EPJ manuscript No.
(will be inserted by the editor)

Price response functions and spread impact in correlated financial markets

Juan C. Henao-Londono [a], Sebastian M. Krause, and Thomas Guhr

Fakultät für Physik, Universität Duisburg-Essen, Lotharstraße 1, 47048 Duisburg, Germany

Received: date / Revised version: date

Abstract Recent research about the response of stock prices to trading activity revealed long lasting effects, even across stocks of different companies. These results implies non-markovian effects in price formation and reveal the consequences of trading many stocks at the same time, from trading costs to price correlations. How the price response is measured depends on the data set and research focus. However, it is still unclear, how the details of the price response definition modify the results. Here, we evaluate different price response implementations for the Trades and Quotes (TAQ) data set from the NASDAQ stock market and find that the results are qualitatevily the same for two different definitions of time scale, but the response can vary by up to a factor of two. Further, we confirm the dominating contribution of immediate price response directly after a trade, as we find that delayed responses are suppresed. Finally, we test the impact of the spread in the price response, spotting that large spreads have larger impact.

PACS. 89.65.Gh Econophysics – 89.75.-k Complex systems – 05.10.Gg Statistical physics

1 Introduction

Financial markets use order books to list the number of shares being bid or offered at each price point. An order book is an electronic list of buy and sell orders for a specific security or financial instrument organized by price level. Agents can place different types of instructions (orders) to buy or to sell a given number of shares.

In general, the dynamics of the prices follow a pure random walk. There are two extreme models that can describe this behavior: the Efficient Market Hypothesis (EMH) and the Zero Intelligence Trading (ZIT). The EMH affirms that all available information is included in prices and price changes can only be the result of unanticipated news, which by definition are totally unpredictable [5,7,21,31]. On the other hand, the ZIT assumes that agents instead of being fully rational, have "zero intelligence" and randomly buy or sell. It is supposed that their actions are interpreted by other agents as potentially containing some information [5,7,31,39]. In both cases the outcome is the same, the prices follow a random walk. In real life, they will not behave as one of these extreme scenarios, but somewhere in between [7,31].

There are diverse studies focused on the price response [3,4,5,7,13,15,16,19,20,21,25,27,28,36,37,38,39], but they concentrate on a general definition and discuss the results without going deep in the specific details of the price response measurements and how these aspects of the price response definition modify the results.

[a] *e-mail:* juan.henao-londono@uni-due.de

Regarding price self-response functions, [4,5,7], Bouchaud et al. found an increase to a maximum followed by a decrease as the time lag grows. In [16], the researchers found that larger sized transactions have a larger absolute impact than smaller sized transactions but a much smaller relative impact. In [28], the results show that the impact of small trades on the price is, in relative terms, much larger than that of large trades and the impact of trading on the price is quasi-permanent.

For price cross-responses functions, [3] found that the diagonal terms are on average larger than the off-diagonal ones by a factor ~ 5. The response at positive times is roughly constant, consistently with the hypothesis of a statistically efficient price. In other words, the current sign does not predict future returns. In [39] the results show the trends in the cross-responses does not depend on whether or not the stock pairs are in the same economic sector or extend over two sectors.

In this paper, we want to discuss, based on a series of detailed empirical results obtained on trade by trade data, that the variation in the details of the parameters used in the price response definition modify the characteristics of the results. Aspects like time scale, time shift, time lag and spread used in the price response calculation have a large impact on the outcomes.

Here, we delve into the key details needed to compute the price response functions, and investigate their corresponding roles. We perform a empirical study in different time scales using financial data and find that using different implementations of the price response, the results are qualitatively the same, but the response can vary up to a factor of two. We show that the order between the trade signs and the returns have a key importance in the price response signal and suggest an interval where the time shift have to be set. We split the time lag to understand the contribution of the immediate returns and the late returns. The price response is highly influenced by the instantaneous returns and as the time lag grows, the influence starts to decrease. We shed light on the spread impact in the response functions for single stocks. We check that when the spread is large, the price response tend to be large.

The paper is organized as follows: in Sect. 2 we present our data set of stocks. We then analyze the definition of the price response functions and describe the physical and trade time in Sect. 3. We implement different price responses for several stocks and pairs of stocks in Sect. 4. In Sect. 5 we show how the relative position between trade signs and returns has a huge impact in the results of the computation of the response functions. In Sect. 6 we explain in detail how the time lag τ behaves in the response functions. Finally, in Sect. 7 we analyze the spread impact in the response functions. Our conclusions follows in Sect. 8.

2 Data set

In a modern financial market, there is a double continuous auction. To find possible buyers and sellers in the market, agents can place different types of orders to buy or to sell a given number of shares, that can be grouped into two categories: market orders and limit orders.

Market orders will go into market to execute at the best available buy or sell price [13,14,31]. Limit orders allow to set a maximum purchase price for a buy order, or a minimum sale price for a sell order. If the market does not reach the limit price, the order will not be executed [13,14,31].

Limit orders often fail to result in an immediate transaction, and are stored in a queue called the limit order book [8,14,28]. An order book is an electronic list of buy and sell orders for a specific security or financial instrument organized by price level. An order book lists the number of shares being bid or offered at each price point. It also identifies the market participants behind the buy and sell orders, although some choose to remain anonymous. The order book is visible for all traders, its main purpose is to ensure that all traders have the information about what is offered on the market.

Buy limit orders are called "bids", and sell limit orders are called "asks". At any given time there is a best (lowest) offer to sell with price $a\left(t\right)$, and a best (highest) bid to buy with price $b\left(t\right)$ [5,9,11,28,31]. The price gap between them is called the spread $s\left(t\right)=a\left(t\right)-b\left(t\right)$ [5,6,7,9,13,31]. Spreads are significantly positively related to price and significantly negatively related to trading volume. Firms with more liquidity tend to have lower spreads. [1,2,9,17].

In this study, we have analyzed trades and quotes (TAQ) data from the NASDAQ stock market. We selected NASDAQ because it is an electronic exchange where stocks are traded through an automated network of computers instead of a trading floor, which makes trading more efficient, fast and accurate. NASDAQ is the second largest stock exchange based on market capitalization in the world.

In the TAQ data set, there are two data files for each stock. One gives the list of all successive quotes. Thus, we have the best bid price, best ask price, available volume and the time stamp accurate to the second. The other data file is the list of all successive trades, with the traded price, traded volume and time stamp accurate to the second. Despite the one second accuracy of the time stamps, in both files more than one quote or trade may be recorded in the same second.

Due to the the time stamp accuracy, it is not possible to match each trade with the directly preceding quote. Hence, we cannot determine the trade sign by comparing the traded price and the preceding midpoint price [39]. In this case we need to do a preprocessing of the data to relate the midpoint prices with the trade signs in trade time scale and in physical time scale.

To analyze the response functions across different liquid stocks in Sects. 4, 5 and 6, we select the six companies with the largest average market capitalization (AMC) (Alphabet Inc., Mastercard Inc., CME Group Inc., Goldman Sachs Group Inc., Transocean Ltd. and Apache Corp.) in three economic sectors (information technology, financials and energy) of the S&P index in 2008. Table 1 shows the companies analyzed with their corresponding symbol and sector, and three average values for a year.

To analyze the spread impact in response functions (Sect. 7), we select 530 stocks in the NASDAQ stock market for the year 2008. The selected stocks are listed in Appendix A.

In order to avoid overnight effects and any artifact due to the opening and closing of the market, we systematically discarded the first ten and the last ten minutes of trading in a given day [7,13,19,39]. Therefore, we only consider trades of the same day from 9:40:00 to 15:50:00 New York local time. We will refer to this interval of time as the "market time".

3 Price response function definitions

In Sect. 3.1 we describe the physical time scale and the trade time scale. We introduce the price response functions used in literature in Sect. 3.2.

3.1 Time definition

A key concept in the analysis of the response functions is the time. Due to the nature of the data, they are several options to define it.

In general, the time series are indexed in calendar time (hours, minutes, seconds, milliseconds). Moreover, tick-by-tick data available on financial markets all over the world is time stamped up to the millisecond, but the order of

Table 1. Analyzed companies.

Company	Symbol	Sector	Quotes[1]	Trades[2]	Spread[3]
Alphabet Inc.	GOOG	Information Technology (IT)	164489	19029	\$0.40
Mastercard Inc.	MA	Information Technology (IT)	98909	6977	\$0.38
CME Group Inc.	CME	Financials (F)	98188	3032	\$1.08
Goldman Sachs Group Inc.	GS	Financials (F)	160470	26227	\$0.11
Transocean Ltd.	RIG	Energy (E)	107092	11641	\$0.12
Apache Corp.	APA	Energy (E)	103074	8889	\$0.13

[1] Average number of quotes from 9:40:00 to 15:50:00 New York time during 2008.
[2] Average number of trades from 9:40:00 to 15:50:00 New York time during 2008.
[3] Average spread from 9:40:00 to 15:50:00 New York time during 2008.

magnitude of the guaranteed precision is much larger, usually one second or a few hundreds of milliseconds [6,10]. In several papers are used different time definitions (calendar time, physical time, event time, trade time, tick time) [10,18,32]. The TAQ data used in the analysis has the characteristic that the trades and quotes can not be directly related due to the time stamp resolution of only one second [39]. Hence, it is impossible to match each trade with the directly preceding quote. However, using a classification for the trade signs, we can compute trade signs in two scales: trade time scale and physical time scale.

The trade time scale is increased by one unit each time a transaction happens. The advantage of this count is that limit orders far away in the order book do not increase the time by one unit. The main outcome of trade time scale is its "smoothing" of data and the aggregational normality [10].

The physical time scale is increased by one unit each time a second passes. This means that computing the responses in this scale involves sampling [18,39], which has to be done carefully when dealing for example with several stocks with different liquidity. This sampling is made in the trade signs and in the midpoint prices.

Facing the impossibility to relate midpoint prices and trade signs with the TAQ data in trade time scale, we will use the midpoint price of the previous second with all the trade signs of the current second. This will be our definition of trade time scale analysis for the response function analysis.

For physical time scale, as we can sampling, we relate the unique value of midpoint price of a previous second with the unique trade sign value of the current second.

Thus, trade sign values will be used in trade time scale and physical time scale and returns will be only used in physical time scale.

3.1.1 Trade time scale

We use the trade sign classification in trade time scale proposed by S. Wang et al. in [39] and used in [34,35,38] that reads

$$\varepsilon^{t}\left(t,n\right)=\begin{cases}\operatorname{sgn}\left(S\left(t,n\right)-S\left(t,n-1\right)\right), & \text{if}\\ \quad S\left(t,n\right)\neq S\left(t,n-1\right) & \\ \quad\varepsilon\left(t,n-1\right), & \text{otherwise}\end{cases}\tag{1}$$

$\varepsilon^{t}\left(t,n\right)=+1$ implies a trade triggered by a market order to buy, and a value $\varepsilon^{t}\left(t,n\right)=-1$ indicates a trade triggered by a market order to sell.

In the second case of the classification, if two consecutive trades with the same trading direction did not exhaust all the available volume at the best quote, the trades would have the same price, and in consequence they will have the same trade sign.

With this classification we obtain trade signs for every single trade in the data set. According to [39], the average accuracy of the classification is 85% for the trade time scale.

TAQ time step is one second, and as it is impossible to find the correspondences between trades and midpoint prices values inside a second step, We used the last midpoint price of every second as the representative value of each second. This introduce an apparent shift between trade signs and returns. In fact, we set the last midpoint price from the previous second as the first midpoint price of the current second, as explained in [39].

As we know the second in which the trades were made, we can relate the trade signs and the midpoint prices as shown in Fig. 1. For the trade time scale, they are in general, several midpoint prices in a second. For each second we select the last midpoint price value, and we relate it to the next second trades. In Fig. 1, the last midpoint price (circle) between the second -1 and 0 is related with all the trades (squares and triangles) in the second 0 to 1, and so on. It is worth to note, in the seconds that there are no changes in the quotes, it is used the value of the previous second (vertical line over the physical time interval). Thus, all the seconds in the open market time have a midpoint price value, and in consequence returns values. We assume that as long as they were not changes in the quotes, the midpoint price remain the same as the one of the previous second.

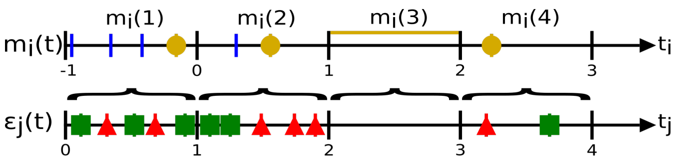

Figure 1. Sketch of data processing for trade time scale. In the midpoint price time line, the vertical lines represent the change in price of the quotes and the circles represent the last price change in a quote in a second. In the trade signs time line, the squares represent the buy market orders and the triangles represent the sell market orders. The midpoint price time line and the trade sign time line are shifted in one second.

We computed all the analysis for the trade time scale using Equations 6 and 1.

The methodology described is an approximation to compute the response in the trade time scale. A drawback in the computation could come from the fact that the return of a given second is composed by the contribution of small returns corresponding to each change in the midpoint price during a second. As we are assuming only one value for the returns in each second, we are supposing all the returns in one second interval to be positive or negative with the same magnitude, which could not be the case. This could increase or decrease the response signal at the end of the computation.

Figure 2 illustrate with one example this point. Suppose one second interval, in which they are three different midpoint prices, and as result, three different returns for this three midpoint price values. Furthermore, consider that the volume of limit orders that have the corresponding midpoint prices are the same in the bid and in the ask (so the returns have the same magnitude). In the case of the top left (top right) sketch, all the changes are due to the rise (decrease) of the midpoint price, that means, consumption of the best ask (bid), so all the contributions of the individual returns in the second are positive (negative), and in consequence, the net return is positive (negative). In the case of the bottom, the changes are due to a combination of increase and decrease of the midpoint price, so in the end the individual returns sum up to a net return, which can be positive or negative, depending of the type of midpoint price values in the interval. Thus, in this case, we are assuming at the end that all the returns were positive or negative, what probably was not the case, and in consequence will increase or decrease the real value of the net return.

In all the cases we choose the last change in the midpoint price in a second interval as described before (Fig. 1). We use this method knowing that the variation in one second of the midpoint price is not large (in average, the last midpoint price of a second differ with the average midpoint of that second in 0.007%), so it can give us valuable information about the response functions.

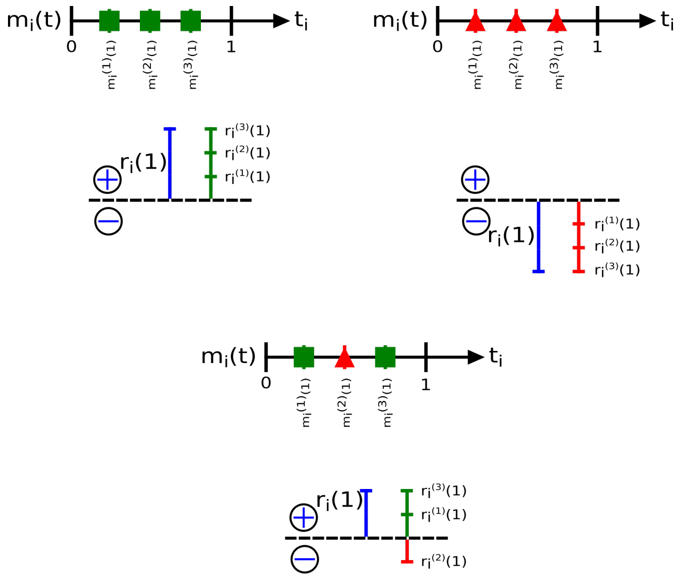

Figure 2. Sketch of the return contributions from every midpoint price change in a second. The squares represent the rise of the price of the midpoint price and the triangles represent the decrease of the price of the midpoint price. We illustrate three cases: (top left) the changes of the midpoint prices and return are due to the rise of the prices, (top right) the changes of the midpoint prices and return are due to the decrease of the prices, and (bottom) the changes of the midpoint prices and return are due to a combination of rise and decrease of the prices.

3.1.2 Physical time scale

We use the trade sign definition in physical time scale proposed by S. Wang et al. in [39] and used in [34,38], that depends on the classification in Eq. 1 and reads

$$\varepsilon^p(t) = \begin{cases} \operatorname{sgn}\left(\sum_{n=1}^{N(t)} \varepsilon^t(t,n)\right), & \text{If } N(t) > 0 \\ 0, & \text{If } N(t) = 0 \end{cases} \qquad (2)$$

Where $N(t)$ is the number of trades in a second interval. $\varepsilon^p(t) = +1$ implies that the majority of trades in second t were triggered by a market order to buy, and a value $\varepsilon^p(t) = -1$ indicates a majority of sell market orders. In this definition, they are two ways to obtain $\varepsilon^p(t) = 0$. One way is that in a particular second there is not trades, and then no trade sign. The other way is that the addition of the trade signs ($+1$ and -1) in a second be equal to zero. In this case, there is a balance of buy and sell market orders.

Market orders show opposite trade directions to limit order executed simultaneously. An executed sell limit order corresponds to a buyer-initiated market order. An executed buy limit order corresponds to a seller-initiated market order.

As in the trade time scale, in the physical time scale we use the same strategy to obtain the midpoint price for every second, so all the seconds in the open market time have a midpoint price value. It is worth to note again, that even if the second does not have a change in quotes, it will has still a midpoint price value and a return value.

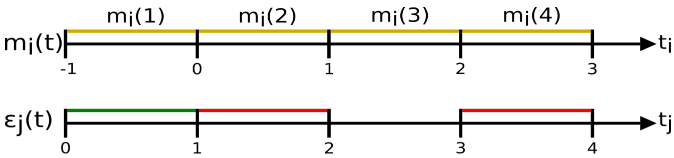

Figure 3. Sketch of data processing for physical time scale. In the midpoint price time line, the horizontal lines between seconds represent the midpoint prices. In the trade signs time line, the horizontal lines between seconds represent the trade sign values. The midpoint price time line and the trade sign time line are shifted in one second.

In this case we do not compare every single trade sign in a second, but the net trade sign obtained for every second with the definition. This can be seen in Fig. 3, we related the midpoint price of the previous second with the trade sign of the current second.

According to [39], this definition has an average accuracy up to 82% in the physical time scale.

3.2 Response function definitions

The average of the best ask and the best bid is the midpoint price, which is defined as [5,7,13,28,31]

$$m(t) = \frac{a(t) + b(t)}{2} \quad (3)$$

As the midpoint price depends on the quotes, it changes if the quotes change. The midpoint price grows if the best ask or the best bid grow. On the other hand, the midpoint price decreases if the best ask or the best bid decrease.

Price changes are typically characterized as returns. If one denotes $S(t)$ the price of an asset at time t, the return $r(t)$, at time t and time lag τ is simply the relative variation of the price from t to $t+\tau$ [5,10,22,23,24,30],

$$r^g(t,\tau) = \frac{S(t+\tau) - S(t)}{S(t)} \quad (4)$$

It is also common to define the returns as [3,5,10,12, 13,15,16,19,26,29]

$$r^l(t,\tau) = \ln S(t+\tau) - \ln S(t) = \ln \frac{S(t+\tau)}{S(t)} \quad (5)$$

Equation 4 and Eq. 5 coincide if τ is small enough [5, 10].

At longer timescales, midpoint prices and transaction prices rarely differ by more than half the spread. The midpoint price is more convenient to study because it avoids problems associated with the tendency of transaction prices to bounce back and forth between the best bid and ask [13].

We define the returns via the midpoint price as

$$r(t,\tau) = \frac{m(t+\tau) - m(t)}{m(t)} \quad (6)$$

The distribution of returns is strongly non-Gaussian and its shape continuously depends on the return period τ. Small τ values have fat tails return distributions [5].

The trade signs are defined for general cases as

$$\varepsilon(t) = \text{sign}(S(t) - m(t-\delta)) \quad (7)$$

where δ is a positive time increment. Hence we have

$$\varepsilon(t) = \begin{cases} +1, \text{ If } S(t) \text{ is higher than the last } m(t) \\ -1, \text{ If } S(t) \text{ is lower than the last } m(t) \end{cases} \quad (8)$$

$\varepsilon(t) = +1$ indicates that the trade was triggered by a market order to buy and a trade triggered by a market order to sell yields $\varepsilon(t) = -1$ [5,7,19,25,33].

It is well-known that the series of the trade signs on a given stock exhibit large autocorrelation. A very plausible explanation of this phenomenon relies on the execution strategies of some major brokers on given markets. These brokers have large transactions to execute on the account of some clients. In order to avoid market making move because of an inconsiderable large order, they tend to split large orders into small ones [10].

In [3] they use a market response R^i_τ defined as

$$R^{ij}_\tau = \mathbf{E}\left[(\ln S_i(t+\tau) - \ln S_i(t))\, \varepsilon_j(t)\right] \quad (9)$$

measures the average price change of contract i at time $t+\tau$, after experiencing a sign imbalance $\varepsilon_j(t)$ in contract j at time t.

The response function is used to study the mutual dependence between stocks. In [4,5,7], Bouchaud et al. use a self-response function that only depends on the time lag τ

$$R(\tau) = \langle (S_{n+\tau} - S_n) \cdot \varepsilon_n \rangle_{trades} \quad (10)$$

Where ε_n is the sign of the n^{th} trade and the price S_n is defined as the midpoint just before the n^{th} trade. The quantity $R(\tau)$ measures how much, on average, the price moves up (down) at time τ conditioned to a buy (sell) order at time zero. They found for France Telecom that $R(\tau)$ increases by a factor 2 between $\tau = 1$ and $\tau \approx 1000$, before decreasing back. Using larger τ, they also confirm that $R(\tau)$ decreases, and even becomes negative beyond $\tau \approx 5000$.

In [16], the price impact function, $f(\cdot)$, is defined as the average price response due to a transaction as a function of the transaction's volume v_i

$$f(v_i) \equiv \mathbf{E}\left[\varepsilon_i r^l_i | v_i\right] \quad (11)$$

Empirically they found the function is highly concave. The curvature of $f(v_i)$ is entirely due to the probability that a transaction causes a nonzero impact - the larger the size of the transaction, the larger the probability. In [28], the response function is given by

$$R(V,\tau) = \langle \varepsilon(t) \cdot [m(t+\tau) - m(t)] | V \rangle \tag{12}$$
$$= R(\tau) \ln V \tag{13}$$

They found that $R(t)$ for three French stocks first increases from $\tau = 10s$ to a few hundred seconds, and then appears to decrease back to a finite value.

In a later work [19,39], S. Wang et al. and Grimm et al. use the logarithmic return for stock i and time lag τ, defined via the midpoint price $m_i(t)$. The cross-response function is then defined as

$$R_{ij}(\tau) = \langle r_i(t-1,\tau) \cdot \varepsilon_j(t) \rangle_t \tag{14}$$

They found that in all cases, an increase to a maximum is followed by a decrease. The trend is eventually reversed.

Finally, in [36], S. Wang et al. define the response function as

$$R_{ij} = \left\langle \left(\ln m_i^{(f)}(t_j) - \ln m_i^{(p)}(t_j) \right) \cdot \varepsilon_j(t_j) \right\rangle_{t_j} \tag{15}$$

For the price change of stock i caused by a trade of stock j.

Here, $m_i^{(p)}(t_j)$ is the midpoint price of stock i previous to the trade of stock j at its event time t_j and $m_i^{(f)}(t_j)$ is the midpoint price of stock i following that trade.

The difference between the definition in [39] and in [36], is that [39] measures how a buy or sell order at time t influences on average the price at a later time $t+\tau$. The physical time scale was chosen since the trades in different stocks are not synchronous (TAQ data). In [36], it was used a response function on a trade time scale (Totalview data), as the interest is to analyze the immediate responses. In [36] the time lag τ is restricted to one, such that the price response quantifies the price impact of a single trade.

4 Price response function implementations

The main objective of this work is to analyze the response functions. In general we define the self- and cross-response functions in a correlated financial market as

$$R_{ij}^{scale}(\tau) = \left\langle r_i^{scale}(t-1,\tau) \cdot \varepsilon_j^{scale}(t) \right\rangle_{scale} \tag{16}$$

where the index i and j correspond to stocks in the market, r_i^{scale} is the return of the stock i in a time lag τ in the corresponding scale and ε_j^{scale} is the trade sign of the stock j in the corresponding scale. The subscript and superscript *scale* refer to the time scale used, whether physical time scale or trade time scale. Finally, we average the product over the physical time or trade time depending on the time scale.

We use the returns and the trade signs to define three response functions: trade time scale response, physical time scale response and activity response.

To compare the three response functions, we define the following quantities

$$E_{j,d}(t) = \sum_{n=1}^{N(t)} \varepsilon_{j,d}^{t}(t,n) \tag{17}$$
$$E_{j,d}(t) = \mathrm{sgn}(E_{j,d}(t)) \cdot |E_{j,d}(t)| \tag{18}$$
$$\varepsilon_{j,d}^{p}(t) = \mathrm{sgn}(E_{j,d}(t)) \tag{19}$$

Where the subscript d refers to the days used in the response computation.

In Sect. 4.1 we analyze the responses functions in trade time scale, in Sect. 4.2 we analyze the responses functions in trade time scale and in Sect. 4.3 we define a response function to analyze the influence of the frequency of trades in a second.

4.1 Response functions in trade time scale

We define the self- and cross-response functions in trade time scale, using the trade signs in trade time scale and the returns in physical time scale. The response is

$$R_{ij}^{t}(\tau) = \left\langle r_i^{p}(t-1,\tau) \cdot \varepsilon_j^{t}(t,n) \right\rangle_t \tag{20}$$

where the superscript t refers to the trade time scale. However, to be explicit with the way the averaging is made, the function reads

$$R_{ij}^{t}(\tau) = \frac{1}{\sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} N_{j,d}(t)} \cdot \sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} \sum_{n=1}^{N(t)} r_{i,d}^{p}(t-1,\tau) \cdot \varepsilon_{j,d}^{t}(t,n) \tag{21}$$
$$= \sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau) \cdot \frac{\sum_{n=1}^{N(t)} \varepsilon_{j,d}^{t}(t,n)}{\sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} N_{j,d}(t)}$$
$$= \sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau) \cdot \mathrm{sgn}(E_{j,d}(t)) \cdot w_{j,d}^{t}(t) \tag{22}$$

Where

$$w_{j,d}^{t}(t) = \frac{|E_{j,d}(t)|}{\sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} N_{j,d}(t)} \tag{23}$$

is a weight function that depends on the normalization of the response.

To compute the response functions in trade time scale, we used all the trade signs during a day in market time. As we can not associate an individual midpoint price with their corresponding trade signs, all the trade signs in one second are associated with the midpoint price of the previous second. As τ depends on the midpoint price, even if we are using trade signs in trade time scale, the value of τ is in seconds.

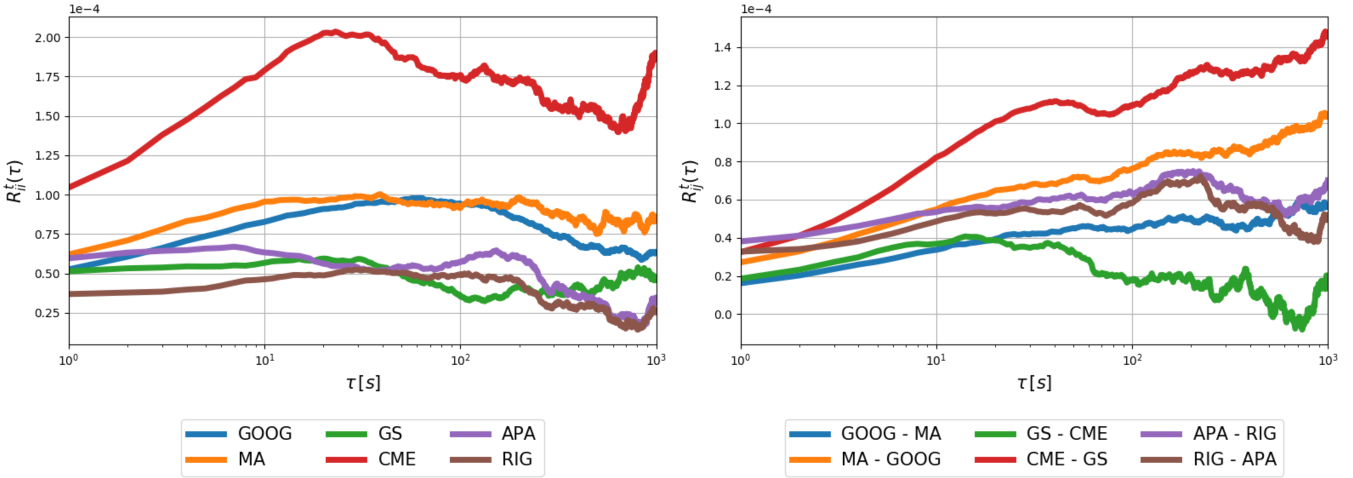

Figure 4. Self- and cross-response functions $R^t_{ij}(\tau)$ in 2008 versus time lag τ on a logarithmic scale in trade time scale. Self-response functions (left) of individual stocks and cross-response functions (right) of stock pairs from the same economic sector.

The results of Fig. 4 show the self- responses of the six stocks used in the analysis and the cross-responses for pairs of stocks representing three different economic sectors.

The self-response functions increase to a maximum and then slowly decrease. In some stocks this behavior is more pronounced than in others. For our selected tickers, a time lag of $\tau = 10^3$ is enough to see an increase to a maximum followed by a decrease. Thus, the trend in the self-response functions is eventually reversed. On the other hand, the cross-response functions have smaller signal strength than the self-response functions. For our cross-response functions of stocks in the same sectors, some couples exhibit the increase-decrease behavior inside a time lag of $\tau = 10^3$. Other couples seems to need a larger time lag to reach the decrease behavior.

4.2 Response functions in physical time scale

One important detail to compute the market response in physical time scale is to define how the averaging of the function will be made. This, because the response functions highly differ when we include or exclude $\varepsilon^p_j(t) = 0$ [39]. The cross-responses including $\varepsilon^p_j(t) = 0$ are weaker than the excluding ones due to the omission of direct influence of the lack of trades. However, either including or excluding $\varepsilon^p_j(t) = 0$ does not change the trend of price reversion versus the time lag, but it does affect the response function strength [38].

Regarding the definition of the response functions in- and excluding $\varepsilon^p_j(t) = 0$, the general averaging is

$$R^{(\text{inc. }0)}_{ij}(\tau) = \frac{\sum_{t=1}^{T_j+T_{j,n}} r^p_i(t-1,\tau)\cdot\varepsilon^p_j(t)}{T_j+T_{j,n}} \tag{24}$$

$$R^{(\text{exc. }0)}_{ij}(\tau) = \frac{\sum_{t=1}^{T_j} r^p_i(t-1,\tau)\cdot\varepsilon^p_j(t)}{T_j} \tag{25}$$

Where the superscript inc. and exc. refers to including and excluding $\varepsilon^p_j(t) = 0$. For stock j, T_j is the total trading time of stock j and $T_{j,n}$ is the total time of lack of trading or buy sell balance. The numerators in Eqs. 24 and 25 are the same, while the denominators differ [38].

Hence,

$$R^{(\text{inc. }0)}_{ij}(\tau) = f_j\cdot R^{(\text{exc. }0)}_{ij}(\tau) \tag{26}$$

Where the relative trading frequency is defined as [38]

$$f_j = \frac{T_j}{T_j+T_{j,n}} \tag{27}$$

The most frequently traded stocks have $f_i = 1$, because the time $T_{j,n}$ is zero. According to Eq. 26, the cross-response including $\varepsilon^p_j(t) = 0$ is the one excluding $\varepsilon^p_j(t) = 0$ scaled by a proper probability.

Then, we will only take in account the cross-response function excluding $\varepsilon^p_j(t) = 0$.

We define the self- and cross-response functions in physical time scale, using the trade signs and the returns in physical time scale. The response is

$$R^p_{ij}(\tau) = \left\langle r^p_i(t-1,\tau)\cdot\varepsilon^p_j(t)\right\rangle_p \tag{28}$$

where the superscript p refers to the physical time scale. The corresponding explicit expression reads

$$R_{ij}^{p}(\tau)=\frac{1}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}\eta\left[\varepsilon_{j,d}^{p}(t)\right]} \cdot\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\varepsilon_{j,d}^{p}(t)\cdot\eta\left[\varepsilon_{j}^{p}(t)\right] \tag{29}$$

$$=\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\frac{\varepsilon_{j,d}^{p}(t)\cdot\eta\left[\varepsilon_{j,d}^{p}(t)\right]}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}\eta\left[\varepsilon_{j,d}^{p}(t)\right]}$$

$$=\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\operatorname{sgn}\left(E_{j,d}(t)\right)\cdot w_{j,d}^{p}(t) \tag{30}$$

Where

$$\eta(x)=\begin{cases}1, & \text{If } x\neq 0\\ 0, & \text{otherwise}\end{cases} \tag{31}$$

take only in account the seconds with trades and

$$w_{j,d}^{p}(t)=\frac{\eta\left[\operatorname{sgn}\left(E_{j,d}(t)\right)\right]}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}\eta\left[\operatorname{sgn}\left(E_{j,d}(t)\right)\right]} \tag{32}$$

is a weight function that depends on the normalization of the response.

The results showed in Figure 5 are the self- and cross-response functions in physical time scale. For the self-response functions we can say again that in almost all the cases, an increase to a maximum is followed by a decrease. Thus, the trend in the self- and cross-response is eventually reversed. In the cross-response functions, we have a similar behavior with the previous subsection, where the time lag in some pairs was not enough to see the decrease of the response.

Compared with the response functions in trade time scale, the response functions in physical time scale are stronger.

4.3 Activity response functions in physical time scale

Finally, we define the activity self- and cross-response functions in physical time scale, using the trade signs and the returns in physical time scale. We add a factor $N_{j,d}(t)$ to check the influence of the frequency of trades in a second in the response functions. The activity response is

$$R_{ij}^{a}(\tau)=\left\langle r_{i}^{p}(t-1,\tau)\cdot\varepsilon_{j}^{p}(t)\cdot N(t)\right\rangle_{p} \tag{33}$$

where the superscript a refers to the activity response function. The corresponding explicit expression reads

$$R_{ij}^{a}(\tau)=\frac{1}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}N_{j,d}(t)} \cdot\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\varepsilon_{j,d}^{p}(t)\cdot N_{j,d}(t) \tag{34}$$

$$=\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\frac{\varepsilon_{j,d}^{p}(t)\cdot N_{j,d}(t)}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}N_{j,d}(t)}$$

$$=\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\operatorname{sgn}\left(E_{j,d}(t)\right)\cdot w_{j,d}^{a}(t) \tag{35}$$

Where

$$w_{j,d}^{a}(t)=\frac{N_{j,d}(t)}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}N_{j,d}(t)} \tag{36}$$

is a weight function that depends on the normalization of the response.

As $E_{j,d}(t)$ is the sum of $+1$ and -1 in one second and $N_{j,d}(t)$ is the number of trades in a second, $N_{j,d}(t)\geq E_{j,d}(t)$. $N_{j,d}(t)=E_{j,d}(t)$ only when all the trades in a second are buys $(+1)$.

The trade weight reduces noises, The physical weight gives every step the same weight, and the activity weight emphasizes seconds with large activity.

In Figure 6, we can see how the three responses have approximately the same shape, but the strength of the signal varies depending on the definition. The frequency of trades have a large influence in the responses.

As predicted by the weights, the event response is weaker than the physical response, and the activity response is the strongest response.

In the three curves in the figure can be seen the increase-decrease behavior of the response functions.

5 Time shift response functions

The relative position between returns and trade signs directly impact the result of the response functions. Shifting the values to the right or to the left either in trade time scale or physical time scale have approximately the same effect.

To test this claim, we used the definition of the response function from [39] that is showed in Equation 14 and added a parameter t_s that shifts the position between returns and trade signs. To see the impact of the time shift we analyzed the stocks showed in Table 1 in the year 2008. We used different time shifts in the response function

$$R_{ij}^{s,scale}(\tau)=\left\langle r_{i}^{scale}(t-t_s,\tau)\cdot\varepsilon_{j}^{scale}(t)\right\rangle_{scale} \tag{37}$$

where the index i and j correspond to stocks in the market, r_{i}^{scale} is the return of the stock i in a time lag

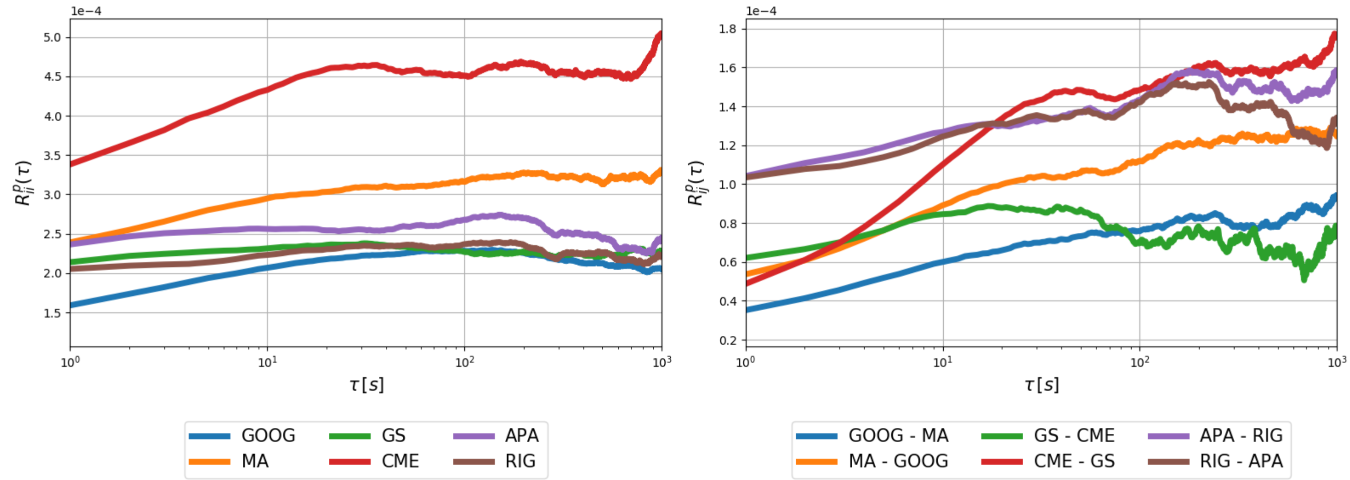

Figure 5. Self- and cross-response functions $R^p_{ij}(\tau)$ excluding $\varepsilon^p_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale in physical time scale. Self-response functions (left) of individual stocks and cross-response functions (right) of stock pairs from the same economic sector.

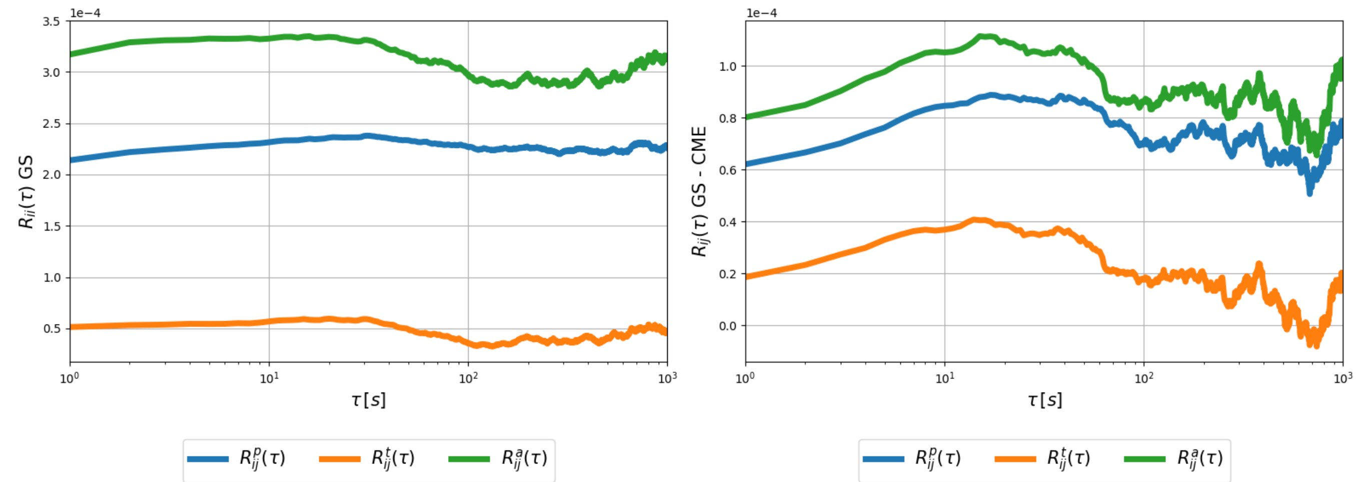

Figure 6. Self- and cross-response functions $R^{scale}_{ij}(\tau)$ excluding $\varepsilon^p_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale. Self-response functions (left) of Goldman Sachs Group Inc. stock and cross-response functions (right) of Goldman Sachs Group Inc.-CME Group Inc. stocks.

τ with a time shift t_{shift} in the corresponding scale and ε^{scale}_j is the trade sign of the stock j in the corresponding scale. The subscript and superscript *scale* refer to the time scale used, whether physical time scale or trade time scale and $R^{s,scale}_{ij}$ is the time shift response function, where the superscript s refers to the time shift. Finally, we average the product over the physical time or trade time depending on the time scale.

We compute the response functions according to two cases. In one case we set τ to a constant value and vary t_s, and in the other case we set t_s to a constant value and vary τ.

In Sect. 5.1 we analyze the influence of the time shift between the trade signs and returns in trade time scale and in Sect. 5.2 we analyze the influence of the time shift between the trade signs and returns in physical time scale.

5.1 Trade time scale shift response functions

In the trade time scale we compute the response function

$$R^{s,t}_{ij}(\tau) = \left\langle r^t_i(t - t_s, \tau) \cdot \varepsilon^t_j(t) \right\rangle_t \tag{38}$$

In this case for r^t_i, we associate all the trade signs to a return value and create pseudo midpoint price values in trade time scale. Then, we shift the trade signs and the returns by trades. Hence, the time lag and time shift are in trade time scale.

In Fig. 7, it can be seen the response functions results for fixed τ values while t_s is variable. In the different τ values figures, the results are almost the same. The response functions are zero either if the time shift is larger than τ, or if the time shift is smaller than zero. For values between zero and τ there is a peak in a position related to τ.

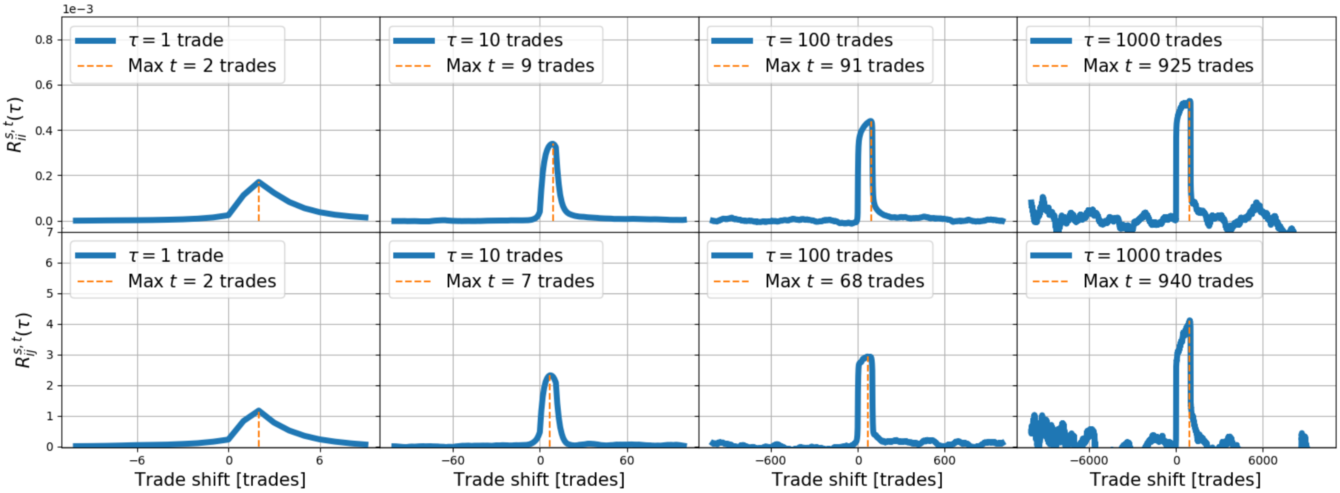

Figure 7. Self-response functions $R^t_{ii}(\tau)$ in 2008 versus shift for the Transocean Ltd. stock (top) and cross-response functions $R^t_{ij}(\tau)$ excluding $\varepsilon^t_j(t)=0$ in 2008 versus shift for the Transocean Ltd.-Apache Corp. stocks (bottom) in trade time scale.

The response function grows and decrease relatively fast. However, related on the time lag, there is a zone where the signal is different to zero.

We tested the response function for fixed time shift values while τ is variable. For every time shift (Fig. 8), both, self- and cross-response results are qualitatively the same. It can be seen that the response functions have a zero signal before the time shift. After the returns and trade signs find their corresponding order the signals grow. In comparison with the values obtained in Fig. 4, it looks like the response function values are stronger. However, this is an effect of the averaging of the functions. As the returns and trade signs are shifted, they are less values to average, and then the signals are stronger. Anyway, the figure shows the importance of the relation between the trade signs and returns to compute the response function.

5.2 Physical time scale shift response functions

In the physical time scale we compute the response function

$$R^{s,p}_{ij}(\tau)=\left\langle r^p_i(t-t_s,\tau)\cdot\varepsilon^p_j(t)\right\rangle_t \tag{39}$$

Similar to the results in Subsect. 5.1, Fig. 9 shows the responses functions for fixed τ values while t_s is variable. Again, the response functions are zero if the time shift is larger than the time lag, or if the time shift is smaller than zero. For every τ value, there is a peak. The peak grows and decay relatively fast. The response signal usually starts to grow in zero or a little bit earlier and grows to a value around to τ. In this zone the response functions are different to zero.

The results for fixed time shift values and variable time lag are shown in Fig. 10. The self- and cross-response results are qualitatively the same. As seen in the previous subsection, the response functions are zero before the time shift value. After the returns and the trade signs reach their order, the signals grow. The same effect of the apparent stronger signal can be seen here, and again, it is due to the averaging values.

The results in trade time scale and physical time scale can be explained understanding the dynamics of the market. A trade can or can not change the price of a ticker. Therefore, when a change in price happens, a change in midpoint price, and consequently in returns happens. Thus, it is extremely important to keep the order of the events and the relation between them. When we shift the trade signs and returns, this order is temporarily lost and as outcome the signal does not have any meaningful information. When the order is recovered during the shift, the signal grows again, showing response function values different to zero. In this section we were interested only in the order (shift) and not in the responses values, which were analyzed in Sect. 4.

Then the question is what is the ideal time shift to compute the response functions. Our approach in Sect. 4 takes in account that the changes in the quotes are the ones that attract the agents to buy or sell their shares. Hence, they directly impact the trade signs. According to the results, the response can take up to two time steps in the corresponding scale to react to the change in quotes. Thus, a time shift larger than two time steps makes no sense. On the other hand, in the case of the physical time scale, where a sampling is used, to assure the selection of a midpoint price at the beginning of a second, it is a good strategy to use the last midpoint of the previous second as the first midpoint price of the current second. In this case an apparent one second shift is used between returns and trade signs.

6 Short and long response functions

Regarding Equation 4, we use a time lag τ in the returns to see the gains or loses in a future time. However, the

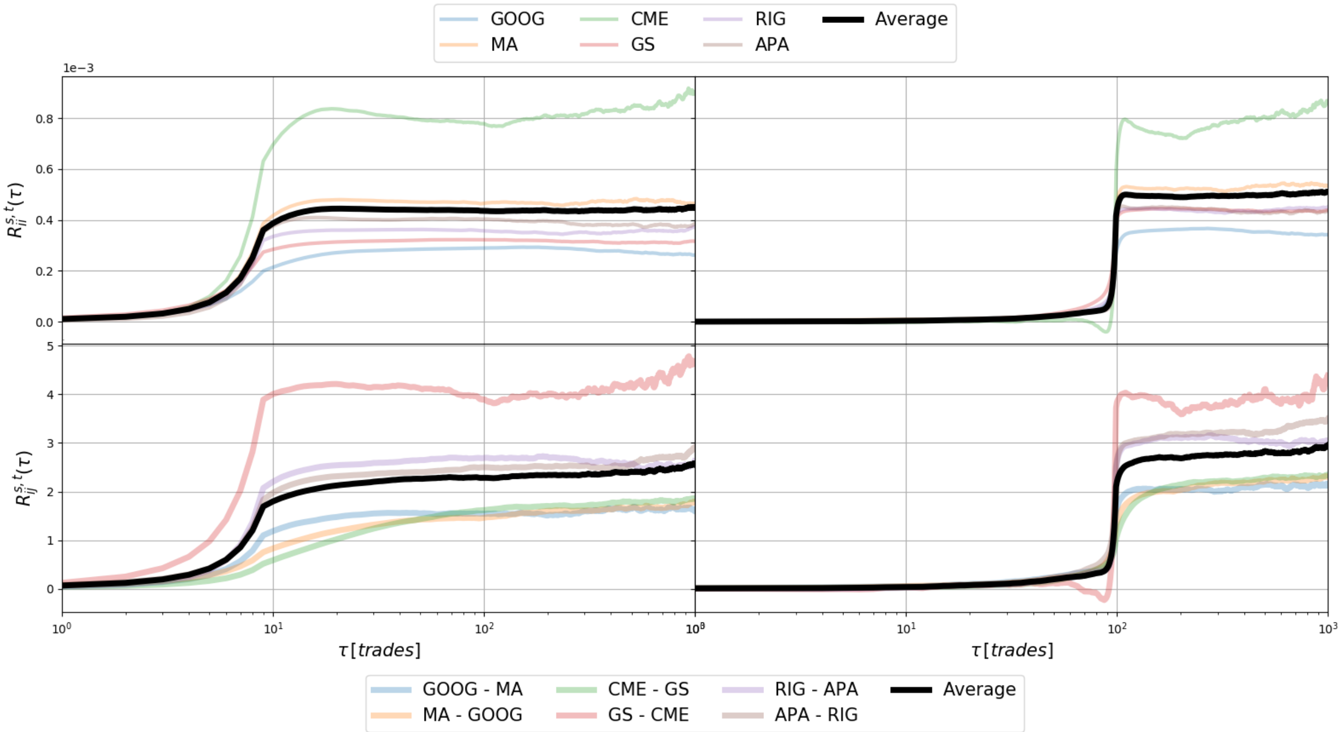

Figure 8. Self- and cross-response functions $R^t_{ij}(\tau)$ in 2008 versus time lag τ on a logarithmic scale for different shifts in trade time scale. Self-response functions (left) of individual stocks and cross-response functions (right) of stocks pairs from the same economic sector.

strength of the return in the time lag should not be equal along its length. Then, we divide the full range time lag τ in an immediate time lag and in a late time lag as show in Fig. 11, where

$$\tau = \tau' + (\tau - \tau') \tag{40}$$

for $\tau' < \tau$. This distinguish the returns depending in the time lag as the short (immediate) return τ' with the long return $\tau - \tau'$.

To use the short and long time lag, we rewrite the returns in physical time scale as

$$\begin{aligned} r_i^{sl,p}(t,\tau) &= \ln\left(\frac{m_i(t+\tau)}{m_i(t)}\right) \\ &= \ln\left(\frac{m_i(t+\tau)}{m_i(t+\tau')}\cdot\frac{m_i(t+\tau')}{m_i(t)}\right) \\ &= \ln\left(\frac{m_i(t+\tau)}{m_i(t+\tau')}\right) + \ln\left(\frac{m_i(t+\tau')}{m_i(t)}\right) \\ &\approx \frac{m_i(t+\tau)-m_i(t+\tau')}{m_i(t+\tau')} + \frac{m_i(t+\tau')-m_i(t)}{m_i(t)} \end{aligned} \tag{41}$$

where sl refers to short-long and the second term of the right part is constant with respect to τ. Replacing Equation 41 in the response function (Eq. 28) we have

$$\begin{aligned} R_{ij}^{sl,p}(\tau) &= \left\langle r_i^{sl,p}(t-1,\tau)\cdot\varepsilon_j^p(t)\right\rangle_p \\ &\approx \left\langle \frac{m_i(t-1+\tau)-m_i(t-1+\tau')}{m_i(t-1+\tau')}\cdot\varepsilon_j(t)\right\rangle_p \\ &+ \left\langle \frac{m_i(t-1+\tau')-m_i(t-1)}{m_i(t-1)}\cdot\varepsilon_j(t)\right\rangle_p \end{aligned} \tag{42}$$

Where the first term in the right side of Equation 42 is the long response and the right term is the short response. Again, the right term of Equation 42 is independent of τ.

The results in Fig. 12 show the short response, the long response, the addition of the short response and long response (Sum), the original response, a random response and the value of τ'.

The main signal of the response function come from the short response. Depending on the stock and the value of τ' the long response can increase or decrease the short response signal, but in general the long response does not give a significant contribution to the complete response.

Before τ', the short response and long response are the same, as the self and cross-response definition do not define values smaller than τ', so it is computed as the original response. In the figure, the curves of the short and long response are under the curve of the original response. After τ', the short response is a strong constant signal. On the other hand, the long response immediately fades, showing the small contribution to the final response. To compare

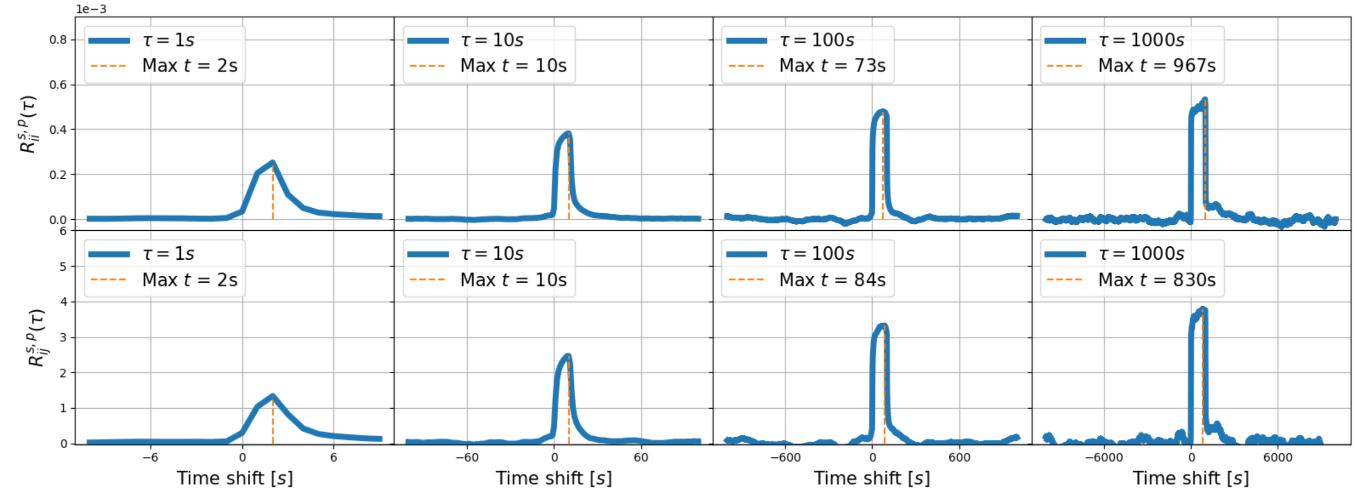

Figure 9. Self-response functions $R_{ii}^{p}(\tau)$ excluding $\varepsilon_{i}^{p}(t) = 0$ in 2008 versus shift for the Transocean Ltd. stock (top) and cross-response functions $R_{ij}^{p}(\tau)$ excluding $\varepsilon_{j}^{p}(t) = 0$ in 2008 versus shift for the Transocean Ltd.-Apache Corp. stocks (bottom) in physical time scale.

the significance of the long response, We added a random response made with the trade signs used to compute the response but with a shuffle order. The long response and the random response are comparable, and show how the long response is not that representative in the final response. If we add the short and long response, we obtain the original response. In Fig. 12, the original response (red line) has the same shape to the addition of the short and long response (green line).

For the response functions that show the increase-decrease behavior in between the time lag $\tau = 10^3$, the peak is usually between $\tau = 10^1$ and $\tau = 10^2$. In these cases the long response are always negative after the τ' value and is comparable in magnitude with the random signal.

On the other hand, the response functions that requires a bigger time lag to show the increase-decrease behavior, have non negative long responses, but still they are comparable in magnitude with the random signal.

7 Spread impact

When we calculate the price response functions, the signal of the response depends directly on the analyzed stock. Thus, even if the responses functions are in the same scale, their values differ from one to another. We choose the spread to group 530 stocks in the NASDAQ stock market for the year 2008, and check if the average strength of the price self-response functions in physical time scale were similar for this groups. For each stock we computed the average spread for a year, and using this value we classified the groups.

We used three intervals to select the stocks groups to average the response functions ($s < 0.05\$$, $0.05\$ \leq s < 0.10\$$ and $0.10\$ \leq s < 0.40\$$). The detailed information of the groups can be seen in Appendix A.

In Fig. 13 can be seen the average response functions for the three groups. The response functions start at the bottom with the average response for the stocks with the smaller spreads (more liquid) and grow to the larger average response functions for the stocks with the larger spreads (less liquid). All the average response functions follows the increment to a maximum followed by a decrease described in Sect. 3 and 4.

The strength of the self-response function signals grouped by spread can be explained knowing that the response functions directly depend on trade signs. As long as the stock is liquid, the number of trade signs grow. Thus, at the moment of the averaging, the large amount of trades, reduces the response function signal. Therefore, the response function decrease as long as the liquidity grows. And as stated in the introduction the spread is negatively related to trading volume, hence, firms with more liquidity tend to have lower spreads.

8 Conclusion

We went into detail about the response functions in correlated financial markets. We define the trade time scale and physical time scale to compute the self- and cross-response functions for six companies with the largest average market capitalization for three different economic sectors of the S&P index in 2008. Due to the characteristics of the data used, we had to classify and sampling values to obtain the corresponding values in different time scales. The classification and sampling of the data had impact on the results, making them smoother or stronger, but always keeping their shape and behavior.

The response functions were analyzed according to the time scales. For trade time scale, the signal is weaker due to the large averaging values from all the trades in a year. In the physical time scale, the response functions had less noise and their signal were stronger. The activity in every second highly impact the responses. As the response

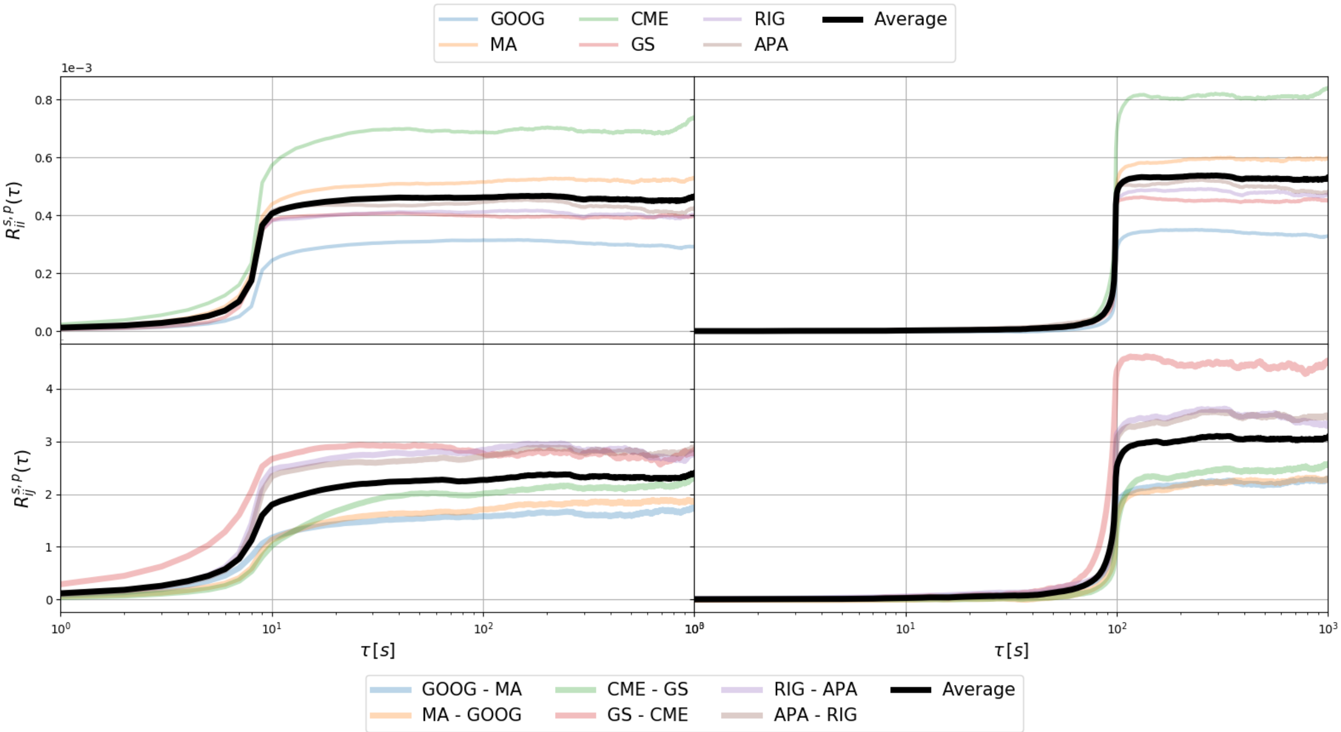

Figure 10. Self- and cross-response functions $R_{ij}^{p}(\tau)$ excluding $\varepsilon_j^p(t) = 0$ in 2008 versus time lag τ on a logarithmic scale for different shifts in physical time scale. Self-responses functions (left) of individual stocks and cross-response functions (right) of stocks pairs from the same economic sector.

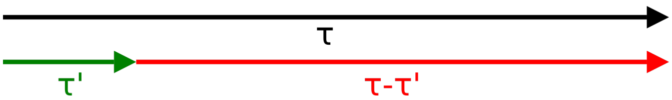

Figure 11. τ value divided in short and long time lag.

functions can not grow indefinitely with the time lag, they increase to a peak, to then decrease. It can be seen that the market needs time to react and revert the growing. In both time scale cases depending on the stocks, two characteristics behavior were shown. In one, the time lag was large enough to show the complete increase-decrease behavior. In the other case, the time lag was not enough, so some stocks only showed the growing behavior.

We modify the response function to add a time shift parameter. With this parameter we wanted to analyze the importance in the order of the relation between returns and trade signs. In trade time scale and physical time scale we found similar results. When we shift the order between returns and trade signs, the information from the relation between them is temporarily lost and as outcome the signal does not have any meaningful information. When the order is recovered, the response function grows again, showing the expected shape. We showed that this is not an isolated conduct, and that all the shares used in our analysis exhibit the same behavior. Thus, even if they are values of time shift that can give a response function signal, from the theory this time shift should be a value between $t_s = [0, 2]$.

We analyzed the impact of the time lag in the response functions. We divided the time lag in a short and long time lag. The response function that depended on the short time lag, showed a stronger response. The long response function depending on the stock could take negative and non-negative values. However, in general the influence were not intense.

Finally, we checked the spread impact in self-response functions. We divided 530 stocks from the NASDAQ stock market in five groups depending on the year average spread of every stock. The response functions signal were stronger for the group of stocks with the larger spreads and weaker for the group of stocks with the smaller spreads.

9 Author contribution statement

TG proposed the research. SMK and JCHL developed the method of analysis. The idea to look the time shift and to analyze the spread impact was due to JCHL, and the idea of the time lag analysis was due to SMK. JCHL carried out the analyisis. All the authors contributed equally to analyze the results and write the paper.

One of us (JCHL) acknowledges financial support from the German Academic Exchange Service (DAAD) with the program "Research Grants - Doctoral Programmes in Germany" (Funding programme 57381412)

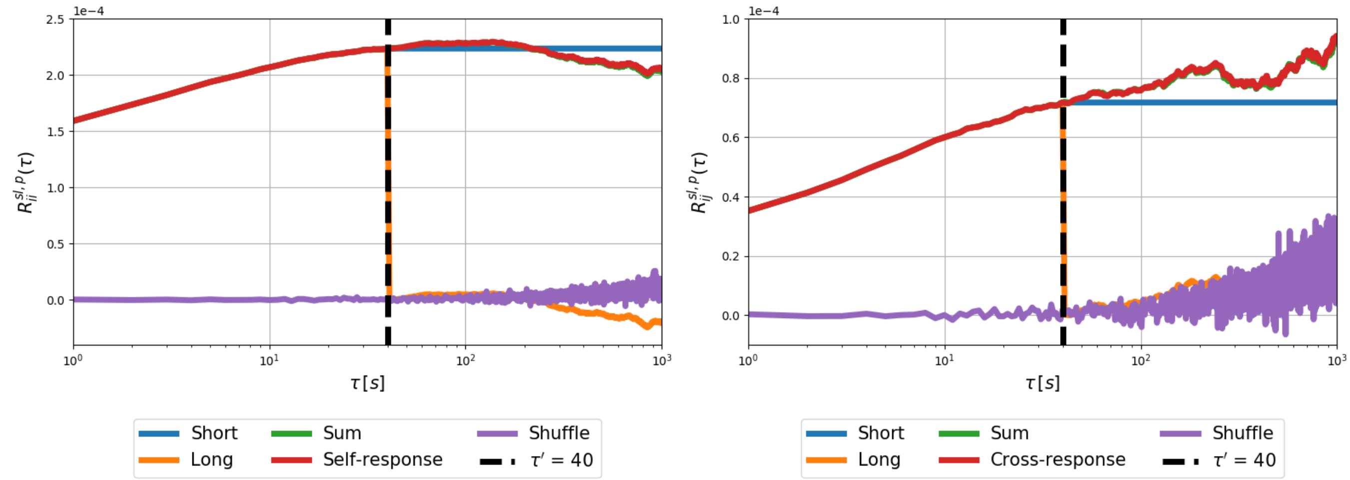

Figure 12. Self- and cross-response functions $R_{ij}^{sl,p}(\tau)$ excluding $\varepsilon_j^p(t) = 0$ in 2008 versus time lag τ on a logarithmic scale using a $\tau' = 40$ in physical time scale. Self-response functions (left) of Alphabet Inc. stock and cross-response functions (right) of Alphabet Inc.-Mastercard Inc. stocks.

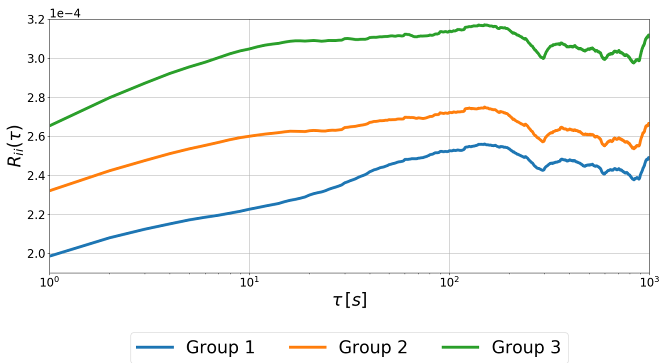

Figure 13. Average price self-response functions $R_{ii}^p(\tau)$ excluding $\varepsilon_i^p(t) = 0$ in 2008 versus time lag τ on a logarithmic scale in physical time scale for 530 stocks divided in three representative groups.

Appendix A NASDAQ stocks used to analyze the spread impact

We analyzed the spread impact in the response functions for 530 stocks from the NASDAQ stock market for the year 2008. In Table PONERREF, we listed the stocks in their corresponding spread group.

References

1. Hee-Joon Ahn, Jun Cai, Yasushi Hamao, and Richard Y.K Ho. The components of the bid–ask spread in a limit-order market: evidence from the tokyo stock exchange. *Journal of Empirical Finance*, 9(4):399 – 430, 2002.
2. Yakov Amihud and Haim Mendelson. The effects of beta, bid-ask spread, residual risk, and size on stock returns. *The Journal of Finance*, 44(2):479–486, 1989.
3. M Benzaquen, I Mastromatteo, Z Eisler, and J-P Bouchaud. Dissecting cross-impact on stock markets: an empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, Feb 2017.
4. J. P. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets, 2004.
5. Jean-Philippe Bouchaud. The subtle nature of financial random walks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026104, 2005.
6. Jean-Philippe Bouchaud, J. Doyne Farmer, and Fabrizio Lillo. How markets slowly digest changes in supply and demand. 2008.
7. Jean-Philippe Bouchaud, Yuval Gefen, Marc Potters, and Matthieu Wyart. Fluctuations and response in financial markets: the subtle nature of "random" price changes. *Quantitative Finance*, 4(2):176–190, Apr 2004.
8. Jean-Philippe Bouchaud, Marc Mézard, and Marc Potters. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, Aug 2002.
9. Carolyn Callahan, Charles Lee, and Teri Yohn. Accounting information and bid-ask spread. *Accounting Horizons*, 11:50–60, 01 1997.
10. Anirban Chakraborti, Ioane Toke, Marco Patriarca, and Frédéric Abergel. Econophysics: Empirical facts and agent-based models. *arXiv.org, Quantitative Finance Papers*, 09 2009.
11. Kee H Chung, Bonnie F [Van Ness], and Robert A [Van Ness]. Limit orders and the bid–ask spread. *Journal of Financial Economics*, 53(2):255 – 287, 1999.
12. Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.
13. J. Doyne Farmer, László Gillemot, Fabrizio Lillo, Szabolcs Mike, and Anindya Sen. What really causes large price changes? *Quantitative Finance*, 4(4):383–397, 2004.
14. J. Doyne Farmer, Paolo Patelli, and Ilija I. Zovko. The predictive power of zero intelligence in financial markets, 2003.

15. Xavier Gabaix, Parameswaran Gopikrishnan, Vasiliki Plerou, and H. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–70, 06 2003.
16. Austin Gerig. A theory for market impact: How order flow affects stock price. *arXiv.org, Quantitative Finance Papers*, 04 2008.
17. Lawrence R Glosten and Lawrence E Harris. Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1):123 – 142, 1988.
18. Jim E. Griffin and Roel C. A. Oomen. Sampling returns for realized variance calculations: Tick time or transaction time? *Econometric Reviews*, 27(1-3):230–253, 2008.
19. Stephan Grimm and Thomas Guhr. How spread changes affect the order book: comparing the price responses of order deletions and placements to trades. *The European Physical Journal B*, 92:1–11, 2018.
20. Fabrizio Lillo, J Farmer, and Rosario Mantegna. Master curve for price-impact function. *Nature*, 421:129–30, 02 2003.
21. Fabrizio Lillo and J. Doyne Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.
22. Michael C. Münnix, Rudi Schäfer, and Thomas Guhr. Compensating asynchrony effects in the calculation of financial correlations. *Physica A: Statistical Mechanics and its Applications*, 389(4):767 – 779, 2010.
23. Michael C. Münnix, Rudi Schäfer, and Thomas Guhr. Impact of the tick-size on financial returns and correlations. *Physica A: Statistical Mechanics and its Applications*, 389(21):4828 – 4843, 2010.
24. Michael C. Münnix, Rudi Schäfer, and Thomas Guhr. Statistical causes for the epps effect in microstructure noise. *International Journal of Theoretical and Applied Finance*, 14(08):1231–1246, 2011.
25. Vasiliki Plerou, Parameswaran Gopikrishnan, Xavier Gabaix, and H. Eugene Stanley. Quantifying stock-price response to demand fluctuations. *Phys. Rev. E*, 66:027104, Aug 2002.
26. Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luís A. Nunes Amaral, Thomas Guhr, and H. Eugene Stanley. Random matrix approach to cross correlations in financial data. *Phys. Rev. E*, 65:066126, Jun 2002.
27. Vasiliki Plerou, H Eugene Stanley, Xavier Gabaix, and Parameswaran Gopikrishnan. On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4(1):11–15, 2004.
28. Marc Potters and Jean-Philippe Bouchaud. More statistical properties of order books and price impact. 2002.
29. Bernd Rosenow. Fluctuations and market friction in financial trading. *International Journal of Modern Physics C*, 13(03):419–425, 2002.
30. Thilo A. Schmitt, Desislava Chetalova, Rudi Schäfer, and Thomas Guhr. Non-stationarity in financial time series: Generic features and tail behavior. *EPL (Europhysics Letters)*, 103(5):58003, sep 2013.
31. Eric Smith, J Doyne Farmer, László Gillemot, and Supriya Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.
32. Ioane Muni Toke. *"Market Making" in an Order Book Model and Its Impact on the Spread*, pages 49–64. Springer Milan, Milano, 2011.
33. Bence Tóth, Imon Palit, Fabrizio Lillo, and J. Doyne Farmer. Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51(C):218–239, 2015.
34. Shanshan Wang. Trading strategies for stock pairs regarding to the cross-impact cost, 2017.
35. Shanshan Wang and Thomas Guhr. Local fluctuations of the signed traded volumes and the dependencies of demands: a copula analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(3):033407, mar 2018.
36. Shanshan Wang, Sebastian Neusüß, and Thomas Guhr. Grasping asymmetric information in price impacts. *The European Physical Journal B*, 91(11):266, Nov 2018.
37. Shanshan Wang, Sebastian Neusüß, and Thomas Guhr. Statistical properties of market collective responses. *The European Physical Journal B*, 91(8):191, Aug 2018.
38. Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Average cross-responses in correlated financial markets. *The European Physical Journal B*, 89(9):207, Sep 2016.
39. Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Cross-response in correlated financial markets: individual stocks. *The European Physical Journal B*, 89(4), Apr 2016.