Quantitative Finance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Are trading invariants really invariant? Trading costs matter

Frédéric Bucci , Fabrizio Lillo , Jean-Philippe Bouchaud & Michael Benzaquen

To cite this article: Frédéric Bucci , Fabrizio Lillo , Jean-Philippe Bouchaud & Michael Benzaquen (2020) Are trading invariants really invariant? Trading costs matter, Quantitative Finance, 20:7, 1059-1068, DOI: 10.1080/14697688.2020.1741667

To link to this article: https://doi.org/10.1080/14697688.2020.1741667

Quantitative Finance, 2020
Vol. 20, No. 7, 1059–1068, https://doi.org/10.1080/14697688.2020.1741667

Are trading invariants really invariant? Trading costs matter

FRÉDÉRIC BUCCI†‡‡, FABRIZIO LILLO ‡§, JEAN-PHILIPPE BOUCHAUD¶‖‡‡ and MICHAEL BENZAQUEN *¶††‡‡

†Scuola Normale Superiore di Pisa, Piazza dei Cavalieri, Pisa 56126, Italy
‡Department of Mathematics, Università di Bologna, Piazza di Porta San Donato 5, Bologna 40126, Italy
§CADS, Human Technopole, Milan, Italy
¶Capital Fund Management, 23 rue de l'Université, Paris 75007, France
‖CFM-Imperial Institute of Quantitative Finance, Department of Mathematics, Imperial College, 180 Queen's Gate, SW7 2RH London, UK
††Ladhyx UMR CNRS 7646 & Department of Economics, Ecole Polytechnique, Palaiseau Cedex 91128, France
‡‡Chair of Econophysics and Complex Systems, Ecole polytechnique, Palaiseau Cedex 91128, France

(*Received 10 February 2019; accepted 6 March 2020; published online 7 April 2020*)

We revisit the trading invariance hypothesis recently proposed by Kyle, A.S. and Obizhaeva, A.A. ['Market microstructure invariance: Empirical hypotheses.' *Econometrica*, 2016, 84(4), 1345–1404] by empirically investigating a large dataset of metaorders provided by ANcerno. The hypothesis predicts that the quantity $I := \mathcal{W}/N^{3/2}$, where $\mathcal{W}$ is the daily exchanged risk (volatility $\times$ volume $\times$ price) and N is the daily number of metaorders, is invariant, either in distribution or in expectation. We find that the 3/2 scaling between $\mathcal{W}$ and N works well and is robust against changes of year, market capitalisation and economic sector. However our analysis shows that I is not invariant, and we find a very high correlation ($R^2 > 0.8$) between I and the trading cost (spread $+$ market impact costs) of the metaorder. Guided by these results we propose new invariants defined as a ratio of I to the aforementioned trading costs and find a large decrease in variance. We show that the small dispersion of the new invariants is mainly driven by (i) the scaling of the spread with the volatility per transaction, (ii) the near invariance, across stocks, of the shape of the distribution of metaorder size and of the volume and number of metaorders normalised to market volume and number of trades, respectively.

Keywords: Trading invariance; Metaorders; Trading costs; Stocks

1. Introduction

Finding universal scaling laws between trading variables is highly valuable to make progress in our understanding of financial markets and market microstructure. Indeed, statistical physics has taught us that scaling laws between physical variables most often reflect the dynamics of complex scale-invariant systems, by that giving precious insights about the mechanisms underlying the phenomena at hand. Benoit Mandelbrot was the first to propose the idea of scaling in a financial and economic context (Mandelbrot and Goldenfeld 1998), and ever since many have capitalised on his ideas, for a review see Gabaix (2009). Relevant to this study, examples of universal laws between trading variables include the square root impact law of meta-orders (Tóth *et al.* 2011) or the relation between spread and volatility per trade (Wyart *et al.* 2008). Recently, Kyle and Obizhaeva posited an intriguing *trading invariance principle* that must be valid for a *bet*, theoretically defined as a sequence of orders with a fixed direction (buy or sell) belonging to a single trading idea (Kyle and Obizhaeva 2016, 2017, 2019). This principle supports the existence of a universal invariant quantity I—expressed in dollars, independent of the asset and constant over time—which represents the average cost of a single *bet*, or *dollar risk transfer* in the words of the authors. In particular, taking the share price P (in dollars per share), the square daily volatility σ_d^2 (in %^{2} per day), the total daily amount traded with *bet*s V (shares per day) and the average volume of an individual *bet* Q (in shares) as relevant variables, dimensional analysis (see

*Corresponding author. Email: michael.benzaquen@polytech-nique.edu

e.g. Buckingham 1914, Pohl *et al.* 2017) suggests a relation of the form:

$$\frac{PQ}{I} = f\left(\sigma_{\mathrm{d}}^2 \frac{Q}{V}\right), \tag{1}$$

where f is a dimensionless function. Invoking the Modigliani–Miller capital structure irrelevance principle, which notably states that capital restructuring might impact the variables P and σ_d, while not affecting the product $P \times \sigma_{\mathrm{d}}$ and the other variables, yields $f(x) \sim x^{-1/2}$ (see Miller and Modigliani 1958, Kyle and Obizhaeva 2017). From equation (1) follows, up to a numerical factor, the 3/2-law:

$$I = \frac{\sigma_{\mathrm{d}} P Q^{3/2}}{V^{1/2}} := \frac{\mathcal{W}}{N^{3/2}}, \tag{2}$$

where $\mathcal{W} := \sigma_{\mathrm{d}} PV$ measures the total dollar amount of risk traded per day (also referred to as total exchanged risk or trading activity) while $N := V/Q$ represents the number of daily *bet*s for a given financial instrument. The above equation can be tested at different levels. First, I is a random variable associated to each day and stock. The original (strong form of) trading invariance states that I has invariant distribution across different stocks and time. A weaker, and more easily testable, form states that only the mean value of I is invariant. Clearly, if this second form is violated (as shown empirically below), *a fortiori* the stronger form is violated. A second, yet related, testing approach investigates the level of universality of the 3/2 scaling of equation (2). Specifically, Benzaquen *et al.* (2016) distinguishes: no universality (the 3/2-law holds for some financial instruments only), weak universality (the 3/2-law holds but with a non-universal value of I), and strong universality (the 3/2-law holds and I is constant across assets and time).†

Let us stress that identifying an elementary *bet* in the market is not a straightforward task. Theoretically, a bet is defined as a trading idea typically executed in the market as many trades over one or several days. As suggested by Kyle and Obizhaeva in their original work (Kyle and Obizhaeva 2016), metaorders, i.e. a bundle of orders corresponding to a single trading decision typically traded incrementally through a sequence of child orders, can be considered a proxy of these *bets*.‡ Beyond the subtleties in the bet's definition, there has been in the past few years empirical evidence that the scaling law discussed above matches patterns in financial data, at least approximately. The 3/2-law was empirically confirmed by Kyle and Obizhaeva using portfolio transition data related to rebalancing decisions made by institutional investors and executed by brokers (Kyle and Obizhaeva 2016). Andersen *et al.* (2016) reformulated suitably the *trading invariance hypothesis* at the single-trade level and showed that the equivalent version of equation (1) in such a setting holds remarkably well using public trade-by-trade data relative to the E-mini S&P 500 futures contracts. Benzaquen *et al.* (2016) substantially extended these empirical results showing that the 3/2-law holds very precisely across 12 futures contracts and 300 single US stocks, and across a wide range of time scales. Amongst others, Bowe *et al.* (2017) examined market microstructure invariance relationships for equity markets using a subset of 25 equities from the FTSE 100 stocks trading on the London Stock Exchange, and Pohl *et al.* (2018) provided additional empirical evidence that the intriguing 3/2-law holds on trades data from the NASDAQ stock exchange. See also Kyle *et al.* (2016), Bae *et al.* (2016).

Notwithstanding, empirical data at the single transaction scale—see in particular (Benzaquen *et al.* 2016)—revealed that while the 3/2-law is very robust, the invariant I is actually quite far from invariant, as it varies from one asset to the other and across time, thus in favour of the *weak universality* degree. Note that this is consistent with the idea that a universal invariant with dollar units would be quite incongruous, given that the value of the dollar is itself stochastically time-dependent.§ Benzaquen *et al.* (2016) showed that a more suitable candidate for an invariant was actually the dimensionless $\mathcal{I} := I/\mathcal{C}$ where $\mathcal{C}$ denotes the spread trading costs.

Yet, single transactions are typically not the same as single bets. Large and medium sized orders are typically split in multiple transactions and traded incrementally over long periods of time. Public market data do not allow to infer the trading decision and to link different transactions to a single execution.¶ In order to test the trading invariance hypothesis at the metaorder level and its relation with trading costs, it is necessary to have a dataset of market-wide (i.e. not from a single institution) metaorders.

This is precisely the aim of the present paper, which leverages on a heterogeneous dataset of metaorders extracted from the ANcerno database.‖ Although from a preliminary research Kyle and Fong found that proxies for *bet*s in ANcerno data have size patterns consistent with the proposed invariance hypothesis (Kyle and Obizhaeva 2016, 2017), to our knowledge such a thorough analysis at the *metaorder* level for a wide range of assets is still lacking.

Our main finding is that, while the scaling law $\mathcal{W} \sim N^{3/2}$ works surprisingly well independently of the chosen asset, the quantity I is not invariant, as pointed out in Benzaquen *et al.* (2016) at the trade-by-trade level. In other words, for a given asset the 3/2-law equation (2) holds, but the invariance principle implying that I is the same for all assets does not. We show that the latter quantity is strongly correlated with transaction costs, including spread and impact costs.

† Note that here we only explore the daily level, time does not mean the same thing as in Benzaquen *et al.* (2016) where we varied the time intervals over which the variables were computed.

‡ In the following we will make use of such an approximation and use the words 'metaorder' instead of 'bet'.

§ Note that Kyle and Obizhaeva commented on how to modify their invariance principle in an international context. In particular they suggested that 'invariance relationships can also be applied to an international context in which markets have different currencies or different real exchange rates' by scaling to 'the nominal cost of financial services calculated from the productivity-adjusted wages of finance professionals in the local currency of the given market during the given time period' (Kyle and Obizhaeva 2017).

¶ In fact, for example, Kyle and Obizhaeva tackled this problem investigating a proprietary dataset of portfolio transitions (Kyle and Obizhaeva 2016).

‖ 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. In Kyle and Obizhaeva (2016) the authors claim that the ANcerno database includes more orders than the data set of portfolio transitions they used in their work.

This leads us to introduce new invariants, obtained by dividing I by the trading costs, and which appear to fluctuate very little across stocks. Finally we show that the observed small dispersion of the new invariants can be connected with three microstructural properties: (i) the linear relation between spread and volatility per transaction; (ii) the near invariance of the metaorder size distribution, and (iii) of the total volume and number fractions of the bets across different stocks.

The paper is organised as follows. In Section 2 we describe the dataset collecting trading decisions of institutional investors operating in the US equity market. In Section 3 we show that the 3/2-law holds surprisingly well at the daily level independently of the time period, of the market capitalisation and of the economic sector. In Section 4 we compute the invariant I and we argue in favour of *weak universality*. We propose a more natural definition for a trading invariant that accounts both for the spread and the market impact costs; and we exhibit the microstructural origin of its small dispersion. Some conclusions and open questions are presented in Section 5.

2. Data

Our analysis relies on a database made available by ANcerno, a leading transaction-cost analysis provider (`www.ancerno.com`). Our dataset counts heterogeneous institutional investors placing large buy or sell orders executed by a broker as a succession of smaller orders belonging to the same trading decision of a single investor (for full details see Zarinelli *et al.* 2015, Bucci *et al.* 2019, Bucci *et al.* 2020). Our sample includes the period January 2007–June 2010 for a total of 880 trading days. Only metaorders completed within at most a single trading day are held. Further, we select stocks belonging to the Russell 3000 index, thereby retaining $\sim$ 8 million metaorders distributed quite uniformly in time and representing $\sim 5\%$ of the total reported market volume, regardless of market capitalisation (large, mid and small) and economical sectors (basic materials, communications, consumer cyclical and non-cyclical, energy, financial, industrial, technology and utilities). As can be seen in Bucci *et al.* (2019), Bucci *et al.* (2020) and Zarinelli *et al.* (2015), which use very similar filtering of the dataset, the distribution of metaorder duration, traded volume, and participation rate are very heterogeneous, spanning several orders of magnitude. When considering the number N of metaorders of a stock traded in a day, left panel of the figure in Appendix 1 shows a quite heterogeneous distribution spanning from 1 to 10^2 and on average approximately five metaorders are executed per day in each asset.

3. The 3/2-law

Here we investigate the trading invariance hypothesis at the daily level. The daily timescale choice avoids an elaborate analysis of when precisely each metaorder starts and ends, thereby averaging out all the non-trivial problems related to the daily simultaneous metaorders executed on the same asset (Zarinelli *et al.* 2015).

3.1. Exchanged risk

From the metaorders executed on the same stock during the same day we compute the total exchanged volume in dollars: $\sum_{i=1}^{N} P_i Q_i$, where N is the number of daily metaorders per asset in the ANcerno database, Q_i and P_i are respectively the number of shares and the volume weighted average price (vwap) of the ith available metaorder. We then define the total daily exchanged ANcerno risk per asset as:

$$\mathcal{W} := \sum_{i=1}^{N} \mathcal{W}_i, \quad \text{with} \quad \mathcal{W}_i = \sigma_{\mathrm{d}} Q_i P_i, \tag{3}$$

and where σ_{d} denotes the daily volatility per asset, computed as $\sigma_{\mathrm{d}} = (P_{\mathrm{high}} - P_{\mathrm{low}})/P_{\mathrm{open}}$ from the high, low, and open daily prices only.† The statistical properties of the bets, in terms of their associated risk $\mathcal{W}_i$ and of their total daily number N per asset are discussed in Appendix 1. The distribution of $\mathcal{W}_i$ and $\mathcal{W}$ span almost eight orders of magnitudes. This is important because a careful testing of the scaling relation predicted by the trading invariance hypothesis requires large variability of the considered variables. Thus the ANcerno database is ideal for this testing exercise.

3.2. Empirical evidence

We introduce the mean daily exchanged risk conditional to the number of metaorders N per day and asset, $\mathbb{E}[\mathcal{W}|N]$. This is empirically estimated by the quantity:

$$\langle \mathcal{W} \rangle_N := \frac{\sum_{j:N^{(j)}=N} \mathcal{W}^{(j)}}{\sum_{j:N^{(j)}=N} 1}, \tag{4}$$

where for each day j the total daily exchanged risk is given by $\mathcal{W}^{(j)} := \sum_{i=1}^{N(j)} \mathcal{W}_i^{(j)}$ with $N^{(j)}$ the number of daily metaorders per asset and $\mathcal{W}_i^{(j)} = \sigma_d^{(j)} Q_i^{(j)} P_i^{(j)}$. To test the 3/2-law we bin the data (one observation per stock per day) depending on N and we plot it against $\langle \mathcal{W} \rangle_N$ in log-log scale. We consider different subsets of stocks, depending on market cap, economic sector, investigated period, and we perform the linear regression of $\log\langle \mathcal{W} \rangle_N$ versus $\log N$.

As shown in the first three panels of figure 1 the scaling $\langle \mathcal{W} \rangle_N \sim N^{3/2}$ holds well independently of the conditioning to market capitalisation, economic sector, and time period. The insets show the estimated exponent which always quite close to 3/2. Slight deviations may have different origins but can mostly be attributed to the heterogeneous sample's composition in terms of stocks for each bucket in N. The 3/2-law is

† We checked that the results discussed in the present work are still valid using other definitions of the daily volatility and of the price in analogy to what done for example in Kyle and Obizhaeva (2016). Specifically, the results are still valid when computing σ_{d} with the Rogers-Satchell volatility estimator (Rogers and Satchell 1991, Benzaquen *et al.* 2016) or as the monthly averaged daily volatility, i.e. $\bar{\sigma}_{\mathrm{d}} = \sum_{m=1}^{25} \sigma_{d,m}$ and/or defining the price P_i as the closing price of the day before the metaorder's execution.

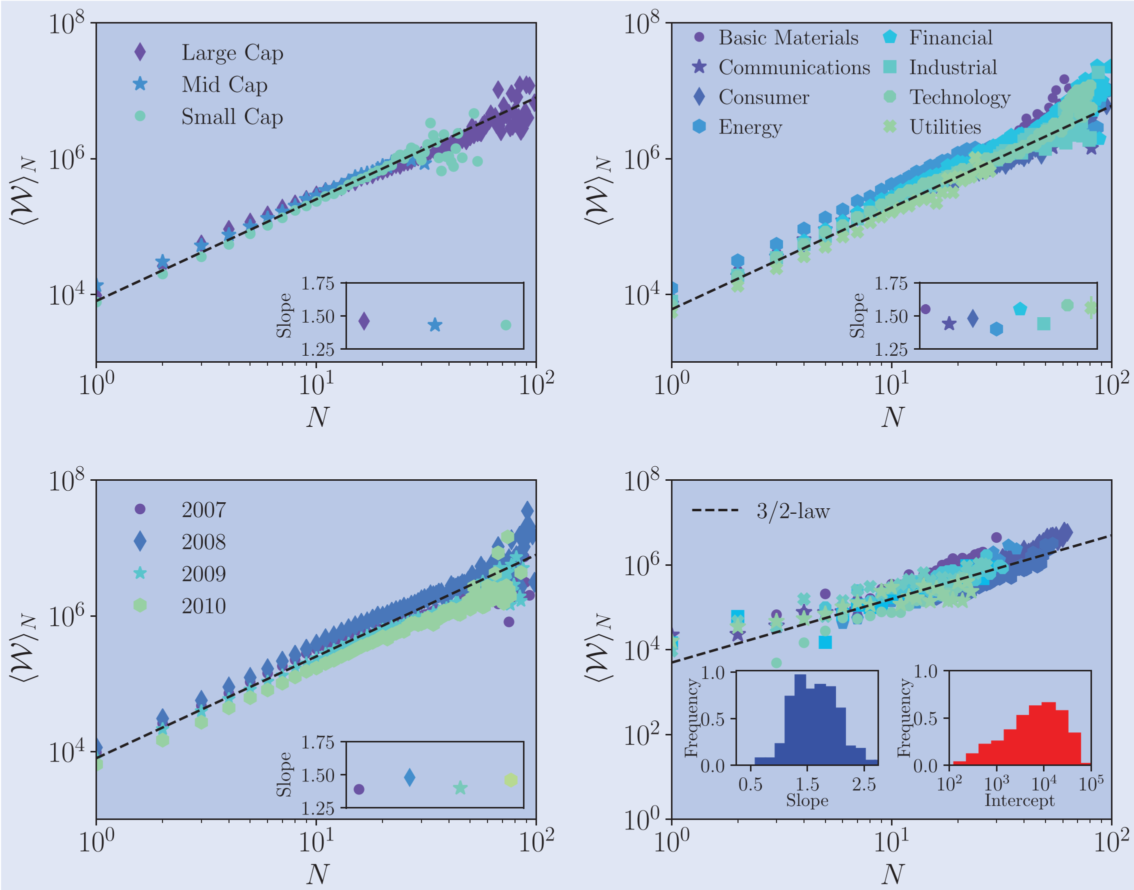

Figure 1. Mean daily exchanged risk $\langle \mathcal{W} \rangle_N$ conditional on the daily number N of metaorders per asset for different market capitalisation (top left panel), economic sector (top right panel), and time period (bottom left panel). The insets show the slopes obtained from linear regression of the data, firstly averaged with respect to N and secondly log-transformed. The bottom right panel shows a plot of $\langle \mathcal{W} \rangle_N$ as function of N for a subset of 10 stocks chosen randomly from the pool of around three thousand US stocks: the two insets represent respectively the empirical distribution of the slopes and of the y-intercept obtained from linear regression of a larger sub-sample of 200 stocks randomly chosen, firstly averaged with respect to N and secondly log-transformed considering each stock separately.

also valid for individual stocks, as shown in the bottom right panel of figure 1, where data from 10 randomly chosen stocks are displayed. We perform the above regression on each of the approximately three thousand stocks and the histogram of the slopes (exponents), shown in the bottom left inset is well centred around the 3/2-value. This shows that the 3/2 exponent works very well in describing the scaling relation between $\langle \mathcal{W} \rangle_N \sim N^{3/2}$ and N.

The bottom right inset in the bottom right panel of figure 1 shows the histogram of the intercept $\langle I \rangle$ of the regression $\log\langle \mathcal{W} \rangle = \log\langle I \rangle + \beta \log N$ of the fits for individual stocks of the binned data as shown in the main panel. It is evident that there is a very large dispersion (the abscissa is in log scale), which indicates that I in equation (2) is not constant across different stocks. More empirical insights on the origin of the 3/2-law are presented in Appendix 2.

4. The trading invariant

The conjecture that the quantity $\langle I \rangle$ is invariant across different contracts is clearly rejected by the empirical analysis performed in the previous section. Indeed, the quantity $\langle I \rangle$ varies by at least one order of magnitude across different stocks. This result goes against the *strong universality* version of the trading invariance hypothesis which states that both the average value $\langle I \rangle$ and the full probability distribution of $I = \mathcal{W}/N^{3/2}$ should be invariant across products. Dimensionally I is a cost (i.e. it is measured in dollars) and indeed the trading invariance hypothesis posits that the cost of a metaorder is invariant. Using the identification of metaorders and bets we can use the ANcerno dataset to estimate the trading cost, including a spread and a market impact component. Below, we show that I and the trading costs are highly correlated, and therefore propose new invariants based on their ratio.

4.1. *Trading costs and trading invariants*

Trading costs are typically divided into fees/commissions, spread, and market impact. For large orders, like those investigated here, fees/commissions typically account for a very small fraction and therefore we will neglect them. We shall however take into consideration both the spread cost (as was done at the single-trade level in Benzaquen *et al.* 2016) and the market impact cost as computed from the well established *square root law* (see e.g. Torre and Ferrari 1998, Almgren *et al.* 2005, Engle *et al.* 2006, Tóth *et al.* 2011, Zarinelli *et*

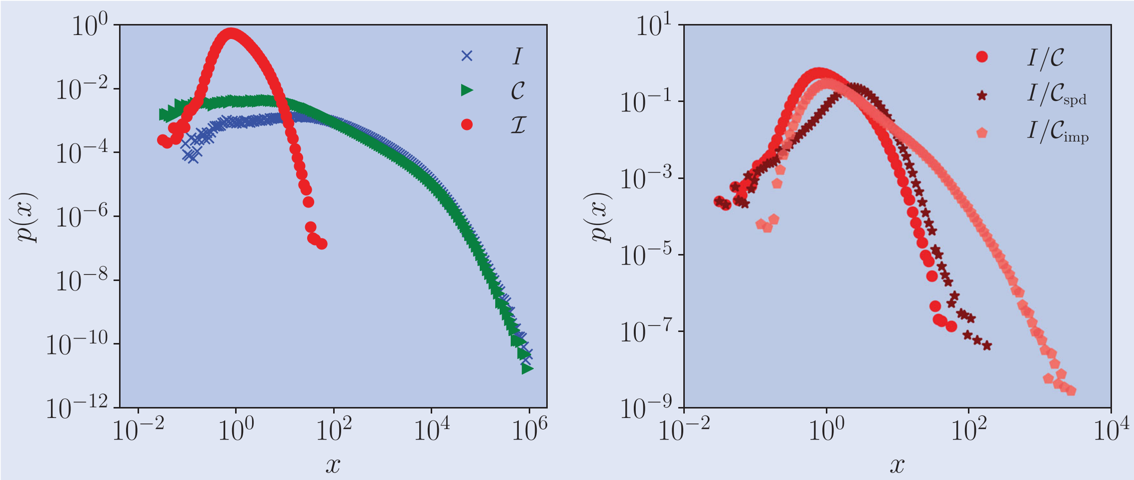

Figure 2. (Left) Empirical distributions of the KO invariant $I = \mathcal{W}/N^{3/2}$, of the daily average metaorder's total trading cost $\mathcal{C}$ (using $Y_{\text{spd}} = 3.5$ and $Y_{\text{imp}} = 1.5$ in equation (5)), and of the dimensionless invariant $\mathcal{I} := I/\mathcal{C}$. (Right) Empirical distributions in log-log scale of the KO invariant I rescaled respectively by the total daily average cost $\mathcal{C}$, by the spread cost $\mathcal{C}_{\text{spd}}$ and by the market impact cost $\mathcal{C}_{\text{imp}}$.

al. 2015, Brokmann *et al.* 2015, Bucci *et al.* 2019, Bucci *et al.* 2020), which states that the price impact of a metaorder of size Q is

$$\Delta \log P \propto \epsilon \sigma_d \sqrt{\frac{Q}{V_d}},$$

where $\epsilon = \pm 1$ is the sign of the trade and V_{d} is the total daily market volume. We thus define the average daily metaorder's trading cost as:

$$\mathcal{C} = \mathcal{C}_{\text{spd}} + \mathcal{C}_{\text{imp}} = Y_{\text{spd}} \times \frac{1}{N} \sum_{i=1}^{N} SQ_i + Y_{\text{imp}} \times \frac{1}{N} \sum_{i=1}^{N} \sigma_{\text{d}} Q_i P_i \times \sqrt{\frac{Q_i}{V_{\text{d}}}} := Y_{\text{spd}} \times \mathcal{C}^0_{\text{spd}} + Y_{\text{imp}} \times \mathcal{C}^0_{\text{imp}}, \tag{5}$$

with S the average daily spread,† and $Y_{\text{spd}}, Y_{\text{imp}}$ two constants to be determined. The factor Y_{spd} depends, among other things, on the fraction of trades of the metaorder executed with market orders, whereas Y_{imp} only weakly depends on the execution algorithm and is typically estimated to be close to unity (Tóth *et al.* 2011, Zarinelli *et al.* 2015, Bouchaud *et al.* 2018). Thus, while $\mathcal{C}_{\text{imp}}$ is a quite faithful estimation of the impact cost of the metaorders in a day and stock, $\mathcal{C}_{\text{spd}}$ is an upper bound, reached if all the considered metaorders are executed with market orders.

The empirical properties of $\mathcal{C}^0_{\text{spd}}$ and $\mathcal{C}^0_{\text{imp}}$ and the relative importance of the two terms as a function of the metaorder size are presented in Appendix 3. As expected, at the single metaorder level, spread cost is dominant for small metaorders, while impact cost is dominant for large ones. At the aggregated level, the average daily market impact cost $\mathcal{C}_{\text{imp}}$ accounts on average for approximately half of the total daily trading average cost.

To determine Y_{spd} and Y_{imp} we perform an ordinary least square regression of the KO invariant I with respect to the daily average cost $\mathcal{C}$ defined for each asset by equation (5). We obtain $Y_{\text{spd}} \simeq 3.5 \pm 0.2$, $Y_{\text{imp}} \simeq 1.5 \pm 0.1$ and a coefficient of determination $R^2 \simeq 0.8$. These results show that the original KO invariant is indeed strongly correlated with the trading cost. Since these costs have no a priori reason to be universal, this explains why I is not invariant.

Guided by such results and by the fact that a market microstructure invariant, if any, should be dimensionless, we define new invariants by dividing the original KO invariant I by the cost of trading. Therefore, we consider three different specifications, namely:

$$\mathcal{I} = \frac{I}{\mathcal{C}}, \quad \mathcal{I}_{\text{spd}} = \frac{I}{\mathcal{C}_{\text{spd}}}, \quad \mathcal{I}_{\text{imp}} = \frac{I}{\mathcal{C}_{\text{imp}}}. \tag{6}$$

The left panel of figure 2 shows the empirical distribution of the original KO invariant I together with that of $\mathcal{I}$, and of the cost $\mathcal{C}$. It is visually quite clear that rescaling by the cost dramatically reduces the dispersion, and that the distribution of I is very similar to that of $\mathcal{C}$, despite some deviation for small values. The right panel compares the distribution of $\mathcal{I}$ with that of the other two new invariants. A quantitative comparison is provided in table 1, which reports the mean, the standard deviation, the coefficient of variation‡ (CV) of I and of the three new invariants. It is clear that, due to the correlation between I and $\mathcal{C}$, the new invariants $\mathcal{I}_{\text{spd}}$ and $\mathcal{I}_{\text{imp}}$ have a much smaller CV than I. Since the distributions have clear fat tails, we also implemented the Gini coefficient, as in Pohl *et al.* (2018). The table indicates that also in this case the new invariants are much more peaked than I.

4.2. *Origin of the small dispersion of the new invariants*

Here we investigate the origin of the small dispersion of the new invariants. Let us first consider the market impact cost normalisation only and rewrite $\mathcal{I}_{\text{imp}}$ with the approximation of $P_i \simeq P$ for all the metaorders executed in a day and on the

† The daily spread is not provided in the ANcerno dataset. We computed it as the time average spread across the day using publicly available market data.

‡ The coefficient of variation is the ratio of standard deviation and mean, an indicator of distribution 'peakedness'.

Table 1. Statistics of the different invariants, namely the original KO invariant I (left), and the three new ones rescaled by cost (right).

	$I \cdot 10^3$ (\$)	$\mathcal{I}$	$\mathcal{I}_{\text{spd}}$	$\mathcal{I}_{\text{imp}}$
Mean	6.33	2.20	4.70	7.8
Std. dev.	11	1.84	3.11	12.2
CV	1.74	0.84	0.66	1.56
Gini Coefficient	0.77	0.33	0.39	0.65

Note: CV stands for coefficient of variation.

same stock as:

$$\mathcal{I}_{\text{imp}} = \frac{N\sum_{i=1}^{N}\sigma_{\text{d}}P_iQ_i}{Y_{\text{imp}}N^{3/2}(\sigma_{\text{d}}\sum_{i=1}^{N}P_iQ_i\sqrt{Q_i/V_{\text{d}}})} = \frac{1}{Y_{\text{imp}}\sqrt{\eta}}\frac{[Q]^{3/2}}{[Q^{3/2}]} = \frac{1}{Y_{\text{imp}}m\sqrt{\eta}}, \tag{7}$$

where $\eta := V/V_{\text{d}}$ with $V := \sum_{i=1}^{N} Q_i$ the total ANcerno metaorder volume, [$\bullet$] a daily average operation per stock, i.e. $[x] := 1/N\sum_{i=1}^{N} x_i$, and $m > 1$ the normalised 3/2th moment of the number of shares of a metaorder, which depends on the *shape* of the distribution of metaorder size. We have checked that m as well as η are, to a first approximation, independent of the stock (see left and central panels in figure 3) indicating that the distribution of metaorder size is, to a large degree, universal and that the ANcerno database is representative of the trading across all stocks. These observations explains why $\mathcal{I}_{\text{imp}}$ is also, to a large degree, stock independent.

For the total cost normalisation, our understanding of the invariance property relies on the following empirical fact. The average spread is proportional to the volatility per trade, that is $S = cP\sigma_{\text{d}}/\sqrt{N_{\text{d}}}$, where N_{d} is the total number of daily transactions per asset and c is a stock independent numerical constant, see Bouchaud *et al.* (2018), Wyart *et al.* (2008). Indeed, the above arguments taken together show that the dimensionless quantity $\mathcal{I}$ can be written as:

$$\mathcal{I} = \frac{1}{Y_{\text{spd}}c\sqrt{\xi} + Y_{\text{imp}}m\sqrt{\eta}}, \tag{8}$$

where $\xi := N/N_{\text{d}}$ is found to be stock independent (see right panel in figure 3). Therefore $\mathcal{I}$ is also stock independent. Finally, the fact that the CV of $\mathcal{I}$ is less than both that of $\mathcal{I}_{\text{spd}}$ and $\mathcal{I}_{\text{imp}}$ suggests that KO's invariant is commensurate to the *total* cost of trading, including both the spread cost and the impact cost.

5. Conclusions

In this work we empirically investigated the *market microstructure invariance hypothesis* recently proposed by Kyle and Obizhaeva (2016, 2017). Their conjecture is that the expected dollar cost of executing a *bet* is constant across assets and time. The ANcerno dataset provides a unique laboratory to test this intriguing hypothesis through its available metaorders which can be treated as a proxy for *bet*s, i.e. a decision to buy or sell a quantity of institutional size generated by a specific trading idea. Let us summarise what we have achieved in this paper:

- Using metaorders issued for around three thousand stocks, we showed that, at the daily timescale interval, the 3/2 scaling law between exchanged risk $\mathcal{W}$ and number of bets, notably resulting from Kyle and Obizhaeva's invariance principle, is observed independently of the year, the economic sector and the market capitalisation.
- The trading invariant $I := \mathcal{W}/N^{3/2}$ proposed by Kyle and Obizhaeva is non-universal: both its average value $\langle I\rangle$ and distribution clearly depend on the considered stocks, in favour of a *weak universality* interpretation. Furthermore, this quantity has dollar units which makes its hypothesised invariance rather implausible.
- On the basis of dimensional and empirical arguments, we propose a dimensionless invariant defined as a ratio of I and of the metaorder's total cost, which includes both spread and market impact costs. We find a variance reduction of more than 50%, qualitatively traceable to the proportionality between spread and volatility per trade, and the near invariance of the distributions of metaorder size, of the volume fraction and number fraction of metaorders across stocks.

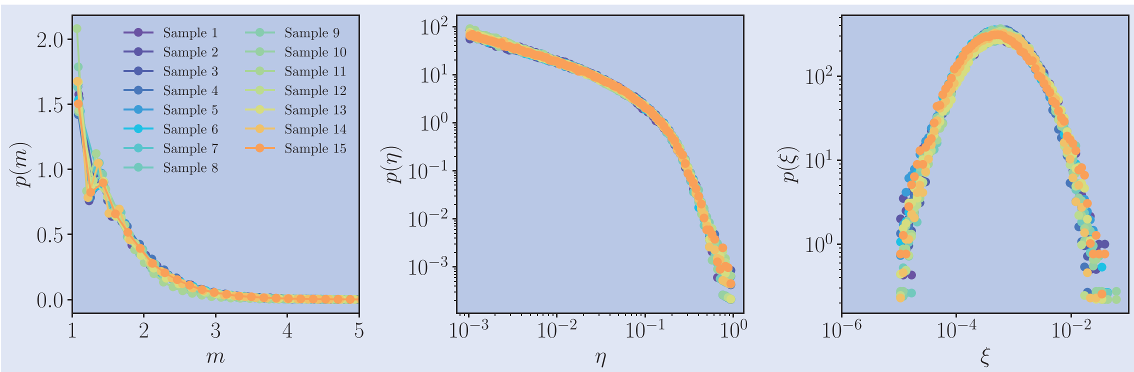

Figure 3. Empirical distribution of the ratio $m = [Q^{3/2}]/[Q]^{3/2}$ (left panel), $\eta = V/V_{\text{d}}$ (central panel) and $\xi = N/N_{\text{d}}$ (right panel), all three computed at the daily level for each asset: we randomly group the stocks in equally sized samples and for each of them we compute the empirical distribution respectively of m, η, and ξ finding that they are, to a first approximation, stock independent.

Our empirical analysis has allowed to show that the trading invariance hypothesis holds at the metaorder level in a strong sense provided one considers the exchanged risk *and* the total trading cost of the metaorders. This is in the spirit of Kyle and Obizhaeva's arguments, but takes into account the fact that transaction costs are both asset and epoch dependent. As anticipated in Benzaquen *et al.* (2016), our results strongly suggest that trading 'invariance' is a consequence of the validity of the square root law for market impact as well as to the proportionality between spread and volatility as discussed in Benzaquen *et al.* (2016), Wyart *et al.* (2008), Madhavan (1997). It would actually be quite interesting to investigate other markets such as bond markets, currency markets or futures markets, for which the Modigliani–Miller theorem is totally irrelevant, while trading invariance still holds—at least at the level of single trades, see Benzaquen *et al.* (2016), Andersen *et al.* (2016). Finally, note that differences in market structure across countries, such as execution mechanisms, fees and regulations could also challenge the validity of the results presented here.

Acknowledgments

We thank Alexios Beveratos, Laurent Erreca, Antoine Fosset, Charles-Albert Lehalle and Amine Raboun for fruitful discussions.

Data availability statement

The data were purchased from the company ANcerno Ltd (formerly the Abel Noser Corporation) which 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. 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was conducted within the Econophysics & Complex Systems Research Chair, under the aegis of the Fondation du Risque, the Fondation de l'Ecole polytechnique, the Ecole polytechnique and Capital Fund Management.

ORCID

Fabrizio Lillo http://orcid.org/0000-0002-4931-4057
Michael Benzaquen http://orcid.org/0000-0002-9751-7625

References

Almgren, R., Thum, C., Hauptmann, E. and Li, H., Direct estimation of equity market impact. *Risk*, 2005, **18**(7), 5862.
Andersen, T.G., Bondarenko, O., Kyle, A.S. and Obizhaeva, A.A., Intraday trading invariance in the E-mini S&P 500 futures market, 2016. Available online at SSRN: https://ssrn.com/abstract=2693810.
Bae, K.-h., Kyle, A.S., Lee, E.J. and Obizhaeva, A.A., Invariance of buy-sell switching points. Robert H. Smith School Research Paper No. RHS, 2730770, 2016.
Benzaquen, M., Donier, J. and Bouchaud, J.-P., Unravelling the trading invariance hypothesis. *Market Microstruct. Liquid.*, 2016, **2**(03n04), 1650009.
Bouchaud, J.-P., Bonart, J., Donier, J. and Gould, M., *Trades, Quotes and Prices Financial Markets under the Microscope*, 2018 (Cambridge University Press: Cambridge).
Bowe, M., Rizopoulos, E. and Sarah Zhang, S., Market microstructure invariance in the FTSE 100, 2017. Available online at: http://www.fmaconferences.org/Norway/Papers/MMI_Paper.pdf.
Brokmann, X., Serie, E., Kockelkoren, J. and Bouchaud, J.-P., Slow decay of impact in equity markets. *Market Microstruct. Liquid.*, 2015, **1**(02), 1550007.
Bucci, F., Benzaquen, M., Lillo, F. and Bouchaud, J.-P., Crossover from linear to square-root market impact. *Phys. Rev. Lett.*, 2019, **122**(10), 108302.
Bucci, F., Mastromatteo, I., Eisler, Z., Lillo, F., Bouchaud, J.-P. and Lehalle, C.-A., Co-impact: Crowding effects in institutional trading activity. *Quant. Finance*, 2020, **20**, 193–205.
Buckingham, E., On physically similar systems; Illustrations of the use of dimensional equations. *Phys. Rev.*, 1914, **4**(4), 345.
Engle, R., Ferstenberg, R. and Russell, J., Measuring and modeling execution cost and risk. Chicago GSB Research Paper, no. 08–09, 2006.
Gabaix, X., Power laws in economics and finance. *Annu. Rev. Econ.*, 2009, **1**(1), 255–294.
Jones, C.M., Kaul, G. and Lipson, M.L., Transactions, volume, and volatility. *Rev. Financ. Stud.*, 1994, **7**(4), 631–651.
Kyle, A.S. and Obizhaeva, A.A., Market microstructure invariance: Empirical hypotheses. *Econometrica*, 2016, **84**(4), 1345–1404.
Kyle, A.S. and Obizhaeva, A.A., Dimensional analysis, leverage neutrality, and market microstructure invariance, 2017. Available online at: SSRN: https://ssrn.com/abstract=2785559.
Kyle, A.S. and Obizhaeva, A.A., Market microstructure invariance: A dynamic equilibrium model, 2019.
Kyle, A.S., Obizhaeva, A.A. and Tuzun, T., Microstructure invariance in US stock market trades, 2016. Available online at SSRN: https://ssrn.com/abstract=1107875.
Madhavan, A., Richardson, M. and Roomans, M.. Why do security prices change? A transaction-level analysis of NYSE stocks. *Rev. Financ. Stud.*, 1997, **10**(4), 1035–1064.
Mandelbrot, B.B. and Goldenfeld, N., Fractals and scaling in finance: Discontinuity, concentration, risk. *Phys. Today.*, 1998, **51**, 59.
Miller, M. and Modigliani, F., The cost of capital. Corporate finance and the theory of investment. *Am. Econ. Rev.*, 1958, **48**, 261–297.
Pohl, M., Ristig, A., Schachermayer, W. and Tangpi, L., The amazing power of dimensional analysis: Quantifying market impact. *Market Microstruct. Liquid.*, 2017, **3**(03n04), 1850004.
Pohl, M., Ristig, A., Schachermayer, W. and Tangpi, L., Theoretical and empirical analysis of trading activity, 2018. arXiv preprint arXiv:1803.04892.
Rogers, L.C.G. and Satchell, S.E., Estimating variance from high, low and closing prices. *Ann. Appl. Probab.*, **1**, 504–512.
Torre, N.G. and Ferrari, M.J., The market impact model. *Horizons, The Barra Newsletter*, 165, 1998.
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**(2), 021006.

Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M. and Vettorazzo, M., Relation between bid–ask spread, impact and volatility in order-driven markets. *Quant. Finance*, 2008, **8**(1), 41–57.
Zarinelli, E., Treccani, M., Doyne Farmer, J. and Lillo, F., Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstruct. Liquid.*, 2015, **1**(02), 1550004.

Appendices

Appendix 1. Statistics of metaorder sample

Here we describe some statistics of the metaorders executed from the main investments funds and brokerage firms gathered by ANcerno. The empirical probability distribution of the number of metaorders N per asset, of the risk $\mathcal{W}_i$ exchanged by a metaorder and of the total daily traded risk $\mathcal{W}$ per asset are illustrated in figure A1. It emerges that both the number of daily metaorders N and the risk measures typically vary over several orders of magnitude. In particular, as evident from the left panel in figure A1, there is a significant number of metaorders active every day, since in average ~ 5 metaorders are executed per day for each asset. Furthermore, as shown in the right panel of figure A1, both the single metaorder's risk $\mathcal{W}_i$ and the total daily exchanged risk $\mathcal{W}$ vary over almost eight decades. Note that these statistical properties are approximately independent from the time period and from the economical sector of the asset exchanged through metaorders.

Appendix 2. The 3/2-law under the microscope

One may rightfully wonder whether it is possible to understand the 3/2-law from the statistical properties of the metaorders. To this purpose we start by investigating the individual metaorder's risk $\mathcal{W}_i$ distribution properties as a function of N. We find that when rescaling the metaorder's risk $\mathcal{W}_i$ by the square root of the number N of daily metaorders per asset one obtains a conditional cumulative distribution $\mathcal{P}(\mathcal{W}_i/\sqrt{N}|N)$ dependent on N but with a mean $\langle \mathcal{W}_i/\sqrt{N}\rangle_N$ invariant on N (see figure A2).† It emerges then that the conditional average metaorder risk $\mathcal{W}_i$ can be predicted from the number N of daily metaorders per asset since $\langle \mathcal{W}_i\rangle_N$ scales as N^{γ} with $\gamma \simeq 0.5$, that is $\langle \mathcal{W}_i\rangle_N \sim \sqrt{N}$.‡ It immediately follows that combining this empirical result and the linearity property of the mean, one recovers the 3/2-law $\langle \mathcal{W}\rangle_N \sim N^{3/2}$, since:

$$\langle \mathcal{W}\rangle_N = \langle \sum_{i=1}^{N} \mathcal{W}_i\rangle_N = \sum_{i=1}^{N} \langle \mathcal{W}_i\rangle_N = N\, \langle \mathcal{W}_i\rangle_N \sim N\sqrt{N} = N^{3/2}. \tag{A1}$$

To explain the scaling $\langle \mathcal{W}_i\rangle_N \sim \sqrt{N}$ through the product $\langle \sigma_\mathrm{d}\rangle_N \times \langle Q_iP_i\rangle_N$ we need to check for the correlation between the daily volatility σ_d and the volume in dollars Q_iP_i of a metaorder, which is found to be $\langle \mathrm{Corr}(\sigma_\mathrm{d}, Q_iP_i)\rangle \approx 3\times 10^{-2}$, where the average $\langle\bullet\rangle$ is done over all the days and stocks. For each stock we regress $\mathcal{W}_i \sim N^{\gamma}$, $\sigma_\mathrm{d} \sim N^{\nu}$, $Q_iP_i \sim N^{\delta}$, and we obtain from the empirical distributions of the exponents in figure A3 that their average values read $\langle\gamma\rangle = 0.5$, $\langle\nu\rangle = 0.25$ and $\langle\delta\rangle = 0.20$, thus $\langle\gamma\rangle \neq \langle\nu\rangle + \langle\delta\rangle$. However, by looking at the scatter plot of the estimated exponent γ as function of the sum $\nu+\delta$ computed separately for each stock (see bottom right panel in figure A3) one observes a clear linear relation.

A possible and intuitive explanation of the non null measured correlation between σ_d and Q_iP_i is that metaorders add up to volume, generate market impact and thus increase price volatility. In this way trading volume increases due to both an increase in the number of *bets* and in their sizes, and so does volatility from the increased market impact as discussed for example in Jones *et al.* (1994). Note that this reasoning is valid even if the metaorders only account for a certain percentage of the total daily market volume $V = \sum_{i=1}^{N} Q_i = \eta V_\mathrm{d}$ with η adjusting for the partial view of the ANcerno sample in terms of volume, and for the non-*bet* traded by intermediaries: from our dataset we measure in average $\langle\eta\rangle \approx 5\times 10^{-2}$.

Appendix 3. Statistics of trading costs

As expected, we find that, for a single metaorder with unsigned volume Q, the spread cost $c_\mathrm{spd} = S\times Q$ is dominant for small volumes, while the market impact cost $c_\mathrm{imp} = \sigma_\mathrm{d}\times QP\times\sqrt{Q/V_\mathrm{d}}$ takes over for large volumes (see left panel of figure A4). Furthermore, as shown in the right panel of figure A4, the average daily market impact cost $\mathcal{C}_\mathrm{imp}$ accounts on average for $\approx 1/2$ of the total daily trading average cost $\mathcal{C} = \mathcal{C}_\mathrm{spd} + \mathcal{C}_\mathrm{imp}$, computed using $Y_0 = 3.5$ and $Y = 1.5$ in equation (5).

† Here $\langle\cdot\rangle$ denotes the average over all days and stocks present in the sample.

‡ In analogy, the variance $\langle \mathcal{W}_i^2\rangle_N - \langle \mathcal{W}_i^2\rangle_N$ scales linearly with N, i.e. $\langle \mathcal{W}_i^2\rangle_N - \langle \mathcal{W}_i^2\rangle_N \approx \langle \mathcal{W}_i\rangle_N^2 \sim N$.

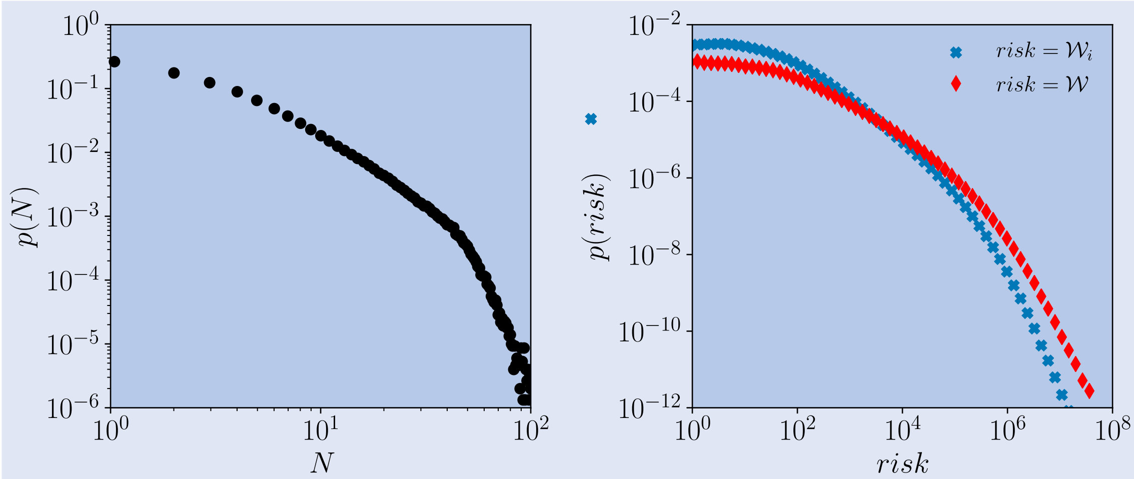

Figure A1. (Left panel) Empirical probability distribution of the daily number N of metaorders per asset: N is broadly distributed over two decades with an average close to 5. (Right panel) Empirical probability distributions of the exchanged risk per metaorder, i.e $\mathcal{W}_i := \sigma_d Q_i P_i$, and of the total daily risk per day/assets, i.e $\mathcal{W} := \sum_{i=1}^{N} \mathcal{W}_i$.

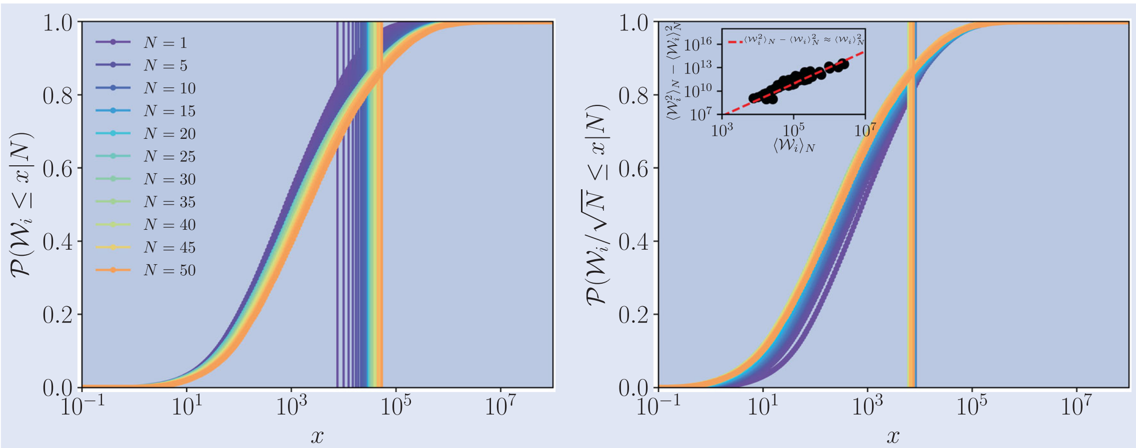

Figure A2. Empirical cumulative distribution of the traded metaorder's risk $\mathcal{W}_i = \sigma_d Q_i P_i$ without (left panel) and with (right panel) rescaling by the square root of the daily number N of metaorders per asset. The coloured vertical lines represent the location of the average for each sample conditional on N. To note that also if the empirical distribution is not an invariant function of N, we observe that $\langle \mathcal{W}_i/\sqrt{N} \rangle_N \simeq$ const., as evident from the vertical lines in the right panel, which is at the origin of the measured 3/2-law. Furthermore, as shown in the inset the variance $\langle \mathcal{W}_i^2 \rangle_N - \langle \mathcal{W}_i^2 \rangle_N$ scales linearly with N, i.e. $\langle \mathcal{W}_i^2 \rangle_N - \langle \mathcal{W}_i^2 \rangle_N \approx \langle \mathcal{W}_i \rangle_N^2 \sim N$.

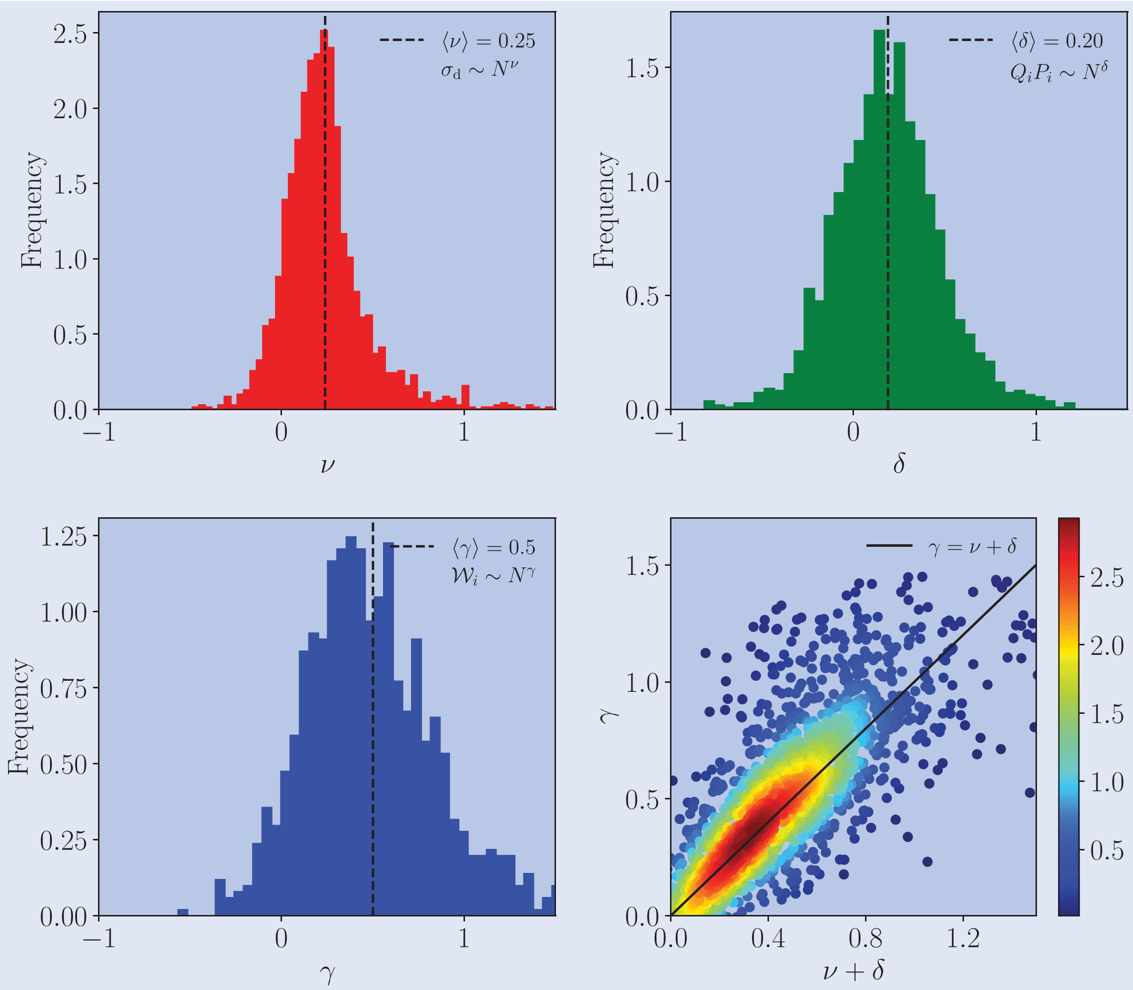

Figure A3. (Top left panel) Empirical distribution of the scaling exponent ν computed for each stock regressing $\sigma_d \sim N^\nu$: in average $\langle\nu\rangle = 0.25$ as shown by the dashed black line. (Top right panel) Empirical distribution of the scaling exponent δ computed for each stock regressing $Q_iP_i \sim N^\delta$: in average $\langle\delta\rangle = 0.20$ as shown by the dashed black line. (Bottom left panel) Empirical distribution of the scaling exponent γ computed for each stock regressing $\mathcal{W}_i \sim N^\gamma$: in average $\langle\gamma\rangle = 0.5$ as shown by the dashed black line. (Bottom right panel) Signature scatter plot (coloured by density of data) of the coefficients $\nu + \delta$ and γ respectively estimated conditioning to each stock.

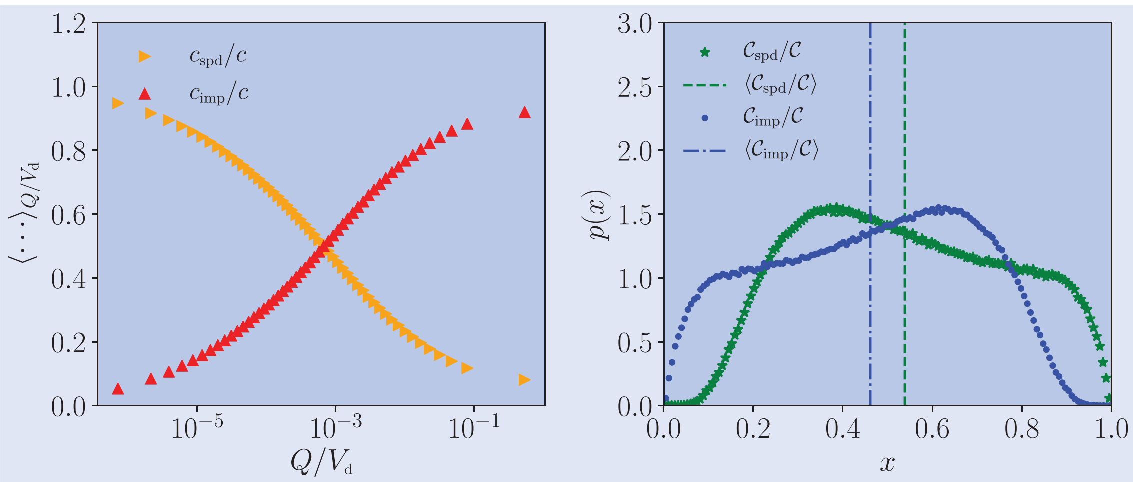

Figure A4. (Left panel) Averaged spread and market impact cost ratios given respectively by $c_{\rm spd}/c$ and $c_{\rm imp}/c$ - with $c_{\rm spd} = S \times Q$ (spread cost), $c_{\rm imp} = \sigma_d \times QP \times \sqrt{Q/V_d}$ (market impact cost) and $c = c_{\rm spd} + c_{\rm imp}$ (total cost per *bet*) - as function of the metaorder's order size Q/V_d: to note that for a metaorder with small (large) order size the spread (market impact) cost is dominant. (Right panel) Empirical distributions of the $\mathcal{C}_{\rm spd}/\mathcal{C}$ and $\mathcal{C}_{\rm imp}/\mathcal{C}$ ratios which give us an idea of the order of magnitude of the different contributions to the total daily average cost per metaorder $\mathcal{C} = \mathcal{C}_{\rm spd} + \mathcal{C}_{\rm imp}$ (computed from equation 5 fixing $Y_{\rm spd} = 3.5$ and $Y_{\rm imp} = 1.5$): the dashed vertical lines represent the location of the mean values equal respectively to $\langle\mathcal{C}_{\rm spd}/\mathcal{C}\rangle = 0.49$ and $\langle\mathcal{C}_{\rm imp}/\mathcal{C}\rangle = 0.51$.