for the Pareto distribution for metaorder size because we believe that the evidence supports this hypothesis. This gives the simple result that the impact is a power law of the form $\mathcal{I}_t \sim t^{\beta-1}$, and the ratio of permanent impact to the temporary impact of the last transaction is $I_N/\mathcal{I}_N = 1/\beta$. However, the bulk of our results do not depend on this assumption. Thus the reader who is skeptical about power laws may simply view the results for the Pareto distribution as a worked example.

The strength of our approach is its empirical predictions. Because these involve explicit functional relationships between observable variables they are strongly falsifiable in the Popperian sense. A preliminary empirical analysis seems to support the theory, but the statistical analysis so far remains inconclusive. We look forward to more rigorous empirical tests.

Acknowledgements

References

Almgren, R.F., Optimal execution with nonlinear impact functions and trading-enhanced risk. *Appl. Math. Finance*, 2003, **10**, 1–18.

Almgren, R.F. and Chriss, N., Value under liquidation. *Risk*, 1999, **12**, 61–63.

Almgren, R.F. and Chriss, N., Optimal execution of portfolio transactions. *J. Risk*, 2000, **3**, 5–39.

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Equity market impact. *Risk*, 2005, **7**, 57–62.

Back, K. and Baruch, S., Working orders in limit order markets and floor exchanges. *J. Finance*, 2007, **62**, 1589–1621.

Beran, J., *Statistics for Long-Memory Processes*, 1994 (Chapman & Hall: New York).

Berk, J.B. and Green, R.C., Mutual fund flows and performance in rational markets. *J. Polit. Econ.*, 2004, **112**, 1269–1295.

Bernhardt, D., Hughson, E. and Naganathan, G., Cream-skimming and payment for order flow. Technical Report, University of Illinois, 2002.

Bershova, N. and Rakhlin, D., The non-linear market impact of large trades: Evidence from buy-side order flow, 2013a. Available online at: http://ssrn.com/abstract=2197534.

Bershova, N. and Rakhlin, D., High-frequency trading and long-term investors: A view from the buy-side, 2013b. Available online at: http://ssrn.com/abstract=2066884.

Bertismas, D. and Lo, A., Optimal control of execution costs. *J. Financ. Markets*, 1998, **1**, 1–50.

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, pp. 57–160, 2009 (Elsevier: Amsterdam).

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.

Challet, D., The demise of constant price impact functions and single-time step models of speculation. *Physica A.*, 2007, **382**, 29–35.

Chan, L.K.C. and Lakonishok, J., Institutional trades and intraday stock price behavior. *J. Financ. Econ.*, 1993, **33**, 173–199.

Chan, L.K.C. and Lakonishok, J., The behavior of stock prices around institutional trades. *J. Finance*, 1995, **50**, 1147–1174.

Chordia, T. and Subrahmanyam, A., Order imbalance and individual stock returns: Theory and evidence. *J. Financ. Markets*, 2004, **72**, 485–518.

Ding, Z., Granger, C.W.J. and Engle, R.F., A long memory property of stock returns and a new model. *J. Empir. Finance*, 1993, **1**, 83–106.

Donier, J., Market impact with autocorrelated order flow under perfect competition, 2012. Available online at: http://arxiv.org/abs/1212.4770.

Eisler, Z. and Kertesz, J., Size matters, some stylized facts of the market revisited. *Eur. J. Phys. B*, 2006, **51**, 145–154.

Engle, R., Ferstenberg, R. and Russel, J., Measuring and modeling execution cost and risk. Technical Report 08-09, University of Chicago, 2008.

Evans, M.D.D. and Lyons, R.K., Order flow and exchange rate dynamics. *J. Polit. Econ.*, 2002, **110**, 170–180.

Farmer, J.D., Gillemot, L., Lillo, F., Mike, S. and Sen, A., What really causes large price changes? *Quant. Finance*, 2004, **4**, 383–397.

Farmer, J.D., Patelli, P. and Zovko, Ilija, The predictive power of zero intelligence in financial markets. *Proc. Nat. Acad. Sci. U. S. A.*, 2005, **102**, 2254–2259.

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.

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H.E., A theory of power-law distributions in financial market fluctuations. *Nature*, 2003, **423**, 267–270.

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H.E., Institutional investors and stock market volatility. *Q. J. Econ.*, 2006, **121**, 461–504.

Gatheral, J., No-dynamic-arbitrage and market impact. *Quant. Finance*, 2010, **10**, 749–759.

Gerig, A., A theory for market impact: How order flow affects stock price. PhD Thesis, University of Illinois, 2007.

Gerig, A., Farmer, J.D. and Lillo, F., How prices respond to worked orders. Working Paper, 2011.

Gillemot, L., Farmer, J.D. and Lillo, F., There's more to volatility than volume. *Quant. Finance*, 2006, **6**, 371–384.

Glosten, L.R., Is the electronic limit order book inevitable? *J. Finance*, 1994, **49**, 1127–1161.

Glosten, L.R., Discriminatory limit order books, uniform price clearing and optimality. Technical Report, Columbia, 2003.

Gomes, C. and Waelbroeck, H., Is market impact a measure of the information value of trades? Market response to liquidity vs informed trades, 2013. Available from: http://ssrn.com/abstract=2291720.

Gopikrishnan, P., Plerou, V., Gabaix, X. and Stanley, H.E., Statistical properties of share volume traded in financial markets. *Phys. Rev. E*, 2000, **62**, R4493–R4496; Part A.

Hasbrouck, J., Measuring the information content of stock trades. *J. Finance*, 1991, **46**, 179–207.

Hausman, J.A., Lo, A.W. and Mackinlay, A.C., An ordered probit analysis of transaction stock prices. *J. Financ. Econ.*, 1992, **31**, 319–379.

Hopman, C., Do supply and demand drive stock prices? *Quant. Finance*, 2007, **7**, 37–53.

Huberman, G. and Stanzl, W., Price manipulation and quasi-arbitrage. *Econometrica*, 2004, **72**, 1247–1275.

Keim, D.B. and Madhavan, A., The upstairs market for large-block transactions: analysis and measurement of price effects. *Rev. Financ. Stud.*, 1996, **9**, 1–36.

Kempf, A. and Korn, O., Market depth and order size. *J. Financ. Markets*, 1999, **2**, 29–48.

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**, 1–33.

Lillo, F., Farmer, J.D. and Mantegna, R.N., Master curve for price impact function. *Nature*, 2003, **421**, 129–130.

Lillo, F., Mike, S. and Farmer, J.D., Theory for long memory in supply and demand. *Phys. Rev. E*, 2005, **7106**, 066122.

Moro, E., Moyano, L.G., Vicente, J., Gerig, A., Farmer, J.D., Vaglica, G., Lillo, F. and Mantegna, R.N., Market impact and trading profile of hidden orders in stock markets. *Phys. Rev. E.*, 2009, **80**, 066102.

Obizhaeva, A.A. and Wang, J., Optimal trading strategy and supply/demand dynamcis. Technical Report. AFA 2006 Boston Meetings Paper, 2005.

Plerou, V., Gopikrishnan, P., Amaral, L.A.N., Meyer, M. and Stanley, H.E., Scaling of the distribution of price fluctuations of individual companies. *Phys. Rev. E*, 1999, **60**, 6519–6529; Part A.

Plerou, V., Gopikrishnan, P., Gabaix, X. and Stanley, H.E., Quantifying stock price response to demand fluctuations. *Phys. Rev. E*, 2002, **66**, article no. 027104.

Potters, M. and Bouchaud, J.-P., More statistical properties of order books and price impact. *Physica A*, 2003, **324**, 133–140.

Racz, E., Eisler, Z. and Kertesz, J., Comment on 'tests of scaling and universality.' by plerou and stanley. Technical Report, 2009.

Schwartzkopf, Y. and Farmer, J.D., Technical Report. For a preliminary report, see Y. Schwartzkopf's Caltech PhD Thesis, Complex Phenomena in Social and Financial Systems: From bird population growth to the dynamics of the mutual fund industry, 2010.

Torre, N., *BARRA Market Impact Model Handbook*, 1997 (BARRA Inc.: Berkeley).

Toth, B., Lillo, F. and Farmer, J.D., Segmentation algorithm for non-stationary compound Poisson processes. With an application to inventory time series of market members in a financial market Eur. *Phys. J. B.*, 2010, **78**, 235–243.

Toth, 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*, 2011a, **1**, 021006.

Toth, B., Palit, I., Lillo, F. and Farmer, J.D., Why is order flow so persistent? arXiv:1108.1632.

Vaglica, G., Scaling laws of strategic behavior and specialization of strategies in agent dynamics of a financial market. PhD Thesis, University of Palermo, 2008.

Vaglica, G., Lillo, F., Moro, E. and Mantegna, R., Scaling laws of strategic behavior and size heterogeneity in agent dynamics. *Phys. Rev. E.*, 2008, **77**, 036110.

Viswanathan, J. and Wang, James J.D., Market architecture: Limit order books versus dealership markets. *J. Financ. Markets*, 2002, **5**, 127–167.

Weber, P. and Rosenow, B., Large stock price changes: Volume or liquidity? *Quant. Finance*, 2006, **6**, 7–14.

Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M. and Vettorazzo, M., Relation between bid-ask spread, impact and volatility in double auction markets. Technical Report, 2006.

Zhang, Y.C., Toward a theory of marginally efficient markets. *Physica A*, 1999, **269**, 30–44.

Appendix A: Proofs of the propositions

PROPOSITION 1 *The martingale condition implies zero overall profits, i.e.*

$$E_1[N\pi_N] \equiv \sum_{N=1}^{M} p_N N\pi_N = 0. \qquad (A1)$$

Proof Given the definition of R_t^+ and R_t^- we can write the prices as

$$S_t = X_0 + \sum_{i=0}^{t-1} R_i^+ \qquad (A2)$$

$$X_t = X_0 + \sum_{i=0}^{t-1} R_i^+ - R_t^- \qquad (A3)$$

With these expressions for $N < M$, π_N can be rewritten as

$$\pi_N = \frac{1}{N} \sum_{t=1}^{N} S_t - X_N = R_N^- - \frac{1}{N} \sum_{i=1}^{N-1} i R_i^+$$

$$= \frac{1}{p_N} R_N^+ \sum_{i=N+1}^{M} p_i - \frac{1}{N} \sum_{i=1}^{N-1} i R_i^+ \qquad (A4)$$

where in the last equality we have used the martingale condition of equation (7). For $N = M$ the profit per share is

$$\pi_M = \frac{1}{M} \sum_{i=1}^{M} S_i - X_M = \frac{1}{M} \sum_{i=1}^{M} S_i - S_M$$

$$= -\frac{1}{M} \sum_{i=1}^{M-1} i R_i^+ \qquad (A5)$$

By substituting these two last expressions in equation (A1), we obtain

$$E_1[N\pi_N] = \sum_{N=1}^{M-1} N R_N^+ \sum_{i=N+1}^{M} p_i - \sum_{N=1}^{M-1} p_N \sum_{i=1}^{N-1} i R_i^+$$

$$- p_M \sum_{i=1}^{M-1} i R_i^+ = \sum_{N=1}^{M-1} N R_N^+ \sum_{j=N+1}^{M} p_j \qquad (A6)$$

$$- \sum_{N=1}^{M} p_N \sum_{i=1}^{N-1} i R_i^+$$

By explicitly computing the coefficients of each R_i^+, it is easy to show they vanish for each i, i.e. $E_1[N\pi_N] = 0$. □

Infinite support. This proposition does not hold when α has infinite support. in order to show this let us consider the expected profit for orders of length between $N = 1$ and $N = \bar{N}$. The following proposition holds:

PROPOSITION 1 *The martingale condition for all intervals implies that for any integer $\bar{N} \geq 1$,*

$$\sum_{N=1}^{\bar{N}} p_N N\pi_N = \left(\sum_{i=\bar{N}+1}^{M} p_i\right) \left(\sum_{i=1}^{\bar{N}} i R_i^+\right) \geq 0. \qquad (A7)$$

This equation holds both for finite and infinite support (i.e. M can be finite or infinite).

Proof From the equation (A4) in the previous proposition, we know that martingale condition allows us to write the profit per lot traded as

$$\pi_N = \frac{1}{p_N} R_N^+ \sum_{i=N+1}^{M} p_i - \frac{1}{N} \sum_{i=1}^{N-1} i R_i^+$$

Therefore, the expected profit for orders shorter or equal to $\bar{N} < M$ is

$$\sum_{N=1}^{\bar{N}} p_N N\pi_N = \sum_{N=1}^{\bar{N}} N R_N^+ \sum_{i=N+1}^{M} p_i - \sum_{N=1}^{\bar{N}} p_N \sum_{i=1}^{N-1} i R_i^+$$

$$= \left(\sum_{i=\bar{N}+1}^{M} p_i\right) \left(\sum_{i=1}^{\bar{N}} i R_i^+\right) \qquad (A8)$$

This is equal to the quantity in equation (A7). Moreover it is clear that both terms in brackets are non-negative. This means that the market maker typically makes profits on short metaorders.

If the support of p_N is infinite then the martingale condition at all intervals implies that

$$E_1[N\pi_N] \equiv \sum_{N=1}^{\infty} p_N N\pi_N = \lim_{\bar{N}\to\infty} \left(\sum_{i=\bar{N}+1}^{\infty} p_i\right) \left(\sum_{i=1}^{\bar{N}} i R_i^+\right). \qquad (A9)$$

In the infinite support case, the behaviour of the limit in the last term of the above expression depends on the asymptotic behaviour of p_N and R_N^+ for large N. This is due to the fact that for large $\bar{N}$ the first term in brackets goes to zero while the second term diverges. It is possible to construct examples where $E_1[N\pi_N]$ goes to zero, to a finite value, or diverges. This result shows that in the infinite support case the martingale condition does not imply zero overall immediate profits. □

PROPOSITION 2 *If the second derivative of the immediate impact $\mathcal{I}_t$ is bounded strictly below zero, in the limit where the number of informed traders $\mathcal{K} \to \infty$, any Nash equilibrium must satisfy the fair pricing condition $\pi_N = 0$ for $1 < N < M$. On average market makers profit from orders of length one and take (equal and opposite) losses from orders of length M.*

The strategy of the proof is to show that if $\pi_N \neq 0$ for some N, the long-term traders would have an incentive to change strategy, so the equilibrium must satisfy $\pi_N = 0$.

A long-term trader k receives the signal α and buys n_k shares at an average price $\sum_{t=1}^{N} \tilde{S}_t/N$. After averaging over η_t the expected average transaction price is $\sum_{t=1}^{N} S_t/N$ and the expected final price is $X_N = X_0 + \alpha$. The expected profit is therefore,

$$\Pi_k(\alpha) = n_k \left(X_0 + \alpha - \frac{1}{N} \sum_{t=1}^{N} S_t\right) = -n_k \pi(\alpha).$$

If the long-term trader k increases her order size by one lot while all others hold their order size constant, then her expected profit becomes

$$\Pi'_k(\alpha) = (n_k + 1) \left(X_0 + \alpha - \frac{1}{N+1} \sum_{t=1}^{N+1} S_t\right),$$

which after a little algebra can be rewritten,

$$\Pi'_k(\alpha) = -(n_k + 1)\pi(\alpha) - \left(\frac{n_k + 1}{N+1}\right) \left(S_{N+1} - \frac{1}{N} \sum_{t=1}^{N} S_t\right).$$

The expected change in profit is therefore,

$$\Delta\Pi = \Pi'_k(\alpha) - \Pi_k(\alpha) = -\pi(\alpha) - \left(\frac{n_k+1}{N+1}\right)$$
$$\times \left(S_{N+1} - \frac{1}{N}\sum_{t=1}^{N} S_t\right). \qquad (A10)$$

The first term on the right $(-\pi(\alpha))$ represents the additional profit to the long-term trader if it were possible to trade one extra lot at the same average price, and the second term represents the reduction in profit because the average price increases due to the extra lot.

In the limit as $\mathcal{K}$ is large the long-term trader's own order is a small fraction of the aggregated trade, $n_k \ll N$. The second term vanishes in the large N limit if

$$\lim_{N\to\infty} \left(\frac{n_k+1}{N+1}\right)\left(S_{N+1} - \frac{1}{N}\sum_{t=1}^{N} S_t\right) = 0.$$

This is true providing the second derivative of the function S_t is bounded strictly below zero.† Thus in this limit $\Delta\Pi = -\pi(\alpha) > 0$ and the candidate equilibrium fails because the long-term traders have an incentive to deviate. Similarly if $\pi(\alpha) > 0$ the long-term traders take a loss which can be reduced by trading less.

When $\pi(\alpha) = 0$ (and as before $1 < N < M$) no long-term trader has an incentive to change her order size. This is clear since in equation (A10) with $\pi(\alpha) = 0$ the change in profit is given by the second term alone, which is always negative. A similar calculation shows that this is also true for decreasing order size, i.e. when $\pi(\alpha) = 0, n_k \to n_k - 1$ causes $\Delta\Pi < 0$.

To show that this implies $\pi_N = 0$ for all N, we first discretize α into bins labelled by the mode of $p(N|\alpha)$: $\alpha \in \alpha_i \iff Mode(p(N|\alpha)) = i$. Then, we invert the set of homogeneous linear equations $\pi(\alpha_i) = \sum_{N=1}^{M} p(N|\alpha_i)\pi_N = 0$ for $1 < i < M$. Since the mode of $p(N|\alpha_i)$ is equal to i, the matrix $M_{iN} = p(N|\alpha_i)$ is diagonally-dominant, therefore by the Levy-Desplanques theorem it is also non-singluar. It follows that $\pi_N = 0$ for all N.

The cases $N = 1$ and $N = M$ have to be examined separately because our equilibrium argument relied on the ability to increase or reduce n_k, which is not possible at the boundaries. Long-term traders will rationally abstain from taking a loss on metaorders of length $N = 1$ by simply not participating when they receive α signals that are too weak; thus, the trading volume at $N = 1$ is due entirely to the day trader. Similarly, although $\pi_M < 0$, the long-term traders are unable to improve their profits by trading more, since we have bounded the total amount an individual can trade at $M/\mathcal{K}$ so they are blocked from further increase.

The fair pricing condition for $N = 1$ and $N = M$ would also be incompatible with the martingale condition and informational efficiency (i.e. with the conditions on the final price). For $N = 1$ the market makers' profit is

$$\pi_1 = S_1 - X_1 = R_1^-,$$

and from (equation (6)) the martingale condition is

$$\mathcal{P}_1(S_2 - S_1) + (1 - \mathcal{P}_1)(X_1 - S_1) = \mathcal{P}_1(S_2 - X_1) + X_1 - S_1 = 0.$$

Thus if $\mathcal{P}_1 \neq 0$, satisfaction of both the martingale condition and the fair pricing condition‡ requires that $S_1 = S_2 = X_1$, or equivalently that $R_1^- = R_1^+ = 0$. In other words, if both conditions are satisfied then both the permanent and the temporary impact on the first step are

identically zero, which would violate informational efficiency since $\alpha > 0$. In section 5 we show by construction that this holds for all N, i.e. it is clear in equations (13) and (14) that the impacts $\mathcal{I}_t$ and I_N are identically zero if $R_1^- = 0$. To have sensible impact functions, we must have $R_1^- = \pi_1 > 0$, which means that market making is profitable on the first timestep.§

Similarly if $N = M$ the martingale condition implies $S_M = X_M$, i.e. no reversion, and since S_t is an increasing function the market maker takes a loss

$$M\pi_M = \sum_{t=1}^{M} S_t - MX_M = \sum_{i=1}^{M}(S_i - S_M) < 0.$$

Assuming $\pi_N = 0$ for $1 < N < M$, the market makers' profit π_1 and loss $M\pi_M$ are related by equation (8) as

$$\pi_1 p_1 + M\pi_M p_M = 0. \qquad (A11)$$

For realistic size distributions, we expect metaorders of size one to be much more common than those of size M, i.e. $p_1 \gg p_M$. The ratio of the total profits is

$$-\frac{M\pi_M}{\pi_1} = \frac{p_1}{p_M} \gg 1.$$

Thus the market maker receives frequent but small profits on metaorders of length one and rare but large losses for metaorders of length M.

PROPOSITION 3 *The system of martingale conditions (equation 6) and fair pricing conditions (equation (10)) has solution*

$$R_t^+ = \frac{1}{t}\frac{p_t}{\sum_{i=t+1}^{M} p_i}\frac{1 - p_1}{\sum_{i=t}^{M} p_i}R_1^+ \quad t = 2, 3, \ldots, M - 1 \qquad (A12)$$

$$R_t^- = \frac{\mathcal{P}_t}{1 - \mathcal{P}_t}R_t^+ \quad t = 1, 2, \ldots, M - 1 \qquad (A13)$$

Proof The solution of equation (A13) is a direct consequence of the martingale conditions (equation 6). The total profit of metaorders of length $N < M$ can be rewritten as (see proof of Proposition 1)

$$N\pi_N = NR_N^- - \sum_{i=1}^{N-1} iR_i^+ = N\frac{\mathcal{P}_N}{1 - \mathcal{P}_N}R_N^+ - \sum_{i=1}^{N-1} iR_i^+ \quad (A14)$$

The fair pricing conditions (equation 10) state that for $1 < N < M$ it is $N\pi_N = 0$, i.e.

$$R_N^+ = \frac{1}{N}\frac{1 - \mathcal{P}_N}{\mathcal{P}_N}\sum_{i=1}^{N-1} iR_i^+ \qquad (A15)$$

This is a recursive equation which determines R_N^+ once R_1^+ is given (note that this equation does not hold for $N = 1$ because we do not have fair pricing for metaorders of length one). The solution of this equation is

$$R_t^+ = \frac{1}{t}\frac{1 - \mathcal{P}_t}{\mathcal{P}_t}\frac{1}{\mathcal{P}_1\mathcal{P}_2\mathcal{P}_3\ldots\mathcal{P}_{t-1}}R_1^+ \quad t > 1 \qquad (A16)$$

and we prove it by induction. We assume that the solution holds for $N = 2, 3, ..., t - 1$ and we prove that it is true for $N = t$. If equation (A16) holds for $N = 2, 3, .., t - 1$ we can rewrite equation (A15) for

---

†When the distribution of metaorders is a power law with Pareto exponent β, Proposition 2 leads to a power law impact function with exponent $\beta - 1$, so our assumption on the second derivative of immediate impact leads to a self-consistent theory as long as $\beta < 2$. For metaorder distributions with thinner tails than this, the fair pricing condition could be imposed as a hypothesis but it would not follow from a Nash equilibrium. Empirical observations cluster near $\beta = 1.5$.
‡In different terms, the incompatibility of fair pricing and martingale condition was pointed out by Glosten (1994).

§One might be tempted to naively conclude that market makers can defect from the equilibrium by simply trading orders only of length 1, so that they always make a profit. This is false: The profit from a metaorder of length one is $\pi_1 = S_1 - (X_0 + \alpha)$, where α is a small number. In contrast, if a market maker participates only in the first trade of a large metaorder, her profit is $\pi'_1 = S_1 - (X_0 + \alpha')$, where α' is a large number. Thus while $\pi_1 > 0$, $\pi'_1 < 0$.

$N = t$ as

$$R_t^+ = \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t} \sum_{i=1}^{t-1} i R_i^+ = \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t}$$

$$\times \left(1 + \sum_{i=2}^{t-1} \frac{1 - \mathcal{P}_i}{\mathcal{P}_i} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{i-1}} \right) R_1^+ \quad (A17)$$

Now by expanding the sum in brackets it is direct to show that

$$\left(1 + \sum_{i=2}^{t-1} \frac{1 - \mathcal{P}_i}{\mathcal{P}_i} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{i-1}} \right) = 1 - \frac{1}{\mathcal{P}_1} + \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{t-1}}$$
$$(A18)$$

Since, by definition, $\mathcal{P}_1 = 1$ the first two terms in the right-hand side cancel and thus one obtains equation (A16). This equation is equivalent to equation (A12). In fact

$$R_t^+ = \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{t-1}} R_1^+$$

$$= \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^{M} p_i} \frac{1}{\frac{\sum_{i=3}^{M} p_i}{\sum_{i=2}^{M} p_i} \frac{\sum_{i=4}^{M} p_i}{\sum_{i=3}^{M} p_i} \dots \frac{\sum_{i=t}^{M} p_i}{\sum_{i=t-1}^{M} p_i}} R_1^+$$
$$(A19)$$

$$= \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^{M} p_i} \frac{\sum_{i=2}^{M} p_i}{\sum_{i=t}^{M} p_i} R_1^+ = \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^{M} p_i} \frac{1 - p_1}{\sum_{i=t}^{M} p_i} R_1^+$$

i.e. our thesis, equation (A12). $\qquad\qquad\qquad\square$

Appendix B: Market impact for other metaorder size distributions: stretched exponential and lognormal

Changing the metaorder distribution has a dramatic effect on the impact. While we believe that the Pareto distribution is empirically the correct functional form for metaorder size, to get more insight into the role of p_N we compute the impact for two alternative functional forms.

The first one is the stretched exponential, which can be tuned from thin tailed to heavy tailed behaviour and contains the exponential distribution as a special case. There is no simple expression for the normalization factor needed for a discrete stretched exponential distribution, so we make a continuous approximation, in which the metaorder size distribution is

$$p_N = \frac{\lambda}{\Gamma(1/\lambda, 1)} e^{-N^\lambda}. \quad (B1)$$

The normalization factor $\Gamma(a, z)$ is the incomplete Gamma function. The shape parameter $\lambda > 0$ specifies whether the distribution decays faster or slower than an exponential. ($\lambda > 1$ implies faster decay and $\lambda < 1$ implies slower decay.) For short data-sets, when λ is small this functional form is easily confused with a power law.

It can be shown that the stretched exponential leads to an immediate impact function that for large t asymptotically behaves as

$$\mathcal{I}_t \sim \frac{e^{t^\lambda}}{t^{2-\lambda}}. \quad (B2)$$

This is the product of a power law and an exponential; for large t the exponential dominates. The permanent impact is

$$I_N \sim \frac{1}{N} \int^N \frac{e^{x^\lambda}}{x^{2-\lambda}} dx \sim \frac{e^{N^\lambda}\lambda - E_{1+1/\lambda}(-N^\lambda)}{N^2 \lambda^2} \sim \frac{e^{N^\lambda}}{N^2 \lambda}, \quad (B3)$$

where $E_\nu(z)$ is the exponential integral function and in the last approximation we have used its asymptotic expansion. The ratio of the permanent to the immediate impact of the last transaction is

$$\frac{I_N}{\mathcal{I}_N} = \frac{1}{\lambda N^\lambda}. \quad (B4)$$

In contrast to the Pareto metaorder size distribution this is not constant. Instead the ratio between permanent and immediate impact decreases with size, going to zero in the limit as $N \to \infty$.

A similar result is obtained for the case in which metaorder size is lognormal distributed. One of the reasons for considering this form is the empirical results reported in Vaglica et al (2008) where it was shown that the power-law behaviour of metaorder size comes from the heterogeneity of market participants and metaorder size distribution of individual institutions is better fit by a lognormal distribution.

Let us consider for simplicity a standardized lognormal distribution

$$p_N = \frac{1}{N\sqrt{2\pi}} e^{\log^2(N)/2}, \quad (B5)$$

where we have again used the continuous approximation. Note that qualitatively similar results are obtained with non standardized lognormal (i.e. with $\mu \neq 0$ and $\sigma \neq 0$), but the expressions are longer and less transparent.

The immediate impact function for large t can be computed and it is

$$\mathcal{I}_t \sim e^{\log^2(t)/2} \frac{\log(t)}{t} \quad (B6)$$

As expected, this is an increasing convex function of t, at odds with empirical data. The permanent impact is

$$I_N \sim \frac{1}{N} \int^N e^{\log^2(x)/2} \frac{\log(x)}{x} dx = \frac{1}{N} e^{\log^2(N)/2} \quad (B7)$$

The ratio of the permanent to the immediate impact of the last transaction is

$$\frac{I_N}{\mathcal{I}_N} = \frac{1}{\log N}. \quad (B8)$$

As for the stretched exponential case the ratio is not constant but decreases with metaorder size N. The fixed ratio of permanent and immediate impact is a rather special property of the Pareto distribution.

Appendix C: Effect of finite M on impact

As already discussed briefly in section 6.3, if the condition $N \ll M$ is violated this has an effect on the impact. In this section we consider the exact case of a finite support Pareto distribution. We show that when $N \ll M$ we obtain the same results of the previous section and we discuss what happens when $N \approx M$.

We assume that the metaorder size distribution is a truncated Pareto distribution for all $N \leq M$, i.e.

$$p_N = \frac{1}{H_M^{(1+\beta)}} \frac{1}{N^{\beta+1}} \quad N \geq 1 \quad (C1)$$

where the normalization constant $H_M^{(1+\beta)}$ is the harmonic number of order $1 + \beta$. For the truncated Pareto distribution the probability $\mathcal{P}_t$ that a metaorder of size t will continue is

$$\mathcal{P}_t = \frac{\zeta(1+\beta, t+1) - \zeta(1+\beta, M+1)}{\zeta(1+\beta, t) - \zeta(1+\beta, 1+M)} \quad (C2)$$

where $\zeta(s, a)$ is the generalized Riemann zeta function (also called the Hurwitz zeta function). For small t the function $\mathcal{P}_t$ increases meaning that it is more and more likely that the order continues. In the regime of $t \ll M$, $\mathcal{P}_t$ is well approximated by the expression of equation (17) for an infinite support Pareto distribution. However, around $t \simeq M/2$, $\mathcal{P}_t$ starts to decrease meaning that it becomes more and more likely that the order is going to stop soon, with a corresponding effect on the impact.

The immediate impact can be easily calculated once the distribution of metaorder size is known by using equation (13). For truncated Pareto distributed metaorder sizes, R_t^+ is (for $t > 2$) is

$$R_t^+ = \left(\frac{H_M^{(1+\beta)} - 1}{(\zeta(1+\beta, t) - \zeta(1+\beta, M+1))(\zeta(1+\beta, t+1) - \zeta(1+\beta, M+1))} \right) \frac{R_1^+}{t^{2+\beta}} \tag{C3}$$

For large t but $t \ll M$ it is

$$R_t^+ \sim \frac{R_1^+}{t^{2-\beta}} \tag{C4}$$

which is the same scaling as the infinite support Pareto distribution (see equation (19)). The same holds true for the permanent impact. We have therefore shown that when $t \ll M$ the finite support of the metaorder size distribution is irrelevant and we obtain approximately the same results as in section 6.2.

The finite size effects and the role of the finiteness of the support becomes relevant when $t \gtrsim M/2$. Figure 3 shows the total impact for $M = 1000$ and different values of β. It is clear that the impact is initially described by a power law, but then it becomes strongly convex when the order length becomes comparable with the maximal length.