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

How to measure response functions in correlated financial markets

Juan Camilo 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 Insert your abstract here.

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

Possible paper names (I will erase this when we choose a name)

The following are possible names for the paper:

- Details that influence the response functions results.
- Influence of the details in the response function measurement.
- Response function measurement in correlated financial markets.
- Response function calculation in correlated financial markets.
- Influence of the methodology in response functions results.

Or we can use another.

1 Introduction

Shares are the equal parts in which the capital of a company is divided. A share is an indivisible unit of capital, expressing the ownership relationship betwen a company and a shareholder. The shareholder owns a percentage of the company depending on the amount of shares he has. Shares are issued in two moments: when the company is created and when the company want to raise funds.

The shares can be taken as an invesment, and receive dividends from them, or they can be traded at any time. This second possibility is the want that concerns us.

To trade the stocks exist markets where the buyers and sellers meet. sellers transfer (in exchange for money) the ownership of equities to buyers. This requires these two parties to agree on a price.

In a modern financial market, there is a double continous auction. To find possible buyers and sellers in the market, agents can place different types of instructions (known as 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, they are executed as fast as possible and only after the purchase of the stock is possible to know the exact price [7].

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 [7].

Limit orders often fail to result in an immediate transaction, and are stored in a queue called the limit order book. 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 $a(t)$, and a best (highest) bid to buy with price $b(t)$ [3]. These are also called the inside quotes or the best prices. The price gap between them is called the spread $s(t) = a(t) - b(t)$ [7,4,3].

The average of the best ask and the best bid is the midpoint price, which is defined as [4,7,3]

$$m(t) = \frac{a(t) + b(t)}{2} \tag{1}$$

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. This happen if someone buys and consumes all the volume of the sell limit order with the price of the best ask, or someone sets a buy limit order with a bigger price than the previous best bid, or there is a cancellation of the best ask.

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

On the other hand, the midpoint price decreases if the best ask or the best bid decrease. This happen if someone sells and consumes all the volume of the buy limit order with the price of the best bid, or someone sets a sell limit order with a lower price than the previous best bid, or there is a cancellation of the best bid.

The midpoint price will not change if there is no activity in the market.

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 $t+\tau$ [3,5],

$$r^{g}(t,\tau)=\frac{S(t+\tau)-S(t)}{S(t)} \tag{2}$$

It is also common to define the returns as [6,12,7,3,8,5,1,10]

$$r^{l}(t,\tau)=\ln S(t+\tau)-\ln S(t)=\ln\frac{S(t+\tau)}{S(t)} \tag{3}$$

Equation 2 and Eq. 3 coincide if τ is small enough [3,5].

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 [7].

We define the returns via the midpoint price as

$$r(t,\tau)=\frac{m(t+\tau)-m(t)}{m(t)} \tag{4}$$

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 [3].

Then we can expect three kind of values of the returns. The returns are positive values, when the midpoint price $m(t+\tau)>m(t)$, hence, there is a buy in the market or there is a cancellation of the best ask or an addition in the best bid during the time lag τ. The returns are negative values, when the midpoint price $m(t+\tau)<m(t)$, thus, there is a sell in the market, or there is a cancellation of the best bid or an addition in the best ask during the time lag τ. The returns are zero when there is no activity during the time lag τ.

The trade signs are defined for general cases as

$$\varepsilon(t)=\mathrm{sign}\left(S(t)-m(t-\delta)\right) \tag{5}$$

where δ is a positive time increment. Hence we have

$$\varepsilon(t)=\begin{cases}+1, \text{ If } S(t) \text{ is higher than the last } m(t)\\ -1, \text{ If } S(t) \text{ is lower than the last } m(t)\end{cases} \tag{6}$$

$\varepsilon(t)=+1$ indicates that the trade was triggered by a market order to buy and a trade triggered by a market order to sell yields $\varepsilon(t)=-1$ [11,4,3,14,10].

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 a given markets. These brokers have large transaction 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 [5].

The response function is used to study the mutual dependence between stocks. In [4,2], Bouchaud et al. use the response function that only depends on the time lag τ

$$R(\tau)=\left\langle\left(S_{n+\tau}-S_{n}\right)\cdot\varepsilon_{n}\right\rangle_{trades} \tag{7}$$

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 ($S_n\equiv m_{n^-}$). 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.

In a later work [19], S. Wang 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

$$R_{ij}(\tau)=\left\langle r_i(t-1,\tau)\cdot\varepsilon_j(t)\right\rangle_t \tag{8}$$

Finally, in [17], S. Wang et al. define the response function as

$$R_{ij}=\left\langle\left(\ln m_i^{(f)}(t_j)-\ln m_i^{(p)}(t_j)\right)\cdot\varepsilon_j(t_j)\right\rangle_{t_j} \tag{9}$$

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 [19] and in [17], is that [19] 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 [17], it was used a response function on a trade time scale (Totalview data), as the interest is to analyze the immediate responses. In [17] the time lag τ is restricted to one, such that the price response quantifies the price impact of a single trade.

The paper is organized as follows: in Sect. 2 we present our data set of stocks and describe the physical and trade time. We then analyze the definition of the response functions in Sect. 3, and compute them for several stocks and pairs of stocks. In Sect. 4 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. Finally, in Sect. 5 we explain in detail how the time lag τ behaves in the response functions. Our conclusions follows in Sect. 6.

2 Data set and time definition

In Sect. 2.1 we introduce the data set used in the paper. In Sect. 2.2 we describe the physical time scale and the trade time scale.

2.1 Data set

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 [19]. 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. Observe that we will not be discussing the returns, but the midpoint price. This because both are intrinsically related, as explained before, and it is more intuitive to understand the changes in midpoint prices than in returns.

To analyze the response functions across different liquid stocks, 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.

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 [10,4,19,7]. Therefore, we only consider trades of the same day from 9:40:00 to 15:50:00 New York local time.

2.2 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 the time.

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 magnitude of the guaranteed precision is much larger, usually one second or a few hundreds of milliseconds [5]. In several papers are used different time definitions (calendar time, physical time, event time, trade time, tick time) [5, 9,13]. 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 [19]. 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 [5].

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 [19,9], 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.

2.2.1 Trade time scale

We use the trade sign classification in trade time scale proposed by S. Wang et al. in [19] and used in [18,15,16] that reads

$$\varepsilon^{t}\left(t,n\right)=\begin{cases}\operatorname{sgn}\left(S\left(t,n\right)-S\left(t,n-1\right)\right), & \text{if } S\left(t,n\right)\neq S\left(t,n-1\right)\\ \varepsilon\left(t,n-1\right), & \text{otherwise}\end{cases} \tag{10}$$

$\varepsilon^{t}\left(t,n\right)=+1$ 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 [19], the average accuracy of the classification is 85% for the trade time scale.

Table 1. Analyzed companies average number of quotes in a day.

Company	Symbol	Sector	Quotes[1]	Trades[2]	Spread[3]
Alphabet Inc.	GOOG	Information Technology (IT)	164489	19029	\$0.04
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.
[2] Average number of trades from 9:40:00 to 15:50:00 New York time.
[3] Average spread from 9:40:00 to 15:50:00 New York time.

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 [19].

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 there was not a change in the quotes, the midpoint price remain the same as the last one.

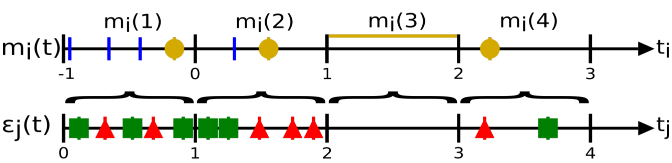

Figure 1. Sketch of data processing for trade time scale. In the midpoint price time line, the vertical lines represent the change in price of the quotes and the circles represent the last price change in a quote in a second. In the trade signs time line, the squares represent the buy market orders and 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 4 and 10.

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, which could not be the case. This could increase or decrease the response signal at the end of the computation.

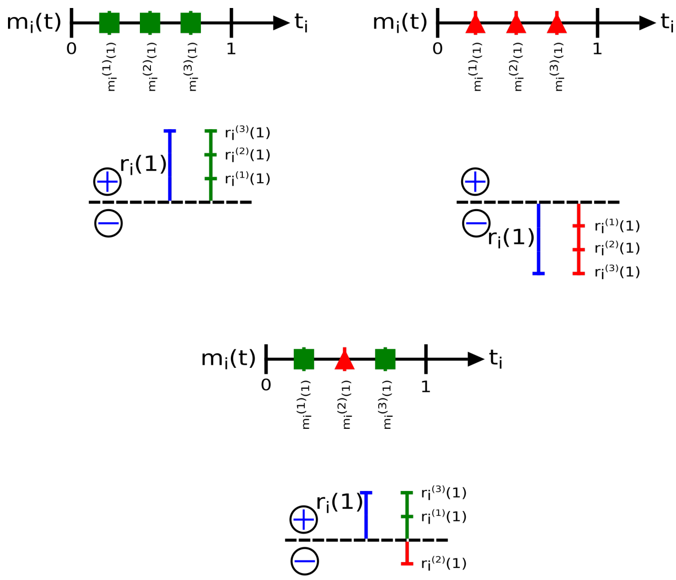

Figure 2. Sketch of the return contributions from every midpoint price change in a second. The squares represent the rise of the price of the midpoint price and the triangles represent the decrease of the price of the midpoint price. We illustrate three cases: (top left) the changes of the midpoint prices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

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 price are the same in the bid and in the ask (so the impact 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 repercussion, 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.

2.2.2 Physical time scale

We use the trade sign definition in physical time scale proposed by S. Wang et al. in [19] and used in [18,15], that depends on the classification in Eq. 10 and reads

$$\varepsilon^{p}\left(t\right)=\begin{cases}\operatorname{sgn}\left(\sum_{n=1}^{N(t)}\varepsilon^{t}\left(t,n\right)\right), & \text{If } N\left(t\right)>0\\ 0, & \text{If } N\left(t\right)=0\end{cases} \tag{11}$$

Where $N\left(t\right)$ is the number of trades in a second interval. $\varepsilon^{p}\left(t\right)=+1$ implies that the majority of trades in second t were triggered by a market order to buy, and a value $\varepsilon^{p}\left(t\right)=-1$ indicates a majority of sell market orders. In this definition, they are two ways to obtain $\varepsilon^{p}\left(t\right)=0$. 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 I 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.

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

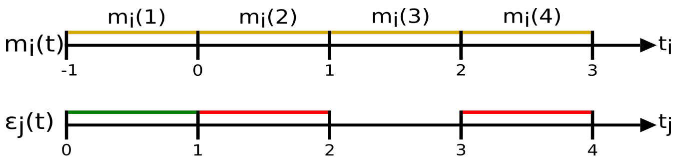

Figure 3. Sketch of data processing for physical time scale. In the midpoint price time line, the horizontal lines between seconds represent the midpoint prices. In the trade signs time line, the horizontal lines between seconds represent the trade sign values. The midpoint price time line and the trade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

3 Response functions

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

$$R_{ij}^{scale}\left(\tau\right)=\left\langle r_{i}^{p}\left(t-1,\tau\right)\cdot\varepsilon_{j}^{scale}\left(t\right)\right\rangle_{scale} \tag{12}$$

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

$$E_{j,d}\left(t\right)=\sum_{n=1}^{N(t)}\varepsilon_{j,d}^{t}\left(t,n\right) \tag{13}$$

$$E_{j,d}\left(t\right)=\operatorname{sgn}\left(E_{j,d}\left(t\right)\right)\cdot\left|E_{j,d}\left(t\right)\right| \tag{14}$$

$$\varepsilon_{j,d}^{p}\left(t\right)=\operatorname{sgn}\left(E_{j,d}\left(t\right)\right) \tag{15}$$

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

In Sect. 3.1 we analyze the responses functions in trade time scale, in Sect. 3.2 we analyze the responses functions in trade time scale and in Sect. 3.3 we define a response function to analyze the influence of the frequency of trades in a second.

3.1 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

$$R_{ij}^{t}\left(\tau\right)=\left\langle r_{i}^{p}\left(t-1,\tau\right)\cdot\varepsilon_{j}^{t}\left(t,n\right)\right\rangle_{t} \tag{16}$$

However, to be explicit with the way the averaging is made, the function reads

$$R_{ij}^{t}(\tau)=\frac{1}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}N_{j,d}(t)} \cdot \sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}\sum_{n=1}^{N(t)} r_{i,d}^{p}(t-1,\tau)\cdot\varepsilon_{j,d}^{t}(t,n) \tag{17}$$

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

$$=\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\operatorname{sgn}\left(E_{j,d}(t)\right)\cdot w_{j,d}^{t}(t) \tag{18}$$

Where

$$w_{j,d}^{t}(t)=\frac{\left|E_{j,d}(t)\right|}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}N_{j,d}(t)} \tag{19}$$

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

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. Compared with the market response in second time scale, the market response in trade time scale is almost one order of magnitude smaller.

3.2 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 $\varepsilon_j^p(t)=0$ [19]. The cross-responses including $\varepsilon_j^p(t)=0$ are weaker than the excluding ones due to the omission of direct influence of the lack of trades. However, either including or excluding $\varepsilon_j^p(t)=0$ does not change the trend of price reversion versus the time lag, but it does affect the response function strength [18].

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

$$R_{ij}^{(\text{inc. }0)}(\tau)=\frac{\sum_{t=1}^{T_j+T_{j,n}} r_i^p(t-1,\tau)\cdot\varepsilon_j^p(t)}{T_j+T_{j,n}} \tag{20}$$

$$R_{ij}^{(\text{exc. }0)}(\tau)=\frac{\sum_{t=1}^{T_j} r_i^p(t-1,\tau)\cdot\varepsilon_j^p(t)}{T_j} \tag{21}$$

Where the superscript inc. and exc. refers to including and excluding $\varepsilon_j^p(t)=0$. 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. 20 and 21 are the same, while the denominators differ [18].

Hence,

$$R_{ij}^{(\text{inc. }0)}(\tau)=f_j\cdot R_{ij}^{(\text{exc. }0)}(\tau) \tag{22}$$

Where the relative trading frequency is defined as [18]

$$f_j=\frac{T_j}{T_j+T_{j,n}} \tag{23}$$

The most frequently traded stocks have $f_i=1$, because the time $T_{j,n}$ is zero. According to Eq. 22, the cross-response including $\varepsilon_j^p(t)=0$ is the one excluding $\varepsilon_j^p(t)=0$ scaled by a proper probability.

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

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

$$R_{ij}^{p}(\tau)=\left\langle r_i^p(t-1,\tau)\cdot\varepsilon_j^p(t)\right\rangle_p \tag{24}$$

And the corresponding explicit expression reads

$$R_{ij}^{p}(\tau)=\frac{1}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}\eta\left[\varepsilon_{j,d}^{p}(t)\right]} \cdot \sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\varepsilon_{j,d}^{p}(t)\cdot\eta\left[\varepsilon_j^p(t)\right] \tag{25}$$

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

$$=\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f} r_{i,d}^{p}(t-1,\tau)\cdot\operatorname{sgn}\left(E_{j,d}(t)\right)\cdot w_{j,d}^{p}(t) \tag{26}$$

Where

$$\eta(x)=\begin{cases}1, & \text{If } x\neq 0\\ 0, & \text{otherwise}\end{cases} \tag{27}$$

take only in account the seconds with trades and

$$w_{j,d}^{p}(t)=\frac{\eta\left[\operatorname{sgn}\left(E_{j,d}(t)\right)\right]}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}\eta\left[\operatorname{sgn}\left(E_{j,d}(t)\right)\right]} \tag{28}$$

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

The results showed in Figure 5 are identical with the ones showed in [19] for the same data. We can say again that in all cases, an increase to a maximum is followed by a decrease, i.e. the trend in the self- and cross-response is eventually reversed.

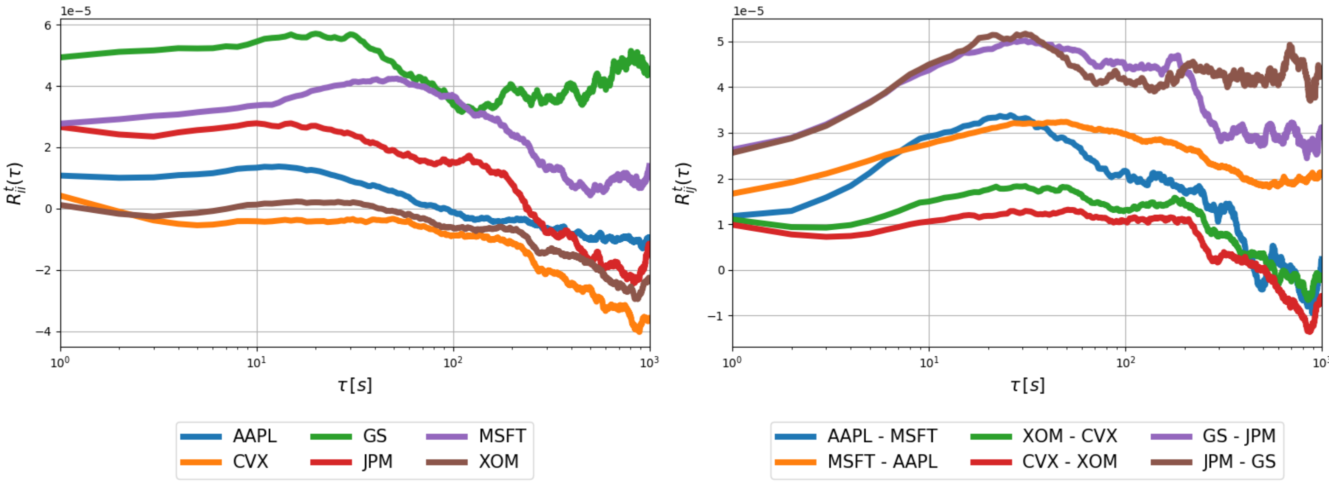

Figure 4. Self- and cross-response functions $R^t_{ij}(\tau)$ excluding $\varepsilon^t_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale in trade time scale. Self- responses (left) of individual stocks and cross-response (right) of stock pairs from the same economic sector.

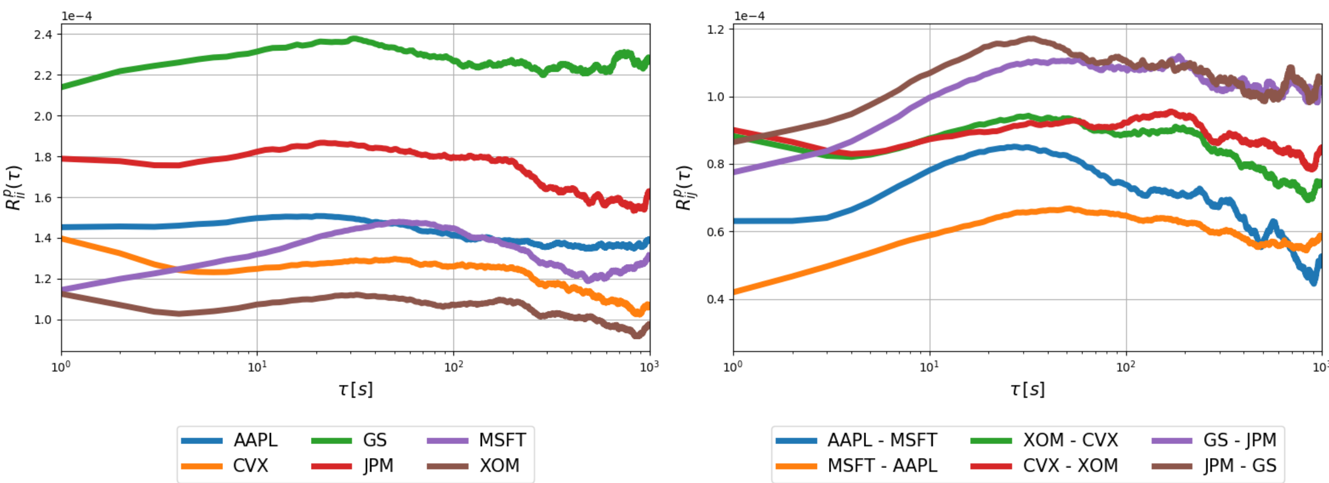

Figure 5. Self- and cross-response functions $R^p_{ij}(\tau)$ excluding $\varepsilon^p_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale in second time scale. Self-responses (left) of individual stocks and cross-response (right) of stock pairs from the same economic sector.

3.3 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

$$R^a_{ij}(\tau) = \left\langle r^p_i(t-1,\tau) \cdot \varepsilon^p_j(t) \cdot N(t) \right\rangle_p \tag{29}$$

And the corresponding explicit expression reads

$$R^a_{ij}(\tau) = \frac{1}{\sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} N_{j,d}(t)} \cdot \sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} r^p_{i,d}(t-1,\tau) \cdot \varepsilon^p_{j,d}(t) \cdot N_{j,d}(t) \tag{30}$$

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

$$= \sum_{d=d_0}^{d_f} \sum_{t=t_0}^{t_f} r^p_{i,d}(t-1,\tau) \cdot \mathrm{sgn}(E_{j,d}(t)) \cdot w^a_{j,d}(t) \tag{31}$$

Where

$$w^{a}_{j,d}\left(t\right)=\frac{N_{j,d}\left(t\right)}{\sum_{d=d_0}^{d_f}\sum_{t=t_0}^{t_f}N_{j,d}\left(t\right)} \tag{32}$$

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

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

4 Time shift response functions

We used the definition of the response function from [19] that is showed in Equation 8. To see the impact of the time shift I analyzed the TAQ data in the year 2008. I used different time shifts in the response function

$$R^{s,escale}_{ij}\left(\tau\right)=\left\langle r_i\left(t-t_s,\tau\right)\cdot\varepsilon^{scale}_{j}\left(t\right)\right\rangle_{scale} \tag{33}$$

Making τ constant, and varying t_s I obtained values for self- and cross-responses. CORREGIR As t_{shift} is related with the response, the shift is in seconds for both scales.

In Sect. 4.1 we analyze the influence of the time shift between the trade signs and returns in trade time scale and in Sect. 4.2 we analyze the influence of the time shift between the trade signs and returns in physical time scale.

4.1 Trade time scale shift response functions

In Fig. 7, for fixed τ values, there is a peak in a position related to τ.

In the trade time scale, I tested the response function for different shifted times (Fig. 8). The one second shift curve is the control curve. From small τ values to a τ_{max} value the response grows. Then it starts to decay, but not under the original value of $t_{shift}=1$.

4.2 Physical time scale shift response functions

For every τ value, there is a peak. The peak grows fast and decay slowly for $\tau<10$, and grows and decays fast for large τ. The response signal usually starts to grow on zero or a little bit earlier and grows to around τ.

I tested the response function for different shifted times (Fig. 10). The one second shift curve is the control curve. From small τ values to a τ_{max} value the response grows. Then it starts to decay, but not under the original value of $t_{shift}=1$.

My interpretation for the rise of the signal in the different cases, is that there is a zone of the beginning of the shifted signals where there is not relevant information, which make the response zero. When the signal reach the one or two event/second shift, the response have again relevant information, and the response is different to zero.

When the τ value grows, the averaging value decrease, hence the apparent response is bigger. This explains that in larger values the signal seems to be large.

5 Short and long response functions

Regarding Equation 2, we use a time lag τ in the returns to see the gains or loses in a future time. We divide the time lag τ as show in Fig. 11. As

$$\tau=\tau'+\left(\tau-\tau'\right) \tag{34}$$

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

To compute the short and long response, we start redefining the returns as

$$\begin{aligned}r^{p}_{i}\left(t,\tau\right)&=\ln\left(\frac{m_i\left(t+\tau\right)}{m_i\left(t\right)}\right)\\&=\ln\left(\frac{m_i\left(t+\tau\right)}{m_i\left(t+\tau'\right)}\cdot\frac{m_i\left(t+\tau'\right)}{m_i\left(t\right)}\right)\\&=\ln\left(\frac{m_i\left(t+\tau\right)}{m_i\left(t+\tau'\right)}\right)+\ln\left(\frac{m_i\left(t+\tau'\right)}{m_i\left(t\right)}\right)\\&\approx\frac{m_i\left(t+\tau\right)-m_i\left(t+\tau'\right)}{m_i\left(t+\tau'\right)}+\frac{m_i\left(t+\tau'\right)-m_i\left(t\right)}{m_i\left(t\right)}\end{aligned} \tag{35}$$

The second term of the right part is constant with respect to τ. That means, the main signal of the return come from the short return. The long return does not give a significant contribution to the complete return. Replacing Equation 35 in the response function we have

$$\begin{aligned}R^{p}_{ij}\left(\tau\right)&=\left\langle r^{p}_{i}\left(t,\tau\right)\cdot\varepsilon^{p}_{j}\left(t\right)\right\rangle_{p}\\&\approx\left\langle\frac{m_i\left(t+\tau\right)-m_i\left(t+\tau'\right)}{m_i\left(t+\tau'\right)}\cdot\varepsilon_j\left(t\right)\right\rangle_{t\in\tau-\tau'}\\&\quad+\left\langle\frac{m_i\left(t+\tau'\right)-m_i\left(t\right)}{m_i\left(t\right)}\cdot\varepsilon_j\left(t\right)\right\rangle_{t\in\tau'}\end{aligned} \tag{36}$$

Where the first term in the right side of Equation 36 is the long response and the right term is the short response. Again, the right term of Equation 36 is independent of τ.

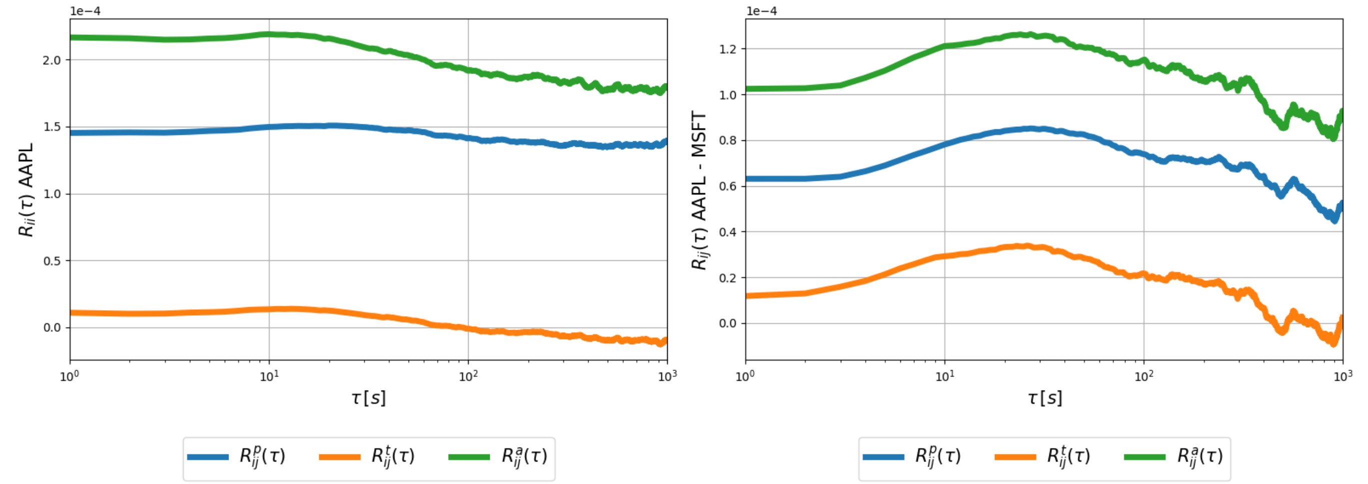

Figure 6. Self- and cross-response functions $R_{ij}^{scale}(\tau)$ excluding $\varepsilon_j^p(t) = 0$ in 2008 versus time lag τ on a logarithmic scale. a) self-response function of Apple Inc. stock, b) cross-response function of Apple Inc-Microsoft Corp. stocks.

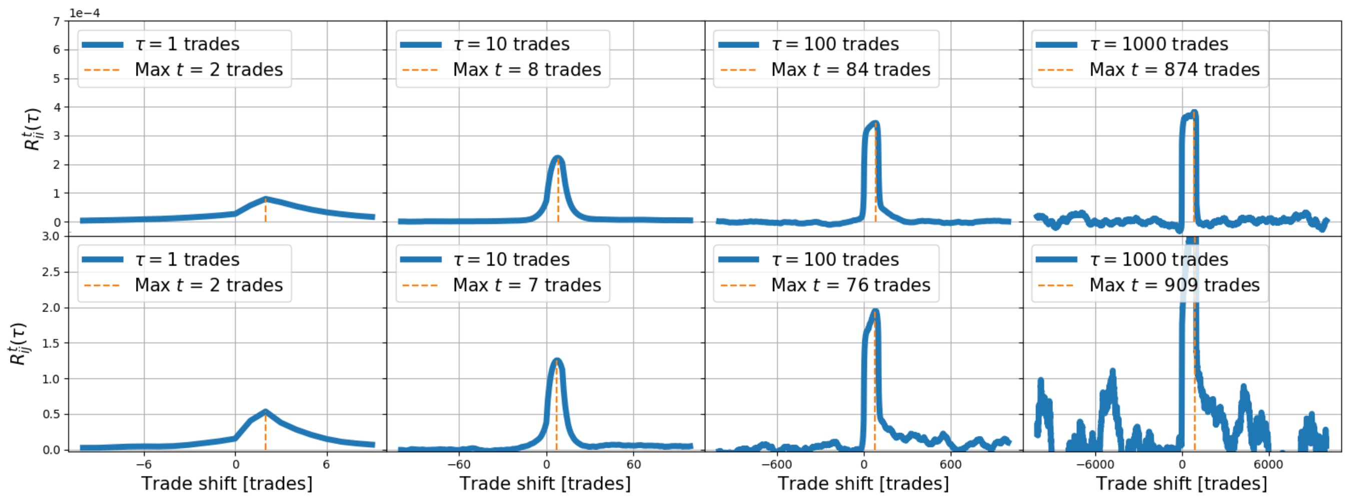

Figure 7. Self-response functions $R_{ii}^{trades}(\tau)$ excluding $\varepsilon_i^{trades}(t) = 0$ in 2008 versus shift for the Goldman Sachs Group stock (top) and cross-response functions $R_{ij}^{trades}(\tau)$ excluding $\varepsilon_j^{trades}(t) = 0$ in 2008 versus shift for the Goldman Sachs Group-JPMorgan Chase stocks (bottom).

Check the legend position of the figure, change the scale to seconds in the axis, find a way to make more general the results. Maybe the average can help. Maybe repeat the analysis for the trade time scale.

The results in Fig. **??** 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'.

Before τ', the short response and long response are the same, as the self and cross-response definition do not define these values for values smaller than τ', so it is computed as 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 the significance of the long response, I 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 representative in the final response. If we add the short and long response, we obtain the original response. In Fig. **??**, the original response (red line) has the same shape to the addition of the short and long response (green line). I analyzed the short and long responses with different τ' values $(10s, 20s, 30s, 40s, 50s, 60s, 70s, 80s, 90s, 100s)$. On average, the maximum response is around $\tau = 40$. That means the long response before this value have some positive values and after this value have all negative values.

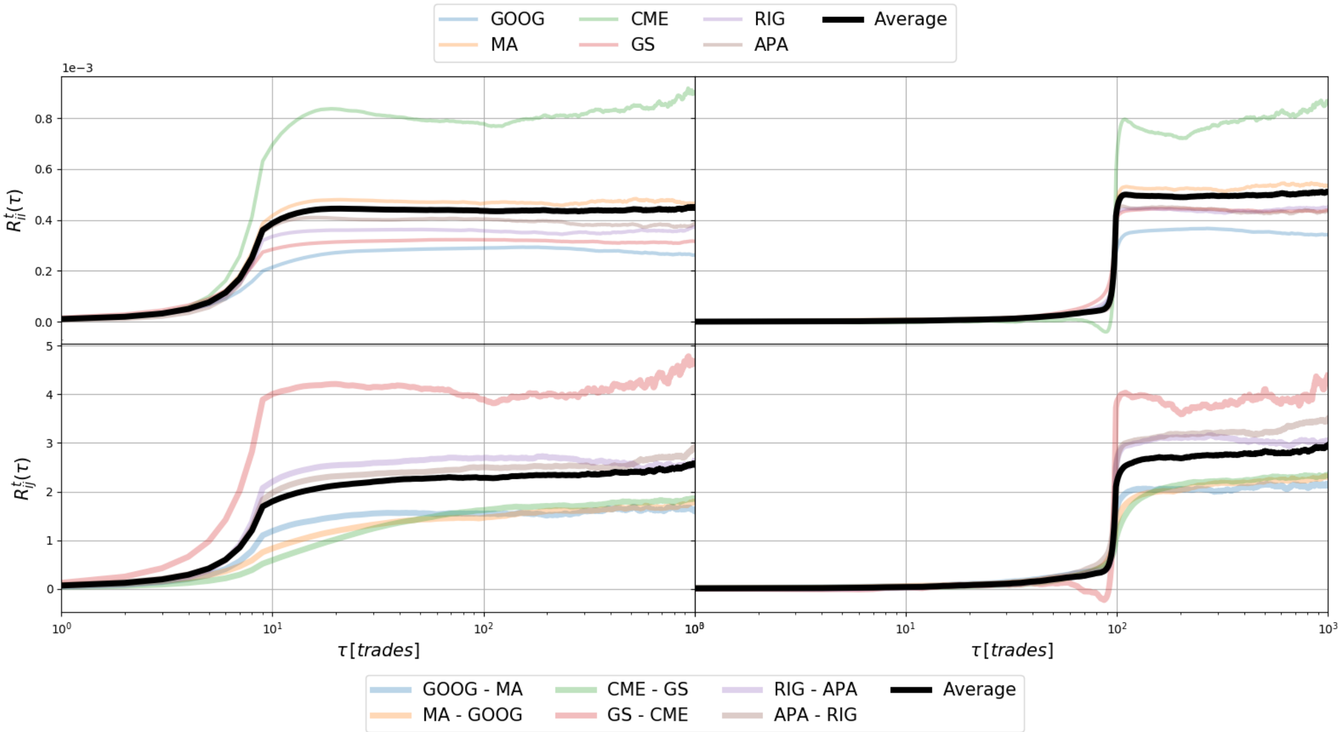

Figure 8. Self- and cross-response functions $R_{ij}^{trades}(\tau)$ excluding $\varepsilon_{j}^{trades}(t)=0$ in 2008 versus time lag τ on a logarithmic scale for different shifts. Self-responses for the Goldman Sachs Group stock in trade time scale (left), and cross-response of Goldman Sachs Group-JPMorgan Chase stocks in trade time scale (right).

6 Conclusion

7 Author contribution statement

All the authors were involved in the preparation of the manuscript. All the authors have read and approved the final manuscript.

One of us (J. C. H. L) acknowledges financial support from the DAAD with the program “Research Grants - Doctoral Programmes in Germany” (Funding programme 57381412)

References

1. 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.
2. J. P. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets, 2004.
3. Jean-Philippe Bouchaud. The subtle nature of financial random walks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026104, 2005.
4. 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.
5. Anirban Chakraborti, Ioane Toke, Marco Patriarca, and Frédéric Abergel. Econophysics: Empirical facts and agent-based models. *arXiv.org, Quantitative Finance Papers*, 09 2009.
6. Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.
7. 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.
8. Austin Gerig. A theory for market impact: How order flow affects stock price. *arXiv.org, Quantitative Finance Papers*, 04 2008.
9. 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.
10. 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.
11. Vasiliki Plerou, Parameswaran Gopikrishnan, Xavier Gabaix, and H. Eugene Stanley. Quantifying stock-price response to demand fluctuations. *Phys. Rev. E*, 66:027104, Aug 2002.
12. Bernd Rosenow. Fluctuations and market friction in financial trading. *International Journal of Modern Physics C*, 13(03):419–425, 2002.
13. Ioane Muni Toke. *“Market Making” in an Order Book Model and Its Impact on the Spread*, pages 49–64. Springer Milan, Milano, 2011.

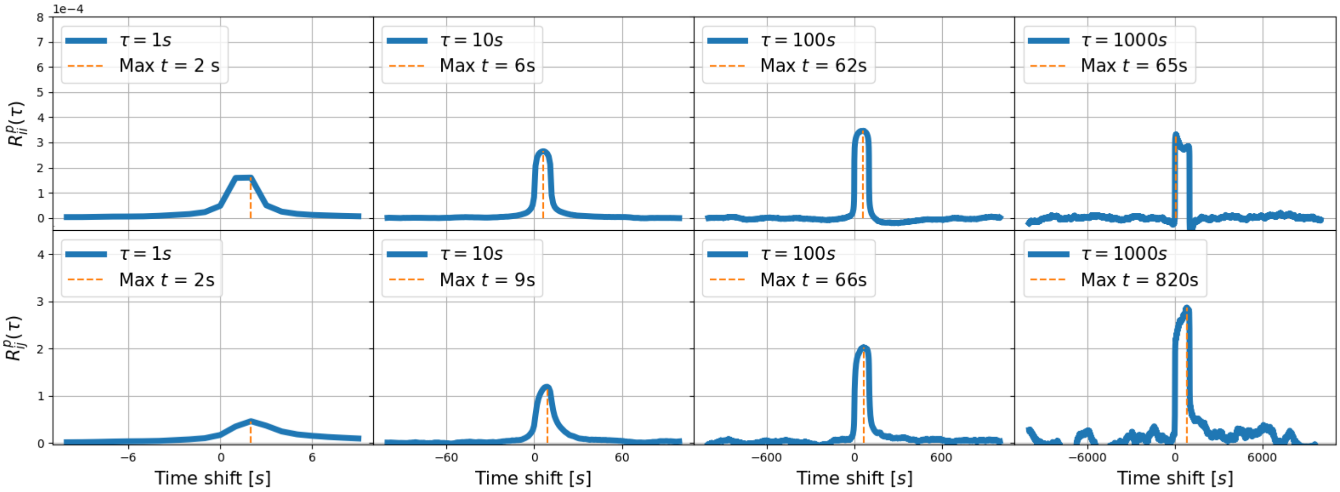

Figure 9. Self-response functions $R^{seconds}_{ii}(\tau)$ excluding $\varepsilon^{seconds}_{i}(t) = 0$ in 2008 versus shift for the Goldman Sachs Group stock (top) and cross-response functions $R^{seconds}_{ij}(\tau)$ excluding $\varepsilon^{seconds}_{j}(t) = 0$ in 2008 versus shift for the Goldman Sachs Group-JPMorgan Chase stocks (bottom).

14. 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.
15. Shanshan Wang. Trading strategies for stock pairs regarding to the cross-impact cost, 2017.
16. 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.
17. Shanshan Wang, Sebastian Neusüß, and Thomas Guhr. Grasping asymmetric information in price impacts. *The European Physical Journal B*, 91(11):266, Nov 2018.
18. Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Average cross-responses in correlated financial markets. *The European Physical Journal B*, 89(9):207, Sep 2016.
19. Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Cross-response in correlated financial markets: individual stocks. *The European Physical Journal B*, 89(4), Apr 2016.

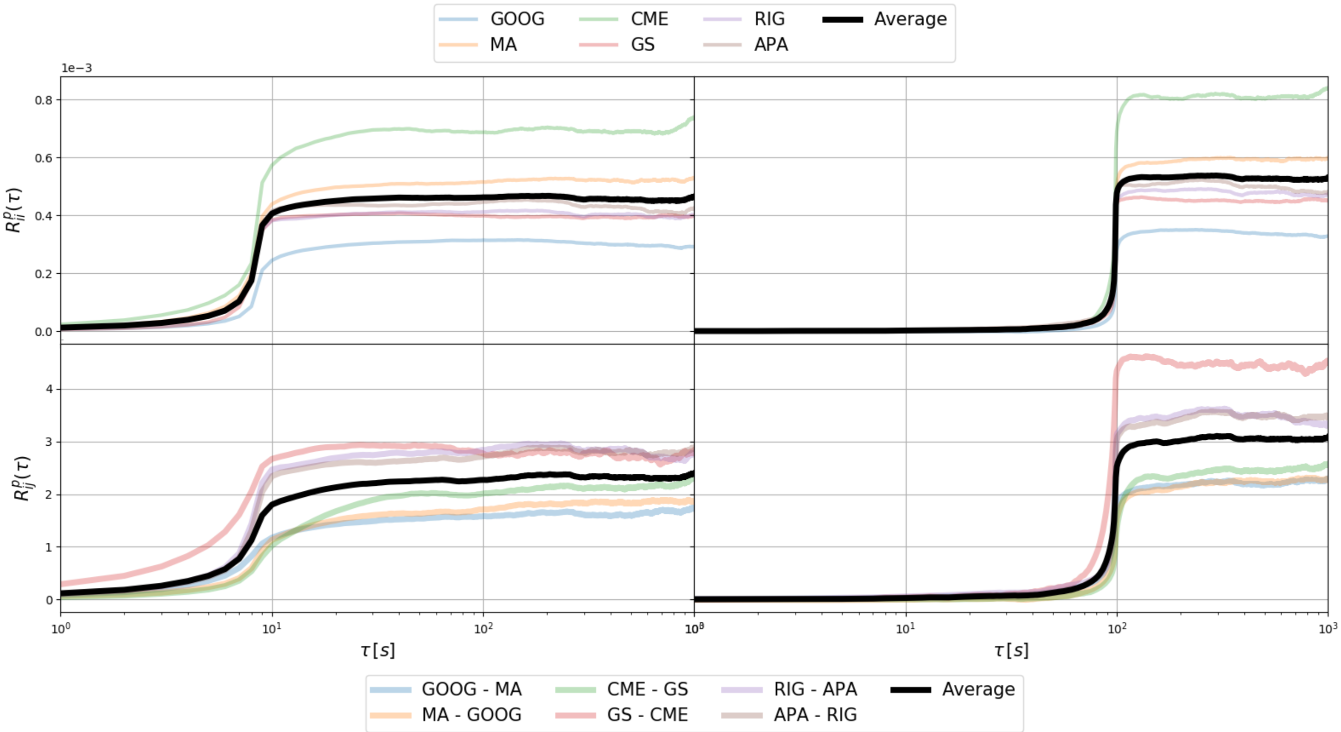

Figure 10. Self- and cross-response functions $R_{ij}^{seconds}(\tau)$ excluding $\varepsilon_j^{seconds}(t) = 0$ in 2008 versus time lag τ on a logarithmic scale for different shifts. Self-responses for the Goldman Sachs Group stock in physical time scale (left), and cross-response of Goldman Sachs Group-JPMorgan Chase stocks in second time scale (right).

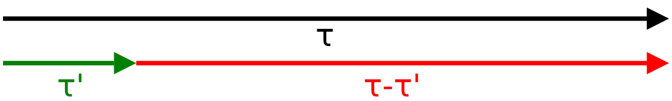

Figure 11. τ value divide in short and long response.

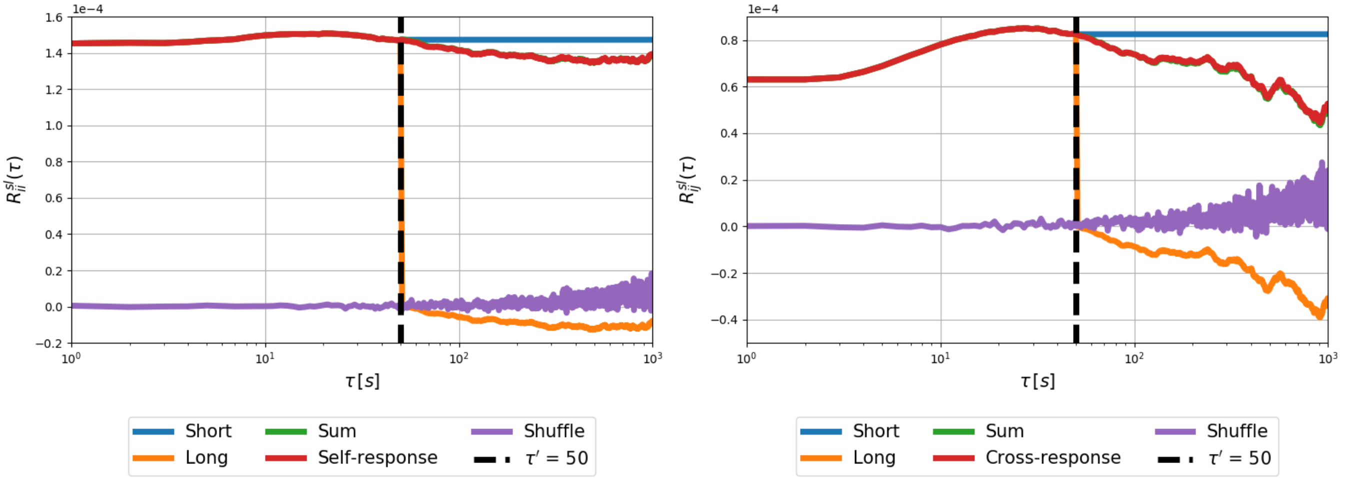

Figure 12. Cross-response