Physica A 508 (2018) 28–34

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Volatility–Trading volume intraday correlation profiles and its nonstationary features

Michelle B. Graczyk, Sílvio M. Duarte Queirós *,[1]

Centro Brasileiro de Pesquisas Físicas, Rua Dr Xavier Sigaud 150, Urca, CEP 22290-180 RJ, Rio de Janeiro, Brazil

HIGHLIGHTS

- Analysis of intraday volatility–volume matrices.
- Correlations increase during the morning and dwindle aftwards.
- MDH and SIAH are both relevant, but in different parts of the day.

ARTICLE INFO

Keywords:
Complex systems
Financial markets
Trading volume
Volatility
Correlation matrices
MDH
SIAH

ABSTRACT

We analyse the statistical properties of volatility–volume cross-correlation matrices of stocks composing the Dow Jones Industrial Average since 2003. Using different definitions of volatility, we verify there is an intraday profile where the average values of the entries significantly increase from the opening of the trading session until its midway and it dwindles therefrom afterwards. Higher-order moments of the correlation matrix are studied and exhibit intraday profiles as well. Within the scope of the (endless) discussion "Mixture of Distributions *versus* Sequential Information Arrival" our results allow us to assert that both seem to be relevant in different parts of the business day.

1. Introduction

In being a complex system, wherein its elements influence both one another and the system as a whole in an intricate way, the understanding about the relations between observables like price (fluctuations) and trading volume of the companies is crucial to the factual description of a financial market. With respect thereto, ever more frequent contributions based on the application of techniques stemming from Physics – and which paved the way to the establishment of Econophysics – have shed light on many properties that were previously ill-characterised [1].

Within a Physical approach to a financial market, the application of Random Matrix Theory (RMT) – in conjunction with Principal Component Analysis (PCA) – have provided important insights into the collective features of the aforementioned quantities with impact in portfolio management and risk assessment [2]. Concerning RMT, that framework was first applied by E.P. Wigner to explain the energy levels of compound atomic nuclei [3]; i.e, Wigner claimed that the complex nature of a physical system like that would be best described by stochastic elements, which has found similarity in a wide range of other physical and non-physical problems [4].

Assuming the technical analysis (chartist) hypothesis that the price contains all the information we need to know, we understand whenever there is a trade we have transfer of information transactions are the form of transferring information.

* Corresponding author.
E-mail addresses: sdqueiro@gmail.com, sdqueiro@cbpf.br (S.M. Duarte Queirós).

[1] Associate to the National Institute of Science and Technology for Complex Systems, Brazil.

https://doi.org/10.1016/j.physa.2018.05.066
0378-4371/

Accordingly, it is possible to assume the volume is a proxy for market information and the volatility as a way of quantifying that tranfer [5]. Recalling the theory of Dow [6], where the volume confirms the trend of the price, it is not much hard to grasp that the price gets a given direction as the percolates across the market agents leading to large trading volume.

The relations between volatility, trading volume and information are the cornerstones of the two leading dynamical hypotheses in finance. On the one hand, there is the Mixture of Distributions Hypothesis (MDH) [7], where the three quantities are intermingled; in other words, heeding that the dynamics of both the volatility and the trading volume depend on latent events (disclosure of information), there is a joint distribution for these two quantities with both quantities marginally follow a log-Normal distribution. Additionally, one has two ways of measuring the time: the physical (clock) and the proper (event) time at which information is input. It is noteworthy that such a difference of time concepts is relevant in the description of many complex systems [8]. As examples of a quantitative approaches assuming the MDH, we point to studies on the stochastic nature of the volatility as well as ARCH-like heteroskedastic models [9,10], which consider the price fluctuations and trading volume share the same underlying. On the other hand, there is the Sequential Arrival of Information Hypothesis (SAIH) [11] that considers the information reaches the market agents at different times so that the final state of the market is attained by a sequence of local stationary states. In that case, the maximum value of the correlation between the volatility and the trading volume is achieved when a time delay is considered.

In previous publications [12–14], we focussed on the survey of seasonalities of the trading volume and volatility and its evolution in the last 15 years, namely collective features measured by correlation matrices, that enlarges the scope of the well-known ∪-shape of these two quantities [15]. The results therein showed interesting properties about how the flux of information evolves within the trading session and to what extent such features changed either by new trading rules or events like the 2008 sub-prime crisis.

With this work, we aim at concluding that description by looking at the volatility–trading volume cross-correlation matrix. That allows us to assess whether the flux of information related to a given company affects the another one and how such relation changes and has modified with time. The remaining of this article is organised as follows: In the next section we characterise our data and set the definitions we use through the paper; in Section 3 we show the results, in Section 4 we elaborate upon our analysis and in Section 5 we introduce a conclusion over our results.

2. Data and methods

Our results are obtained from 1-minute frequency data of price and trading volume of the 30 companies composing the Dow Jones Industrial Average spanning the period between the 4th January 2004 and the 30th December 2013 provided by the Chair of Quantitative Finance of the École CentraleSupélec. There are 28 companies traded at NYSE and 2 at NASDAQ. Both markets open at 9:30 and close 16:00 and have pre- and post-market periods that we do not take into account.[2] Despite being a pivotal quantity in finance, the volatility is not directly observed and therefore its value can be sensitive to its several definitions. In order to compute the volatility in different ways, we coarse-grained these data to a 5-minute scale so that we have 78 time ticks in a trading session. That said, we employ the following notation:

- $\upsilon_i(d, t; s)$: 5-minute trading volume of company, i, at the intraday time stamp, t – defined as an integer number format with $t = 1$ representing 9:35 in clock time, $t = 2$ corresponds to 9:40 and so forth –, on the day, d; moreover, we make explicit the semester, s, to which d belongs because we divide our data into contiguous semesters to appraise the nonstationarity of the intraday properties. Such segmentation yields a good balance between quasi-stationarity and a statistically significant number of days within each span so that a higher order statistical analysis can be implemented. The first semester of 2004 (1S04) corresponds to $s = 1$, the second semester of 2004 (2S04) to $s = 2$ until the second semester of 2013 (2S13), i.e., $s = 19$.

Next, we consider different forms of computing the volatility, $\mathcal{V}$, namely:

- $\sigma_i(d, t; s)$: standard deviation of the (5-minute) log-prices, $\ln S_i(d, t; s)$;
- $\omega_i(d, t; s)$: the High–Low–Close–Open, Garman–Klass [16], volatility $\omega_i(d, t; s) \equiv \sqrt{\frac{1}{2}\left(\ln \frac{H_i(d,t;s)}{L_i(d,t;s)}\right)^2 - (2\ln 2 - 1)\left(\ln \frac{C_i(d,t;s)}{O_i(d,t;s)}\right)^2}$;
- $\psi_i(d, t; s)$: the absolute price fluctuation within the 5-minute spell,

$$\psi_i(d, t; s) \equiv |\ln S_i(d, \mathrm{hh} : \mathrm{mm}; s) - \ln S_i(d, \mathrm{hh} : (\mathrm{mm} - 5); s)| .$$

With this quantities in hand we can compute the cross-correlation matrix using the following definition for its entries:

$$C_{ij}(t; s) \equiv \frac{\overline{\upsilon_i(d, t; s)\,\mathcal{V}_j(d, t; s)} - \overline{\upsilon_i(d, t; s)}\;\overline{\mathcal{V}_j(d, t; s)}}{\sqrt{\overline{\upsilon_i(d, t; s)^2} - \overline{\upsilon_i(d, t; s)}^2}\sqrt{\overline{\mathcal{V}_j(d, t; s)^2} - \overline{\mathcal{V}_j(d, t; s)}^2}}, \tag{1}$$

where the overline indicates that we averaged over the days of semester s.

[2] Nonetheless, we can mention we have not found any pattern during either off-market period.