

# How efficiency shapes market impact

J. Doyne Farmer,<sup>1,2</sup> Austin Gerig,<sup>3,1</sup> Fabrizio Lillo,<sup>1,4</sup> and Henri Waelbroeck<sup>5</sup>

<sup>1</sup>*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501*

<sup>2</sup>*LUISS Guido Carli, Viale Pola 12, 00198, Roma, Italy*

<sup>3</sup>*School of Finance and Economics, University of Technology,  
Sydney, Broadway, New South Wales 2007, Australia*

<sup>4</sup>*Dipartimento di Fisica e Tecnologie Relative,  
viale delle Scienze I-90128, Palermo, Italy*

<sup>5</sup>*Pipeline Financial Group, Inc., 60 E. 42<sup>nd</sup> St., Suite 624, New York NY 10165*

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. Within a framework in which informed traders are competitive we 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 temporary and permanent components of market impact. Based on a commonly observed asymptotic distribution for the volume of large trades, it says that market impact should increase asymptotically roughly as the square root of size, with average permanent impact relaxing to about two thirds of peak impact.

## Contents

<b>I. Introduction</b>	3
<b>II. Model description</b>	5
A. Framework	5
B. Simplifications and empirical interpretation	7
<b>III. Preliminaries</b>	8
A. Expectations	8
B. Metaorder size and the persistence of order flow	8
<b>IV. Martingale condition</b>	8
A. Asymmetric price response	9
B. Zero overall immediate profits	9
<b>V. Fair pricing</b>	10
A. Derivation of Nash equilibrium	10
B. Discussion	12
<b>VI. General expressions for impact</b>	13
A. General solution for immediate impact	14
B. General solution for permanent impact	14
C. Setting the scale	14
D. Relation between volume and information	15
<b>VII. Dependence on the metaorder size distribution</b>	15
A. Empirical evidence supporting the Pareto distribution	15
1. Direct observation	15
2. Connection to long-memory in order flow	16
3. Reconstruction from membership codes	17
B. Market impact for Pareto metaorder size	17
1. Immediate impact	17
2. Permanent impact	18
C. Effect of maximum order size	18
D. Market impact for stretched exponential metaorder size	20
<b>VIII. Discussion</b>	20
A. Empirical implications	20
B. Information revelation	21
C. Final thoughts	21
<b>Acknowledgments</b>	22
<b>References and Notes</b>	22
<b>A. Proofs of the propositions</b>	26
<b>B. Effect of finite <math>M</math> on impact</b>	28

## I. 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<sup>1</sup>. 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<sup>2</sup>. 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*<sup>3</sup>. 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 analyzed by Kyle (1985), who developed a model for an inside trader with information about future prices. Kyle was able to show that such a trader will 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.

Real data, however, does not show linear impact. Empirical studies consistently find that impact is concave, i.e. impact per share is a decreasing function of size<sup>4</sup>. It is possible to reconcile the empirical findings and the Kyle model by making the additional hypothesis that larger trades contain less information per share than smaller trades. This might be true for behavioral or institutional reasons, e.g. because more informed traders prefer a particular size trade for tactical reasons. A drawback of this approach 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.

We derive how prices respond to metaorders by assuming two conditions. The first condition is that transaction prices are a martingale with respect to information about the number of trades a metaorder has so far executed (we assume this information is public

---

<sup>1</sup> See Bertismas and Lo (1998), Almgren and Chriss (1999,2000), Almgren (2003), Almgren, Thum, and Hauptmann (2005), and Obizhaeva and Wang (2005).

<sup>2</sup> See for example Berk and Green (2004). Schwartzkopf and Farmer (2010) have shown that managers of large funds offset increases in market impact by lowering fees, and diversifying assets.

<sup>3</sup> Other names used in the literature are large trades, packages, or hidden orders.

<sup>4</sup> An early study of institutional trades is Chan and Lakonishok (1993, 1995). Later studies by Torre (1997), Almgren et al. 2005, and Moro et al. (2009) find concave impacts roughly consistent with a square root functional form. These should be distinguished from the large number of studies that either look at 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 trades of a given institution. See Hasbrouck (1991), Hausman, Lo and MacKinlay (1992), Keim and Madhavan (1996), Torre (1997), Kempf and Korn (1999), Plerou et al. (2002), Evans and Lyons (2002), Lillo, Farmer, and Mantegna (2003), Potters and Bouchaud (2003), Gabaix et al. (2003, 2006), Chordia and Subrahmanyam 2004, Farmer, Patelli and Zovko (2005), Weber and Rosenow (2006), and Hopman (2006).

knowledge). If the martingale condition did not hold, then profitable opportunities would exist that could be exploited by market participants. The second condition is that the final price after a metaorder completes must equal the average transaction price of the metaorder. We call this condition *fair pricing*. This is “fair” in the sense that the transaction price of the metaorder is equal to the price of the security given that the order was placed. When this condition is violated, there is incentive for investors to seek other execution services where they can obtain a fair price. Although we assume these conditions when deriving impact, we also show they result from the equilibrium strategies of market participants. The key difference with Kyle is that informed trading is competitive and multiple informed traders act at the same time.

Our theory makes several strong predictions based on a simple set of hypotheses. For a given metaorder size distribution it predicts impact as a function of time both during and after execution. We thus make sharp predictions relating the functional form of two observable quantities with no a priori relationship, making our theory falsifiable in a strong sense. This is in contrast to theories that make assumptions about the functional form of utility and/or behavioral or institutional assumptions about the informativeness of trades, which typically leave room for interpretation and require auxiliary assumptions to make empirical tests.

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 similar to one of our results here. However, in contrast to our model, their prediction for market impact depends on the functional form for risk aversion. They argue that if risk aversion is proportional to  $\sigma^\delta$ , where  $\sigma$  is the standard deviation of profits, the impact will increase with the size  $N$  of the metaorder as  $N^{\delta/2}$ . The impact will be concave if  $\delta < 2$ , whereas if  $\delta = 2$  (risk proportional to variance) the impact is linear. Our theory, in contrast, is based entirely on market efficiency and does not depend on the functional form of utility.

Another important point of comparison is Huberman and Stanzl (2004). They derive a linear impact function based on the assumption that impact is permanent and time independent, and that there is no arbitrage by splitting or combining orders. Like them, we assume no-arbitrage. However we argue here that due to the long-memory of order flow, which is a robust and well-established empirical fact<sup>5</sup>, impact must be either time dependent and/or purely temporary, violating their assumptions, as also pointed out by Challet (2007). A similar point has been made recently by Gatheral (2010) who showed that a non-permanent (and non exponentially decaying) impact function does not violate the no-arbitrage condition when the impact is a non linear function of the volume.

Our approach here was originally motivated by the discovery of long-memory order flow by Bouchaud et al. (2004) and by Lillo and Farmer (2004). This led to the development of efficiency-based theories for market impact of individual transactions<sup>6</sup>. However, 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.

Finally, this work is related to several papers that study market design. Viswanathan and

---

<sup>5</sup> See the discussion in Section VIA2.

<sup>6</sup> See Bouchaud et al. (2006), Farmer et al. (2006), Wyart et al. (2006), and Bouchaud, Farmer and Lillo (2008). 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)

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 setup 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.”<sup>7</sup> 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, and they have no incentive to search out arrangements for better execution. 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.

In Section II we give a description of the model and present an outline of the approach we use here. In Section III we give several useful preliminary results, deriving the persistence of order flow from the metaorder size distribution and illustrating how we will decompose profits. In Section IV 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 V we show that any Nash equilibrium must satisfy the fair pricing condition. In Section VI we derive in general terms what this implies about market impact. In Section VII we introduce specific functional forms for the metaorder size distribution and explicitly compute the impact for these cases. Finally, in Section VIII we discuss the empirical implications of the model and make some concluding remarks.

## II. MODEL DESCRIPTION

### A. Framework

The three types of agents in the model are informed traders, market makers, and liquidity traders. The *informed traders* can be thought of as long-term investors such as portfolio managers. Their decisions are based on a perfect information signal  $\alpha$  of the future price  $S_0 + \alpha$ , where  $S_0$  is the initial price. The signal  $\alpha$  is drawn from a distribution  $p(\alpha)$ , which has nonzero support over a continuous interval  $0 \leq \alpha < \alpha_{max}$ , where  $0 < \alpha_{max} \leq \infty$ . The information distribution  $p(\alpha)$  is given exogenously. For convenience we assume that  $\alpha$  is positive. The informed traders are rational, so they either buy or do nothing.

The orders of the informed traders are aggregated together and executed in a package. The bundling process can be thought of as a broker or an algorithmic trading firm. There are  $\mathcal{K}$  informed traders, labeled by an index  $k = 1, 2, \dots, \mathcal{K}$ , where  $\mathcal{K}$  is a large number. All informed traders receive the same  $\alpha$ . When the information is received each trader submits an order of size  $n_k(\alpha)$  and the individual orders are bundled together into a metaorder of size  $N = \sum_{i=1}^{\mathcal{K}} n_k$ . Each individual order  $n_k(\alpha)$  is made without knowing the other order sizes.

The metaorder is split into pieces and executed in increments of one share at successive times  $t = 1, \dots, N$  at transaction prices  $\tilde{S}_t$ , where  $1 < N < M$ . For any realistic application

---

<sup>7</sup> For example, see Bernhardt, Hughson, and Naganathan (2002)

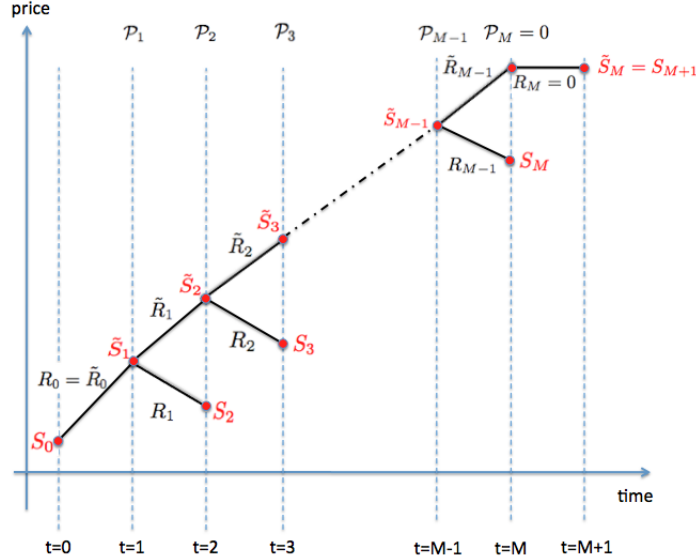


FIG. 1: The tree of possible price paths for a buy metaorder for different sizes  $N$ . The price is initially  $S_0$ ; after the first increment is executed it rises to  $\tilde{S}_1 = S_0 + R_0$ . If  $N = 1$  it is finished and on average the price reverts to  $S_2 = \tilde{S}_1 - R_1$ , but if  $N > 1$  another increment is executed and it rises to  $\tilde{S}_2 = \tilde{S}_1 + \tilde{R}_1$ . This proceeds similarly until the execution of the metaorder is completed. At any given point the probability that the metaorder has size  $N > t$ , and therefore that the order continues, is  $\mathcal{P}_t$ . The maximum size is  $N = M$ ; for reasons that we will explain later, the martingale condition requires that  $R_M = \tilde{R}_M = 0$ . The *immediate impact* is  $\mathcal{I}_t \equiv \tilde{S}_t - S_0$ , and the *permanent impact* is  $I_N \equiv S_{N+1} - S_0$ .

we assume that  $M$  is very large<sup>8</sup>. The execution process is illustrated in Figure 1. The counterparties of the trades are *market makers*<sup>9</sup>, who provide liquidity. At each time step the market makers competitively make bids and offers in order to maximize their profits. For convenience we assume that only one metaorder is executed at a time.

There is also an uninformed liquidity trader who places a single order<sup>10</sup> without regard to  $\alpha$ . This order is bundled together with those of the informed traders. Thus, the overall process is as follows: The informed traders and the liquidity trader place their orders, the metaorder is executed over the subsequent  $N$  steps, and then there is another round of order placement and metaorder execution.

We assume informational efficiency. When the execution of the metaorder is completed,

<sup>8</sup> We have imposed a bound on  $M$  to avoid mathematical problems that occur in the limit  $M \rightarrow \infty$ . For simplicity assume the bound on  $M$  is enforced by placing a limit  $(M - 1)/\mathcal{K}$  on the number of shares a single informed trader can trade. For the typical situations we have in mind  $M$ ,  $\mathcal{K}$ , and  $(M - 1)/\mathcal{K}$  are all large numbers.

<sup>9</sup> Note that we are not assuming any special institutional role for market makers, such as the specialists in the NYSE. Also we are not making any assumptions about market microstructure, e.g. the use of limit vs. market orders.

<sup>10</sup> We could have more realistically allowed liquidity trades with random size but having a single background order simplifies the derivation of the Nash equilibrium and is sufficient to illustrate the basic idea.

at timestep  $t = N + 1$  the price goes to the *reversion price*  $S_{N+1} \equiv S_0 + \alpha$ . The *immediate impact* is  $\mathcal{I}_t \equiv \tilde{S}_t - S_0$ , and the *permanent impact* is  $I_N \equiv S_{N+1} - S_0$ .

The market makers know the initial price  $S_0$ , the distribution  $p_N$  of metaorder sizes, and the distribution  $p(\alpha)$ , but they do not know the informed traders signal  $\alpha$  and hence do not know the metaorder size  $N$ . We assume the number of informed traders  $\mathcal{K}$  is fixed and is common knowledge.

Our main result is to derive an equilibrium between the size distribution  $p_N$ , and the immediate and permanent impacts  $\mathcal{I}_t$  and  $I_N$ . The key new result is the fair pricing condition, which states that the average transaction price is equal to the reversion price.

## B. Simplifications and empirical interpretation

Although we have designed this model to make realistic predictions of market impact we have made several simplifications that require some interpretation, particularly to understand the empirical implications.

Real prices are noisy due to the background order flow. We are only interested in the price effects due to a given metaorder, and so we assume there is only one active metaorder at a time. To make a correspondence with real prices, the impacts  $\mathcal{I}_t$  and  $I_N$  should be viewed as ensemble averages over the impact of many different metaorders of a given sign, each of which generates a different price path<sup>11</sup>.

We have not specified the time period  $\tau$  that elapses between each interval  $t$ , but from the informational assumptions of the model  $\tau$  is effectively defined as the time needed for the market maker to detect the presence or absence of a given metaorder. The market maker might infer this through analysis of order flow imbalance, or through idiosyncrasies of order placement, such as a tendency to use a particular order size. The existence of algorithms for detecting metaorders from order flow suggests that this is feasible, at least when brokerage codes are known<sup>12</sup>. Detecting that a metaorder is finished might be more difficult than detecting it is still present, so in real time the last interval (from  $N$  to  $N + 1$ ) might be considerably longer than the others. There is nothing in our analysis that requires that  $\tau$  be fixed in real time.

Another simplification we have made in the model is the assumption that  $M$  is finite. As we discuss in Section 5, this avoids technical problems that occur when  $M$  is infinite. In reality  $M$  is bounded by, for example, the market capitalization or the leveraged wealth of the richest investor, and thus is usually a large number. In almost any realistic setting we expect that  $N \ll M$ , and unless otherwise noted we will assume this throughout the paper.

<sup>11</sup> To make the model more realistic we could have added noise traders, who generate random order flow, then averaged over the random order flow to compute the impact. This would give the same solution we have derived here, but just add one more element of clutter in the model. Thus we have not included noise traders.

<sup>12</sup> See Vaglica et al. (2008) and Toth et al. (2010).

### III. PRELIMINARIES

#### A. Expectations

The crux of our argument hinges around the market maker's ignorance of the metaorder's true size  $N$ , which is the only source of uncertainty. If the market maker has already observed  $t$  time intervals during which the metaorder was active then she knows that the size  $N$  of the metaorder is larger or equal to  $t$ . The probability that the metaorder has size  $N \geq t$  is  $p_N / \sum_{N=t}^M p_N$ . We will use the notation  $E_t$  to represent an average over all metaorders of size  $N \geq t$ . More specifically, for a generic function  $f_N$  this average is

$$E_t[f_N] = \frac{\sum_{N=t}^M p_N f_N}{\sum_{N=t}^M p_N} = \frac{\sum_{i=0}^{M-t} p_{t+i} f_{t+i}}{\sum_{i=0}^{M-t} p_{t+i}}, \quad (1)$$

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

#### B. Metaorder size and the persistence of order flow

The likelihood that a metaorder 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  $t$ . This is equivalent to the probability that it is at least of size  $t$ . This is

$$\mathcal{P}_t = \frac{\sum_{i=t+1}^M p_i}{\sum_{i=t}^M p_i}. \quad (2)$$

If  $p_N$  is an exponential distribution, it is possible to show that for  $t \ll M$ ,  $\mathcal{P}_t$  is approximately independent of  $t$ . In the same approximation, if  $p_N$  has heavier tails than an exponential, then as  $t$  increases it becomes increasingly more likely to continue, whereas if it has thinner tails than an exponential it becomes increasingly more likely to stop. This plays an important role in determining the shape of market impact.

The incremental transactions caused by an active buy metaorder impart a positive bias to order flow. As long as a metaorder is present, all else being equal, order flow will be positively autocorrelated. As discussed in more detail in Section VII A 2, empirical data from equity markets consistently shows that real order flow is a long-memory process, which has been shown to be mainly due to order splitting of the type we are modeling here. This suggests that the underlying  $p_N$  is very heavy tailed.

### IV. MARTINGALE CONDITION

In this section we discuss the relationship between the size distribution  $p_N$  and the persistence of order flow. We introduce a martingale condition, derive the implications of the persistence of order flow on liquidity, and investigate the implications of the martingale condition on the immediate profits.



### A. Asymmetric price response

Since the market maker does not know the metaorder size, from her point of view, or that of any external observer other than the informed trader, at equilibrium the price process is a martingale. This is necessary to prevent arbitrage by an uninformed trader.

The time varying persistence of order flow and the martingale condition combine to induce time varying differences in the response to buy vs. sell orders. At each time step there are two possibilities: The metaorder continues with probability  $\mathcal{P}_t$  or it stops with probability  $1 - \mathcal{P}_t$ . Arbitrage efficiency implies the *martingale property*

$$\mathcal{P}_t(\tilde{S}_{t+1} - \tilde{S}_t) + (1 - \mathcal{P}_t)(S_{t+1} - \tilde{S}_t) = 0. \quad (3)$$

This is true for  $t = 1, 2, \dots, M - 1$ . If the metaorder has maximal length at the end of the  $M$ th interval by definition  $\mathcal{P}_M = 0$ , which implies that  $\tilde{S}_M = S_{M+1}$ .

Equation (3) can be trivially rewritten in the form

$$\frac{\tilde{R}_t}{R_t} = \frac{1 - \mathcal{P}_t}{\mathcal{P}_t}, \quad (4)$$

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

To complete the calculation we need another condition to set the scale of the price responses  $R_t$  and  $\tilde{R}_t$ , which may change as  $t$  varies<sup>13</sup>. Such a condition is introduced in Section V.

### B. Zero overall immediate profits

The martingale condition alone requires that the market maker has to break even overall, i.e. that their total profits when added across all sizes must be zero.

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

$$\Pi = E_1[\pi_N] \equiv \sum_{N=1}^M N\pi_N p_N = 0, \quad (5)$$

where

$$\pi_N \equiv \frac{1}{N} \sum_{t=1}^N \tilde{S}_t - S_{N+1} \quad (6)$$

is the *immediate profit per share* of the market maker, who sells  $N$  shares at prices  $\tilde{S}_t$ , under the assumption that they are valued at the reversion price  $S_{N+1}$ . The proof of proposition 1 is given in Appendix A. The phrase “overall profits” emphasizes that the martingale condition

<sup>13</sup> Various authors have used alternative conditions. For example, Gerig (2007) used 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.

only implies zero immediate profits when averaged over metaorders of all sizes: It allows for the possibility that the market maker may make immediate profits on metaorders in a given size range, as long as she takes corresponding losses in another size range.

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 zero overall immediate profits holds are more complicated, as discussed in Appendix A.

## V. FAIR PRICING

We now derive the *fair pricing condition*, which states that for any  $N$

$$\pi_N = \frac{1}{N} \sum_{i=1}^N \tilde{S}_i - S_{N+1} = 0. \quad (7)$$

This says that the average price is equal to the reversion price, or in other words, the average price that the informed trader pays to execute the metaorder is equal to the final price after it has been executed. We call this fair pricing for the obvious reason that one would naturally regard this as “fair”: It means that the current value of the asset, as priced by the market, is equal to the price that was paid.

### A. Derivation of Nash equilibrium

**Proposition 2.** *If the immediate impact  $\mathcal{I}_t$  is concave, in the limit where the number of informed traders  $\mathcal{K} \rightarrow \infty$ , any Nash equilibrium must satisfy the fair pricing condition  $\pi_N = 0$  for  $1 < N < M$ .*

This result is driven by competition between informed traders. All informed traders receive the same information signal  $\alpha$ , and the strategy of informed trader  $k$  is simply the choice of the order size  $n_k(\alpha)$ . The orders are then bundled together (with the order of the liquidity trader) 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, 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.

First consider a candidate equilibrium with  $\pi_N < 0$  for some value of  $N$ , where  $1 < N < M$ . Assume informed trader  $k$  buys  $n_k$  shares at an average price  $(\sum_{t=1}^N \tilde{S}_t)/N$ . Since we have assumed informational efficiency, the shares are subsequently valued at price  $S_{N+1} = S_0 + \alpha$ . Her profit is

$$\Pi_k(\alpha) = n_k \left( S_0 + \alpha - \frac{1}{N} \sum_{t=1}^N \tilde{S}_t \right) = -n_k \pi_N > 0$$

If she increases her order size by one share while all other traders hold their order size the same, her profit becomes

$$\Pi'_k(\alpha) = (n_k + 1) \left( S_0 + \alpha - \frac{1}{N+1} \sum_{t=1}^{N+1} \tilde{S}_t \right).$$

After some algebraic manipulation the change in profit can be written

$$\Delta\Pi = \Pi'_k(\alpha) - \Pi_k(\alpha) = -\pi_N - \left( \frac{n_k + 1}{N+1} \right) \left( \tilde{S}_{N+1} - \frac{1}{N} \sum_{t=1}^N \tilde{S}_t \right). \quad (8)$$

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

At an equilibrium, since there is nothing to distinguish the informed traders, they all make the same decision. Thus if there are  $\mathcal{K}$  traders buying  $n_k$  shares, plus a liquidity trader buying one share, at an equilibrium  $N = \mathcal{K}n_k + 1$ . Thus if  $\mathcal{K}$  is large,  $N$  is also large. In the limit as  $N$  is large the second term vanishes for fixed  $n_k$  if

$$\lim_{N \rightarrow \infty} \left( \frac{n_k + 1}{N+1} \right) \left( \tilde{S}_{N+1} - \frac{1}{N} \sum_{t=1}^N \tilde{S}_t \right) = 0.$$

This is true providing the function  $\tilde{S}_t$  is concave. Thus in this limit  $\Delta\Pi = -\pi_N > 0$  and the candidate equilibrium fails because the informed trader has an incentive to deviate. Similarly if  $\pi_N > 0$  the informed traders take a loss which can be reduced by trading less.

When  $\pi_N = 0$  (and as before  $1 < N < M$ ) no informed trader has an incentive to change her order size. This is clear since in Eq. (8) with  $\pi_N = 0$  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  $\pi_N = 0$ ,  $n_k \rightarrow n_k - 1$  causes  $\Delta\Pi < 0$ .

The cases  $N = 1$  and  $N = M$  have to be examined separately because in these cases the fair pricing condition is incompatible with the martingale condition and informational efficiency. For  $N = 1$  the market maker's profit is

$$\pi_1 = \tilde{S}_1 - S_2 = R_1,$$

and from (Eq. 3) the martingale condition is

$$\mathcal{P}_1(\tilde{S}_2 - \tilde{S}_1) + (1 - \mathcal{P}_1)(S_2 - \tilde{S}_1) = \mathcal{P}_1(\tilde{S}_2 - S_2) + S_2 - \tilde{S}_1 = 0.$$

Thus if  $\mathcal{P}_1 \neq 0$ , satisfaction of both the martingale condition and the fair pricing condition<sup>14</sup> requires that  $\tilde{S}_1 = \tilde{S}_2 = S_2$ , or equivalently that  $R_1 = \tilde{R}_1 = 0$ . 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 VI we show by construction that this holds for all  $N$ , i.e. it is clear in Eq. (12) and (13) that

<sup>14</sup> In different terms, the incompatibility of fair pricing and the martingale condition was pointed out by Glosten (1994).

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  $\tilde{S}_M = S_{M+1}$ , i.e. no reversion, and since  $\tilde{S}_t$  is an increasing function the market maker takes a loss

$$M\pi_M = \sum_{t=1}^M \tilde{S}_t - MS_{M+1} = \sum_{i=1}^M (\tilde{S}_i - \tilde{S}_M) < 0.$$

Assuming  $\pi_N = 0$  for  $1 < N < M$ , the market makers' profit  $\pi_1$  and loss  $M\pi_M$  are related by Eq. (5) as

$$\pi_1 p_1 + M\pi_M p_M = 0. \quad (9)$$

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  $M$ .

Informed traders will rationally abstain from taking a loss on metaorders of length  $N = 1$  by simply not participating when they receive  $\alpha$  signals that are too weak; thus, the trading volume at  $N = 1$  is due entirely to the liquidity trader. Similarly, although  $\pi_M < 0$ , informed traders are unable to improve their profits by trading more, since we have bounded the total amount an individual can trade at  $(M - 1)/\mathcal{K}$  so they are blocked from further increase. Thus the violations of the fair pricing condition when  $N = 1$  and  $N = M$  occur naturally due to the institutional constraints that we have assumed and do not invalidate the equilibrium.

In conclusion, at the equilibrium informed traders make a small profit at the expense of the liquidity traders, while market makers break even on average<sup>15</sup>, profiting from metaorders of length one and taking occasional losses on metaorders of length  $M$ . The fair pricing condition holds for  $1 < N < M$ .

## B. Discussion

The setup we have used is a bit contrived in order to simplify the proof of the Nash equilibrium, but we think the basic idea will hold in a more general setting, e.g. with a more realistic size distribution of liquidity trades. In a more realistic setting where the number of informed traders is not so large, deviations from the Nash equilibrium might occur. In this case  $\pi_N$  is not necessarily zero, but it is still bounded by the same basic reasoning, and should be a reasonable approximation.

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

<sup>15</sup> 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  $\pi_1 = \tilde{S}_1 - (S_0 + \alpha)$ , where  $\alpha$  is a small number. In contrast, if a market maker participates only in the first trade of a large metaorder, her profit is  $\pi'_1 = \tilde{S}_1 - (S_0 + \alpha')$ , where  $\alpha'$  is a large number. Thus while  $\pi_1 > 0$ ,  $\pi'_1 < 0$ .

size. It implies that no one pays any costs or makes any profits simply by trading in any particular size range.

Although this derivation is based on rationality, the fair pricing condition potentially stands on its own, even if other aspects of rationality and efficiency are violated. In modern markets portfolio managers routinely receive trade cost analysis reports that compare their execution prices relative to the close, which for a trade that takes place over one day is a good proxy for  $S_{N+1}$ . Such reports are typically broken down into size bins, making persistent inconsistencies across sizes clear. Execution times for metaorders range from less than a day to a few months (Vaglica et al., 2008), and are much shorter than typical holding times, which for mutual funds are on average a year and are often much more (Schwartzkopf and Farmer, 2010). Thus the statistical fluctuations for assessing whether fair pricing holds are much smaller than those for assessing informational efficiency. Since  $\pi_N$  is fairly well determined, portfolio managers will exert pressure on their brokers to provide them with good execution. As a result we expect the fair pricing condition to be obeyed to a higher degree of accuracy than the informational efficiency condition.

## VI. GENERAL EXPRESSIONS FOR IMPACT

Although we have derived the Nash equilibrium only in the case where the immediate impact is concave, for the remainder of the paper we will simply 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 (Eq. 3) and the fair pricing condition (Eq. 7) define a system of linear equations for  $\tilde{S}_t$  and  $S_t$  at each value of  $t$ , which we can alternatively express in terms of the price differences  $\tilde{R}_t = \tilde{S}_{t+1} - \tilde{S}_t$  and  $R_t = \tilde{S}_t - S_{t+1}$ , where  $t = 1, \dots, M$ . The martingale condition holds for  $t = 1, \dots, M$  and the fair pricing condition holds for  $t = 2, \dots, M - 1$ . There are thus  $2M - 2$  homogeneous linear equations with  $2M - 1$  unknowns<sup>16</sup>. Because the number of unknowns is one greater than the number of conditions there is necessarily an undetermined constant, which we choose to be  $\tilde{R}_1$ .

**Proposition 3.** *The system of martingale conditions (Eq. 3) and fair pricing conditions (Eq. 7) has solution*

$$\tilde{R}_t = \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^M p_i} \frac{1 - p_1}{\sum_{i=t}^M p_i} \tilde{R}_1 \quad t = 2, 3, \dots, M - 1 \quad (10)$$

$$R_t = \frac{\mathcal{P}_t}{1 - \mathcal{P}_t} \tilde{R}_t \quad t = 1, 2, \dots, M - 1 \quad (11)$$

The proof is given in Appendix A.

Note that an important property of the solution is equivalence of the impact  $\mathcal{I}_t$  as a function of either time or size. That is, the immediate impact as a function of time is  $\mathcal{I}_t = \sum_{i=1}^t \tilde{R}_i$ , and as a function of size is  $\mathcal{I}_N = \sum_{t=1}^N \tilde{R}_t$ ; in either case this is the same function. This means that up until the execution of a metaorder is completed, the impact from the first  $t$  steps of a metaorder of length  $N > t$  is the same, regardless of  $N$ .

<sup>16</sup> The price  $\tilde{S}_{M+1}$  does not exist, so  $\tilde{R}_M$  is not needed. This reduces the number of unknowns by one.

### A. General solution for immediate impact

Summing Eq. 10 implies that for  $N > 2$  the immediate impact is

$$\mathcal{I}_t = \tilde{S}_t - S_0 = R_0 + \tilde{R}_1 \left( 1 + \sum_{k=2}^{t-1} \frac{1}{k} \frac{p_k}{\sum_{i=k+1}^M p_i} \frac{1-p_1}{\sum_{i=k}^M p_i} \right), \quad (12)$$

For  $t = 1$  the immediate impact is  $\mathcal{I}_1 = \tilde{S}_1 - S_0 = \tilde{R}_0 = R_0$  and for  $t = 2$  it is  $\mathcal{I}_2 = \tilde{S}_2 - S_0 = R_0 + \tilde{R}_1$ . (The meaning of the undetermined constants  $\tilde{R}_1$  and  $R_0$  is discussed in a moment).

### B. General solution for permanent impact

The permanent impact  $S_{N+1} - \tilde{S}_1$  is easily obtained. Making some simple algebraic manipulations

$$S_{N+1} = S_{N+1} - \tilde{S}_N + \tilde{S}_N = \tilde{S}_N - R_N = \tilde{S}_N - \frac{\mathcal{P}_N}{1 - \mathcal{P}_N} \tilde{R}_N.$$

By combining Eqs. (12) and (2) we get

$$I_N = S_{N+1} - S_0 = R_0 + \tilde{R}_1 \left( 1 + \sum_{k=2}^{N-1} \frac{1}{k} \frac{p_k}{\sum_{i=k+1}^M p_i} \frac{1-p_1}{\sum_{i=k}^M p_i} - \frac{1-p_1}{N \sum_{i=0}^{M-N} p_{N+i}} \right). \quad (13)$$

### C. 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  $\tilde{R}_1$ . 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  $\bar{\alpha}_1$ , and signals in the range  $\alpha \in (\alpha_1, \alpha_2]$  will be assigned to metaorders of size two, with an average size  $\bar{\alpha}_2$ . Thus the scale is set by the relations

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

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

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

The power of the theory developed here is that the impact is predicted in terms of  $p_N$ , which is directly measurable (at least with the proper data). Barring the ability to independently measure information, the constants  $R_0$  and  $\tilde{R}_1$  remain undetermined parameters.  $\tilde{R}_1 > 0$  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. It never enters the analysis, it is simply something that has to be added to include the impact of the first trade.

### D. Relation between volume and information

There is a simple relationship between the distribution of information and the distribution of volume. The relationship  $\bar{\alpha}_N = I_N/q$  approximately maps the trading size  $N$  onto the signal  $\alpha$ . In the continuum limit the relationship takes on the simple form  $\alpha(N) = I(N)/q$ , with a unique inverse  $qN(\alpha)$ . The two distributions are related by conservation of probability as

$$p(\alpha) = qp_N \frac{dN}{d\alpha}. \quad (14)$$

So for example, if  $p_N$  is Pareto distributed, as studied in the next section,  $P_N \sim N^{-(\beta+1)}$ ,  $\alpha = N^{\beta-1}$ , and  $p(\alpha) = \alpha^{(1-2\beta)/(\beta-1)}$ . Based on the empirically estimated value  $\beta \approx 1.5$ , this gives  $p(\alpha) = \alpha^{-4}$ , which means that the cumulative scales as  $P(\alpha > x) \sim x^{-3}$ , which is typically what is observed for price returns in American stock markets (Plerou et al., 1999).

## VII. 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 two examples. The first of these is the Pareto distribution,

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

which we argue is well-supported by empirical data<sup>17</sup>. The second is the stretched exponential distribution, which is not supported by data, but provides a useful point of comparison.

### A. Empirical evidence supporting the Pareto distribution

There is now considerable evidence that in the large size limit in most major equity markets the metaorder size  $V$  is distributed as  $P(V > v) \sim v^{-\beta}$ , i.e. that it is a power law with tail exponent  $\beta \approx 1.5$ . This evidence takes several forms: (1) Direct observations of trade size, (2) observations of long-memory in correlated order flow, and (3) reconstruction of large metaorders using data with exchange membership codes. As discussed below, all three of these indicate power law behavior and yield similar values of  $\beta$ .

#### 1. Direct observation

A power law distribution for the trading volume  $V$  was originally observed for trades in the American stock markets by Gopikrishnan et al. (2000), and also for the Paris and London stock markets by Gabaix et al. (2006)<sup>18</sup>. There is a problem with these data,

<sup>17</sup> The notation  $f(x) \sim g(x)$  means that there exists a constant  $C \neq 0$  such that in the limit  $x \rightarrow \infty$ ,  $f(x)/g(x) \rightarrow C$ . We use it to indicate that this relationship is only valid in the limit of large metaorder size  $N$ .

<sup>18</sup> The value of  $\beta$  is somewhat controversial, however: Eisler and Kertesz (2006) and Racz et al. (2009) have argued that the correct value of  $\beta > 2$ .

however, because they mix together order book trades and block trades, whereas block trades are generally larger and are believed to be a better proxy for metaorders. Lillo et al. (2005) analyzed block trade sizes in the London market and showed that they are consistent with a power law, with a tail exponent  $\beta \approx 1.6$ , in contrast to order book trades, whose tails are much thinner.

## 2. Connection to long-memory in order flow

Positive serial autocorrelation of signed order flow has been observed by many authors in many different markets, including the Paris Bourse by Biais, Hillion and Spratt (1995), in foreign exchange markets by Danielsson and Payne (2001), and the NYSE by Ellul et al. (2005) and Yeo (2006). (Signs of orders are based on the initiator of the trade, e.g. trades initiated by buyers are given a positive sign). These papers all emphasized the first autocorrelation, which is only a small part of the story. If one carefully examines the whole autocorrelation function it becomes clear that *all* the coefficients of the autocorrelation function are positive statistically significant out to large lags. In the London, Paris, New York and Spanish stock exchanges it is empirically observed that the signs of trades in stock markets obey a long-memory process<sup>19</sup> [Bouchaud et al. (2004), Lillo and Farmer (2004), Vaglica (2008)]. This means that the transaction sign autocorrelation function  $C(\tau)$  decays in time as  $C(\tau) \sim \tau^{-\gamma}$ , where  $0 < \gamma < 1$ . The observation of long-memory in stock markets is very robust; in London, for example, every stock examined shows long-memory under strict statistical tests (Lillo and Farmer, 2004). Thus we believe that for these stock markets the long-memory of order flow is a well established fact.

Lillo et al. (2005) have postulated that the long-memory of order flow comes from order splitting and presented a model whose predictions appear to agree well with the data. Under their model, in the limit  $\tau \rightarrow \infty$  the resulting autocorrelation  $C(\tau)$  of the order signs of the individual trades used to execute metaorders is a power law  $C(\tau) \sim \tau^{-\gamma}$ , with

$$\beta = 1 + \gamma. \tag{16}$$

They tested this hypothesis for 20 stocks from the London Stock Exchange by comparing their tail exponents  $\beta$  to their correlation scaling exponents  $\gamma$  and showed that it is well satisfied.

A final piece of evidence that the source of long-memory in order flow is due to order splitting comes from an empirical study by Gerig (2007) in the London Stock Exchange using data with brokerage codes. Trades coming from the same brokerage have long-memory, whereas trades from different brokerages do not (see also Bouchaud, Farmer, and Lillo (2008)).

---

<sup>19</sup> A standard example of a long-memory process is a fractionally integrated Brownian motion. We use the term 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, Engle and Granger (1993).



### 3. Reconstruction from membership codes

Further supporting evidence comes from a study of the Spanish stock exchange by Vaglica et al. (2008), who have reconstructed metaorders using data with brokerage codes. They find that  $N$  is distributed as a power law for large  $N$  with  $\beta \approx 1.7$ , which is consistent with the observations of the distribution of block market sizes to within statistical errors.

An interesting point that is relevant for the theory developed here is that the power law behavior 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.

## B. Market impact for Pareto metaorder size

In this section we derive the functional form of the impact for Pareto metaorder size, which from the three types of evidence discussed in the previous section appears to be a good assumption. We do this by using Eq. 12 in the limit as  $M \rightarrow \infty$ , and return in Section VII C 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 that it is exactly Pareto for all  $N$ , i.e.

$$p_N = \frac{1}{\zeta(\beta)} \frac{1}{N^{\beta+1}}, \quad N \geq 1 \quad (17)$$

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

$$\mathcal{P}_t = \frac{\zeta(1 + \beta, t + 1)}{\zeta(1 + \beta, t)} \simeq \left( \frac{t}{t + 1} \right)^\beta \sim 1 - \frac{\beta}{t}. \quad (18)$$

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.

### 1. Immediate impact

The immediate impact can be easily calculated once the distribution of metaorder size is known by using Eq. (12). For Pareto distributed metaorder sizes, using Eqs. (11) and (17),  $\tilde{R}_t$  is

$$\tilde{R}_t = \left( 1 + \frac{1}{t^{2+\beta}} \frac{\zeta(1 + \beta) - 1}{\zeta(1 + \beta, t)\zeta(1 + \beta, t + 1)} \right) \tilde{R}_1 \sim \frac{1}{t^{2-\beta}}. \quad (19)$$

Thus the immediate impact  $\mathcal{I}_t = \tilde{S}_t - S_0$  behaves asymptotically for large  $t$  as

$$\mathcal{I}_t \sim \begin{cases} t^{\beta-1} & \text{for } \beta \neq 1 \\ \log(t + 1) & \text{for } \beta = 1 \end{cases} \quad (20)$$

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  $\beta$  the impact goes from concave to convex, with  $\beta = 2$  as the borderline case<sup>20</sup>. Thus the exponent  $\beta$  has a dramatic effect on the shape of the impact. Figure 2 illustrates the reversion process for  $\beta = 1.5$  and shows how the shape of the impact varies with  $\beta$ .

## 2. Permanent impact

The permanent impact under the Pareto assumption is easily computed using Eq. (13). A direct calculation shows that<sup>21</sup>

$$I_N \sim \frac{1}{N} \int^N x^{\beta-1} dx = \frac{1}{\beta} N^{\beta-1}. \quad (21)$$

Eqs. (21) and (20) imply that the ratio of the permanent to the immediate impact is

$$\frac{I_N}{\mathcal{I}_N} = \frac{1}{\beta}. \quad (22)$$

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. Moreover the maximum fraction of informed trades such that the informed trader trades at prices where she is taking a loss is  $q_N \leq 1/\beta$ . For  $\beta = 1.5$  less than two third of the trades must be informed.

## C. 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 behavior in the limit  $M \rightarrow \infty$ , 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  $N = M$ . Since by definition  $\mathcal{P}_M = 0$ , the martingale condition requires that  $S_{N+1} = \tilde{S}_N$ , 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.

<sup>20</sup> 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  $\beta < 2$ .

<sup>21</sup> 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 also do not make a prediction about reversion.

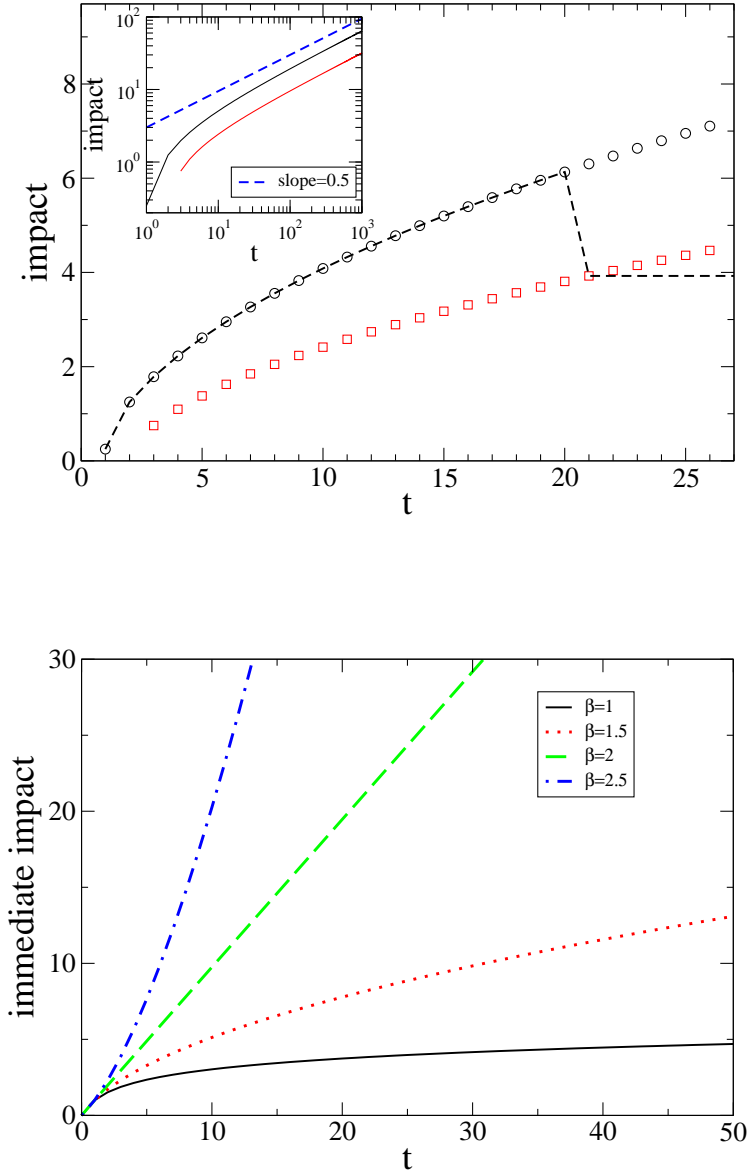


FIG. 2: An illustration of predicted market impact for Pareto distributed metaorder size. *Top panel:* Immediate impact  $\mathcal{I}_t$  (black circles) and permanent impact  $I_t$  (red squares) for  $\beta = 1.5$ . ( $I_t$  here is the permanent impact if the order were to end at  $t = N$ ). The dashed 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 metaorder sizes from  $N = 1$  to  $N = 1000$ . The blue dashed line is a guide for the eye indicating the asymptotic square root behavior. *Bottom panel:* Expected immediate impact  $\mathcal{I}_t$  as a function of time  $t$  for Pareto distributed metaorder size for tail exponents  $\beta = 1, 1.5, 2$  and  $2.5$ . This illustrates how the impact goes from concave to convex as  $\beta$  increases.

### D. Market impact for stretched exponential metaorder size

Changing the metaorder distribution has a dramatic effect on the impact. As we have already shown, as we increase the tail exponent  $\beta$  of the Pareto distribution upward from one, the impact function goes from a logarithm to a concave power law to a linear function (at  $\beta = 2$ ) to a convex power law for  $\beta > 2$ . 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 an alternative functional form.

For this purpose we choose the stretched exponential, which can be tuned from thin tailed to heavy tailed behavior, 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

$$p_N = \frac{\lambda}{\Gamma(1/\lambda, 1)} e^{-N^\lambda}. \quad (23)$$

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  $\lambda$  is small this functional form is easily confused with a power law. It can be shown that this leads to an immediate impact function that for large  $t$  asymptotically behaves as

$$\mathcal{I}_t \sim \frac{e^{t^\lambda}}{t^{2-\lambda}}. \quad (24)$$

Thus the power law scaling of the impact is multiplied by an exponential, which asymptotically dominates. The permanent impact is

$$I_N \sim \frac{1}{N} \int^N \frac{e^{x^\lambda}}{x^{2-\lambda}} dx \sim \frac{e^{N^\lambda} \lambda - E_{1+1/\lambda}(-N^\lambda)}{N^2 \lambda^2} \sim \frac{e^{N^\lambda}}{N^2 \lambda}, \quad (25)$$

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

$$\frac{I_N}{\mathcal{I}_N} = \frac{1}{\lambda N^\lambda}. \quad (26)$$

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  $N \rightarrow \infty$ . Thus the fixed ratio of permanent and immediate impact seems to be a rather special property of the Pareto distribution.

## VIII. DISCUSSION

### A. Empirical implications

The theory presented here makes several strong 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.

- The fair pricing condition, Eq. 7, 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).
- The asymmetric price response predicted by Eq. 4 is testable. However, this only tests the martingale condition.
- 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.
- Prediction of immediate and permanent impact based on  $p_N$ . Equations 12 and 13 are directly testable.
- 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 VII B.

Preliminary results seem to support, or at least not contradict, the prediction of the last bullet. 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, and Moro et al. (2009); they both found permanent 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. Better tests are needed.

## B. Information revelation

One of the key assumptions that we make here is that it is possible for market makers to detect when the execution of a given metaorder begin and ends. The ability to do this from imbalances in order flow using brokerage codes has been demonstrated [(Vaglica et al. 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<sup>22</sup>, for metaorders whose average size was just under 100 in both markets, making such metaorders difficult to hide. Nonetheless, the detection problem introduces uncertainties in starting and stopping times that may affect the price impact.

## C. 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

---

<sup>22</sup> Participation rate is defined as the fraction of trades that a given agent participates in.

is determined by behavioral and institutional factors, such as how informed the agents who use a given trading size are. This hypothesis is difficult to test because information is difficult to measure independently of impact. Furthermore, within the framework developed here, this would violate market efficiency.

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 behavioral 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.

### Acknowledgments

This work was supported by PRIN project 2007TKLTSR “Computational markets design and agent-based models of trading behavior” and National Science Foundation grant 0624351.

- 
- [1] Almgren, R.F., 2003. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10, 1–18.
  - [2] Almgren, R.F. and Chriss, N., 1999. Value under liquidation. *Risk*, 12, 61–63.
  - [3] Almgren, R.F. and Chriss, N., 2000. Optimal execution of portfolio transactions. *Journal of Risk*, 3, 5–39.
  - [4] Almgren, R.F., Thum, C. and Hauptmann, H. L., 2005. Direct estimation of equity market impact. *Risk*.
  - [5] Back, K. and Baruch, S., 2007. Working orders in limit order markets and floor exchanges. *Journal of Finance*, 62, 1589–1621.
  - [6] Beran, J., 1994. *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

- [7] Berk, J.B. and Green, R.C, 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112, 1269–1295.
- [8] Bernhardt, D., Hughson, E. and Naganathan, G., 2002. Cream-skimming and payment for order flow. Technical Report. University of Illinois.
- [9] Bertismas, D and Lo, A., 1998. Optimal control of execution costs. *Journal of Financial Markets*, 1, 1–50.
- [10] Biais, B., Hillion, P. and Spatt, C., 1995. An empirical analysis of the limit order book and order flow in the paris bourse. *The Journal of Finance*, 50, 1655–1689.
- [11] Bouchaud, J-P, Farmer, J. Doyne and Lillo, F., 2008. How markets slowly digest changes in supply and demand, in: Hens, T. and Schenk-Hoppe, K. (Eds.), *Handbook of Financial Markets: Dynamics and Evolution*. Elsevier. In press.
- [12] Bouchaud, J-P., Gefen, Y., Potters, M. and Wyart, M., 2004. Fluctuations and response in financial markets: The subtle nature of “random” price changes. *Quantitative Finance*, 4, 176–190.
- [13] Bouchaud, J-P., Kockelkoren, J. and Potters, M., 2006. Random walks, liquidity molasses and critical response in financial markets. *Quantitative Finance*, 6, 115–123.
- [14] Challet, 2007. The demise of constant price impact functions and single-time step models of speculation. *Physica A.*, 382, 29–35.
- [15] Chan, L. K.C. and Lakonishok, J., 1993. Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, 33, 173–199.
- [16] Chan, L. K.C. and Lakonishok, J., 1995. The behavior of stock prices around institutional trades. *The Journal of Finance*, 50, 1147–1174.
- [17] Chordia, T. and Subrahmanyam, A., 2004. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Markets*, 72, 485–518.
- [18] Danielsson, J. and Payne, R., 2001. Measuring and explaining liquidity on the limit order book: Evidence from reuters d2000-2. Technical Report. London School of Economics.
- [19] Ding, Z., Granger, C. W. J. and Engle, R. F., 1993. A long memory property of stock returns and a new model. *Journal of Empirical Finance*, 1, 83–106.
- [20] Eisler, Z. and Kertecz, J., 2006. Size matters, some stylized facts of the market revisited. *European Journal of Physics B*, 51, 145–154.
- [21] Ellul, A., Holden, C., Jain, P. and Jennings, R., 2005. A comprehensive test of order choice theory: Recent evidence from the nyse. Technical Report. Indiana University.
- [22] Evans, M. D. D. and Lyons, R. K., 2002. Order flow and exchange rate dynamics. *Journal of Political Economy*, 110, 170–180.
- [23] Farmer, J. D., Patelli, P. and Zovko, Ilija, 2005. The predictive power of zero intelligence in financial markets. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2254–2259.
- [24] Farmer, J.D., Gerig, A., Lillo, F. and Mike, S., 2006. Market efficiency and the long-memory of supply and demand: Is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6, 107–112.
- [25] Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H. E., 2003. A theory of power-law distributions in financial market fluctuations. *Nature*, 423, 267–270.
- [26] Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H.E., 2006. Institutional investors and stock market volatility. *Quarterly Journal of Economics*, 121, 461–504.
- [27] Gatheral, J., 2010. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10, 749–759.

- [28] Gerig, A.N., 2007. A theory for market impact: How order flow affects stock price. Ph.D. Thesis. University of Illinois.
- [29] Glosten, L. R., 2003. Discriminatory limit order books, uniform price clearing and optimality. Technical Report. Columbia.
- [30] Glosten, L.R., 1994. Is the electronic limit order book inevitable? *Journal of Finance*, 49, 1127–1161.
- [31] Gopikrishnan, P., Plerou, V., Gabaix, X. and Stanley, H. E., 2000. Statistical properties of share volume traded in financial markets. *Physical Review E*, 62, R4493–R4496; Part A.
- [32] Hasbrouck, J., 1991. Measuring the information content of stock trades. *The Journal of Finance*, 46, 179–207.
- [33] Hausman, J. A., Lo, A. W. and Mackinlay, A. C., 1992. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31, 319–379.
- [34] Hopman, C., 2006. Do supply and demand drive stock prices? *Quantitative Finance*; To appear.
- [35] Huberman, G. and Stanzl, W., 2004. Price manipulation and quasi-arbitrage. *Econometrica*, 72, 1247–1275.
- [36] Keim, D. B. and Madhavan, A., 1996. The upstairs market for large-block transactions: analysis and measurement of price effects. *The Review of Financial Studies*, 9, 1–36.
- [37] Kempf, A. and Korn, O., 1999. Market depth and order size. *Journal of Financial Markets*, 2, 29–48.
- [38] Kyle, A. S., 1985. Continuous auctions and insider trading. *Econometrica*, 53, 1315–1335.
- [39] Lillo, F. and Farmer, J. D., 2004. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8,.
- [40] Lillo, F., Farmer, J. D. and Mantegna, R. N., 2003. Master curve for price impact function. *Nature*, 421, 129–130.
- [41] Lillo, F., Mike, S. and Farmer, J. D., 2005. Theory for long memory in supply and demand. *Physical Review E*, 7106, 066122.
- [42] Moro, E., Moyano, L. G., Vicente, J., Gerig, A., Farmer, J. D., vaglica, G., Lillo, F. and Mantegna, R.N., 2009. Market impact and trading protocols of hidden orders in stock markets. Technical Report.
- [43] Obizhaeva, A.A. and Wang, J., 2005. Optimal trading strategy and supply/demand dynamics. Technical Report. AFA 2006 Boston Meetings Paper.
- [44] Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Meyer, M. and Stanley, H. E., 1999. Scaling of the distribution of price fluctuations of individual companies. *Physical Review E*, 60, 6519–6529; Part A.
- [45] Plerou, V., Gopikrishnan, P., Gabaix, X. and Stanley, H. E., 2002. Quantifying stock price response to demand fluctuations. *Physical Review E*, 66, article no. 027104.
- [46] Potters, M. and Bouchaud, J-P., 2003. More statistical properties of order books and price impact. *Physica A*, 324, 133–140.
- [47] Racz, E., Eisler, Z. and Kertesz, J., 2009. Comment on ‘tests of scaling and universality.’ by plerou and stanley. Technical Report.
- [48] Schwartzkopf, Y. and Farmer, J.D., 2010. Technical Report. For a preliminary report, see Y. Schwartzkopf’s Caltech Ph.D thesis, *Complex Phenomena in Social and Financial Systems: From bird population growth to the dynamics of the mutual fund industry*.
- [49] Torre, N., 1997. *BARRA Market Impact Model Handbook*. Berkeley: BARRA Inc.
- [50] Toth, B., Lillo, F. and Farmer, J.D., 2010. Segmentation algorithm for non-stationary com-



- pound poisson processes. Technical Report. Santa Fe Institute.
- [51] Vaglica, G., 2008. Scaling laws of strategic behavior and specialization of strategies in agent dynamics of a financial market. Ph.D. Thesis. University of Palermo.
  - [52] Vaglica, G., Lillo, F., Moro, E. and Mantegna, R., 2008. Scaling laws of strategic behavior and size heterogeneity in agent dynamics. *Physical Review E*, 77, 0036110.
  - [53] Viswanathan, S. and Wang, J. J. D., 2002. Market architecture: Limit order books versus dealership markets. *Journal of Financial Markets*, 5, 127–167.
  - [54] Weber, P. and Rosenow, B., 2006. Large stock price changes: volume or liquidity? *Quantitative Finance*, 6, 7–14.
  - [55] Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M. and Vettorazzo, M., 2006. Relation between bid-ask spread, impact and volatility in double auction markets. Technical Report.
  - [56] Yeo, Wee Yong, 2006. Persistence in limit order flow. Technical Report. National University of Singapore.
  - [57] Zhang, Y. C., 1999. Toward a theory of marginally efficient markets. *Physica A*, 269, 30–44.

### Appendix A: Proofs of the propositions

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

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

**Proof.** Given the definition of  $\tilde{R}_t$  and  $R_t$  we can write the prices as

$$\tilde{S}_t = S_0 + \sum_{i=0}^{t-1} \tilde{R}_i \quad (\text{A2})$$

$$S_{t+1} = S_0 + \sum_{i=0}^{t-1} \tilde{R}_i - R_t \quad (\text{A3})$$

With these expressions for  $N < M$  the profit per share  $\pi_N$  can be rewritten as

$$\begin{aligned} \pi_N &= \frac{1}{N} \sum_{t=1}^N \tilde{S}_t - S_{N+1} = R_N - \frac{1}{N} \sum_{i=1}^{N-1} i\tilde{R}_i = \\ &= \frac{1}{p_N} \tilde{R}_N \sum_{i=N+1}^M p_i - \frac{1}{N} \sum_{i=1}^{N-1} i\tilde{R}_i \end{aligned} \quad (\text{A4})$$

where in the last equality we have used the martingale condition of Eq. (4). For  $N = M$  the profit per share is

$$\begin{aligned} \pi_M &= \frac{1}{M} \sum_{i=1}^M \tilde{S}_i - S_{M+1} = \frac{1}{M} \sum_{i=1}^M \tilde{S}_i - \tilde{S}_M = \\ &= -\frac{1}{M} \sum_{i=1}^{M-1} i\tilde{R}_i \end{aligned} \quad (\text{A5})$$

By substituting these two last expressions in Eq. (A1) we obtain

$$\begin{aligned} E_1[N\pi_N] &= \sum_{N=1}^{M-1} N\tilde{R}_N \sum_{i=N+1}^M p_i - \sum_{N=1}^{M-1} p_N \sum_{i=1}^{N-1} i\tilde{R}_i - p_M \sum_{i=1}^{M-1} i\tilde{R}_i = \\ &= \sum_{N=1}^{M-1} N\tilde{R}_N \sum_{j=N+1}^M p_j - \sum_{N=1}^M p_N \sum_{i=1}^{N-1} i\tilde{R}_i \end{aligned} \quad (\text{A6})$$

By computing explicitly the coefficients of each  $\tilde{R}_i$ , it is direct to find that they vanish for each  $i$ , i.e.  $E_1[N\pi_N] = 0$ .

It is worth noticing that this proposition does not hold in the case of infinite support. in order to show this let us consider the expected profit for orders of length between  $N = 1$  and  $N = \bar{N}$ . Independently on the finiteness of the support, it holds the following proposition.

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

$$\sum_{N=1}^{\bar{N}} p_N N \pi_N = \left( \sum_{i=\bar{N}+1}^M p_i \right) \left( \sum_{i=1}^{\bar{N}} i \tilde{R}_i \right) \geq 0. \quad (\text{A7})$$

*. 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 share as

$$\pi_N = \frac{1}{p_N} \tilde{R}_N \sum_{i=N+1}^M p_i - \frac{1}{N} \sum_{i=1}^{N-1} i \tilde{R}_i$$

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

$$\begin{aligned} \sum_{N=1}^{\bar{N}} p_N N \pi_N &= \sum_{N=1}^{\bar{N}} N \tilde{R}_N \sum_{i=N+1}^M p_i - \sum_{N=1}^{\bar{N}} p_N \sum_{i=1}^{N-1} i \tilde{R}_i = \\ & \left( \sum_{i=\bar{N}+1}^M p_i \right) \left( \sum_{i=1}^{\bar{N}} i \tilde{R}_i \right) \end{aligned} \quad (\text{A8})$$

This is equal to the quantity in Eq. (A7). Moreover it is clear that both terms in brackets are non negative. The non negativity 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

$$E_1[N \pi_N] \equiv \sum_{N=1}^{\infty} p_N N \pi_N = \lim_{\bar{N} \rightarrow \infty} \left( \sum_{i=\bar{N}+1}^{\infty} p_i \right) \left( \sum_{i=1}^{\bar{N}} i \tilde{R}_i \right). \quad (\text{A9})$$

In the infinite support case the behavior of the limit in the last term of the above expression depends on the asymptotic behavior of  $p_N$  and  $\tilde{R}_N$  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 3.** *The system of martingale conditions (Eq. 3) and fair pricing conditions (Eq. 7) has solution*

$$\tilde{R}_t = \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^M p_i} \frac{1-p_1}{\sum_{i=t}^M p_i} \tilde{R}_1 \quad t = 2, 3, \dots, M-1 \quad (\text{A10})$$

$$R_t = \frac{\mathcal{P}_t}{1-\mathcal{P}_t} \tilde{R}_t \quad t = 1, 2, \dots, M-1 \quad (\text{A11})$$

**Proof.** The solution of Eq. (A11) is a direct consequence of the martingale conditions (Eq. 3). The total profit of metaorders of length  $N < M$  can be rewritten as (see proof of Proposition 1)

$$N \pi_N = N R_N - \sum_{i=1}^{N-1} i \tilde{R}_i = N \frac{\mathcal{P}_N}{1-\mathcal{P}_N} \tilde{R}_N - \sum_{i=1}^{N-1} i \tilde{R}_i \quad (\text{A12})$$

The fair pricing conditions (Eq. 7) state that for  $1 < N < M$  it is  $N\pi_N = 0$ , i.e.

$$\tilde{R}_N = \frac{1}{N} \frac{1 - \mathcal{P}_N}{\mathcal{P}_N} \sum_{i=1}^{N-1} i \tilde{R}_i \quad (\text{A13})$$

This is a recursive equation which determines  $\tilde{R}_N$  once  $\tilde{R}_1$  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

$$\tilde{R}_t = \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{t-1}} \tilde{R}_1 \quad t > 1 \quad (\text{A14})$$

and we prove it by induction. We assume that the solution holds for  $N = 2, 3, \dots, t-1$  and we prove that it is true for  $N = t$ . If Eq. (A14) holds for  $N = 2, 3, \dots, t-1$  we can rewrite Eq. (A13) for  $N = t$  as

$$\tilde{R}_t = \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t} \sum_{i=1}^{t-1} i \tilde{R}_i = \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t} \left( 1 + \sum_{i=2}^{t-1} \frac{1 - \mathcal{P}_i}{\mathcal{P}_i} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{i-1}} \right) \tilde{R}_1 \quad (\text{A15})$$

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

$$\left( 1 + \sum_{i=2}^{t-1} \frac{1 - \mathcal{P}_i}{\mathcal{P}_i} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{i-1}} \right) = 1 - \frac{1}{\mathcal{P}_1} + \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{t-1}} \quad (\text{A16})$$

Since, by definition,  $\mathcal{P}_1 = 1$  the first two terms in the right hand side cancel and thus one obtains Eq. (A14). This equation is equivalent to Eq. (A10). In fact

$$\begin{aligned} \tilde{R}_t &= \frac{1}{t} \frac{1 - \mathcal{P}_t}{\mathcal{P}_t} \frac{1}{\mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3 \dots \mathcal{P}_{t-1}} \tilde{R}_1 = \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^M p_i} \frac{1}{\frac{\sum_{i=3}^M p_i}{\sum_{i=2}^M p_i} \frac{\sum_{i=4}^M p_i}{\sum_{i=3}^M p_i} \dots \frac{\sum_{i=t}^M p_i}{\sum_{i=t-1}^M p_i}} \tilde{R}_1 = \\ &= \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^M p_i} \frac{\sum_{i=2}^M p_i}{\sum_{i=t}^M p_i} \tilde{R}_1 = \frac{1}{t} \frac{p_t}{\sum_{i=t+1}^M p_i} \frac{1 - p_1}{\sum_{i=t}^M p_i} \tilde{R}_1 \quad (\text{A17}) \end{aligned}$$

i.e. our thesis, Eq. (A10).

## Appendix B: Effect of finite $M$ on impact

As already discussed briefly in Section VII C, 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.

$$p_N = \frac{1}{H_M^{(1+\beta)}} \frac{1}{N^{\beta+1}} \quad N \geq 1 \quad (\text{B1})$$

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

$$\mathcal{P}_t = \frac{\zeta(1 + \beta, t + 1) - \zeta(1 + \beta, M + 1)}{\zeta(1 + \beta, t) - \zeta(1 + \beta, 1 + M)} \quad (\text{B2})$$

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 Eq. (18) 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 Eq. (12). For truncated Pareto distributed metaorder sizes,  $\tilde{R}_t$  is (for  $t > 2$ ) is

$$\tilde{R}_t = \left( \frac{H_M^{(1+\beta)} - 1}{(\zeta(1 + \beta, t) - \zeta(1 + \beta, M + 1))(\zeta(1 + \beta, t + 1) - \zeta(1 + \beta, M + 1))} \right) \frac{\tilde{R}_1}{t^{2+\beta}} \quad (\text{B3})$$

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

$$\tilde{R}_t \sim \frac{\tilde{R}_1}{t^{2-\beta}} \quad (\text{B4})$$

which is the same scaling as the infinite support Pareto distribution (see Eq. (20)). 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 VII B.

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  $\beta$ . 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.

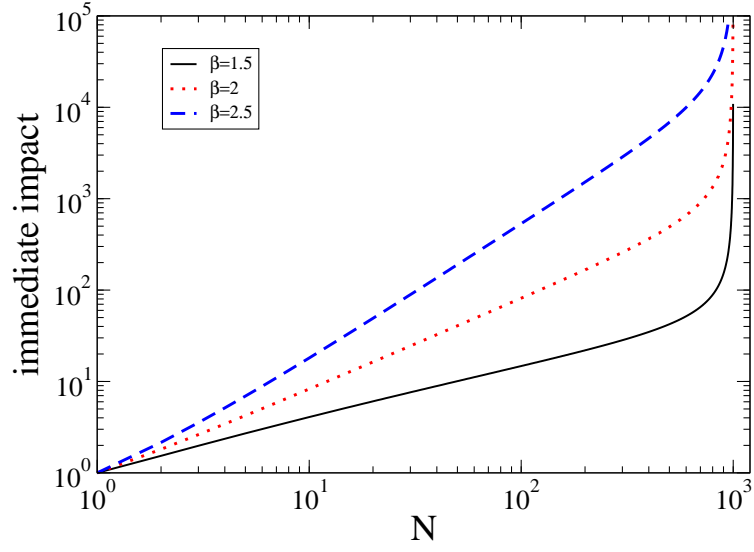


FIG. 3: Immediate impact when the metaorder size has finite support, i.e when  $N$  has a maximum value  $M$ . Plot is in log log scale with  $M = 1000$  and  $\beta = 1.5, 2,$  and  $2.5$ .