Co-impact: crowding effects in institutional trading activity

Co-impact: Crowding Effects in Institutional Trading Activity

Frédéric Bucci[1], Iacopo Mastromatteo[2], Zoltán Eisler[2],
Fabrizio Lillo[3], Jean-Philippe Bouchaud[2,4] and Charles-Albert Lehalle[2,4]

[1] *Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy*
[2] *Capital Fund Management, 23-25, Rue de l'Université 75007 Paris, France*
[3] *Department of Mathematics, University of Bologna, Piazza di Porta San
Donato 5, 40126 Bologna, Italy*

[4] *CFM-Imperial Institute of Quantitative Finance, Department of Mathematics,
Imperial College, 180 Queen's Gate, London SW7 2RH*

July 10, 2018

Abstract

This paper is devoted to the important yet unexplored subject of *crowding* effects on market impact, that we call "co-impact". Our analysis is based on a large database of metaorders by institutional investors in the U.S. equity market. We find that the market chiefly reacts to the net order flow of ongoing metaorders, without individually distinguishing them. The joint co-impact of multiple contemporaneous metaorders depends on the total number of metaorders and their mutual sign correlation. Using a simple heuristic model calibrated on data, we reproduce very well the different regimes of the empirical market impact curves as a function of volume fraction ϕ: square-root for large ϕ, linear for intermediate ϕ, and a finite intercept I_0 when $\phi \to 0$. The value of I_0 grows with the sign correlation coefficient. Our study sheds light on an apparent paradox: How can a non-linear impact law survive in the presence of a large number of simultaneously executed metaorders?

Contents

1

1 Introduction

The market impact of trades, i.e. the change in price conditioned on signed trade size, is a key quantity characterizing market liquidity and price dynamics [1, 2]. Besides being of paramount interest for any economic theory of price formation, impact is a major source of transaction costs, which often makes the difference between a trading strategy that is profitable, and one that is not. Hence the interest in this topic is not purely academic in nature.

One of the most surprising empirical finding in the last 25 years is the fact that the impact of a so-called "metaorder" of total size Q, executed incrementally over time, increases approximately as the *square root* of Q, and not linearly in Q, as one may have naively expected and as indeed predicted by the now-classic Kyle model [3]. Since impact is non-additive, a natural question concerns the *interaction* of different metaorders executed simultaneously – possibly with different signs and sizes. In particular, one may wonder whether the simultaneous impact of different metaorders could substantially alter the square root law; or conversely whether the square root law might itself result from the interaction of different metaorders.

Metaorder information is, however, not publicly available, and earlier analyses were mostly based on (often proprietary) data from single financial institutions. These studies give little insight about effects due to the simultaneous execution of metaorders from different investors, which we will call *co-impact* hereafter. Indeed, even if investors individually decide about their metaorders, they might do so based on the same trading signal. Prices can thus be affected by emergent effects such as crowding. What is the right way to model the total market impact of simultaneous metaorders on the same asset on the same day? In order to answer this question, we will use a rich dataset concerning the execution of metaorders issued by a heterogeneous set of investors.

The paper is organized as follows. In Sec. 2 we introduce the Ancerno dataset we used empirically. In Sec. 3 we discuss the limits of the validity of the square root law on the daily level. In Sec. 4 we find that the market impact of simultaneous daily metaorders is proportional to the square root of their net order flow. This means that the market does not distinguish the different individual metaorders. We then construct a theoretical framework to understand the impact of correlated metaorders in Sec. 5. This allows us to understand when a single asset manager will observe a square-root impact law, and when *crowding effects* will lead to deviations from such a behaviour. We also compare in Sec. 5 the results of our simple mathematical model with empirical data, with very satisfactory results. Sec. 6 concludes.

2   The Ancerno Database

Our analysis relies on a database made available by *Ancerno*, a leading transaction-cost analysis provider[1]. The unique advantage of working with such institutional data is that one can simultaneously analyze the trading of many investors. The main caveat though is that one has little knowledge about the motives and style behind the observed portfolio transitions. For example a given metaorder can be part of a longer execution over multiple days. Another possibility is that the final investor may decide to stop a metaorder execution midway if prices move unfavourably. Such effects can potentially bias our results, but we believe that they do not change the qualitative conclusions below.

In the following we will define as a metaorder a series of jointly reported executions performed by a single investor, through a single broker within a single day, on a given stock and in a given direction (buy/sell). However, contrarily to the version of the database used in Ref. [4], available labels do not allow us to relate different metaorders executed on behalf of the same final investor by the same or different brokers during the same day. These should ideally be counted as a single metaorder. We will comment later on the biases induced by such a lack of information. Thus each metaorder is characterized by a broker label, the stock symbol, the total volume of the metaorder $|Q|$ and its sign $\epsilon = \pm 1$, and the start time t_s and the end time t_e of the execution.

Our dataset includes the period January 2007 – June 2010 for a total of 880 trading days. Following the procedure introduced in Ref. [4] we use the following filters to remove possibly erroneous data:

- **Filter 1**: We select the stocks which belong to the Russell 3000 index discarding metaorders executed on highly illiquid stocks.

- **Filter 2**: We select metaorders ending before 4:01 p.m.

- **Filter 3**: We select metaorders whose duration $D = t_e - t_s$ is longer than 2 mins.

- **Filter 4**: We select metaorders whose participation rate (the ratio between their quantity and the volume traded by the market between t_s and t_e) is smaller than 30%.

---

[1]Ancerno Ltd. (formerly the Abel Noser Corporation) is a widely recognised consulting firm that works with institutional investors to monitor their equity trading costs. Its clients include many pension funds and asset managers. Previous academic studies that use Ancerno data include [4, 5, 6, 7, 8, 9, 10, 11]. See www.ancerno.com for details.

Finally we retain around 7.7 million metaorders distributed quite uniformly in time and across market capitalizations. These filtered metaorders represent around the 5% of the total reported market volume independently of the year and of the stock capitalization.[2] The statistical properties of the metaorders, in terms of volume fraction, duration, etc., are broadly in line with Ref. [4] even though their data was aggregated at the level of brokers – for more details see Appendix A. A particularly important statistic for the following analyses is the number N of simultaneous metaorders in the database, executed on the same stock during the same day. The probability distribution $p(N)$ is shown in the left panel Fig. 1, indicating that N is broadly distributed with an average close to 5.

The right panel of Fig. 1 shows the probability distribution of the absolute value of the volume fraction ϕ of the metaorders. This variable plays a key role in the following and is defined as

$$\phi := Q/V$$

where V is the total volume traded during that day. The figure shows that the volume fraction distribution is independent of the metaorder side ($\text{sign}(\phi) = \pm 1$, i.e. buy or sell) and is also very broad.
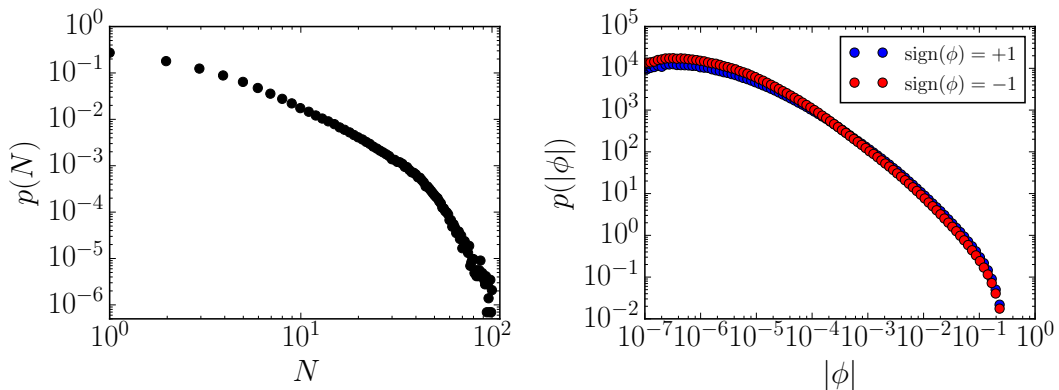
Figure 1: (Left panel) Empirical probability distribution of the number N of daily metaorders per asset. (Right panel) Empirical probability distribution of the absolute value of the volume fraction ϕ per metaorder, separately for signs ± 1, i.e. buy/sell.

3   The Square Root Law and its Domain of Validity

We will quantify market impact in terms of the rescaled log-price $s = \log(S)/\sigma$, where S is the market mid-price, which we normalize by the daily volatility of the asset defined as $\sigma = (S_{\text{high}} - S_{\text{low}})/S_{\text{open}}$ based on the daily high, low and open prices. In this paper we will define impact as the expected change of s between the open and the close of the day. This choice will avoid an elaborate analysis of when

---

[2]Without the above filters, this number would rise to about 10%.

precisely each metaorder starts and ends, how they overlap and which reference prices to take in each case. When a metaorder of total volume Q is executed, its impact will be defined as

$$I(\phi) := \mathbb{E}[s_{\text{close}} - s_{\text{open}}|\phi], \qquad (1)$$

for a given metaorder signed volume fraction ϕ.

Empirically, impact is found to be an odd function of ϕ, displaying a concave behavior in $|\phi|$. It is well described by the square root law [12, 13, 14, 15, 16, 17, 18, 19]

$$I(\phi) = Y \times \phi^{\bullet\delta}, \qquad (2)$$

where here and throughout the paper we will denote the sign-power operation by $x^{\bullet\delta} := \text{sign}(x) \times |x|^{\delta}$. The dimensionless coefficient Y (called the *Y-ratio*) is of order unity and the exponent δ is in the range 0.4–0.7. It is interesting to note that in Eq. (2) only the volume fraction ϕ matters, the time taken to complete execution or the presence of other active metaorders is not directly relevant (remember that the volatility of the instrument has been subsumed in the definition of the rescaled price s). This formula is surprisingly universal across financial products, market venues, time periods and the strategies used for execution.
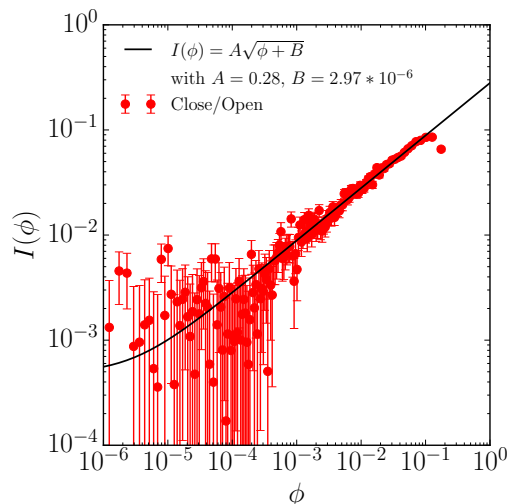
Figure 2: Market impact curve $I(\phi) = \mathbb{E}[s_{\text{close}} - s_{\text{open}}|\phi]$ as a function of the metaorder size ratio $\phi = Q/V$ computed using the filtered metaorders from the Ancerno dataset in the period from January 2007 – June 2010. We also show the simple fit $I(\phi) = A\sqrt{\phi + B}$, which captures some – but not all – of the discrepancy with the square root law at small ϕ.

We first check this empirical result on our dataset. In Fig. 2 we show the market impact curve obtained by dividing the data into evenly populated bins according to the volume fraction ϕ and computing the conditional expectation of impact for each bin. Here and in the following, error bars are determined as standard errors. Note

5

that in all the following empirical plots the price impact curves are normalized by their *Y-ratio* and we will abuse the notation $I(\phi)$ in order to denote the symmetrized measure $I(\phi) = \mathbb{E}[\epsilon(s_{\text{close}} - s_{\text{open}})||\phi|]$, with $\epsilon = \text{sign}(\phi)$, due to the antisymmetric nature of $I(\phi)$.

While the square root law holds relatively well when $10^{-3} \lesssim \phi \lesssim 10^{-1}$, three other regimes seem to be present:

1. For very small volume fractions up to $\phi \lesssim 10^{-4}$, impact appears to saturate to a finite, positive value.

2. In the intermediate regime $10^{-4} \lesssim \phi \lesssim 10^{-3}$, impact is closer to a linear function, although the data is very noisy.

3. In the large ϕ regime $\phi \gtrsim 10^{-1}$, impact seems to saturate, or even to decrease with increasing ϕ.

These results are robust across time periods and market capitalizations, and consistent with Ref. [4], where regimes 2. and 3. were also clearly observed. In the following, we will discard altogether the last, large ϕ regime, which is most likely affected by conditioning effects (for example buying more when the price moves down and less when the price moves up). We will on the other hand seek to understand the other three regimes within a consistent mathematical framework.

Intuitively, the breakdown of the square root law for small ϕ comes from the fact that the signs of the metaorders in our dataset are correlated – particularly so because some metaorders are originating from the same final investor. Let us illustrate the effect of correlations on a simplistic example: Imagine that simultaneously to the considered buy metaorder (with volume fraction $\phi > 0$), another metaorder with the same sign and volume fraction $\phi_m > 0$ is also traded. Assuming that the square root law applies for the combined metaorder (a hypothesis that we will confirm on data), the observed impact should read

$$I(\phi + \phi_m) = Y \times \sqrt{\phi + \phi_m}. \tag{3}$$

This tends to the value $Y\sqrt{\phi_m}$ when $\phi \to 0$, behaves linearly when $\phi \ll \phi_m$ and as a square root when $\phi \gg \phi_m$. We show in Fig. 2 that this simple fit captures some, but not all, of the discrepancy with the square root law at small ϕ. In particular the intermediate linear region is not well accounted for. We will develop in the following a mathematical model that reproduces all these effects.

A way to minimize the effect of correlations is to restrict to days/assets where there is a unique metaorder in the dataset ($N = 1$). As shown in Fig. 3, impact in this case is almost perfectly fitted by a square root law. Fig. 3, also shows that as N increases, significant departures from the square root law can be observed for small ϕ, as suggested by our simple model Eq. (3). An alert reader may however object that the Ancerno database represents a small fraction ($\sim 5\%$) of the total volume. Even when a single metaorder is reported, many other metaorders are likely to be simultaneously present in the market. So why does one observe a square root law at all, even for single metaorders? The solution to precisely this paradox is one of the main messages of our paper.

4   How Do Impacts Add Up?

In the previous section we showed that the number of metaorders in the market strongly influences how price impact behaves, but we have yet to provide insight
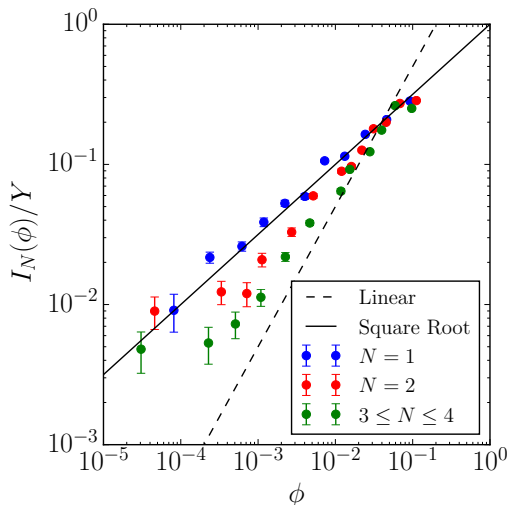
6

Figure 3: Empirical evidence on the effect of the number of metaorders N on the daily price impact curves $I_N(\phi)$ normalized by the prefactor Y-*ratio*: significant departures from the square root law can be observed for small ϕ increasing the number N of daily metaorders per asset.

into why this is the case. As a first step, we want to determine an explicit functional form of the aggregated market impact of N simultaneous metaorders. As we have emphasized, impact is non-linear, so aggregation is a priori non-trivial. Should one add the square root impact of each metaorder, or should one first add the signed volume fractions before taking the square root? Since orders are anonymous and indistinguishable, the second procedure looks more plausible. This is what we test now. Consider the average aggregate impact conditioned to the co-execution of N metaorders:

$$\mathcal{I}(\boldsymbol{\varphi}_N) = \mathbb{E}[s_{\text{close}} - s_{\text{open}} | \boldsymbol{\varphi}_N], \tag{4}$$

where $\boldsymbol{\varphi}_N := (\phi_1, \cdots, \phi_N)$. We make the following parametric ansatz for this quantity:

$$\mathcal{I}(\boldsymbol{\varphi}_N) = Y \times \left(\sum_{i=1}^{N} \phi_i^{\bullet\alpha} \right)^{\bullet\delta/\alpha}, \tag{5}$$

where, again, $x^{\bullet\alpha}$ is the signed power of x. By construction this formula is invariant under the permutation of metaorders, as it should be since they are indistiguishable. Y and δ set, respectively, the scale and the exponent of the impact function. The free parameter α interpolates between the case when impacts add up ($\alpha = \delta$) and when only the net traded volume is relevant ($\alpha = 1$).

Fig. 4 shows the quality of the fit obtained by least squares regressions of Eq. (5) for a grid of (α, δ) pairs. We find that the coefficient of determination $r^2(\alpha, \delta)$ of the fit is maximized close to the point $\alpha = 1.0$ and $\delta = 0.5$, which suggests that the aggregated price impact $\mathcal{I}(\boldsymbol{\varphi}_N)$ of N metaorders at the daily scale only depends on

7

the total net order flow, i.e.[3]

$$\mathcal{I}(\varphi_N) \approx Y \times \Phi^{\bullet 1/2},\tag{6}$$

where $\Phi = \sum_{i=1}^N \phi_i$. In other words, the market only reacts to the net order flow, not to the way in which this order flow is distributed across investors.
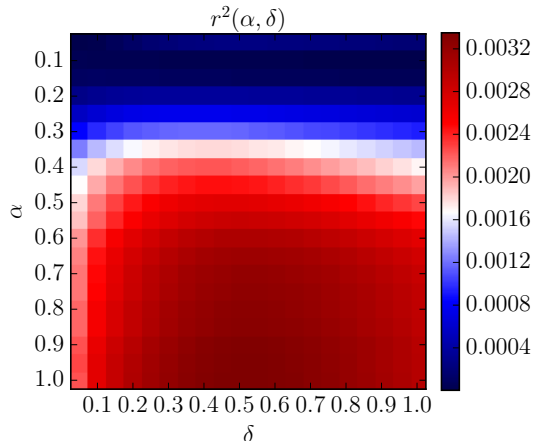
Figure 4: The computed coefficient of determination $r^2(\alpha, \delta)$ of the least squares regressions of Eq. (5) for a grid of (α, δ) pairs. The coefficient of determination is maximized ($r^2 = 0.0035$) in the vicinity of the point $\alpha = 1.0$ and $\delta = 0.5$.

One can now plot $\mathcal{I}(\Phi)$ as a function of Φ for various N, see Fig. 5. One clearly sees that provided Φ is not too small, impact is independent of N and crosses over from linear to square root behavior as Φ increases. The linear behavior is in fact more pronounced at smaller values of N. We now turn to a theoretical analysis that will allow us to quantify more precisely the co-impact problem, and how the square root law can survive at large N.

5 Correlated Metaorders and Co-Impact

5.1 The Mathematical Problem

Even if an asset manager knows the average impact formula Eq. (6), this may not be sufficient to estimate his actual impact which depends strongly on the presence of other contemporaneous metaorders.

Suppose the manager k wants to execute a volume fraction $\phi_k = \phi$. If all the other $N - 1$ metaorders were known, the daily price impact would be given

---

[3]We have in fact tested that the assumption $\alpha = 1$ is also favoured for a general, non parametric shape for the impact function $\mathcal{I}(\Phi)$.
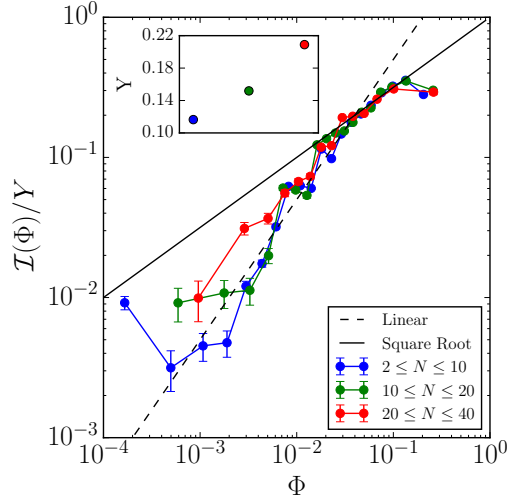
Figure 5: Global market impact $\mathcal{I}(\Phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N)|\sum_{i=1}^{N}\phi_i = \Phi]$ normalized by the prefactor Y as a function of the net order flow Φ for various buckets in N and symmetrized as explained in Sec. 3. The inset shows the normalization Y for the different curves.

by the global impact function $\mathcal{I}(\Phi)$ determined in the previous section, with $\Phi = \phi_k + \sum_{i\neq k}^{N}\phi_i$.

However, this information is obviously not available to the manager k. His best estimate of the average impact given N is the conditional expectation

$$I_N(\phi) = \mathbb{E}[\mathcal{I}(\Phi)|\phi_k = \phi] = \mathbb{E}\Big[\mathcal{I}\Big(\phi_k + \sum_{i\neq k}^{N}\phi_i\Big)\Big|\phi_k = \phi\Big] \qquad (7)$$

over the conditional distribution $P(\boldsymbol{\varphi}_N|\phi_k = \phi)$ of the metaorders. Since the number of metaorders is in general not known either, the expected individual market impact is given by

$$I(\phi) = Y \times \sum_{N} p(N) \int \mathrm{d}\phi_1 \ldots \mathrm{d}\phi_N P(\boldsymbol{\varphi}_N|\phi_k = \phi)\Big(\phi_k + \sum_{i\neq k}^{N}\phi_i\Big)^{\bullet 1/2}, \qquad (8)$$

where we have used Eq. (6). In such a way to compute $I_N(\phi)$ and $I(\phi)$ we need to know the joint probability density function $P(\boldsymbol{\varphi}_N) := P(\phi_1, \ldots, \phi_N)$, which is in general a complicated and high-dimensional object. Then to create a tractable model that can be calibrated on data, we must make some reasonable assumption on the dependence structure of the ϕ_i. In the next subsection we investigate the simple case where the ϕ_i are all independent, and then turn to an empirical characterization of the correlations between metaorders. We finally provide the results of our empirically inspired model and compare them with the empirical market impact curves.

9
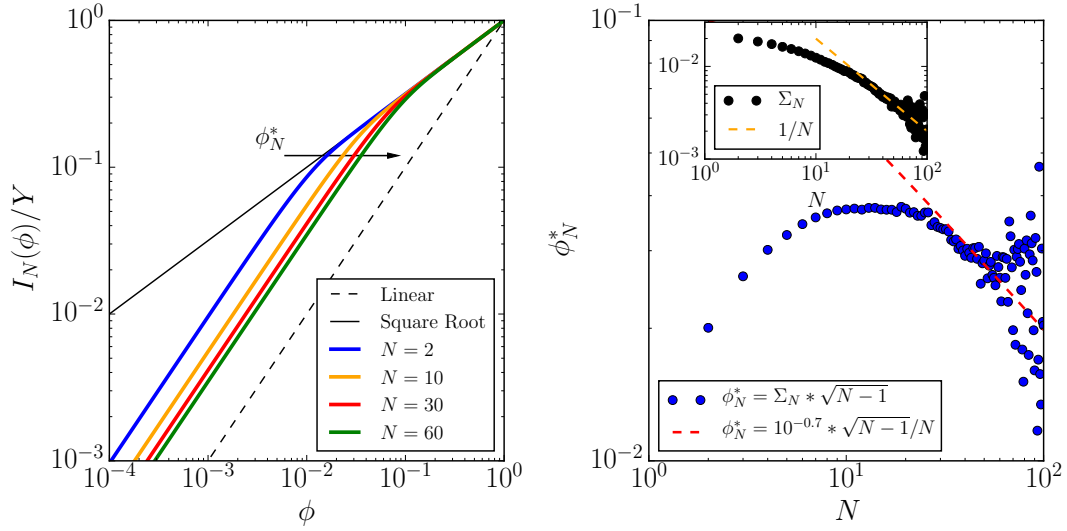
Figure 6: (Left panel) Market impact curves $I_N(\phi)/Y$ for i.i.d. Gaussian metaorders for different $N \in \{2, 10, 30, 60\}$ and *fixed* $\Sigma_N = \Sigma = 0.8\%$ computed from empirical data. The transition from the square root to the linear regime takes place for $\phi_N^* \simeq \Sigma\sqrt{N-1}$. (Right panel) Empirically estimated ϕ_N^* as a function of the number N of daily metaorders per asset. This is obtained by computing the empirical order size's standard deviation $\Sigma_N = \sqrt{\mathbb{V}[\phi|N]}$ conditioned to N. The dashed red line shows the case in which $\Sigma_N \sim 1/N$ for each N: in the inset we report the standard deviation Σ_N of the order size ϕ as a function of the number N of daily metaorders per asset.

5.2 Independent Metaorders

The simplest assumption about the form of $P(\boldsymbol{\varphi}_N)$ is that metaorder volumes are i.i.d., meaning

$$P(\boldsymbol{\varphi}_N) = \prod_{i=1}^{N} p(\phi_i). \tag{9}$$

Assuming for simplicity that each ϕ_i is a Gaussian random variable with zero mean and variance Σ_N^2, where the lower index indicates an explicit dependence on N. Thus $N-1$ simultaneous metaorders generate a Gaussian noise contribution of amplitude $\Sigma_N\sqrt{N-1}$ on top of $\phi_k = \phi$. In Appendix B.1.1 we show analytically that:

- For small metaorders the noise term dominates, leading to

$$I_N(\phi) \propto \phi \quad \text{when} \quad \phi \ll \phi_N^* := \Sigma_N\sqrt{N-1}.$$

- For large metaorders the $N-1$ other simultaneous metaorders can be neglected and thus

$$I_N(\phi) \propto \sqrt{\phi} \quad \text{when} \quad \phi \gg \phi_N^*.$$

In Appendix B.1.2 we show that the above results remain valid in the limit of large N independently of the shape of the volume distribution provided its variance is finite.

The full analytical solutions for different N values, but *fixed* $\Sigma_N = \Sigma$, are shown in the left panel of Fig. 6. One clearly sees the cross-over from a linear behavior at small ϕ to a square root at larger ϕ. However, interestingly, one expects Σ_N to *decrease* with N, simply because as the number of metaorders increases, the volume fraction represented by each of them must decrease[4]. As shown in the right panel of Fig. 6, this is the case empirically since for $N \gtrsim 10$, Σ_N indeed decays as N^{-1} (as also suggested by Fig. 7 below). Hence, for large N, the crossover value ϕ_N^* *decreases* with N as $N^{-1/2}$. This explains why the square root law can at all be observed when a large number of metaorders are present. If these metaorders are independent, their net impact on the price averages out, leaving the considered metaorder as if it was alone in a random flow, as assumed in theoretical models [12, 13]. We now turn to the effect of correlations between metaorders.

5.3 Metaorder Correlations

In order to build a sensible model of $P(\boldsymbol{\varphi}_N)$ we consider separately the size distribution and the size cross-correlations. From the right panel of Fig. 1 we observe that the marginals $p(\phi_i)$ are to a good approximation independent of the direction, buy or sell, and moderately fat tailed. The latter observation suggests that the total net order flow $\Phi = \sum_{i=1}^{N} \phi_i$ is *not* dominated by a single metaorder. A way to quantify this is through the Herfindahl index (or "inverse participation ratio") ζ, defined as:

$$\zeta := \frac{\sum_{i=1}^{N} \phi_i^2}{\left(\sum_{i=1}^{N} |\phi_i|\right)^2}. \tag{10}$$

---

[4]For example, the variance of a flat Dirichlet random variable $(X_1, ..., X_N) \sim \text{Dir}(N)$ describing fractions is $\mathbb{V}[X_i|N] = (N-1)/(N^2(N+1)) \sim N^{-2}$.
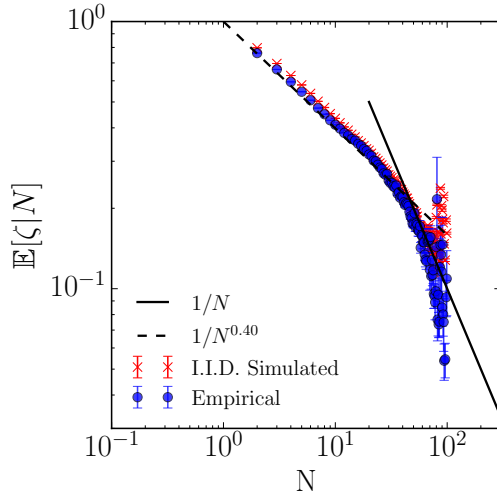
Figure 7: Average value of the Herfindahl index (or "inverse participation ratio") $\mathbb{E}[\zeta|N]$ as a function of the number N of daily metaorders per asset computed from the empirical data (blue dot symbols) and simulating i.i.d. $|\phi_i|$ (red star symbols) extracted from the empirical distribution illustrated in the right panel of Fig. 1.

This quantity is of order $1/N$ if all metaorders are of comparable size, and of order 1 if one metaorder dominates. In Fig. 7 we show the dependence of $\mathbb{E}[\zeta|N]$ as a function of N, which clearly decays with N. It also compares very well to the result obtained assuming the absolute volume fractions $|\phi_i|$ to be independent, identically distributed variables, drawn according to the empirical distribution shown in Fig. 1. We therefore conclude that (a) metaorders in the Ancerno database are typically of comparable relative sizes ϕ and (b) absolute volume correlations do not play a major role, and we will neglect them henceforth.

Sign correlations, on the other hand, *do* play an important role in determining the impact of simultaneous metaorders. The empirical average sign correlation of metaorders simultaneously executed on the same asset is defined as

$$\mathbb{C}_\epsilon(N) := \frac{\mathbb{E}[\epsilon_i \epsilon_j|N] - \mathbb{E}[\epsilon_i|N]^2}{\mathbb{E}[\epsilon_i^2|N] - \mathbb{E}[\epsilon_i|N]^2} \,, \tag{11}$$

where $\mathbb{E}[\cdots|N]$ is the average over all days and assets such that exactly N metaorders were executed. Fig. 8 shows the dependence of $\mathbb{C}_\epsilon$ on N. We clearly see that on average the daily metaorders executed on the same asset are positively correlated. Furthermore, $\mathbb{C}_\epsilon(N)$ is seen to decrease as N increases. This is likely due to the fact that there are multiple concurrent metaorders submitted by the same manager, an effect that becomes less prominent as N increases. The plateau value $\mathbb{C}_\epsilon \approx 0.025$ at large N is, we believe, a reasonable proxy for the correlation of orders submitted by different asset managers.[5]

---

[5]This value is indeed compatible with the value of $\mathbb{C}_\epsilon$ obtained with the version of the database used
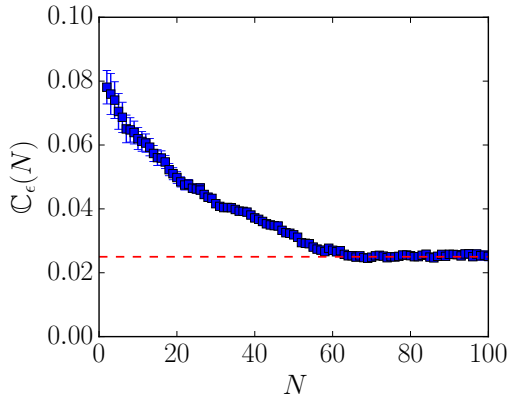
Figure 8: Empirical average sign correlation $\mathbb{C}_\epsilon(N)$ as a function of the number N of daily metaorders: the dashed red line represents the plateau value $\mathbb{C}_\epsilon \approx 0.025$ at large N, which, we believe, is a reasonable proxy for the correlation of orders submitted by different asset managers.

5.4 Market Impact with Correlated Metaorders

A natural model would be to consider the ϕ_i's as exchangeable multivariate Gaussian variables of zero mean, variance Σ_N^2 and cross-correlation coefficient $\mathbb{C}_\phi(N)$. Appendix B.2 shows that the qualitative behavior for independent metaorders remains the same when $\mathbb{C}_\phi(N) > 0$. Specifically, one finds that the average impact $I_N(\phi)$ can be obtained by making the substitution

$$\phi \to \phi[1 + (N-1)\mathbb{C}_\phi(N)]. \tag{12}$$

in the expression of $I_N(\phi)$ for independent Gaussians. This is expected, as $(N-1)\mathbb{C}_\phi(N)$ gives the effective number of additional volume-weighted metaorders correlated to the original one. By the same token though, $I_N(\phi)$ still vanishes linearly for small ϕ, whereas empirical data suggests a positive intercept when $\phi \to 0$.

As an alternative model that emphasizes sign-correlations, let us assume that the joint distribution of the ϕ_i's can be written as

$$P(\boldsymbol{\varphi}_N) = \mathcal{P}(\boldsymbol{\epsilon}_N) \prod_{i=1}^N p(|\phi_i|), \tag{13}$$

meaning that metaorder sizes are independent, while the signs are possibly correlated. This specific form is motivated by the observation that the size of a metaorder is mainly related to the assets under management of the corresponding financial institution, while the sign is related to the trading signal. One can expect that different investors use correlated information sources, while the size of the trades is idiosyncratic.

---

by [4], which identified metaorders coming from the same investor.

13
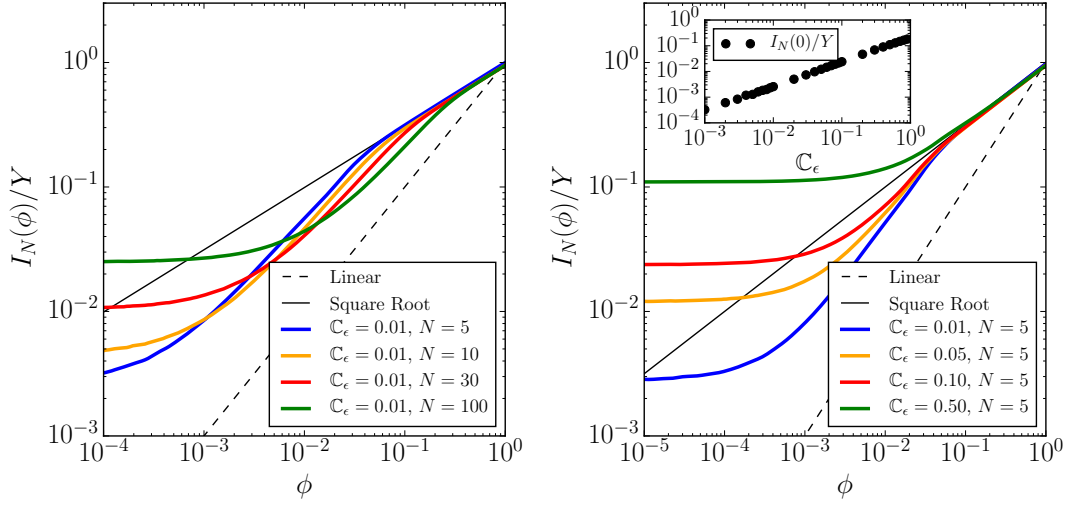
Figure 9: Market impact curves $I_N(\phi)/Y$ computed for the correlated signs model, where volumes are drawn according to a half-normal distribution with $\mathbb{E}[|\phi_i|] = \Sigma\sqrt{2/\pi}$ and $\Sigma = 0.8\%$ is set equal to its empirical value averaged over N. (Left panel) Numerical simulations for fixed sign correlation $\mathbb{C}_\epsilon = 0.01$ but varying number of metaorders $N \in \{5, 10, 30, 100\}$. (Right panel) Numerical simulations for fixed number of metaorders $N = 5$ but with varying sign correlation $\mathbb{C}_\epsilon \in \{0.01, 0.05, 0.1, 0.5\}$: as shown in the inset the intercept $I_N(0)/Y$ decreases linearly with the sign correlation for $\mathbb{C}_\epsilon \to 0$. Note that the individual price impact $I_N(\phi)$ values converges to a positive constant when $\phi \to 0$. For intermediate ϕ, $I_N(\phi)$ is linear and crosses-over at larger ϕ to a square root.

14

We further assume that there is a unique common factor determining the sign of the metaorders. In other words, the statistical model for the signs is the following:

$$\mathbb{P}(\epsilon_i = +1|\tilde{\epsilon}) = \frac{1}{2}(1 + \gamma_\epsilon\tilde{\epsilon}); \qquad \mathbb{P}(\epsilon_i = -1|\tilde{\epsilon}) = \frac{1}{2}(1 - \gamma_\epsilon\tilde{\epsilon}), \qquad (14)$$

where $\tilde{\epsilon}$ is the hidden sign factor, such that $\mathbb{P}(\tilde{\epsilon} = \pm 1) = 1/2$, and γ_ϵ is the sign correlation between each sign ϵ_i and the hidden sign factor $\tilde{\epsilon}$. A simple calculation leads to

$$\mathbb{C}_\epsilon(N) = \mathbb{P}(\epsilon_i = \epsilon_j) - \mathbb{P}(\epsilon_i = -\epsilon_j) = \gamma_\epsilon^2. \qquad (15)$$

where we omitted the γ_ϵ's explicit dependence on N. Contrarily to the Gaussian case, we have not been able to obtain analytical formulas, but instead relied on numerical simulations to obtain $I_N(\phi)$ for different combinations of $\mathbb{C}_\epsilon(N)$ and N, reported in Fig. 9. Results for unsigned volumes generated from a half-normal distribution calibrated on data are shown in Fig. 9. We observe that the individual price impact $I_N(\phi)$ converges to a positive constant $I_N(0) > 0$ when $\phi \to 0$, despite $I_N(0) = 0$ for a Gaussian model. For intermediate ϕ, $I_N(\phi)$ is linear and it crosses over at larger ϕ to a square root. For fixed N the intercept value increases with the sign correlation $\mathbb{C}_\epsilon$, see the right panel of Fig. 9. The intuition is that conditioned to the fact that I buy, and independently of the size of my trade, the order flow of other actors will be biased towards buy as well, and I will suffer from the impact of their trades. In fact, subtracting the non-zero intercept of $I_N(\phi)$ leads to impact curves that look almost identical to those of Fig. 6, i.e. a linear region for small ϕ followed by a square root region beyond a crossover value $\phi_N^* \sim \Sigma_N\sqrt{N-1}$. Since for large N $\phi_N^* \to 0$, one simply expects a square-root law, shifted by the intercept $I_N(0)$.

5.5 Empirical Calibration of the Model

With the aim to compare the model prediction with empirical data, we propose a calibration method described in Appendix B.4.1. This is based on the assumption that metaorder signs are independent random variables sampled from a half-Gaussian calibrated on empirical data. The metaorder sign correlation structure can be estimated by introducing a realized sign correlation

$$\rho_\epsilon := \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \epsilon_i\epsilon_j, \qquad (16)$$

which is then used to estimate the sign correlation $\mathbb{C}_\epsilon(N)$ of Eq. (15). Once the model is calibrated, we use numerical simulations to compute the expected market impact $I(\phi)$, see Appendix B.4.1 for the precise details of the procedure.

Fig. 10 shows that imposing correlation only between the signs leads to a very good prediction of the empirical curves, justifying the adoption of the sign correlated model. All the features of the empirical impact curves are qualitatively well reproduced, in line with Fig. 2. This includes the clear deviations from the square root law for $\phi \leq 10^{-3}$ with both a linear regime and a constant price impact I_0 when $\phi \to 0$.
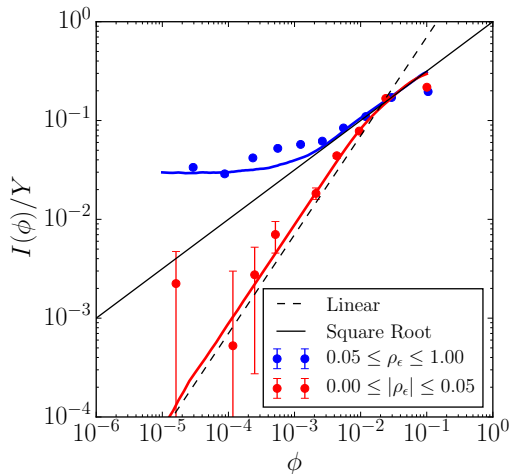
Figure 10: Comparison between calibrated sign-correlated model (colored lines) and empirical data (circles) for the case $N \in [2, 10]$: The sample is split into two sub-samples respectively with $\rho_\epsilon \geq 0.05$ and $0 \leq |\rho_\epsilon| \leq 0.05$. The theoretical curves are calculated through numerical simulations as explained in the main text (see Appendix B.4.1 for all the details of the procedure).

6 Conclusions

It is a commonly acknowledged fact that market prices move during the execution of a trade – they increase (on average) for a buy order and decrease (on average) for a sell order. This is, loosely stated, the phenomenon known as market impact. In this paper we have presented one of the first studies breaking down market impact of metaorders executed by different investors, and taking into account interaction/correlation effects. We investigated how to aggregate the impact of individual actors in order to best explain the daily price moves. The large number and heterogeneity of the metaorders traded by financial institutions allows precise measurements of price impact in different conditions with reduced uncertainty. We found that both the number of actors simultaneously trading on a stock and the *crowdedness* of their trade (measured by the correlation of metaorder signs) are important factors determining the impact of a given metaorder.

Our main conclusions are as follows:

- The market chiefly reacts to the total net order flow of ongoing metaorders, the functional form being well approximated by a square root at least in a range of volume fraction ϕ. As expected in anonymous markets, impact is insensitive to the way order flow is distributed across different investors.

- The number N of executed metaorders and their mutual sign correlations $\mathbb{C}_\epsilon$ are relevant parameters when an investor wants to precisely estimate the market impact of their own metaorders.

- Using a simple heuristic model calibrated on data, we are able to reproduce to

16

a good level of precision the different regimes of the empirical market impact curves, as a function of ϕ, N and $\mathbb{C}_\epsilon$.

- When the number of metaorders is not large, and when $\mathbb{C}_\epsilon > 0$, a small investor will observe *linear* impact with a non-zero intercept I_0, crossing over to a square-root law at larger ϕ. I_0 grows with $\mathbb{C}_\epsilon$ and can be interpreted as the average impact of all other metaorders.

- When the number of metaorders is very large and the investor has no correlation with their average sign, they should expect on a given day a square-root impact randomly shifted upwards or downwards by I_0. Averaged over all days, a pure square-root law emerges, which explains why such behavior has been reported in many empirical papers.

On the last point, we believe that our study sheds light on an apparent paradox: How can a non-linear impact law survive in the presence of a large number of simultaneously executed metaorders? As we have seen, the reason is that for a metaorder uncorrelated with the rest of the market, the impacts of other metaorders cancel out on average. Conversely, any intercept of the impact law can be interpreted as a non-zero correlation with the rest of the market.

Given the importance of the subject, our results present several interesting applications. Our aggregated price impact model should be of interest both to practitioners trying to monitor and reduce their trading costs, and also to regulators that seek to improve the stability of markets.

Acknowledgments

A   Metaorder Statistics

We now describe some statistics of the metaorders. The participation rate π is defined as the ratio between the number of shares $|Q|$ traded by the metaorder and the whole market during the execution interval $[t_\mathrm{s}, t_\mathrm{e}]$

$$\pi = \frac{|Q|}{V(t_\mathrm{e}) - V(t_\mathrm{s})}. \tag{17}$$

The duration D expressed in volume time is equal to

$$D = \frac{V(t_e) - V(t_s)}{V(t_c)}, \tag{18}$$

while the unsigned daily fraction is defined as the ratio between the metaorder unsigned volume $|Q|$ and the volume $V(t_c)$ traded by the market in the whole day:

$$|\phi| = |Q|/V(t_c) = \pi \cdot D. \tag{19}$$

We find that the participation rate π and the duration D are both well approximated by truncated power-law distributions over several orders of magnitude. The
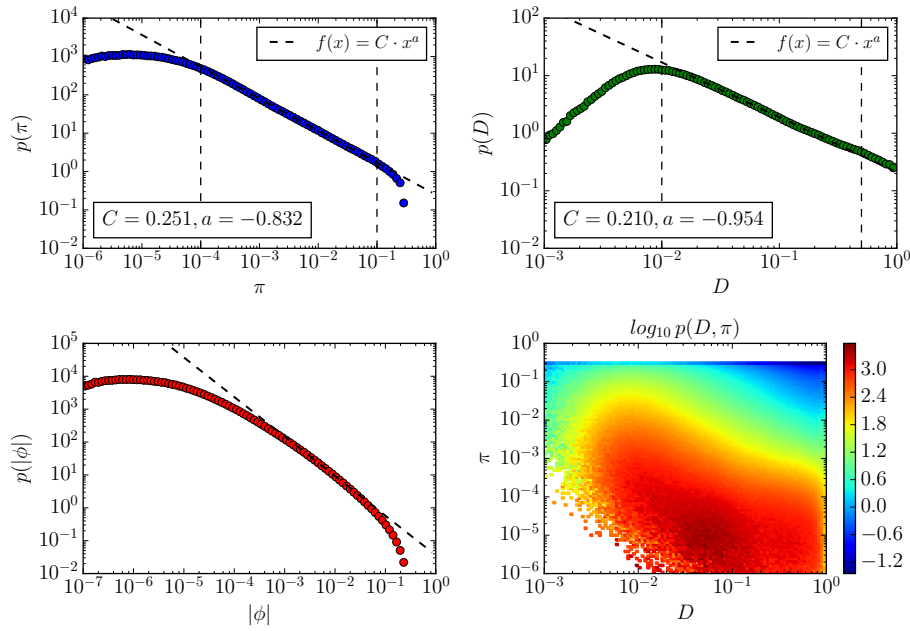
Figure 11: Estimation of the probability density function of the participation rate π (top left), duration D (top right) and unsigned daily fraction $|\phi|$ (bottom left). All these panels are in log-log scale. The first two also show the best fit with a power-law function in the region bounded by the vertical dashed lines. The bottom right panel shows the logarithm of the estimated joint probability density function $p(D, \pi)$ in double logarithmic scale of the duration D and of the participation rate π. These statistics are in agreement with the ones obtained in Ref. [4].

estimated probability density function of the participation rate π is shown in log-log scale in the top left panel of Fig. 11. A power-law fit in the region $10^{-4} \leq \pi \leq 10^{-1}$, i.e. over three orders of magnitude gives a best fit exponent $a = -0.832 \pm 0.001$. The top right panel of Fig. 11 shows the estimated probability density function of the duration D of a metaorder. A power-law fit in the region bounded by vertical dashed lines $(0.01 \leq D \leq 0.5)$ gives a power-law exponent $a = -0.954 \pm 0.002$. These power laws are very heavy-tailed, meaning that there is substantial variability in both the participation rate and the duration. Note that in both cases the variability is intrinsically bounded, and therefore the power law is automatically truncated: by definition $\pi \leq 1$ and $D \leq 1$. In addition, for $p(D)$ there is a small bump on the right extreme of the distribution corresponding to all-day metaorders. The deviation from a power law for small D is a consequence of our Filter 3 retaining only metaorders lasting at least 2 min, which in volume time corresponds to $2/390 \simeq 0.005$ on average. The bottom left panel shows the probability density function of the unsigned daily fraction $|\phi|$. In this case the distribution is less fattailed, and clearly not a power law. This is potentially an important result, as the predictions of some theories for market impact depend on this, and have generally assumed power-law behavior. In conclusion, the bottom right panel of Fig. 11 shows the logarithm of the estimated joint probability density function $p(D, \pi)$ in double logarithmic scale. We measure a very low linear correlation (-0.08) between the two variables, the main contribution coming from the extreme regions, i.e. very large π implies very small D and vice versa. In other words, as expected, very aggressive metaorders are typically short and long metaorders more often have a small participation rate.

B   From the Bare to the Market Impact Function

The expected individual market impact of a metaorder with signed volume ϕ is estimated by

$$I(\phi) = \sum_N p(N) I_N(\phi) \tag{20}$$

where $p(N)$ is the probability distribution of the daily number N of metaorders per asset and

$$I_N(\phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N)|\phi_k = \phi] = \int \mathrm{d}\phi_1 \cdots \mathrm{d}\phi_N P(\boldsymbol{\varphi}_N|\phi_k = \phi) \left(\phi_k + \sum_{i \neq k}^N \phi_i \right)^{\bullet 1/2} \tag{21}$$

is the market impact computed from the bare impact function $\mathcal{I}(\boldsymbol{\varphi}_N) := (\sum_{i=1}^N \phi_i)^{\bullet 1/2}$ with fixed N. Assuming that $p(N)$ is kwown a priori the market impact $I_N(\phi)$ is given from the expectation in Eq. (21) done over the conditional probability distribution $P(\boldsymbol{\varphi}_N|\phi_k = \phi)$ of the volume metaorders. However, in such a way to do analytical computation it is necessary to assume reasonable hypothesis for the joint distribution function $P(\boldsymbol{\varphi}_N) := P(\phi_1, \cdots, \phi_N)$.

For this reason, we start considering the case of i.i.d. metaorders and we firstly show analytically how the transition from the square root to a linear market impact is possible in the i.i.d. Gaussian framework. Secondly, we generalize these results in the limit of large N for any symmetric volume distribution which sastifies the Central

Limit Theorem assumptions. Thirdly, we show that the same results continue to be valid introducing correlation between the signed volumes in a Gaussian framework. For last, we describe how to compute numerically the market impact in Eq. (21) in the case of metaorder signs correlated and i.i.d. unsigned volumes.

B.1 Market Impact with IID Metaorders

In the case of i.i.d. signed metaorders, the conditional joint distribution factorizes as

$$P(\boldsymbol{\varphi}_N|\phi_k = \phi) = \prod_{i \neq k}^{N} p(\phi_i); \tag{22}$$

this implies that the price impact $I_N(\phi)$ of a single metaorder $\phi_k = \phi$ out of N is given by

$$I_N(\phi) = \int d\phi_m \underbrace{\frac{1}{2\pi} \int d\lambda e^{-i\lambda\phi_m} \hat{p}(\lambda)^{N-1}}_{p(\phi_m)} (\phi + \phi_m)^{\bullet 1/2} = \int d\phi_m p(\phi_m)(\phi + \phi_m)^{\bullet 1/2}$$
$$\tag{23}$$

where $\phi_m = \sum_{i \neq k}^{N} \phi_i$ is the net order flow executed simultaneously to the metaorder ϕ_k and $\hat{p}(\lambda) = \mathbb{E}[e^{i\lambda\phi_i}]$ is the characteristic function of the signed volume distribution $p(\phi_i)$. Although the introduction of the characteristic function $\hat{p}(\lambda)$ in Eq. (23) will be a convenient way to exploit the convergence of the net order flow distribution $p(\phi_m)$ as discussed in Appendix B.1.2, the computation of the market impact $I_N(\phi)$ in a analytical closed form is possible only in the Gaussian case as shown in the next Appendix B.1.1.

B.1.1 Independent Gaussian Metaorders

In the gaussian case, i.e. $p(\phi_i) \sim \mathcal{N}(0, \Sigma_N^2)$ with $\Sigma_N^2 = \mathbb{V}[\phi_i|N]$, we can go further analytically in Eq. (23) since $\hat{p}(\lambda) = e^{-\Sigma_N^2 \lambda^2/2}$. In fact the integral representation of the price impact

$$I_N(\phi) = \frac{1}{\Sigma_N \sqrt{2\pi(N-1)}} \int_{-\infty}^{\infty} d\phi_m e^{-\phi_m^2/(2\Sigma_N^2(N-1))}(\phi + \phi_m)^{\bullet 1/2} =$$

$$= \frac{1}{\Sigma_N \sqrt{2\pi(N-1)}} \int_0^{\infty} dx \sqrt{x} \left[e^{-(x-\phi)^2/(2\Sigma_N^2(N-1))} - e^{-(x+\phi)^2/(2\Sigma_N^2(N-1))} \right] \quad (24)$$

can be expressed in the following analytical way

$$I_N(\phi) = \frac{\Gamma(1/4)}{2\sqrt{\pi}} \frac{\phi}{(2(N-1)\Sigma_N^2)^{1/4}} e^{-\frac{\phi^2}{2(N-1)\Sigma_N^2}} {}_1F_1\left(\frac{5}{4}, \frac{3}{2}, \frac{\phi^2}{2(N-1)\Sigma_N^2}\right) \quad (25)$$

where $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$ is the Gamma function and

$$_1F_1\left(\frac{5}{4}, \frac{3}{2}, z\right) = \frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{5}{4})} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{5}{4}+j)}{\Gamma(\frac{3}{2}+j)} \frac{z^j}{j!} \tag{26}$$

is the Kummer confluent hypergeometric function with $z = \frac{\phi^2}{(2(N-1)\Sigma_N^2)}$ [20].

The price impact $I_N(\phi)$ in Eq. (25) is shown in the left panel of Fig. 6 for different N and the parameter Σ_N *fixed*. If the metaorder volume ϕ is smaller than the sum of the other $N-1$ metaorders, i.e. $\phi \ll \phi_m$, then price impact is linear. Instead, when our metaorder dominates, i.e. $\phi \gg \phi_m$, the price impact follows a square root function. The transition from the linear to the square root regime takes place around $\phi_N^* \simeq \Sigma_N \sqrt{N-1}$, where $\Sigma_N \sqrt{N-1}$ is naturally interpreted as a measure for the market noise, in agreement with the change of the functional shape of the rescaled market impact function $y(\tilde{\phi}) = I_N(\phi)/((N-1)\Sigma_N^2)^{1/4}$ represented in Fig. 12 in function of the adimensional parameter $\tilde{\phi} := \phi/(\sqrt{N-1}\Sigma_N)$. To note furthermore that the linear regime comes out immediately from the expansion of $I_N(\phi)$ in Eq. (25) around $\phi = 0$ as follows

$$I_N(\phi) = \frac{1}{\Sigma_N \sqrt{2\pi(N-1)}} \int_0^\infty \mathrm{d}x \sqrt{x} e^{-x^2/(2\Sigma_N^2(N-1))} \left[e^{-(\phi^2-2x\phi)/(2\Sigma_N^2(N-1))} - e^{-(\phi^2+2x\phi)/(2\Sigma_N^2(N-1))} \right] \simeq$$

$$\simeq \sqrt{\frac{2}{\pi}} \frac{\phi}{(\Sigma_N^2(N-1))^{3/2}} \int_0^\infty \mathrm{d}x\, x^{3/2} e^{-x^2/(2\Sigma_N^2(N-1))} = 2^{3/4} \frac{\Gamma(5/4)}{(\pi^2 \Sigma_N^2(N-1))^{1/4}} \phi. \tag{27}$$

Remark 1. It follows that for i.i.d. Gaussian metaorders the slope of the linear price impact region decreases with N (as shown explicitly in Eq. (27)) and the crossover to the square root region happens in ϕ_N^* obtained by solving

$$\xi \frac{\phi_N^*}{(\Sigma_N^2(N-1))^{1/4}} \simeq (\phi_N^*)^{1/2}, \tag{28}$$

i.e. $\phi_N^* \simeq \xi^{-1}\Sigma_N \sqrt{N-1}$ with $\xi = 2^{3/4}\Gamma(5/4)/\sqrt{\pi}$.

B.1.2 Limit of Large N for Generally Distributed Independent Metaorders

The previous conclusions discussed in the i.i.d. Gaussian framework are also valid for others i.i.d. volume distributions as discussed in this Appendix: in fact, we can generalize them in the limit of large N for any symmetric volume distribution $p(\phi_i)$ which sastifies the Central Limit Theorem assumptions.

In the limit of large N, for which the Central Limit Theorem applied if certain conditions (discussed below) on $p(\phi_i)$ are sastified, the symmetric volume distribution $p(\phi_m)$ introduced in Eq. (23) converges to a stable law $\mathcal{G}_\alpha$ described by a characteristic function

$$\hat{p}(\lambda) = \mathbb{E}[e^{i\lambda\phi_m}] = e^{-c|\lambda|^\alpha} \tag{29}$$

where $c \in (0, \infty)$ is the scale parameter and $\alpha \in (0, 2]$ is the stability exponent. In other words we say that the volume distribution $p(\phi_i)$ belongs to the *domain of attraction* of the stable distribution $\mathcal{G}_\alpha$ if there exist constants $a_m \in \mathbb{R}$, $b_m > 0$ such that

$$b_m^{-1}(\phi_m - a_m) \to \mathcal{G}_\alpha, \tag{30}$$

i.e. the renormalized and recentred sum $\phi_m = \sum_{i \neq k}^N \phi_i$ converges in distribution to $\mathcal{G}_\alpha$. The Central Limit Theorem gives the conditions such that this convergence in distribution to a stable law $\mathcal{G}_\alpha$ is guaranted:
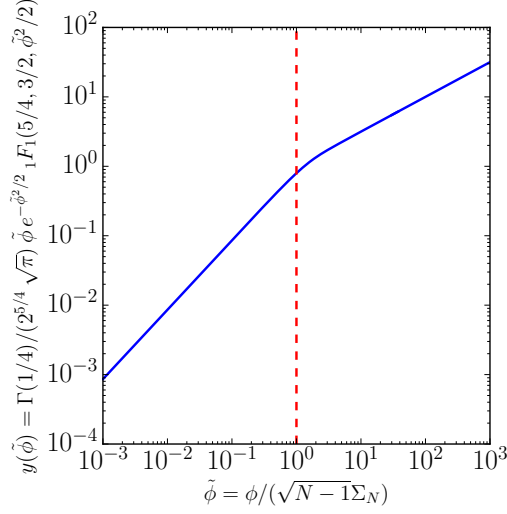
Figure 12: Rescaled market impact function $y(\tilde{\phi}) = I_N(\phi)/((N-1)\Sigma_N^2)^{1/4}$ in function of the dimensionless parameter $\tilde{\phi} := \phi/(\sqrt{N-1}\Sigma_N)$ with $I_N(\phi)$ given by Eq. (25): the vertical dashed red line represents the transition from the linear impact (left side) to the square-root one (right side).

- $p(\phi_m)$ converges to a Gaussian distribution ($\alpha = 2$ in Eq. (29)) if and only if

$$\int_{|\phi_i| \leq x} \phi_i^2 p(\phi_i) \mathrm{d}\phi_i \tag{31}$$

  is a slowly varying function $L(x)$, i.e. $\lim_{x \to \infty} L(tx)/L(x) = 1$ for all $t > 0$; then it follows that if $\Sigma_N^2 = \mathbb{V}[\phi_i|N] < \infty$

$$((N-1)^{1/2}\Sigma_N)^{-1}\phi_m \longrightarrow \mathcal{N}(0,1), \tag{32}$$

  while if $\mathbb{V}[\phi_i|N] = \infty$

$$((N-1)^{1/2}L_1)^{-1}\phi_m \longrightarrow \mathcal{N}(0,1) \tag{33}$$

  with L_1 a slowly varying function and $\mathcal{N}(0,1)$ a gaussian distribution with mean zero and variance 1.

- $p(\phi_m)$ converges to a Lévy distribution (for some $\alpha < 2$ in Eq. (29)) if and only if

$$\int_{-\infty}^{-x} p(\phi_i)\mathrm{d}\phi_i = \frac{c_1 + o(1)}{x^\alpha}L(x), \quad 1 - \int_{-\infty}^{x} p(\phi_i)\mathrm{d}\phi_i = \frac{c_2 + o(1)}{x^\alpha}L(x), \quad x \to \infty \tag{34}$$

  where $L(x)$ is a slowly varying function and c_1, c_2 are non-negative constants such that $c_1 + c_2 > 0$; then it follows that

$$((N-1)^{1/\alpha}L_2)^{-1}\phi_m \longrightarrow \mathcal{L}(c,\alpha) \tag{35}$$

22

with L_2 a slowly varying function and $\mathcal{L}(c, \alpha)$ a Lévy distribution with scale parameter $c \in (0, \infty)$ and stability exponent $\alpha \in (0, 2)$.

Remark 2. It follows immediately that for any volume distributions $p(\phi_i)$ belonging to the *domain of attraction* of the normal law, i.e. satisfying the condition in Eq. (31), the price impact $I_N(\phi)$ in the limit of large N is described by Eq. (25) and the transition from a linear to a square root price impact discussed in the Remark 1 continue to be valid: to note that in the case of $\mathbb{V}[\phi_i | N] = \infty$ it is sufficient to substitute Σ_N in Eq. (28) with the appropriate slowly varying function L_1.

Moreover we can show that the transition from a linear to a square root regime is still present for volume distributions $p(\phi_i)$ belonging to the *domain of attraction* of the Lévy distribution, i.e. that sastify the condition in Eq. (34) with $\alpha < 2$ and then $p(\phi_m) \to \mathcal{L}(c(N-1)^{1/\alpha}, \alpha)$. In fact, expanding to first order the bare impact function $\mathcal{I}(\phi, \phi_m) = (\phi + \phi_m)^{\bullet 1/2}$ around $\phi = 0$ in Eq. (23)

$$I_N(\phi) \simeq \frac{\phi}{2} \int \mathrm{d}\phi_m p(\phi_m) |\phi_m|^{-1/2} = \frac{\phi}{4\pi} \int \mathrm{d}\lambda \hat{p}(\lambda) \int \mathrm{d}\phi_m e^{-i\lambda\phi_m} |\phi_m|^{-1/2}, \quad (36)$$

introducing the characteristic function of the net order flow ϕ_m given by

$$\hat{p}(\lambda) = e^{-c(N-1)^{1/\alpha}|\lambda|^\alpha} \quad (37)$$

for large N and taking present that the last integral in Eq. (36) is a known Fourier transform

$$\int_{-\infty}^{+\infty} \mathrm{d}\phi_m e^{-i\lambda\phi_m} |\phi_m|^{-1/2} = \frac{\sqrt{2\pi}}{|\lambda|^{1/2}} \quad (38)$$

we obtain a linear price impact

$$I_N(\phi) \simeq \frac{\phi}{2\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{p}(\lambda) |\lambda|^{-1/2} \mathrm{d}\lambda. \quad (39)$$

Though it is not possible to analytically compute the last integral, its behavior for large N is known using the saddle point approximation or Perron's method: since all the conditions of the theorem at page 105 of Ref. [21] are satisfied, one can approximate the integral in Eq. (39) as follows

$$\int_{-\infty}^{+\infty} \hat{p}(\lambda) |\lambda|^{-1/2} \mathrm{d}\lambda \sim \Gamma\left(\frac{1}{2\alpha}\right) \frac{1}{\alpha c^{1/(2\alpha)} (N-1)^{1/(2\alpha)}}. \quad (40)$$

Remark 3. In the limit of large N it is then possible to show analytically that for volume distributions $p(\phi_i)$ belonging to the *domain of attraction* of the Lévy distribution the price impact is characterized by a linear regime described by

$$I_N(\phi) \simeq \frac{1}{2\sqrt{2\pi}} \Gamma\left(\frac{1}{2\alpha}\right) \frac{\phi}{\alpha[c(N-1)]^{1/(2\alpha)}}, \quad (41)$$

followed by a transition to a square root one around $\phi_N^* \simeq (c(N-1))^{1/\alpha}$.

B.2 Market Impact with Correlated Gaussian Metaorders

With the aim to introduce correlations between the metaorder volumes $\boldsymbol{\varphi}_N = (\phi_1, \cdots, \phi_N)$ in the Gaussian framework it is useful to define the following joint probability distribution

$$P(\boldsymbol{\varphi}_N) = \frac{1}{Z_N} \exp\left(-\frac{A_N}{2}\sum_{i=1}^N \phi_i^2 + \frac{B_N}{N}\sum_{i<j}^N \phi_i\phi_j + \mu\sum_{i=1}^N \phi_i\right), \qquad (42)$$

where Z_N is a normalization function, A_N and B_N are parameters depending on N and μ is an external field.

B.2.1 Calibration from Data: Means and Correlations

The first step to calibrate $P(\boldsymbol{\varphi}_N)$ in Eq. (42) is to express the model parameters A_N, B_N and μ in terms of observable quantities, namely $\mathbb{E}[\phi_i\phi_j|N]$ and $\mathbb{E}[\phi_i|N]$. Due to the presence of an interaction term, the computation of Z_N requires the use of a Hubbard-Stratonovich transformation (valid only for $B_N > 0$):

$$\exp\left(\frac{B_N}{2N}\sum_{i,j}^N \phi_i\phi_j\right) = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi/NB_N}} \exp\left(-\frac{NB_Ny^2}{2} + B_N\sum_{i=1}^N \phi_i y\right). \quad (43)$$

This allows us to rewrite the probability distribution in Eq. (42) as

$$P(\boldsymbol{\varphi}_N) = \frac{1}{Z_N}\sqrt{\frac{NB_N}{2\pi}}\int dy \prod_{i=1}^N \exp\left[-\frac{1}{2}\left(A_N + \frac{B_N}{N}\right)\phi_i^2 + (\mu + B_Ny)\phi_i - \frac{NB_N}{2}y^2\right].$$
$$(44)$$

The partition function then reads

$$Z_N = \left[\frac{2\pi}{\left(A_N + \frac{B_N}{N}\right)}\right]^{N/2}\sqrt{\frac{A_NN + B_N}{A_NN + B_N(N-1)}}\exp\left[\frac{N^2\mu^2}{2(A_NN + B_N(N-1))}\right]$$
$$(45)$$

valid only for $B_N < A_N + B_N/N$. Eq. (45) can be used to derive the following relations:

$$\frac{\partial \log Z_N}{\partial \mu} = N\mathbb{E}[\phi_i|N], \qquad (46)$$

$$\frac{\partial \log Z_N}{\partial A_N} = -\frac{N}{2}\mathbb{E}[\phi_i^2|N], \qquad (47)$$

$$\frac{\partial \log Z_N}{\partial B_N} = \left(\frac{N-1}{2}\right)\mathbb{E}[\phi_i\phi_j|N] \quad \text{with} \quad (i \neq j) \qquad (48)$$

which assuming symmetric volumes ($\mathbb{E}[\phi_i|N] = 0$, i.e. $\mu = 0$) are equivalent to

$$\mathbb{E}[\phi_i^2|N] = \frac{A_N + 2B_N/N - B_N}{(A_N + B_N/N - B_N)(A_N + B_N/N)} \qquad (49)$$

and

$$\mathbb{E}[\phi_i\phi_j|N] = \frac{B_N/N}{(A_N + B_N/N - B_N)(A_N + B_N/N)}. \qquad (50)$$

Furthermore, combining Eqs. (49) and (50) we can derive the volume correlation

$$\mathbb{C}_\phi(N) = \frac{\mathbb{E}[\phi_i\phi_j|N] - \mathbb{E}[\phi_i|N]^2}{\mathbb{E}[\phi_i^2|N] - \mathbb{E}[\phi_i|N]^2} = \frac{\mathbb{E}[\phi_i\phi_j|N]}{\mathbb{E}[\phi_i^2|N]} = \frac{B_N/N}{A_N + 2B_N/N - B_N}. \tag{51}$$

Vice versa, from Eqs. (49) and (51) we can obtain for the model parameters

$$A_N = \frac{1 - 2\mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N)}{(1 - \mathbb{C}_\phi(N))(1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N))\mathbb{E}[\phi_i^2|N]} \tag{52}$$

and

$$B_N = \frac{N\mathbb{C}_\phi(N)}{(1 - \mathbb{C}_\phi(N))(1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N))\mathbb{E}[\phi_i^2|N]}, \tag{53}$$

which are useful to fit the Gaussian model to data. We will use Eq. (52) and Eq. (53) to estimate A_N and B_N, replacing correlations and expectations by their empirical natural counterpart. The properties of this kind of estimators, belonging to the GMM (Generalized Method of Moments) is for instance discussed in Ref. [22].

B.3 Analytical Computation of Market Impact

To compute analytically the market impact function $I_N(\phi)$ from Eq. (21) in the Gaussian correlated framework we adopt the following strategy

1. firstly we factorize the joint probability distribution $P(\boldsymbol{\varphi}_N)$ in Eq. (42) with a not-null external field $\mu \neq 0$;

2. secondly we use the trick of the previous point to compute the market impact $I_N(\phi)$ in presence of an effective field $\tilde{\mu}$ induced by the correlation of the net order flow $\phi_m = \sum_{i \neq k}^N \phi_i$ with the known metaorder of size ϕ.

Step 1. The joint probability distribution in Eq. (42) can be written in the following matrix form

$$P(\boldsymbol{\varphi}_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}_N^T \mathbb{M}\boldsymbol{\varphi}_N + \boldsymbol{\mu}^T\boldsymbol{\varphi}_N\right), \tag{54}$$

where

- $\mathbb{M}$ is a NxN real and symmetric matrix with the elements on the principal diagonal equal to A_N and the ones elsewhere equal to $-B_N/N$,

- $\boldsymbol{\mu}$ is a N-dimensional vector with all the elements equal to the scalar $\mu \neq 0$.

Through the orthogonal transformation $\tilde{\boldsymbol{\varphi}}_N = \mathbb{O}\boldsymbol{\varphi}_N$ which diagonalizes the matrix $\mathbb{M}$, i.e. $\mathbb{O}^T\mathbb{M}\mathbb{O} = \text{diag}(\lambda_1, \cdots, \lambda_N)$, the joint probability distribution $P(\boldsymbol{\varphi}_N)$ factorizes as

$$P(\tilde{\boldsymbol{\varphi}}_N) = \frac{1}{Z_N}\prod_{m=2}^N \exp\left[-\frac{\lambda_2}{2}\tilde{\phi}_m^2\right]\exp\left[-\frac{1}{2}\lambda_1\tilde{\phi}_1^2 + \mu\sqrt{N}\tilde{\phi}_1\right] \tag{55}$$

where the N eigenvalues of the matrix $\mathbb{M}$

$$\lambda_1 = A_N - (N-1)\frac{B_N}{N} = \frac{1}{\mathbb{E}[\phi_i^2|N](1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N))} \tag{56}$$

and

$$\lambda_2 = \lambda_3 = \cdots = \lambda_N = A_N + \frac{B_N}{N} = \frac{1}{\mathbb{E}[\phi_i^2|N](1 - \mathbb{C}_\phi(N))} \tag{57}$$

allow us to rewrite the partition function as

$$Z_N = \sqrt{\frac{2\pi}{\lambda_1}} \left[\sqrt{\frac{2\pi}{\lambda_2}} \right]^{N-1} \exp\left[\frac{N\mu^2}{2\lambda_1} \right]. \tag{58}$$

In particular, it emerges that the first component of $\tilde{\boldsymbol{\varphi}}_N = \mathbb{O}\boldsymbol{\varphi}_N$ is equal to

$$\tilde{\phi}_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_i, \tag{59}$$

which put in evidence that the orthogonal basis change $\boldsymbol{\varphi}_N \to \tilde{\boldsymbol{\varphi}}_N$ is a useful trick to compute the market impact in the context of correlated Gaussian metaorders.

Step 2. To calculate the price impact $I_N(\phi)$ defined in Eq. (21) with N overall correlated Gaussian metaorders and in absence of an external field it is necessary to explicit the conditional probability distribution

$$P(\boldsymbol{\varphi}_N|\phi_k = \phi) = \frac{P(\phi_1, \cdots, \phi, \cdots, \phi_N)}{p(\phi)} \tag{60}$$

where $P(\phi_1, \cdots, \phi, \cdots, \phi_N)$ is given by Eq. (42) setting $\mu = 0$ while the marginal one is equal to

$$p(\phi) = \frac{1}{\sqrt{2\pi\tilde{\lambda}_1/(\lambda_1\lambda_2)}} \exp\left[-\frac{\phi^2}{2} \frac{\lambda_1\lambda_2}{\tilde{\lambda}_1} \right], \tag{61}$$

with λ_1 and λ_2 respectively given by Eqs. (56) and (57) and

$$\tilde{\lambda}_1 = A_N - \frac{N-2}{N}B_N = \frac{1}{\mathbb{E}[\phi_i^2|N][1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N)][1 - \mathbb{C}_\phi(N)]}. \tag{62}$$

It follows that the conditional probability distribution is equal to

$$P(\boldsymbol{\varphi}_N|\phi_k = \phi) = \frac{\exp\left[-\frac{(N-1)B_N^2\phi^2}{2N^2\tilde{\lambda}_1} \right]}{\underbrace{(2\pi/\tilde{\lambda}_1)^{1/2}(2\pi/\lambda_2)^{\frac{N-2}{2}}}_{\Theta_N^{-1}(\phi)}} \times \tag{63}$$

$$\exp\left[-\frac{A_N}{2} \sum_{\substack{i \neq k}}^N \phi_i^2 + \frac{B_N}{N} \sum_{\substack{i<j \\ i,j \neq k}}^N \phi_i\phi_j + \underbrace{\frac{\phi B_N}{N}}_{\tilde{\mu}} \sum_{\substack{i \neq k}}^N \phi_i \right] \tag{64}$$

where it emerges that the conditioning to the metaorder with volume $\phi_k = \phi$ is equivalent to the introduction of an effective field $\tilde{\mu}$ proportional to ϕ. This implies that the price impact in Eq. (21) is given solving the following conditional

expectation

$$I_N(\phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N)|\phi_k = \phi] \tag{65}$$

$$= \int_{-\infty}^{\infty} \prod_{i \neq k}^{N} d\phi_i \frac{P(\phi_1, \cdots, \phi, \cdots, \phi_N)}{p(\phi)} \left(\phi + \sum_{i \neq k}^{N} \phi_i \right)^{\bullet 1/2} \tag{66}$$

$$= \frac{1}{\Theta_N(\phi)} \int_{-\infty}^{\infty} \prod_{i \neq k}^{N} d\phi_i \exp \left[-\frac{A_N}{2} \sum_{i \neq k}^{N} \phi_i^2 + \frac{B_N}{N} \sum_{\substack{i < j \\ i,j \neq k}}^{N} \phi_i \phi_j + \tilde{\mu} \sum_{i \neq k}^{N} \phi_i \right] \left(\phi + \sum_{i \neq k}^{N} \phi_i \right)^{\bullet 1/2} \tag{67}$$

$$= \frac{1}{\Theta_N(\phi)} \int_{-\infty}^{\infty} \prod_{i \neq k}^{N} d\phi_i \exp \left[-\boldsymbol{\varphi}^{*T} \mathbb{M}^* \boldsymbol{\varphi}^* + \tilde{\mu} \boldsymbol{\varphi}^* \right] \left(\phi + \sum_{i \neq k}^{N} \phi_i \right)^{\bullet 1/2} . \tag{68}$$

Herein

- $\boldsymbol{\varphi}^* = \{\phi_i\}_{i=1,\cdots,N}^{i \neq k}$ is a vector that contains the $N-1$ unknown metaorders volumes simultaneously executed with the one known $\phi_k = \phi$,

- $\mathbb{M}^*$ is a $(N-1) \times (N-1)$ symmetric and real matrix with A_N on the principal diagonal and $-B_N/N$ elsewhere: it is easy to check that its eigenvalues are respectively $\lambda_1^* = \tilde{\lambda}_1$ as in Eq. (62) and $\lambda_m^* = \lambda_2$ as in Eq. (57) for $m = 2, \cdots, N-1$.

As mentioned before, to solve Eq. (68) it is useful to use the trick described in Step 1.: we apply in Eq. (68) the orthogonal transformation $\tilde{\boldsymbol{\phi}}^* = \mathbb{H} \boldsymbol{\varphi}^*$ that diagonalizes the matrix $\mathbb{M}^*$ (i.e. $\mathbb{H}^T \mathbb{M}^* \mathbb{H} = \mathrm{diag}(\tilde{\lambda}_1, \cdots, \tilde{\lambda}_{N-1})$) and since the determinant of the Jacobian matrix associated to this transformation is equal to one, we obtain that

$$I_N(\phi) = \frac{1}{\Theta_N(\phi)} \int_{-\infty}^{\infty} \prod_{m=2}^{N} d\tilde{\phi}_m^* \exp \left[-\frac{\lambda_2}{2} (\tilde{\phi}_m^*)^2 \right] \times$$

$$\times \int_{-\infty}^{\infty} d\tilde{\phi}_1^* \exp \left[-\frac{1}{2} \tilde{\lambda}_1 (\tilde{\phi}_1^*)^2 + \tilde{\mu} \sqrt{N-1} \tilde{\phi}_1^* \right] (\phi + \sqrt{N-1} \tilde{\phi}_1^*)^{\bullet 1/2}; \quad (69)$$

then integrating in $\tilde{\phi}_m^*$ for $m = 2, \cdots, N-1$, completing the square in the argument of the $\exp[-\frac{1}{2} \tilde{\lambda}_1 (\tilde{\phi}_1^*)^2 + \tilde{\mu} \sqrt{N-1} \tilde{\phi}_1^*]$ and doing the variable change $x = \phi + \sqrt{N-1} \tilde{\phi}_1^*$ we derive the following final expression

$$I_N(\phi) = \frac{1}{\sqrt{2\pi(N-1)/\tilde{\lambda}_1}} \int_0^{\infty} dx \sqrt{x} \left(\exp \left[-\frac{\tilde{\lambda}_1}{2(N-1)} (x - \phi(1 + (N-1)\mathbb{C}_\phi(N)))^2 \right] - \right.$$

$$\left. \exp \left[-\frac{\tilde{\lambda}_1}{2(N-1)} (x + \phi(1 + (N-1)\mathbb{C}_\phi(N)))^2 \right] \right) . \tag{70}$$

Remark 4. From Eq. (70) it follows that the price impact $I_N(\phi)$ in the correlated Gaussian framework is equivalent to the one computed in the independent Gaussian case (see Eqs. (24) and (25)) substituting

$$\phi \longrightarrow \phi[1 + (N-1)\mathbb{C}_\phi(N)] \tag{71}$$

and

$$\Sigma_N^2 \longrightarrow 1/\tilde{\lambda}_1. \tag{72}$$

B.4 Market Impact with Correlated Signs and IID Unsigned Volumes

In Section 5.4 we have presented a general model in which the metaorder signs $\epsilon_i = \pm 1$ are correlated while the unsigned volumes $|\phi_i|$ are i.i.d. and described by a generic distribution $p(|\phi_i|)$ defined on the finite positive support $(0, 1)$. This means that the joint probability distribution is factorizable as

$$P(\boldsymbol{\varphi}_N) = \mathcal{P}(\boldsymbol{\epsilon}_N) \prod_{i=1}^{N} p(|\phi_i|). \tag{73}$$

In this theoretical setup we are not able to compute analytically the price impact $I_N(\phi)$ and we will use numerical simulations. To this aim we introduce a latent discrete variable $\tilde{\epsilon}$ in order to simulate N correlated signs with the following statistical model

$$\mathbb{P}(\epsilon_i|\tilde{\epsilon}) = \frac{1}{2}(1 + \gamma_\epsilon \epsilon_i \tilde{\epsilon}) \tag{74}$$

where γ_ϵ can be estimated from data by averaging the realized sign correlation ρ_ϵ appearing in Eq. (16), i.e.

$$\mathbb{E}[\rho_\epsilon|N] = \gamma_\epsilon^2. \tag{75}$$

For clarity we explicitly omit the γ_ϵ's dependence on N. Thus to simulate the model we fix $\epsilon_k = +1$, we draw a hidden factor $\tilde{\epsilon}$ from

$$\mathbb{P}(\tilde{\epsilon}|\epsilon_k = +1) = \frac{1}{2}(1 + \gamma_\epsilon \tilde{\epsilon}) \tag{76}$$

and then we sample $N-1$ other correlated signs $\{\epsilon_i\}_{i=1,\cdots,N}^{i \neq k}$ with probability

$$\mathbb{P}(\epsilon_i|\tilde{\epsilon}) = \frac{1}{2}(1 + \gamma_\epsilon \epsilon_i \tilde{\epsilon}). \tag{77}$$

B.4.1 Numerical Computation of Market Impact

We summarize the main steps for the numerical calibration of the price impact $I_N(\phi)$ from data. Given the number N of metaorders per stock/day pair and fixing $|\phi_k| = \phi > 0$:

1. We compute the average sign correlation $\mathbb{C}_\epsilon(N) = \mathbb{E}[\rho_\epsilon|N]$ as to obtain γ_ϵ through Eq. (75).

2. After fixing the direction $\epsilon_k = +1$ we simulate $N-1$ correlated signs using Eq. (77).

3. We sample $N-1$ random variables $|\phi_i|$ from an half-normal distribution with mean $\Sigma_N \sqrt{2/\pi}$ and standard deviation $\Sigma_N \sqrt{1 - 2/\pi}$ where Σ_N represents the empirical standard deviation of signed volumes.

4. We compute numerically the price impact $I_N(\phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N)|\epsilon_k = +1, |\phi_k| = \phi]$ where $\mathcal{I}(\boldsymbol{\varphi}_N) = \left(\sum_{i=1}^{N} \phi_i\right)^{\bullet 1/2}$ and $\phi_i = \epsilon_i|\phi_i|$, as defined in Eq. (6).

5. Finally, we compute $I(\phi)$ averaging $I_N(\phi)$ over the empirical distribution $p(N)$ shown in the left panel of Fig 1.

References

[1] Jean-Philippe Bouchaud, J Doyne Farmer, and Fabrizio Lillo. *How markets slowly digest changes in supply and demand.* Handbook of Finance, 2009:57–160, 2009.

[2] Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, Quotes and Prices Financial Markets Under The Microscope.* Cambridge University Press, 2018.

[3] Albert S. Kyle. *Continuous auctions and insider trading.* Econometrica: Journal of the Econometric Society, pages 1315–1335, 1985.

[4] Elia Zarinelli, Michele Treccani, J. Doyne Farmer, and Fabrizio Lillo. *Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate.* Market Microstructure and Liquidity, 1(02):1550004, 2015.

[5] Andy Puckett and Xuemin Sterling Yan. *Short-term institutional herding and its impact on stock prices.* Unpublished working paper, University of Missouri, 2008.

[6] Michael A Goldstein, Paul Irvine, Eugene Kandel, and Zvi Wiener. *Brokerage commissions and institutional trading patterns.* The Review of Financial Studies, 22(12):5175–5212, 2009.

[7] Thomas J Chemmanur, Shan He, and Gang Hu. *The role of institutional investors in seasoned equity offerings.* Journal of Financial Economics, 94(3):384–411, 2009.

[8] Russell Jame. *Organizational structure and fund performance: pension funds vs. mutual funds.* Mutual Funds (January 22, 2010), 2010.

[9] Michael A. Goldstein, Paul Irvine, and Andy Puckett. *Purchasing ipos with commissions.* Journal of Financial and Quantitative Analysis, 46(5):1193–1225, 2011.

[10] Andy Puckett and Xuemin Sterling Yan. *The interim trading skills of institutional investors.* The Journal of Finance, 66(2):601–633, 2011.

[11] Jeffrey A. Busse, T. Clifton Green, and Narasimhan Jegadeesh. *Buy-side trades and sell-side recommendations: Interactions and information content.* Journal of Financial Markets, 15(2):207–232, 2012.

[12] Bence Tóth, Yves Lemperiere, Cyril Deremble, Joachim De Lataillade, Julien Kockelkoren, and Jean-Philippe Bouchaud. *Anomalous price impact and the critical nature of liquidity in financial markets.* Physical Review X, 1(2):021006, 2011.

[13] Iacopo Mastromatteo, Bence Tóth, and Jean-Philippe Bouchaud. *Agent-based models for latent liquidity and concave price impact.* Physical Review E, 89(4):042805, 2014.

[14] Nicolo G. Torre and Mark J. Ferrari. *The market impact model.* Horizons, The Barra Newsletter, 165, 1998.

[15] Robert Almgren, Chee Thum, Emmanuel Hauptmann, and Hong Li. *Direct estimation of equity market impact.* Risk, 18(7):5862, 2005.

[16] Robert Engle, Robert Ferstenberg, and Jeffrey Russell. *Measuring and modeling execution cost and risk.* Chicago GSB Research Paper, no. 08-09, 2006.

[17] Xavier Brokmann, Emmanuel Sérié, Julien Kockelkoren, and Jean-Philippe Bouchaud. *Slow decay of impact in equity markets.* Market Microstructure and Liquidity, 1(02):1550007, 2015.

[18] Emmanuel Bacry, Adrian Iuga, Matthieu Lasnier, and Charles-Albert Lehalle. *Market impacts and the life cycle of investors orders.* Market Microstructure and Liquidity, 1(02):1550009, 2015.

[19] Esteban Moro, Javier Vicente, Luis G. Moyano, Austin Gerig, J. Doyne Farmer, Gabriella Vaglica, Fabrizio Lillo, and Rosario N. Mantegna. *Market impact and trading profile of hidden orders in stock markets.* Physical Review E, 80(6):066102, 2009.

[20] Stephen Wolfram and Thierry Dubois. *Mathematica*, volume 3. Addison-Wesley Redwood City (Calif.) etc., 1991.

[21] Roderick Wong. *Asymptotic approximations of integrals.* Society for Industrial and applied mathematics, 2001.

[22] Hansen Lars Peter. *Large sample properties of generalized method of moments estimators.* Econometrica, 50(4):1029–1054, 1982.