Co-impact: crowding effects in institutional trading activity

F. Bucci, I. Mastromatteo, Z. Eisler, F. Lillo, J.-P. Bouchaud & C.-A. Lehalle

Co-impact: crowding effects in institutional trading activity

F. BUCCI[*†], I. MASTROMATTEO[‡], Z. EISLER[‡], F. LILLO[§¶], J.-P. BOUCHAUD[‡‖] and
C.-A. LEHALLE[‡‖]

†Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
‡Capital Fund Management, 23-25, Rue de l'Université, 75007 Paris, France
§Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy
¶Center for Analysis, Decisions, and Society, Human Technopole, Milano, Italy
‖CFM-Imperial Institute of Quantitative Finance, Department of Mathematics, Imperial College, 180 Queen's Gate,
SW7 2RH London, UK

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?

Keywords: Market microstructure; Market impact model; Large trade execution; Institutional trading activity

JEL Classification: D4, C6

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 (Bouchaud *et al.* 2009, 2018). 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 findings 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 (Kyle 1985). 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.

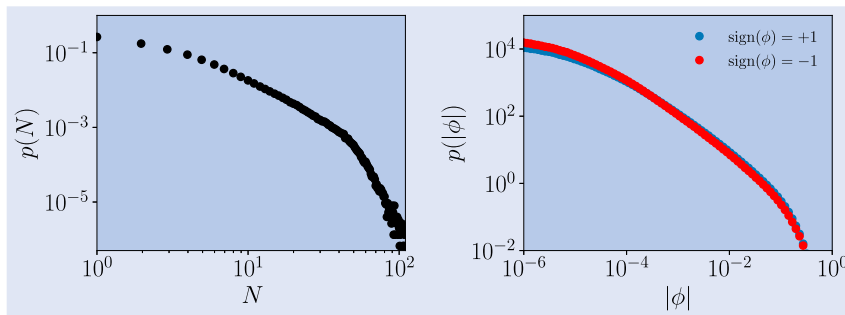*Corresponding author. Email: frederic.bucci@sns.it

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.

The paper is organized as follows. In Section 2 we introduce the ANcerno dataset we used empirically. In Section 3 we discuss the limits of the validity of the square root law on the daily level. In Section 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 Section 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 behavior. We also compare in Section 5 the results of our simple mathematical model with empirical data, with very satisfactory results. Section 6 concludes.

2. The ANcerno database

Our analysis relies on a database made available by *ANcerno*, a leading transaction-cost analysis provider.† 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 unfavorably. 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 Zarinelli *et al.* (2015), 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 Zarinelli *et al.* (2015) 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 $T = t_e - t_s$ 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%.

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.‡§ The statistical properties of the metaorders, in terms of volume fraction, duration, etc., are broadly in line with (Zarinelli *et al.* 2015) even though their data was aggregated at the level of brokers (not shown). 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 figure 1, indicating that N is broadly distributed with an average close to 5.

The right panel of figure 1 shows the probability distribution of the absolute value of the volume fraction ϕ of the

---

† ANcerno Ltd. (formerly the Abel Noser Corporation) is a widely recognized 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 (Chemmanur *et al.* 2009, Goldstein *et al.* 2009, 2011, Puckett and Yan 2011, Busse *et al.* 2012, Zarinelli *et al.* 2015, Busse *et al.* 2016, Chakrabarty *et al.* 2017, Jame 2017). See www.ancerno.com for details.

‡ Without the above filters, this number would rise to about 10%. Although we believe that the 90% of missing volume is not 'noise', the ANcerno database is representative of the full set of metaorders. In fact, many of the results discussed below are quantitatively similar to those observed on much smaller datasets, such as the CFM database of executed metaorders.
§ We checked that the results discussed in the present work are still valid independently of the use of Filter 3 and Filter 4.

metaorders. This variable plays a key role in the following and is defined as $\phi := Q/V$ where V is the total daily market volume averaged over the 25 days before the day in which ϕ is computed. 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.

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 and σ is the daily volatility averaged over the 25 days before S is measured. Daily volatility at day d is estimated as $(S_{\text{high},d} - S_{\text{low},d})/S_{\text{open},d}$, a simple metric based on the high, low and open prices of the day d. 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 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

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

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 (Torre and Ferrari 1998, Almgren *et al.* 2005, Moro *et al.* 2009, Tóth *et al.* 2011, Engle *et al.* 2012, Mastromatteo *et al.* 2014, Bacry *et al.* 2015, Brokmann *et al.* 2015)

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

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 equation (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.

We first check again this empirical result on our dataset. In figure 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 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 $10^{-3} \lesssim \phi \lesssim 10^0$, two 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.
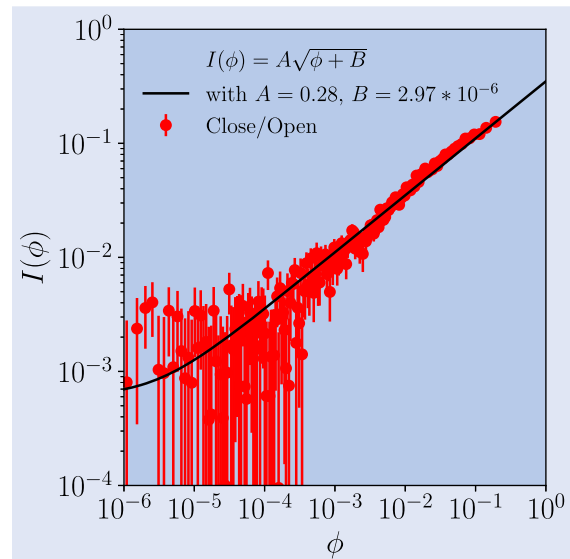

Figure 2. Market impact curve $I(\phi) = \mathbb{E}[s_{\text{close}} - s_{\text{open}}|\phi]$ as a function of the metaorder size ratio $\phi = Q/V$ computed using the filtered metaorders from the ANcerno dataset in the period from January 2007 – June 2010. We also show the simple fit $I(\phi) = A\sqrt{\phi + B}$, which captures some—but not all—of the discrepancy with the square root law at small ϕ.

(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.

These results are robust across time periods and market capitalizations, and consistent with (Zarinelli *et al.* 2015), where regimes 1. and 2. were also clearly observed. In the following we will seek to understand these 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 figure 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 figure 3, impact in this case is almost perfectly fitted by a square root law. Figure 3, also shows that as N increases, significant departures from the square root
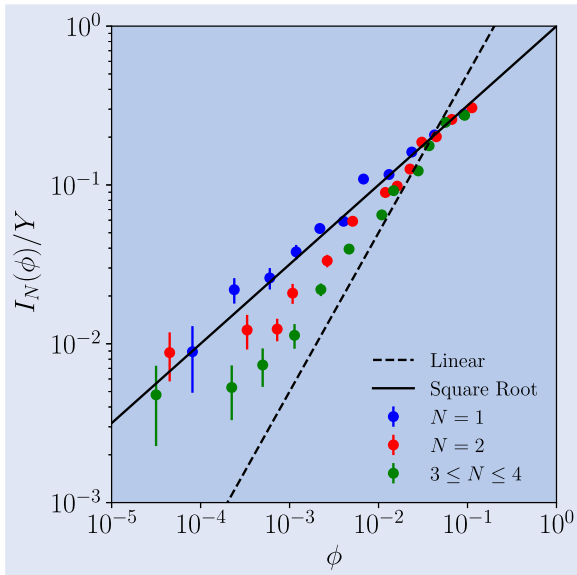
Figure 3. Empirical evidence on the effect of the number of metaorders N on the daily price impact curves $I_N(\phi)$ normalized by the prefactor *Y-ratio*: significant departures from the square root law can be observed for small ϕ increasing the number N of daily metaorders per asset.
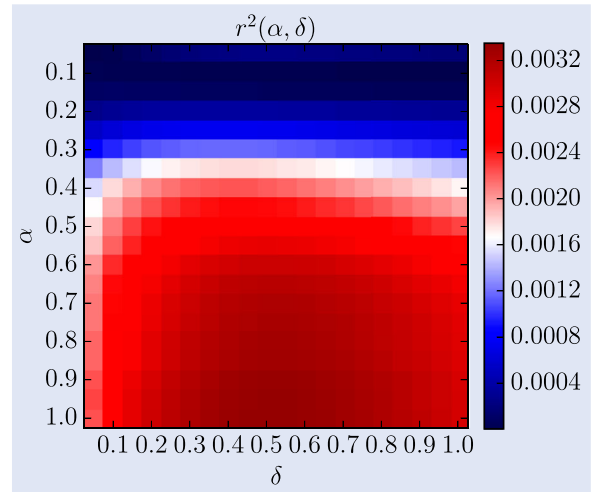

Figure 4. The computed coefficient of determination $r^2(\alpha, \delta)$ of the least squares regressions of equation (5) for a grid of (α, δ) pairs. The coefficient of determination is maximized ($r^2 = 0.0035$) in the vicinity of the point $\alpha = 1.0$ and $\delta = 0.5$.

law can be observed for small ϕ, as suggested by our simple model equation (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 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:

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

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

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

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 indistinguishable. 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 ($\alpha = 1$).

Figure 4 shows the quality of the fit obtained by least squares regressions of equation (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 the total net order flow, i.e.†

$$\mathcal{I}(\boldsymbol{\varphi}_N) \approx Y \times \Phi^{\bullet 1/2}, \qquad (6)$$

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.

One can now plot $\mathcal{I}(\Phi)$ as a function of Φ for various N, see figure 5. One clearly sees that when normalized by the corresponding average $\langle \sigma/\sqrt{V} \rangle_N$ (conditioned on the same range of N), impact is independent of N and crosses over from linear to square root behavior as Φ increases. 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 equation (6), this may not be sufficient to estimate his

---

† We have in fact tested that the assumption $\alpha = 1$ is also favored for a general, non parametric shape for the impact function $\mathcal{I}(\Phi)$.
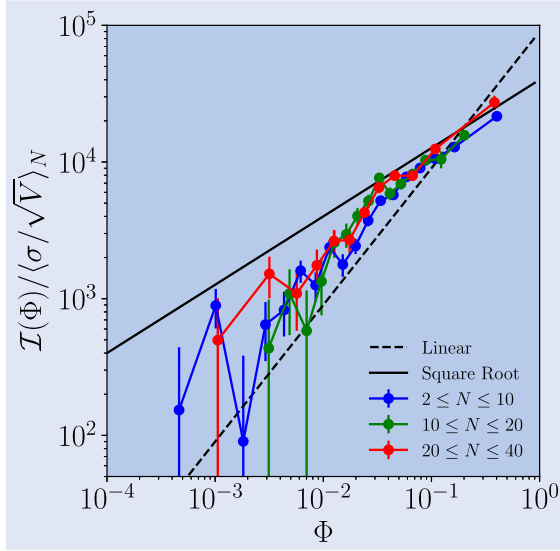
Figure 5. Global market impact $\mathcal{I}(\Phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N)| \sum_{i=1}^{N} \phi_i = \Phi]$ normalized by the average $\langle \sigma/\sqrt{V} \rangle_N$ conditioned on the same range of N, and symmetrized as explained in Section 3.

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 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

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

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

$$I(\phi) = Y \times \sum_N p(N) \int d\phi_1 \ldots d\phi_N P(\boldsymbol{\varphi}_N|\phi_k = \phi)$$

$$\times \Big(\phi_k + \sum_{i \neq k}^{N} \phi_i\Big)^{\bullet 1/2}, \tag{8}$$

where we have used equation (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.

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 A.1.1 we show analytically that:

- For small metaorders the noise term dominates, leading to
  $$I_N(\phi) \propto \phi \quad \text{when } \phi \ll \phi_N^* := \Sigma_N\sqrt{N-1}.$$

- For large metaorders the $N-1$ other simultaneous metaorders can be neglected and thus
  $$I_N(\phi) \propto \sqrt{\phi} \quad \text{when } \phi \gg \phi_N^*.$$

In appendix A.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 figure 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.† As shown in the right panel of figure 6, this is the case empirically since for $N \gtrsim 10$, Σ_N indeed decays as N^{-1} (as also suggested in figure 7). 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 (Tóth *et al.* 2011, Mastromatteo *et al.* 2014). 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 figure 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,

---

† For example, the variance of a flat Dirichlet random variable $(X_1, \ldots, X_N) \sim \text{Dir}(N)$ describing fractions is $\mathbb{V}[X_i|N] = (N-1)/(N^2(N+1)) \sim N^{-2}$.
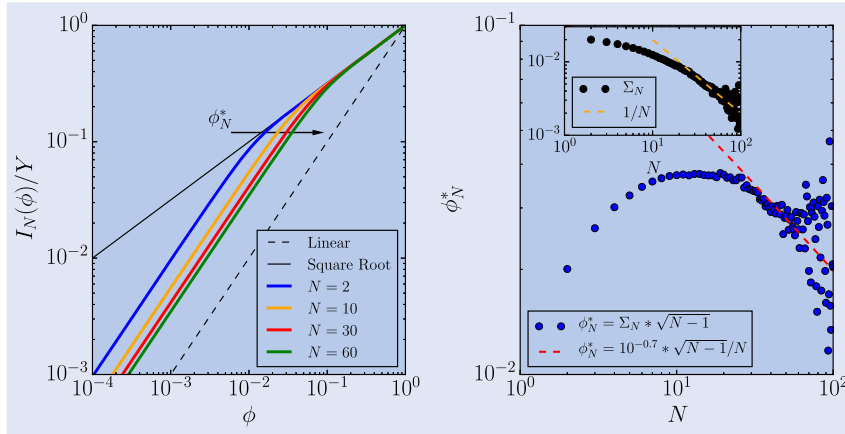
Figure 6. (Left panel) Market impact curves $I_N(\phi)/Y$ for i.i.d. Gaussian metaorders for different $N \in \{2, 10, 30, 60\}$ and *fixed* $\Sigma_N = \Sigma = 0.8\%$ computed from empirical data. The transition from the square root to the linear regime takes place for $\phi_N^* \simeq \Sigma\sqrt{N-1}$. (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.
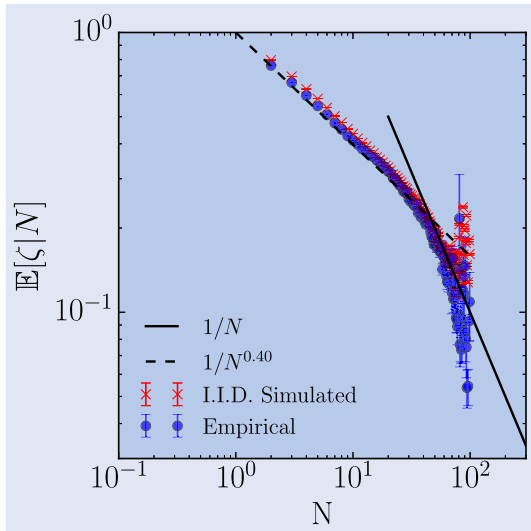

Figure 7. Average value of the Herfindahl index (or 'inverse participation ratio') $\mathbb{E}[\zeta|N]$ as a function of the number N of daily metaorders per asset computed from the empirical data (blue dot symbols) and simulating i.i.d. $|\phi_i|$ (red star symbols) extracted from the empirical distribution illustrated in the right panel of figure 1.

defined as:

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

This quantity is of order $1/N$ if all metaorders are of comparable size, and of order 1 if one metaorder dominates. In figure 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 figure 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.
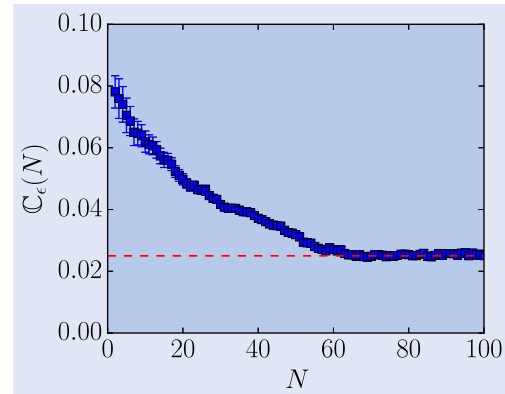

Figure 8. Empirical average sign correlation $\mathbb{C}_\epsilon(N)$ as a function of the number N of daily metaorders: the dashed red line represents the plateau value $\mathbb{C}_\epsilon \approx 0.025$ at large N, which, we believe, is a reasonable proxy for the correlation of orders submitted by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

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. Figure 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.†

---

† This value is indeed compatible with the value of $\mathbb{C}_\epsilon$ obtained with the version of the database used by Zarinelli *et al.* (2015), which identified metaorders coming from the same investor.

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 A.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

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

in the expression of $I_N(\phi)$ for independent Gaussian. 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

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

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.

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:

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

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

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 figure 9. Results for unsigned volumes generated from a half-normal distribution calibrated on data are shown in figure 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 figure 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 figure 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 A.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

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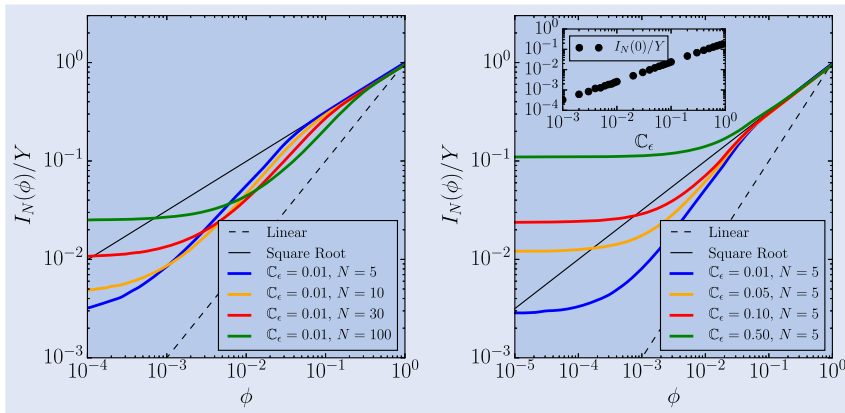

Figure 9. Market impact curves $I_N(\phi)/Y$ computed for the correlated signs model, where volumes are drawn according to a half-normal distribution with $\mathbb{E}[|\phi_i|] = \Sigma\sqrt{2/\pi}$ and $\Sigma = 0.8\%$ is set equal to its empirical value averaged over N. (Left panel) Numerical simulations 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.
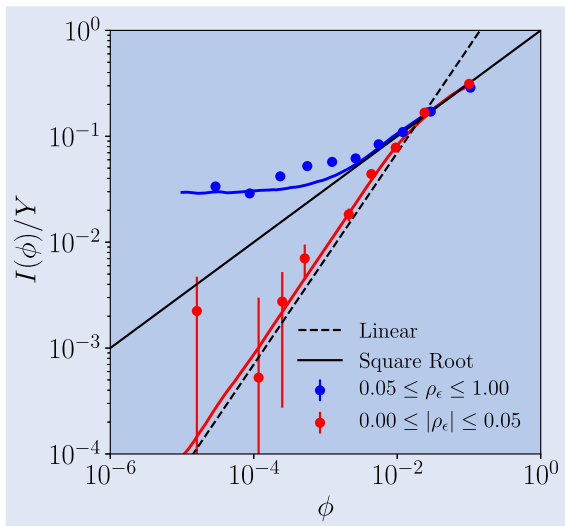
Figure 10. Comparison between calibrated sign-correlated model (colored lines) and empirical data (circles) for the case $N \in [2, 10]$: The sample is split into two sub-samples respectively with $\rho_\epsilon \geq 0.05$ and $0 \leq |\rho_\epsilon| \leq 0.05$. The theoretical curves are calculated through numerical simulations as explained in the main text (see appendix A.4.1 for all the details of the procedure).

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

Figure 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 figure 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 $\phi \to 0$.

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 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.

Data availability statement

The data were purchased from the company ANcerno Ltd (formerly the Abel Noser Corporation) which is a widely recognized 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 Chemmanur *et al.* (2009), Goldstein *et al.* (2009, 2011), Puckett and Yan (2011), Busse *et al.* (2012, 2016), Zarinelli *et al.* (2015), Chakrabarty *et al.* (2017), and Jame (2017). The authors do not have permission to redistribute them, even in aggregate form. Requests for this commercial dataset can be addressed directly to the data vendor. See www.ancerno.com for details.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Direct estimation of equity market impact. *Risk*, 2005, **18**, 5862.

Bacry, E., Iuga, A., Lasnier, M. and Lehalle, C.A., Market impacts and the life cycle of investors orders. *Market Microstruct. Liquidity*, 2015, **1**, 1550009.

Bouchaud, J.P., Bonart, J., Donier, J. and Gould, M., *Trades, Quotes and Prices Financial Markets under the Microscope*, 2018 (Cambridge University Press: Cambridge).

Bouchaud, J.P., Farmer, J.D. and Lillo, F., How markets slowly digest changes in supply and demand. In *Handbook Financial Markets: Dynamics and Evolution*, Vol. 2009, edited by Thorsten Hens and Klaus Schenk-Hoppe, pp. 57–160, 2009 (Elsevier: North Holland).

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

Busse, J., Green, T.C. and Jegadeesh, N., Buy-side trades and sell-side recommendations: Interactions and information content. *J. Financ. Markets*, 2012, **15**, 207–232.

Busse, J., Tarun, C., Lei, J. and Yuechua, T., Mutual fund transaction costs, Working Paper, 2016. Available online at: https://www.ilpost.it/wp-content/uploads/2016/12/mutual-fund-transaction-costs-SSRN-id2350583-1.pdf.

Chakrabarty, B., Moulton, P.C. and Trzcinka, C., The performance of short-term institutional trades. *J. Financ. Quant. Anal.*, 2017, **52**, 1403–1428.

Chemmanur, T.J., He, S. and Hu, G., The role of institutional investors in seasoned equity offerings. *J. Financ. Econ.*, 2009, **94**, 384–411.

Engle, R., Ferstenberg, R. and Russell, J., Measuring and modeling execution cost and risk. *J. Portfolio Manage*, 2012, **32**, 14–28.

Goldstein, M.A., Irvine, P., Kandel, E. and Wiener, Z., Brokerage commissions and institutional trading patterns. *Rev. Financ. Stud.*, 2009, **22**, 5175–5212.

Goldstein, M.A., Irvine, P. and Puckett, A., Purchasing IPOs with commissions. *J. Financ. Quant. Anal.*, 2011, **46**, 1193–1225.

Hansen, L.P., Large sample properties of generalized method of moments estimators. *Econometrica*, 1982, **50**, 1029–1054.

Jame, R., Liquidity provision and the cross-section of hedge fund returns. *Manage. Sci.*, 2017, **64**, 1–25.

Kyle, A.S., Continuous auctions and insider trading. *Econ.: J. Econ. Soc.*, 1985, **53**, 1315–1335.

Mastromatteo, I., Tóth, B. and Bouchaud, J.P., Agent-based models for latent liquidity and concave price impact. *Phys. Rev. E*, 2014, **89**, 042805.

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

Puckett, A. and Yan, X.S., The interim trading skills of institutional investors. *J. Finance.*, 2011, **66**, 601–633.

Roderick, W., *Asymptotic Approximations of Integrals*, 2001 (Society for Industrial and Applied Mathematics: Philadelphia, PA).

Torre, N.G. and Ferrari, M.J., The market impact model. *Horizons, The Barra Newsletter*, 1998, **165**.

Tóth, B., Lemperiere, Y., Deremble, C., De Lataillade, J., Kockelkoren, J. and Bouchaud, J.P., Anomalous price impact and the critical nature of liquidity in financial markets. *Phys. Rev. X*, 2011, **1**, 021006.

Wolfram, S. and Dubois, T., *Mathematica*, 1991 (Addison-Wesley: Redwood City, CA).

Zarinelli, E., Treccani, M., Farmer, J.D. and Lillo, F., Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstruct. Liquidity*, 2015, **1**, 1550004.

A. Appendix 1. From the bare to the market impact function.

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

$$I(\phi) = \sum_N p(N) I_N(\phi), \tag{A1}$$

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

$$I_N(\phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N) | \phi_k = \phi] = \int d\phi_1 \cdots d\phi_N P(\boldsymbol{\varphi}_N | \phi_k = \phi)$$
$$\times \left(\phi_k + \sum_{i \neq k}^N \phi_i \right)^{\bullet 1/2} \tag{A2}$$

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 known a priori the market impact $I_N(\phi)$ is given from the expectation in equation (A2) 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 $P(\boldsymbol{\varphi}_N) := P(\phi_1, \ldots, \phi_N)$.

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 satisfies 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 equation (A2) in the case of metaorder signs correlated and i.i.d. unsigned volumes.

A.1. Market impact with i.i.d. metaorders.

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

$$P(\boldsymbol{\varphi}_N | \phi_k = \phi) = \prod_{i \neq k}^N p(\phi_i); \tag{A3}$$

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

$$I_N(\phi) = \int d\phi_m \underbrace{\frac{1}{2\pi} \int d\lambda e^{-i\lambda\phi_m} \hat{p}(\lambda)^{N-1}}_{p(\phi_m)} (\phi + \phi_m)^{\bullet 1/2}$$
$$= \int d\phi_m p(\phi_m)(\phi + \phi_m)^{\bullet 1/2}, \tag{A4}$$

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 equation (A4) will be a convenient way to exploit the convergence of the net order flow distribution $p(\phi_m)$ as discussed in appendix A.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 A.1.1.

A.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 equation (A4) since $\hat{p}(\lambda) = \mathrm{e}^{-\Sigma_N^2 \lambda^2/2}$. In fact the integral representation of the price impact

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

$$= \frac{1}{\Sigma_N \sqrt{2\pi(N-1)}} \int_0^{\infty} \mathrm{d}x \sqrt{x} \Big[\mathrm{e}^{-(x-\phi)^2/(2\Sigma_N^2(N-1))}$$

$$- \mathrm{e}^{-(x+\phi)^2/(2\Sigma_N^2(N-1))} \Big] \tag{A5}$$

can be expressed in the following analytical way:

$$I_N(\phi) = \frac{\Gamma(1/4)}{2\sqrt{\pi}} \frac{\phi}{(2(N-1)\Sigma_N^2)^{1/4}} \, \mathrm{e}^{-\frac{\phi^2}{2(N-1)\Sigma_N^2}} \, {}_1F_1$$

$$\times \left(\frac{5}{4}, \frac{3}{2}, \frac{\phi^2}{2(N-1)\Sigma_N^2} \right), \tag{A6}$$

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

$${}_1F_1 \left(\frac{5}{4}, \frac{3}{2}, z \right) = \frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{5}{4})} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{5}{4}+j)}{\Gamma(\frac{3}{2}+j)} \frac{z^j}{j!} \tag{A7}$$

is the Kummer confluent hypergeometric function with $z = \phi^2/(2(N-1)\Sigma_N^2)$ (Wolfram and Dubois 1991).

The price impact $I_N(\phi)$ in equation (A6) is shown in the left panel of figure 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 figure 1 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 equation (A6) around $\phi = 0$ as follows:

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

$$\times \left[\mathrm{e}^{-(\phi^2-2x\phi)/(2\Sigma_N^2(N-1))} - \mathrm{e}^{-(\phi^2+2x\phi)/(2\Sigma_N^2(N-1))} \right] \simeq$$

$$\simeq \sqrt{\frac{2}{\pi}} \frac{\phi}{(\Sigma_N^2(N-1))^{3/2}} \int_0^{\infty} \mathrm{d}x x^{3/2} \mathrm{e}^{-x^2/(2\Sigma_N^2(N-1))}$$

$$= 2^{3/4} \frac{\Gamma(5/4)}{(\pi^2 \Sigma_N^2(N-1))^{1/4}} \phi. \tag{A8}$$

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 equation (A8)) and the crossover to the square root region happens in ϕ_N^* obtained by solving

$$\xi \frac{\phi_N^*}{(\Sigma_N^2(N-1))^{1/4}} \simeq (\phi_N^*)^{1/2}, \tag{A9}$$

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

A.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 satisfies 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 satisfied, the symmetric volume distribution $p(\phi_m)$ introduced in equation (A4) converges to a stable law $\mathcal{G}_\alpha$ described by a characteristic function

$$\hat{p}(\lambda) = \mathbb{E}[\mathrm{e}^{i\lambda\phi_m}] = \mathrm{e}^{-c|\lambda|^\alpha}, \tag{A10}$$

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

$$b_m^{-1}(\phi_m - a_m) \to \mathcal{G}_\alpha, \tag{A11}$$

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 guaranteed:

- $p(\phi_m)$ converges to a Gaussian distribution ($\alpha = 2$ in equation (A10)) if and only if

$$\int_{|\phi_i| \leq x} \phi_i^2 p(\phi_i) \mathrm{d}\phi_i \tag{A12}$$

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 $\Sigma_N^2 = \mathbb{V}[\phi_i|N] < \infty$

$$((N-1)^{1/2}\Sigma_N)^{-1}\phi_m \longrightarrow \mathcal{N}(0,1), \tag{A13}$$

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

$$((N-1)^{1/2}L_1)^{-1}\phi_m \longrightarrow \mathcal{N}(0,1) \tag{A14}$$

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 equation (A10)) if and only if

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

$$= \frac{c_2 + o(1)}{x^\alpha} L(x), \quad x \to \infty, \tag{A15}$$

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

$$((N-1)^{1/\alpha}L_2)^{-1}\phi_m \longrightarrow \mathcal{L}(c,\alpha) \tag{A16}$$

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 equation (A12), the price impact $I_N(\phi)$ in the limit of large N is described by equation (A6) 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 equation (A9) 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 satisfy the condition in equation (A15) 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 equation (A4)

$$I_N(\phi) \simeq \frac{\phi}{2} \int d\phi_m p(\phi_m) |\phi_m|^{-1/2}$$
$$= \frac{\phi}{4\pi} \int d\lambda \hat{p}(\lambda) \int d\phi_m e^{-i\lambda\phi_m} |\phi_m|^{-1/2}, \qquad \text{(A17)}$$

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

$$\hat{p}(\lambda) = e^{-c(N-1)^{1/\alpha}|\lambda|^\alpha} \qquad \text{(A18)}$$

for large N and taking present that the last integral in equation (A17) is a known Fourier transform

$$\int_{-\infty}^{+\infty} d\phi_m e^{-i\lambda\phi_m} |\phi_m|^{-1/2} = \frac{\sqrt{2\pi}}{|\lambda|^{1/2}} \qquad \text{(A19)}$$

we obtain a linear price impact

$$I_N(\phi) \simeq \frac{\phi}{2\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{p}(\lambda)|\lambda|^{-1/2} d\lambda. \qquad \text{(A20)}$$

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 Roderick (2001) are satisfied, one can approximate the integral in equation (A20) as follows:

$$\int_{-\infty}^{+\infty} \hat{p}(\lambda)|\lambda|^{-1/2} d\lambda \sim \Gamma\left(\frac{1}{2\alpha}\right) \frac{1}{\alpha c^{1/(2\alpha)}(N-1)^{1/(2\alpha)}}. \quad \text{(A21)}$$

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

$$I_N(\phi) \simeq \frac{1}{2\sqrt{2\pi}} \Gamma\left(\frac{1}{2\alpha}\right) \frac{\phi}{\alpha[c(N-1)]^{1/(2\alpha)}}, \qquad \text{(A22)}$$

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

A.2. Market impact with correlated Gaussian metaorders

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

$$P(\boldsymbol{\varphi}_N) = \frac{1}{Z_N} \exp\left(-\frac{A_N}{2}\sum_{i=1}^N \phi_i^2 + \frac{B_N}{N}\sum_{i<j}^N \phi_i\phi_j + \mu\sum_{i=1}^N \phi_i\right),$$
$$\text{(A23)}$$

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

A.2.1. Calibration from data: means and correlations.
The first step to calibrate $P(\boldsymbol{\varphi}_N)$ in equation (A23) 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 $B_N > 0$):

$$\exp\left(\frac{B_N}{2N}\sum_{i,j}^N \phi_i\phi_j\right) = \int_{-\infty}^\infty \frac{dy}{\sqrt{2\pi/NB_N}} \exp$$
$$\times \left(-\frac{NB_N y^2}{2} + B_N\sum_{i=1}^N \phi_i y\right). \quad \text{(A24)}$$

This allows us to rewrite the probability distribution in equation (A23) as

$$P(\boldsymbol{\varphi}_N) = \frac{1}{Z_N}\sqrt{\frac{NB_N}{2\pi}} \int dy \prod_{i=1}^N \exp\left[-\frac{1}{2}\left(A_N + \frac{B_N}{N}\right)\phi_i^2\right.$$
$$\left. + (\mu + B_N y)\phi_i - \frac{NB_N}{2}y^2\right]. \qquad \text{(A25)}$$

The partition function then reads

$$Z_N = \left[\frac{2\pi}{\left(A_N + \frac{B_N}{N}\right)}\right]^{N/2} \sqrt{\frac{A_N N + B_N}{A_N N + B_N(N-1)}} \exp$$
$$\times \left[\frac{N^2\mu^2}{2(A_N N + B_N(N-1))}\right] \qquad \text{(A26)}$$

valid only for $B_N < A_N + B_N/N$. equation (A26) can be used to derive the following relations:

$$\frac{\partial \log Z_N}{\partial \mu} = N\mathbb{E}[\phi_i|N], \qquad \text{(A27)}$$

$$\frac{\partial \log Z_N}{\partial A_N} = -\frac{N}{2}\mathbb{E}[\phi_i^2|N], \qquad \text{(A28)}$$

$$\frac{\partial \log Z_N}{\partial B_N} = \left(\frac{N-1}{2}\right)\mathbb{E}[\phi_i\phi_j|N] \quad \text{with } (i \neq j) \qquad \text{(A29)}$$

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

$$\mathbb{E}[\phi_i^2|N] = \frac{A_N + 2B_N/N - B_N}{(A_N + B_N/N - B_N)(A_N + B_N/N)} \qquad \text{(A30)}$$

and

$$\mathbb{E}[\phi_i\phi_j|N] = \frac{B_N/N}{(A_N + B_N/N - B_N)(A_N + B_N/N)}. \qquad \text{(A31)}$$

Furthermore, combining equations (A30) and (A31) we can derive the volume correlation

$$\mathbb{C}_\phi(N) = \frac{\mathbb{E}[\phi_i\phi_j|N] - \mathbb{E}[\phi_i|N]^2}{\mathbb{E}[\phi_i^2|N] - \mathbb{E}[\phi_i|N]^2} = \frac{\mathbb{E}[\phi_i\phi_j|N]}{\mathbb{E}[\phi_i^2|N]}$$
$$= \frac{B_N/N}{A_N + 2B_N/N - B_N}. \qquad \text{(A32)}$$

Vice versa, from equations (A30) and (A32) we can obtain for the model parameters

$$A_N = \frac{1 - 2\mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N)}{(1 - \mathbb{C}_\phi(N))(1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N))\mathbb{E}[\phi_i^2|N]} \qquad \text{(A33)}$$

and

$$B_N = \frac{N\mathbb{C}_\phi(N)}{(1 - \mathbb{C}_\phi(N))(1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N))\mathbb{E}[\phi_i^2|N]}, \qquad \text{(A34)}$$

which are useful to fit the Gaussian model to data. We will use equations (A33) and (A34) 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 Hansen (1982).

A.3. *Analytical computation of market impact.*

To compute analytically the market impact function $I_N(\phi)$ from equation (A2) in the Gaussian correlated framework we adopt the following strategy:

(1) firstly we factorize the joint probability distribution $P(\boldsymbol{\varphi}_N)$ in equation (A23) 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 equation (A23) can be written in the following matrix form:

$$P(\boldsymbol{\varphi}_N) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}_N^T \mathbb{M}\boldsymbol{\varphi}_N + \boldsymbol{\mu}^T \boldsymbol{\varphi}_N\right), \quad (A35)$$

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. $\mathbb{O}^T\mathbb{M}\mathbb{O} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, the joint probability distribution $P(\boldsymbol{\varphi}_N)$ factorizes as

$$P(\tilde{\boldsymbol{\varphi}}_N) = \frac{1}{Z_N} \prod_{m=2}^{N} \exp\left[-\frac{\lambda_2}{2}\tilde{\phi}_m^2\right] \exp\left[-\frac{1}{2}\lambda_1\tilde{\phi}_1^2 + \mu\sqrt{N}\tilde{\phi}_1\right], \quad (A36)$$

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

$$\lambda_1 = A_N - (N-1)\frac{B_N}{N} = \frac{1}{\mathbb{E}[\phi_i^2|N](1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N))} \quad (A37)$$

and

$$\lambda_2 = \lambda_3 = \cdots = \lambda_N = A_N + \frac{B_N}{N} = \frac{1}{\mathbb{E}[\phi_i^2|N](1 - \mathbb{C}_\phi(N))} \quad (A38)$$

allow us to rewrite the partition function as

$$Z_N = \sqrt{\frac{2\pi}{\lambda_1}}\left[\sqrt{\frac{2\pi}{\lambda_2}}\right]^{N-1} \exp\left[\frac{N\mu^2}{2\lambda_1}\right]. \quad (A39)$$

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

$$\tilde{\phi}_1 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi_i, \quad (A40)$$

which put in evidence that the orthogonal basis change $\boldsymbol{\varphi}_N \rightarrow \tilde{\boldsymbol{\varphi}}_N$ 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 equation (A2) with N overall correlated Gaussian metaorders and in absence of an external field it is necessary to explicit the conditional probability distribution

$$P(\boldsymbol{\varphi}_N|\phi_k = \phi) = \frac{P(\phi_1, \ldots, \phi, \ldots, \phi_N)}{p(\phi)}, \quad (A41)$$

where $P(\phi_1, \ldots, \phi, \ldots, \phi_N)$ is given by equation (A23) setting $\mu = 0$ while the marginal one is equal to

$$p(\phi) = \frac{1}{\sqrt{2\pi\tilde{\lambda}_1/(\lambda_1\lambda_2)}}\exp\left[-\frac{\phi^2}{2}\frac{\lambda_1\lambda_2}{\tilde{\lambda}_1}\right], \quad (A42)$$

with λ_1 and λ_2 respectively given by equations (A37) and (A38) and

$$\tilde{\lambda}_1 = A_N - \frac{N-2}{N}B_N$$

$$= \frac{1}{\mathbb{E}[\phi_i^2|N][1 - \mathbb{C}_\phi(N) + N\mathbb{C}_\phi(N)][1 - \mathbb{C}_\phi(N)]}. \quad (A43)$$

It follows that the conditional probability distribution is equal to

$$P(\boldsymbol{\varphi}_N|\phi_k = \phi) = \frac{\exp\left[-\frac{(N-1)B_N^2\phi^2}{2N^2\tilde{\lambda}_1}\right]}{\underbrace{(2\pi/\tilde{\lambda}_1)^{1/2}(2\pi/\lambda_2)^{\frac{N-2}{2}}}_{\Theta_N^{-1}(\phi)}} \quad (A44)$$

$$\times \exp\left[-\frac{A_N}{2}\sum_{i\neq k}^{N}\phi_i^2 + \frac{B_N}{N}\sum_{\substack{i<j \\ i,j\neq k}}^{N}\phi_i\phi_j + \underbrace{\frac{\phi B_N}{N}}_{\tilde{\mu}}\sum_{i\neq k}^{N}\phi_i\right], \quad (A45)$$

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 equation (A2) is given solving the following conditional expectation:

$$I_N(\phi) = \mathbb{E}[\mathcal{I}(\boldsymbol{\varphi}_N)|\phi_k = \phi] \quad (A46)$$

$$= \int_{-\infty}^{\infty}\prod_{i\neq k}^{N}\mathrm{d}\phi_i\frac{P(\phi_1, \ldots, \phi, \ldots, \phi_N)}{p(\phi)}\left(\phi + \sum_{i\neq k}^{N}\phi_i\right)^{\bullet 1/2} \quad (A47)$$

$$= \frac{1}{\Theta_N(\phi)}\int_{-\infty}^{\infty}\prod_{i\neq k}^{N}\mathrm{d}\phi_i\exp$$

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

$$\times\left(\phi + \sum_{i\neq k}^{N}\phi_i\right)^{\bullet 1/2} \quad (A48)$$

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

$$\times\left(\phi + \sum_{i\neq k}^{N}\phi_i\right)^{\bullet 1/2}. \quad (A49)$$

Herein

- $\boldsymbol{\varphi}^* = \{\phi_i\}_{i=1,\ldots,N}^{i\neq k}$ 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 equation (A43) and $\lambda_m^* = \lambda_2$ as in equation (A38) for $m = 2, \ldots, N-1$.

As mentioned before, to solve equation (A49) it is useful to use the trick described in Step 1.: we apply in equation (A49) the orthogonal transformation $\tilde{\boldsymbol{\phi}}^* = \mathbb{H}\boldsymbol{\varphi}^*$ that diagonalizes the matrix $\mathbb{M}^*$ (i.e. $\mathbb{H}^T\mathbb{M}^*\mathbb{H} = \mathrm{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{N-1})$) and since the determinant of the Jacobian matrix associated to this transformation is equal to one, we

obtain that

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

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

$$\times (\phi + \sqrt{N-1}\tilde{\phi}_1^*)^{\bullet 1/2}; \tag{A50}$$

then integrating in $\tilde{\phi}_m^*$ for $m = 2, \ldots, N-1$, 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

$$I_N(\phi) = \frac{1}{\sqrt{2\pi(N-1)/\tilde{\lambda}_1}} \int_0^{\infty} dx\sqrt{x}$$

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

$$\left. - \exp\left[-\frac{\tilde{\lambda}_1}{2(N-1)}\left(x + \phi(1 + (N-1)\mathbb{C}_\phi(N))\right)^2\right]\right). \tag{A51}$$

REMARK 4 From equation (A51) 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 equations (A5) and (A6)) substituting

$$\phi \longrightarrow \phi[1 + (N-1)\mathbb{C}_\phi(N)] \tag{A52}$$

and

$$\Sigma_N^2 \longrightarrow 1/\tilde{\lambda}_1. \tag{A53}$$

A.4. Market impact with correlated signs and i.i.d. 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

$$P(\boldsymbol{\varphi}_N) = \mathcal{P}(\boldsymbol{\epsilon}_N) \prod_{i=1}^{N} p(|\phi_i|). \tag{A54}$$

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

$$\mathbb{P}(\epsilon_i|\tilde{\epsilon}) = \frac{1}{2}(1 + \gamma_\epsilon \epsilon_i \tilde{\epsilon}), \tag{A55}$$

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

$$\mathbb{E}[\rho_\epsilon|N] = \gamma_\epsilon^2. \tag{A56}$$

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

$$\mathbb{P}(\tilde{\epsilon}|\epsilon_k = +1) = \frac{1}{2}(1 + \gamma_\epsilon \tilde{\epsilon}) \tag{A57}$$

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

$$\mathbb{P}(\epsilon_i|\tilde{\epsilon}) = \frac{1}{2}(1 + \gamma_\epsilon \epsilon_i \tilde{\epsilon}). \tag{A58}$$

A.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 equation (A56).
(2) After fixing the direction $\epsilon_k = +1$ we simulate $N-1$ correlated signs using equation (A58).
(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 $\mathcal{I}(\boldsymbol{\varphi}_N) = (\sum_{i=1}^{N} \phi_i)^{\bullet 1/2}$ and $\phi_i = \epsilon_i|\phi_i|$, as defined in equation (6).
(5) Finally, we compute $I(\phi)$ averaging $I_N(\phi)$ over the empirical distribution $p(N)$ shown in the left panel of figure A11.

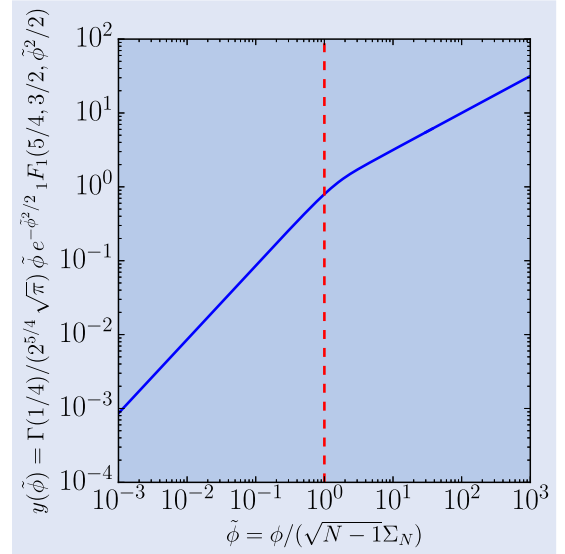

Figure A11. Rescaled market impact function $y(\tilde{\phi}) = I_N(\phi)/((N-1)\Sigma_N^2)^{1/4}$ in function of the dimensionless parameter $\tilde{\phi} := \phi/(\sqrt{N-1}\Sigma_N)$ with $I_N(\phi)$ given by equation (A6): the vertical dashed red line represents the transition from the linear impact (left side) to the square-root one (right side).