Universal scaling and nonlinearity of aggregate price impact in financial markets

Felix Patzelt[1,*] and Jean-Philippe Bouchaud[1,2]
[1]*Capital Fund Management, 23 Rue de l'Université, 75007 Paris, France*
[2]*Ecole Polytechnique, 91120 Palaiseau, France*

(Received 9 August 2017; published 9 January 2018)

How and why stock prices move is a centuries-old question still not answered conclusively. More recently, attention shifted to higher frequencies, where trades are processed piecewise across different time scales. Here we reveal that price impact has a universal nonlinear shape for trades aggregated on any intraday scale. Its shape varies little across instruments, but drastically different master curves are obtained for order-volume and -sign impact. The scaling is largely determined by the relevant Hurst exponents. We further show that extreme order-flow imbalance is not associated with large returns. To the contrary, it is observed when the price is pinned to a particular level. Prices move only when there is sufficient balance in the local order flow. In fact, the probability that a trade changes the midprice falls to zero with increasing (absolute) order-sign bias along an arc-shaped curve for all intraday scales. Our findings challenge the widespread assumption of linear aggregate impact. They imply that market dynamics on all intraday time scales are shaped by correlations and bilateral adaptation in the flows of liquidity provision and taking.

DOI: 10.1103/PhysRevE.97.012304

I. INTRODUCTION

Markets allow different sources of information to be processed and transformed into a single number: the price. Since these market prices in turn play an important signaling role for the rest of the economy, the efficiency of the price formation process is a highly relevant question. Equilibrium models explain some general features of asset prices in a formally elegant way without considering the detailed price formation process [1–3]. There is, however, growing evidence that financial markets are almost never in equilibrium and that prices reflect more than fundamental information [4–7]. Instead, the flow of demand and supply, information and opinions is only slowly digested, one transaction at a time [8]. Understanding such dynamics is of great importance for practitioners optimizing their trading strategies, as well as for exchanges and regulators interested in improving market efficiency and stability.

In modern electronic markets, participants interact through a limit order book (LOB) in a continuous double auction. Some market participants act as liquidity providers by placing limit orders (buy or sell) in the LOB. Other market participants act as liquidity takers: They need to execute their trades immediately and correspondingly trigger transactions by sending market orders. These market orders tend to impact prices: Statistically, a buy (sell) market order pushes the price upward (downward).

While the average price impact of single market orders is relatively well understood, the impact of a series of market orders is much more complex. For example, a perplexing empirical result is the square-root volume dependence of the impact of a metaorder, i.e., a sequence of individual orders belonging to the same trading decision that cannot be executed in a single transaction but must instead be fragmented (see, e.g., [8,9] and references therein).[1] This result is at odds with, e.g., the classical Kyle model of impact, which predicts a linear dependence on volume [11]. The empirical analysis of metaorders is difficult since it requires a proprietary database, where the trades belonging to a given trading decision can be identified. When such data are available, the square-root impact law seems to be universally vindicated, for a wide variety of markets, epochs, and trading styles.

Most available data sets are, however, anonymized: While the sign ϵ and volume v of each market order can be reconstructed (see Sec. II), the identity of the trader (or of the trading institution) at the origin of the market order is usually unknown. One can nevertheless define the aggregate impact $\mathcal{R}_N$ over N consecutive trades as the average price return, conditioned to a certain total volume imbalance $\mathcal{Q}_N$ defined as

$$\mathcal{Q}_N = \sum_{i=1}^{N} q_i, \quad q_i := \epsilon_i v_i, \tag{1}$$

where q_i is the signed volume of the ith trade [see Eq. (2) below for complete definitions]. Although the impact of a single trade is well known to be a strongly concave function of its volume, it is reported that the aggregate impact of N trades becomes linear in $\mathcal{Q}$ as N increases [8,12].

This picture, however, is quite incomplete, as we reveal in this empirical paper. We show that once correctly rescaled, and for $N \gtrsim 10$, the aggregate impact function exhibits a nonlinear sigmoidal shape that is approximately independent of the number of transactions N and of the chosen asset (e.g., large-tick stocks, small-tick stocks, and futures).

We also study the aggregate-sign impact, where the conditioning variable is not $\mathcal{Q}_N$ but rather the sign imbalance

*felix@neuro.uni-bremen.de

[1]Apparent deviations for very short or long time scales, possibly due to conditioning or undersampling, are still debated. See, e.g., [10].

$\mathcal{E}_N = \sum_{i=1}^{N} \epsilon_i$. Scaling is again observed in this case, now with an impact function that reverts to zero at both extremes.

After quantifying the rescaling of the aggregate impact curves over different time horizons, we investigate why the price impact for extreme order-sign imbalances reverts towards zero. We find that the local bias of the order signs and the probability that an order changes the price compensate each other to a very high degree. Although possibly anticipated on general grounds, this effect does not seem to have been quantitatively reported so far and has very fundamental and important consequences on the dynamics of markets.

The present paper is mostly about empirical observations. Nevertheless, we illustrate how the aforementioned vanishing price-change probability can determine the observed sigmoidal impact curves in a qualitative toy model. The ability of currently available models to describe quantitatively the scaling properties of the nonlinear aggregate impact curves is the topic of another paper [13].

II. DATA

Our data set contains the highest turnover instruments on three different platforms: (i) 12 technology stocks on the US primary NASDAQ market, for the years 2011–2016, which includes some of the most traded stocks in the world like Apple (AAPL) and Microsoft; (ii) the 13 highest turnover stocks on NASDAQ OMX NORDIC (called just OMX in the following), which covers the Nordic markets Stockholm, Helsinki, and Copenhagen for October 2011 until the end of September 2015 (OMX is the primary market for the selected stocks); and (iii) 6 futures on EUREX EBS (BOBL, BUND, DAX, EUROSTOXX, SCHATZ, and SMI) for October 2014 until the end of 2015.

We chose to analyze three different platforms in order to have some variability in terms of market microstructure in the sample while keeping the complexity of the data preparation manageable. The instruments were selected for their high turnover, reasonable concentration on their primary markets, quality of the data, and availability via the same provider for the entire period that was analyzed: WOMBAT for NASDAQ, NOMURA for OMX, and the exchange itself for the EUREX data.

The NASDAQ stocks are also traded on different US markets and trades are routed automatically to the best offer. Nevertheless, we chose to not aggregate several US markets, because they are frequently desynchronized at the millisecond scale [14,15], leading to inconsistent aggregate bid and ask prices, that is, the best visible buy and sell limit orders, respectively, as reported by the market just before each a transaction is executed. We found that the microstructural parameter $\eta \in [0,1]$ appears to be a good measure of the importance of price discretization.[2] Prices on NASDAQ are discretized with a fixed tick size of 0.01, which can be considered very small ($\eta = 0.73$) to medium ($\eta = 0.49$) for the analyzed stocks. Up to roughly one-third of the transactions were executed against hidden liquidity.

Stocks on OMX are only traded on one of the Nordic markets at a time and are much less fragmented than US stocks. Tick sizes vary with price and are effectively larger ($0.24 \leqslant \eta \leqslant 0.50$) than for NASDAQ. Here hidden liquidity represents a vanishingly small fraction of all traded volume and seems to be concentrated on the mid. Finally, the EUREX futures are not traded on other platforms at all. Tick sizes vary considerably between moderately large ($\eta = 0.44$) and extremely large ($\eta = 0.03$).

In the following, we calculate price returns $r_t = \ln m_{t+1} - \ln m_t$ from the midprices m defined as the average of the bid price and the ask price just before each transaction. We constructed order signs by labeling all trades above the midprice as $\epsilon = +1$ and all trades below as $\epsilon = -1$. Trades exactly at the midprice were discarded. We decided not to use the signs provided by the exchanges themselves because hidden liquidity is not correctly labeled on NASDAQ for a part of the analyzed period. Nevertheless, we confirmed all the following results using the exchange-provided signs, with only very minor quantitative differences. Trade IDs were only available from EUREX. Therefore, we merged all transactions based on the timestamps, which were reported with millisecond precision for all three platforms (see also Sec. A in [17]).

Trading volumes vary considerably over time. To control for extremely active days, we normalized aggregate transaction volumes $\mathcal{Q} = \langle Q_D \rangle / Q_D \sum_i q_i$ by the daily volume Q_D relative to its average. This global normalization will be omitted in the following equations for notational simplicity.

The first 30 min after opening and before closing on each day are discarded, as well as all days with shortened trading hours. Obviously irregular entries were discarded too, such as transactions labeled as irregular by the exchange or provider, transactions outside the aforementioned hours, or transactions with nonfinite prices (including bid and ask prices).

III. RESULTS

A. Aggregate impact

As mentioned in the Introduction, we measure the aggregate-volume impact as

$$\mathcal{R}_N(\mathcal{Q}) := \left\langle \ln m_{t+N} - \ln m_t \,\middle|\, \mathcal{Q} = \sum_{i=0}^{N-1} q_{t+i} \right\rangle, \qquad (2)$$

where m_t is the midprice immediately before the tth transaction, q_t is the signed volume of the tth transaction, and $\langle \cdots \rangle$ denotes an empirical average over all time windows containing N successive trades, executed the same day. In addition, $\mathcal{R}_1(\mathcal{Q})$ corresponds to the average impact of a single market order of signed volume $\mathcal{Q}$ as studied in, e.g., [8,18].

As expected, both the width and height of the function $\mathcal{R}_N(\mathcal{Q})$ increase with N. However, if one rescales the $\mathcal{Q}$ axis with an N-dependent volume scale Q_N and the $\mathcal{R}$ axis with an N-dependent return scale R_N, all curves for $N \gtrsim 10$ collapse to a single master curve, as shown in Fig. 1 for AAPL and in Sec. B in [17] for a variety of other assets. More precisely, one

[2]Here $\eta := N_c/2N_a$, where N_c is the number of subsequent price movements in the same direction (continuations) and N_a the number of price movements in alternating directions. It measures the effect of discretization of a diffusion process. In addition, $\eta > 0.5$ corresponds to small-tick instruments and $\eta < 0.5$ to large-tick instruments [16].

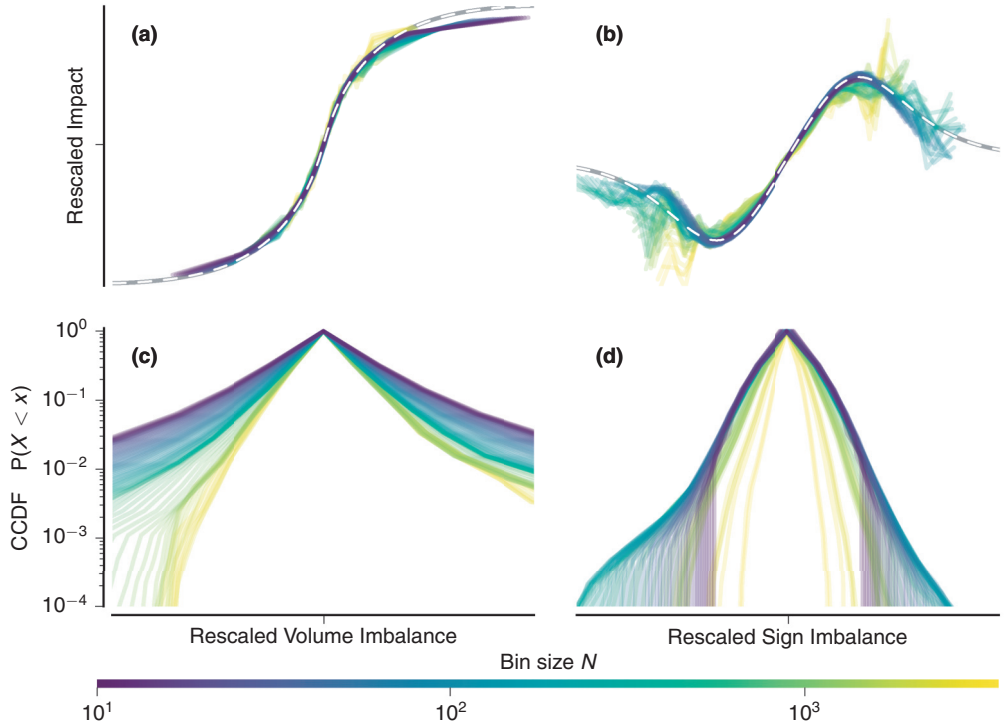

FIG. 1. AAPL on NASDAQ in 2016. (a) Rescaled expected return $\mathcal{R}_N(\mathcal{Q}/N^{\xi})/N^{\psi}$ conditioned on the volume imbalance $\mathcal{Q}$ for different bin sizes N in arbitrary units [see Eq. (2) ff]. The x- and y-axis rescaling exponents are $\xi = 0.84$ and $\psi = 0.53$. (b) Rescaled mean return $\mathcal{R}_N(\mathcal{E}/N^{\psi})/N^{\psi}$ conditioned on the sign imbalance $\mathcal{E}$ [see Eq. (9)]. Here $\xi_\epsilon = 0.69$ and $\psi_\epsilon = 0.48$. (c) and (d) Corresponding complementary cumulative distributions. The positive and negative halves were calculated independently and then binned to smooth out noise and discretization steps for small N. The largest shown N corresponds to the shortest day in the sample.

finds that, empirically,

$$\mathcal{R}_N(\mathcal{Q}) \approx R_N \mathscr{F}\left(\frac{\mathcal{Q}}{Q_N}\right), \tag{3}$$

where Q_N and R_N both obey power-law scaling with N,

$$Q_N \approx Q_1 N^{\xi}, \tag{4}$$

$$R_N \approx R_1 N^{\psi}, \tag{5}$$

and the scaling function $\mathscr{F}(x)$ is a sigmoidal function parametrized as

$$\mathscr{F}(x) = \frac{x}{(1+|x|^{\alpha})^{\beta/\alpha}}, \tag{6}$$

where α and β are fitting parameters that describe the shape of $\mathscr{F}(x)$. Note that for $x \to 0$, the leading behavior is

$$\mathscr{F}(x) = x - \frac{\beta}{\alpha}\mathrm{sgn}(x)|x|^{1+\alpha} + \cdots, \tag{7}$$

i.e., a linear behavior with possibly nonanalytic corrections.

For $x \to \infty$, on the other hand, one has

$$\mathscr{F}(x) = \mathrm{sgn}(x)|x|^{1-\beta} + \cdots. \tag{8}$$

Hence $\beta = 1$ corresponds to saturation for large volumes, $\beta < 1$ to continued growth, and $\beta > 1$ to reversal towards lower impacts.

In order to determine the rescaling exponents ξ and ψ, the shape of $\mathcal{R}_N(Q)$ is fitted for each N using the scaling form (3) with $\mathscr{F}(x)$ given by Eq. (6), keeping the same values of α and β for all N.[3] We obtained $\alpha = 1.2 \pm 0.6$ and $\beta = 1.3 \pm 0.7$ for the mean and standard deviation of the fitted $\mathcal{R}_N(Q)$ across all instruments in the sample. The corresponding scaling function for AAPL is shown as a dashed line in Fig. 1.[4]

Once $\mathscr{F}(x)$ is fixed, one can map out the scale factors Q_N and R_N as a function of N, which are described very accurately by power laws of N as shown in Fig. 2(a).[5] The final rescaled impact functions are shown in Sec. B in [17] for other stocks and futures. All scaling curves look remarkably similar, as indicated by the similar values of α and β in all cases. Any theoretical approach will have to explain the value not only of the exponents ξ and ψ, but also of the full master curve $\mathscr{F}(x)$.

[3]Technically, this was achieved by alternating between fitting either the scales or the shape parameters and using nonlinear regression. Only 80% of all N were randomly included in each pass.

[4]Some instruments exhibit a slight reversal of the aggregate-volume impact $\mathcal{R}(\mathcal{Q})$ for very large arguments. These are sometimes fitted with quite large β, but the fitted curve only strongly reverts outside of the observed range of $\mathcal{Q}$. This is in very different from $\mathcal{R}(\mathcal{E})$ discussed below, which strongly reverts close to zero impact within the frequently observed range of sign imbalances $\mathcal{E}$.

[5]This was done using robust regression. We also tried to fit the power-law rescaling without using parametric curves as an in-between step, but failed to achieve the same level of reliability across instruments and time periods.

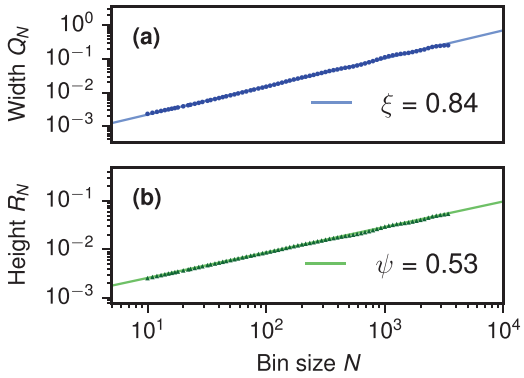

FIG. 2. AAPL in 2016. (a) Volume scales Q_N for the impact curves [see Eq. (3)] that were fitted to each bin size for the same data shown in Fig. 1(a). The solid line shows the power-law fit used for the rescaling along the volume-imbalance axis. (b) Same analysis as in (a) but for the return scale R_N, i.e., the scaling of the impact curve along the return axis. Note that the previous fitting of the impact curves, yielding Q_N and R_N for each N, did not impose any assumptions on their scaling. Relative fitting errors are below 1%, but variability across instruments and periods is much larger (discussed below).

Together with the rescaled aggregate-volume impact, Fig. 1 shows the corresponding cumulative distribution of volume, rescaled by Q_N. Events far in the saturation regime occur with probability $\sim 10^{-2}$ on a daily basis. This must be compared with the typical number of trades per day, which is of the order of 10^4 for AAPL. For example, there about 100 events per day at the end of a bin of size $N = 100$ and within the saturation regime. Events contributing to the saturation regime are relatively frequent and the effect is therefore not anecdotal.

Figure 1(b) shows the rescaled aggregate-sign impact, defined as

$$\mathcal{R}_N(\mathcal{E}) := \left\langle \ln m_{t+N} - \ln m_t \,\middle|\, \mathcal{E} = \sum_{i=0}^{N-1} \epsilon_{t+i} \right\rangle. \tag{9}$$

Here the impact for small sign imbalances is more linear than for the volume imbalance, corresponding to a larger value of the effective parameter α. Around a sign imbalance of 50%, the impact saturates sharply and reverts towards zero at the extremes. This may come as a surprise since it means that a very strong imbalance in the order signs is associated with a very small price change on average. This effect is found for all instruments and also for the trade imbalance, as shown in Sec. B in [17]. The reason for this highly peculiar behavior is investigated below. First, however, we have a closer look at the scaling exponents ξ and ψ (and their counterpart for the aggregate-sign impact ξ_ϵ and ψ_ϵ).

B. Scaling and Hurst exponents

Figure 3 shows the means and standard deviations for several scaling exponents. The scaling exponent of the width Q_N of the aggregate-volume impact is found close to $\xi \approx 0.75$, while the exponent governing the height R_N is $\psi \approx 0.5$. In other words, the width of the impact curve grows faster than its height when the bin size is increased. Very similar values are found for the aggregate-sign impacts (exponents ξ_ϵ and ψ_ϵ). Note that using Eqs. (3) and (6), the slope of the linear region of impact follows as $\partial \mathcal{R}_N(\mathcal{Q})/\partial \mathcal{Q}|_{\mathcal{Q}=0} = R_N/Q_N$. It scales as $N^{-\kappa}$ with $\kappa = \xi - \psi \approx 0.25$, i.e., it decreases with N as a power law N.[6]

Since both $\mathcal{R}_N$ and $\mathcal{Q}_N$ are sums over random variables (returns and signed volumes), one expects that their natural scaling with N is governed by the Hurst exponents of the underlying variables r and q. Here we define the Hurst exponent H of a zero-mean random variable x from the scaling of the standard deviation of sums of N successive events

$$\left\langle \left(\sum_{i=1}^{N} x_i \right)^2 \right\rangle := D N^{2H_x}, \tag{10}$$

where D is a constant and we take advantage of the fact that returns and signed volumes have a zero average over long time

[6]The cross-instrument dispersion of κ around its average $\langle \kappa \rangle = 0.23$ is actually relatively small: $s(\kappa) = 0.10$, where s denotes the standard deviation.

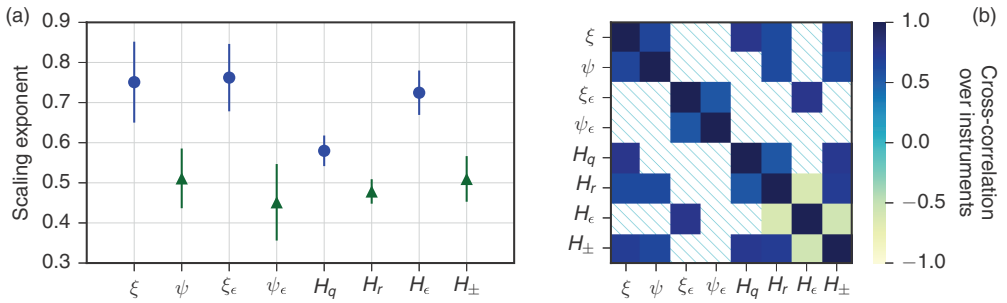

FIG. 3. (a) Analysis of the distributions of several scaling exponents. Plotted symbols denote the means and vertical lines the standard deviations. Both were calculated over instruments, for each of which the exponents were calculated for all 1-yr periods before averaging. Blue circles correspond to the scaling of a variable that a price impact is typically conditioned on, i.e., the width of an impact curve along the x axis. Green triangles correspond to the scaling of a variable measuring price changes, i.e., the height along the y axis. The variables, from left to right, are volume imbalance and corresponding return rescaling ξ and ψ, respectively, sign imbalance and corresponding return rescaling ξ_ϵ and ψ_ϵ, respectively, and Hurst exponents of the volume imbalance H_q, returns H_r, order signs H_ϵ, and return signs $H_\pm$ [see Eq. (10)]. (b) Cross correlation calculated over all instruments for the same variables. Only correlations differing from zero by more than three standard deviations are shown in solid colors.

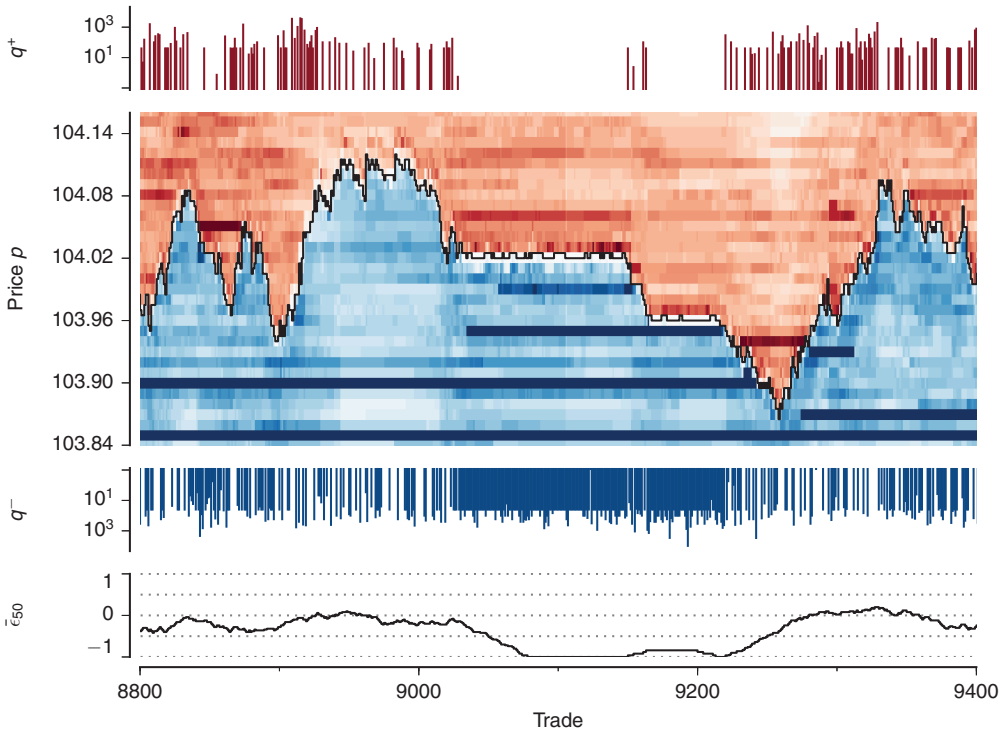

FIG. 4. AAPL, 1 May 2016. A typical example of an extreme order-sign imbalance and a "sticky price." The visible liquidity at each price level just before each trade is drawn in red above and in blue below the midprice (upper solid line). Full saturation corresponds to the 90th percentile for the day. The buy volumes q^+ for each trade are drawn as bars above the order book reconstruction and the sell volumes q^- below. The volume axis is logarithmically scaled. A lower bound for the true number of trades is shown because transactions with the same sign and timestamp were merged (see Sec. II). Therefore, a series of 100 successive market buys drawn above took at least 100 ms in wall time and may have consisted of more than 100 independent market orders submitted to the market. The lower solid curve denotes the 50 trade sign imbalance (causal).

horizons.[7] As usual, $H = 0.5$ corresponds to regular diffusion, $H < 0.5$ to sudiffusion, and $H > 0.5$ to superdiffusion.

Returns are almost diffusive with a very slight tendency for mean reversion (particularly for large-tick instruments). This is consistent with ψ and ψ_ϵ. Volume signs are only slightly positively correlated, at least when measured through H_q. Order signs are (as is well known) strongly correlated with $H_\epsilon > 0.7$, close to the values of ξ and ξ_ϵ. This implies that the scales of the aggregate impact curves are mostly determined by the accumulated variation in the return and sign time series. Interestingly, the scaling of the impact curves is not trivially related to the Hurst exponent of volume fluctuations, but rather to sign fluctuations. This is expected from the fact that the volume of individual orders exhibits extreme variability that mostly reflects the available liquidity at the best price [19–21]. Large fluctuations of order volumes v introduce an independent source of noise that masks part of the order-sign correlations when measuring H_q in the way described above (see [22] for a related discussion).

Figure 3(b) shows the significant cross correlations across instruments between the different scaling exponents.[8] We find positive correlations between ξ and H_q as well as between ψ and H_r, offering some reassurance that the similar average values in Fig. 3(a) are more than a pure coincidence. We also find positive correlations between ξ_ϵ and H_ϵ. However, there are no significant correlations between ψ_ϵ and H_r or between H_ϵ and H_q. This hints at a quite complex interplay of several factors driving the variability in the different scaling behaviors across instruments. Here H_ϵ is negatively correlated with H_r, which implies more order splitting on more mean-reverting, larger-tick instruments. Remember, though, that H_r only varies very little across instruments [Fig. 3(a)].[9,10]

[7]This method turned out to be more robust than the standard rescaled range analysis. The reason is that returns and volumes follow very heavy-tailed distributions, to which the range is much more sensitive than the sum. It is also simpler than detrended fluctuation analysis, which did not seem to be beneficial in this particular use case.

[8]Note that the square of the cross correlation can also be interpreted as the coefficient of determination R^2 of a linear regression with intercept.

[9]We found consistent results measuring order-sign autocorrelations directly (not shown). The latter decay with an exponent $\gamma \approx 0.5$ for long lags. They often exhibit a steeper initial decay, however. Correlations for γ and other exponents are similar to H_ϵ, but are generally weaker even though γ varies more across assets.

[10]We also found that the "implied spread" η (see [16] and footnote 2) seems to be informative (not shown). It varies considerably across instruments (see Sec. II) and is strongly correlated with several exponents. It shares these correlations with the return-sign Hurst exponent. This finding seems to be related to the co-occurrence of larger $H_\pm$ and a larger fraction of trend continuations N_c.

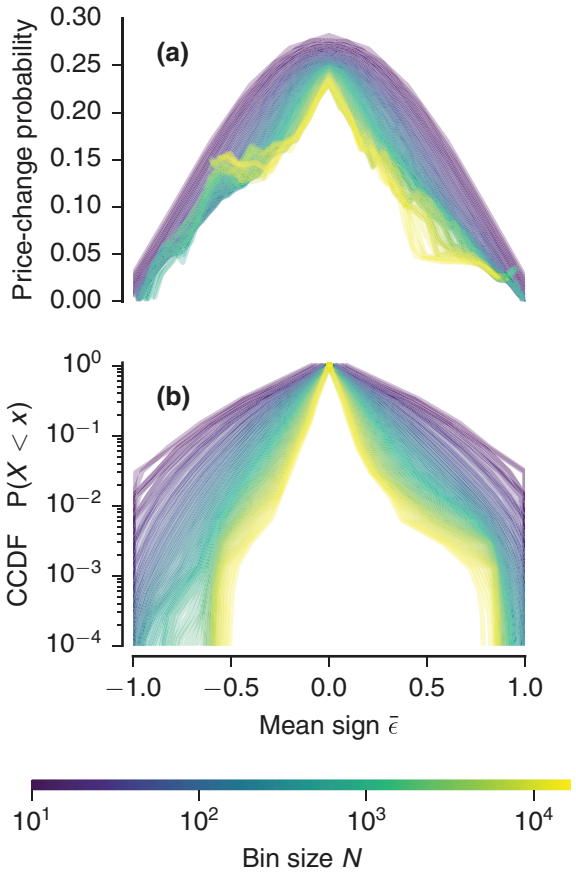

FIG. 5. EUROSTOXX on EUREX in 2015. (a) Probability $P(r \neq 0|N^{-1}\mathcal{E})$ that the midprice changes from one trade to the next conditioned on the mean trade sign in the bin containing N trades. (b) Corresponding complementary cumulative distributions.

C. Large order imbalances and pinned prices

As stated before, extreme order-sign imbalance is not associated with large returns. To the contrary, such large imbalances are observed when the price is pinned to a particular level. A typical example is shown in Fig. 4. Around trade 9100, a series of more than 100 market sell orders arrives, yet the midprice only bounces up and down half a tick as liquidity providers replenish limit buy orders at the best price or inside of the spread (the distance between the best bid and the ask price). Many of these trades are quite small, but some are significantly larger as liquidity takers adapt to the available volume at the best.

Figure 4 shows that the midprice only diffuses freely when trades happen on both sides of the book. A more general and systematic quantification of this effect is shown in Fig. 5(a): the probability of a midprice change between two subsequent trades as a function of the average order sign in the bin, $\bar{\epsilon} := \mathcal{E}/N$. One of the central results of this paper is that biased order signs lead to a lower probability for price changes on any intraday time scale. Note in particular that the price-change probability has an almost invariant triangular shape for all bin sizes of $N \gtrsim 50$. It approaches zero for highly biased order signs. The qualitative behavior is the same for smaller N, although the curve is broader and smoother and has a flatter maximum around $\bar{\epsilon} = 0$. This behavior is universal across all instruments (see Sec. C in [17] for more examples). The corresponding cumulative distribution is shown in Fig. 5(b), confirming that strong order-sign imbalances happen relatively frequently on a daily basis. The nonintuitive negative correlation between sign imbalance and return is therefore not an artifact due to a lack of data.

To understand how the conditional probability for price changes can explain the sigmoidal price impact shown in Fig. 1(b), let us consider a simple toy model. As a caricature of Fig. 5(a), assume that $p_r := P(r \neq 0 \,|\, \bar{\epsilon}) \propto 1 - |\bar{\epsilon}|$. Furthermore, let $\rho := \langle r \,|\, r \neq 0 \rangle \propto \epsilon$. That is, assume prices moved in constant steps in the direction of the last order sign. Then the average price impact for a trade in the bin follows as $\langle r \,|\, \bar{\epsilon} \rangle = \rho\, p_r \propto \bar{\epsilon}\ (1 - |\bar{\epsilon}|)$, which is a sigmoidal curve quite similar to the empirical observations. A stylized N-trade impact $\mathcal{R}_N(\mathcal{E})$ may be obtained, e.g., by rescaling according to the Hurst exponents H_r and H_ϵ as discussed above.

IV. DISCUSSION

We investigated how prices are impacted by the flow of market orders and found a universal behavior on all intraday time scales. We have shown that the impact curves, once correctly rescaled, are remarkably stable across time scales (from bins of $N = 10$ trades up to an entire day of trading, beyond which overnight effects would have to be taken into a account) and instruments (e.g., large- and small-tick US stocks, Nordic stocks, and EUREX futures). This illustrates how measuring master curves instead of either scaling laws for scalar quantities or conditional expectation curves on individual time scales can substantially improve the insights gained into market dynamics. To fully appreciate the master curves' robustness, we encourage the reader to read the Supplemental Material [17]. Our results suggest that the price formation process in financial markets is the result of some general universal mechanism. While the latter is still to be elicited, our findings do provide some hints.

We find that the aggregate-volume impact saturates for large (rescaled) imbalances, on all time scales. The behavior of the aggregate-sign impact is even more striking: Highly biased order flows are associated with very small price changes. More precisely, we find that the probability for an order to change the price decreases with the local imbalance and vanishes when the order signs are locally strongly biased in one direction. At high frequencies, extreme order-sign imbalances occur when a very large volume is available on the opposite side of the order book, resulting in prices being temporarily pinned to a certain level. These large volumes manifest themselves either as visible large limit orders or as repeated refills near a particular price level.

Qualitatively, this dependence of the price-change probabilities on average order signs is consistent with models and empirical results in the literature. For example, the Madhavan-Richardson-Roomans model [23] postulates that the change of price is proportional to the sign surprise, i.e., the difference between the sign ϵ of a market order and its expected value, based on previous signs. When it is extremely likely that the next trade is a buy and a buy trade indeed materializes, then the price change is small. Empirically, many studies have reported that returns in the direction of a particular trade-sign predictor

are on average lower than those in the opposite direction (see, e.g., [24–26]). This effect is attributed to liquidity takers adjusting their market-order volume at the outstanding liquidity, while liquidity providers revise their limit orders and refill to match the incoming order bias. Our results show that the nonlinearity of the aggregate price impact is a consequence of this previously hypothesized bilateral order-flow adaptation. The latter is measured directly by the probability of price changes as a function of the local sign imbalance, which is (for large enough N) a tent-shaped function: It has a discontinuous slope for zero imbalance and vanishes for strong imbalances. This observation is consistent with [12], where a similar shape was reported for the standard deviation of price changes in 15-min windows for US stocks traded in 1994–1995, i.e., before electronic markets and high-frequency algorithms.

Taken together, our findings suggest that markets generally operate in a state where traders collectively counterbalance the impact of predictable events to a very high degree and on all intraday time scales. Since market-order volumes are known to be highly conditioned on visible liquidity (see, e.g., [21]), the dependence of market-order signs on repeated refills (as shown in Fig. 4) should not come as a surprise: These observations simply confirm that liquidity takers pay attention to the currently available liquidity. Reciprocally, liquidity providers observe the flow of market orders and adapt their behavior to the well-known long-range correlations of order signs (see [22] for a related discussion).

In this scenario, price fluctuations mostly reflect the lack of predictability, or surprise, of an event. Dynamics of this type have previously been shown to be capable of generating clustered volatility and extreme price jumps in stylized multiagent systems [27] and in highly adaptive control systems in [28]. Therefore, the present work provides a first step towards more directly testable models along these lines and suggests that the classical notion of market efficiency [29] should be extended to include endogenous information on top of exogeneous news.

In another paper [13], we will investigate in detail how accurately the nonlinear master curves and rescaling exponents described above can be reproduced using propagator models [30] and their generalizations [31,32]. Our results provide important constraints for such models and for realistic market models in general, since they quantify how the market reacts to both order bias and price-change probability on all intraday time scales. Naturally, such improved impact models are of interest both for practitioners trying to reduce their trading costs and for regulators trying to understand the stability of markets.

ACKNOWLEDGMENTS

This work was inspired by a preliminary unpublished study by Julius Bonart. The toy model at the end of Sec. III C was developed in a discussion with Stanislao Gualdi. We thank them, and M. Benzaquen, G. Bormetti, F. Bucci, J. De Lataillade, J. Donier, Z. Eisler, M. Gould, S. Hardiman, J. Kockelkoren, F. Lillo, I. Mastromatteo, D. Taranto, and B. Toth for many inspiring discussions. We also thank G. Bolton, J. Lafaye, and C. A. Lehalle for support and advice during data preparation. We finally thank the DFG for supporting F.P. with a research fellowship (Grant No. PA 2666/1-1).

[1] KVA, Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences, Understanding Asset Prices, nobelprize.org, http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/advanced.html (2013).
[2] J. D. Farmer and J. Geanakoplos, Complexity **14**, 11 (2009).
[3] R. K. Lyons, *The Microstructure Approach to Exchange Rates* (MIT Press, Cambridge, 2000), Chap. 1.
[4] R. J. Shiller, Am. Econ. Rev. **71**, 421 (1981).
[5] D. M. Cutler, J. M. Poterba, and L. H. Summers, J. Portfolio Manage. **15**, 4 (1989).
[6] C. Hopman, Quant. Financ. **7**, 37 (2007).
[7] A. Joulin, A. Lefevre, D. Grunberg, and J.-P. Bouchaud, Wilmott **9–10**, 1 (2008).
[8] J.-P. Bouchaud, J. D. Farmer, and F. Lillo, in *Handbook of Financial Markets: Dynamics and Evolution*, edited by T. Hens and K. R. Schenk-Hoppe (North-Holland, Amsterdam, 2009).
[9] J. Donier, J. Bonart, I. Mastromatteo, and J.-P. Bouchaud, Quant. Financ. **15**, 1109 (2015).
[10] E. Zarinelli, M. Treccani, J. D. Farmer, and F. Lillo, Market Microstruct. Liquidity **01**, 1550004 (2015).
[11] A. S. Kyle, Econometrica **53**, 1315 (1985).
[12] V. Plerou, P. Gopikrishnan, X. Gabaix, and H. E. Stanley, Phys. Rev. E **66**, 027104 (2002).
[13] F. Patzelt and J.-P. Bouchaud, J. Stat. Mech. (2017) 123404.
[14] P. Mackintosh and K. W. Chen, *The Need for Speed III: Physics and a Little Trigonometry* (KCG Holdings, Inc., Chicago, 2016), available at https://www.virtu.com/uploads/documents/Speed_III_Final_January-26.pdf
[15] P. Mackintosh and K. W. Chen, *The Need for Speed V: How Important is 1 ms?* (KCG Holdings, Inc., Chicago, 2016), available at https://www.virtu.com/uploads/documents/the-need-for-speed-v-how-important-is-ms.pdf
[16] K. Dayri and M. Rosenbaum, Market Microstruct. Liquidity **1**, 1550003 (2015).
[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.97.012304 for details.
[18] F. Lillo, J. D. Farmer, and R. N. Mantegna, Nature (London) **421**, 129 (2003).
[19] C. M. Jones, G. Kaul, and M. L. Lipson, Rev. Financ. Stud. **7**, 631 (1994).
[20] J. D. Farmer, L. Gillemot, and F. Lillo, Quant. Financ. **4**, 383 (2004).
[21] P. Gomber, U. Schweickert, and E. Theissen, Eur. Financ. Manage. **21**, 52 (2015).
[22] J.-P. Bouchaud, J. Kockelkoren, and M. Potters, Quant. Financ. **6**, 115 (2006).
[23] A. Madhavan, M. Richardson, and M. Roomans, Rev. Financ. Stud. **10**, 1035 (1997).
[24] D. Taranto, G. Bormetti, and F. Lillo, J. Stat. Mech. (2014) P06002.

[25] F. Lillo and J. D. Farmer, Stud. Nonlinear Dyn. Econometrics **8**, 1 (2004).
[26] A. Gerig, Ph.D. thesis, University of Illinois, 2007.
[27] F. Patzelt and K. Pawelzik, Sci. Rep. **3**, 2784 (2013).
[28] F. Patzelt and K. Pawelzik, Phys. Rev. Lett. **107**, 238103 (2011).
[29] E. Fama, J. Finance **25**, 383 (1970).
[30] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart, Quant. Financ. **4**, 176 (2004).
[31] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren, Quant. Financ. **12**, 1395 (2012).
[32] D. E. Taranto, G. Bormetti, J.-P. Bouchaud, F. Lillo, and B. Toth, arXiv:1602.02735.