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

1 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.

2 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],

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

It is also common to define the returns as [6,11,7,3,8,5,1,9]

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

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

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

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 $\varepsilon(t) = -1$ [10,4,3,12,9].

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

$$R(\tau) = \langle (S_{n+\tau} - S_n) \cdot \varepsilon_n \rangle_{trades} \tag{7}$$

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 [17], 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

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

Finally, in [15], 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 [17] and in [15], is that [17] 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 [15], it was used a response function on a trade time scale (Totalview data), as the interest is to analyze the immediate responses. In [15] 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. 3 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. 4, and compute them for several stocks and pairs of stocks. 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. Finally, in Sect. 6 we explain in detail how the time lag τ behaves in the response functions. Our conclusions follows in Sect. 7.

3 Data set and time definition

In Sect. 3.1 we introduce the data set used in the paper. In Sect. 3.3 we describe the physical time scale and in Sect. 3.2 we describe the trade time scale.

3.1 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 [17]. 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 [9,4,17,7]. Therefore, we only consider trades of the same day from 9:40:00 to 15:50:00 New York local time.

3.2 Trade time scale

We use the trade sign classification in trade time scale proposed by S. Wang et al. in [17] and used in [16,13,14] that reads

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

With this classification we obtain trade signs for every single trade in the data set. According to [17], the average accuracy of the classification is 85% for the trade time scale.

For the trade time scale, as the 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 [17].

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 some cases, 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 between the second −1 and 0 is related with all the trades in the second 0 to 1, and so on. In the seconds that there is no change in the quotes, it is used the value of the previous second. In consequence, all the seconds in the open market time have a midpoint price value.

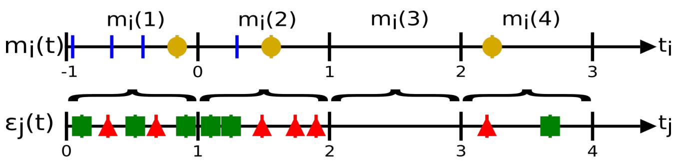

Figure 1. Sketch of data processing for trade time scale. Each trade sign is related with the last value of the midpoint price of the previous second.

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 illustrate with one example this point. Suppose one second interval, in which they are three different midpoint prices, and in consequence, three different returns. 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 left (center), all the changes are due to buys (sells), 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 return is positive (negative). Finally,

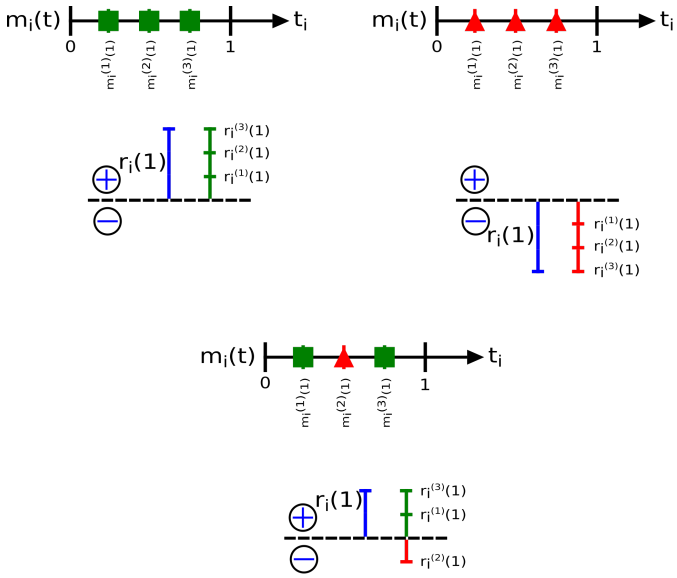

Figure 2. Sketch of the return contributions from every midpoint price change in a second. We illustrate three cases: (left) the changes of the returns and midpoint price are due to buys, (center) the changes of the returns and midpoint prices are due to sells, and (right) the changes of the returns and midpoint prices are due to a combination of buys and sells.

in the case of the right, the changes are due to a combination of buys and sells, 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 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 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, so it can give us valuable information about the responses.

Por hacer: - Revisar en promedio cuantos midpoint price por segundo en los datos TAQ, - Hallar el porcentaje de cuanto cambia el valor del último midpoint price en un segundo con respecto al promedio de ese segundo

3.3 Physical time scale

We use the trade sign definition in physical time scale proposed by S. Wang et al. in [17] and used in [16,13], that depends on the classification in Eq. 10 and reads

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

Where $N\left(t\right)$ is the number of trades in a second interval. For this classification, they are two ways to obtain $\varepsilon^{physical}\left(t\right)=0$. The first way is that in a particular second there is not trades, and then no trade sign. The second way is that the addition of the trade signs ($+1$ and -1) be equal to zero. In this case, could be important to see how is the distribution of the trade scale trades signs for the particular second.

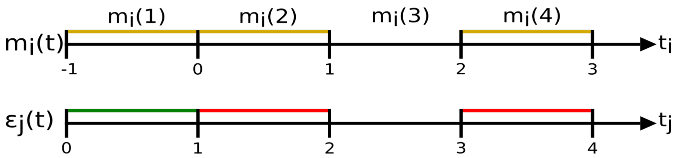

Figure 3. Sketch of data processing for trade time scale. Each trade sign is related with the last value of the midpoint price of the previous second.

$\varepsilon^{physicals}\left(t\right)=+1$ implies that the majority of trades in second t were triggered by a market order to buy, and a value $\varepsilon^{physicals}\left(t\right)=-1$ indicates a majority of 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.

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

According to [17], this classification has an average accuracy up to 82% in the physical time scale.

4 Response functions

5 Time shift

6 Short- and long-response

7 Conclusion

8 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

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. 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.
10. Vasiliki Plerou, Parameswaran Gopikrishnan, Xavier Gabaix, and H. Eugene Stanley. Quantifying stock-price response to demand fluctuations. *Phys. Rev. E*, 66:027104, Aug 2002.
11. Bernd Rosenow. Fluctuations and market friction in financial trading. *International Journal of Modern Physics C*, 13(03):419–425, 2002.
12. 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.
13. Shanshan Wang. Trading strategies for stock pairs regarding to the cross-impact cost, 2017.
14. 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.
15. Shanshan Wang, Sebastian Neusüß, and Thomas Guhr. Grasping asymmetric information in price impacts. *The European Physical Journal B*, 91(11):266, Nov 2018.
16. Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Average cross-responses in correlated financial markets. *The European Physical Journal B*, 89(9):207, Sep 2016.
17. Shanshan Wang, Rudi Schäfer, and Thomas Guhr. Cross-response in correlated financial markets: individual stocks. *The European Physical Journal B*, 89(4), Apr 2016.