Physica A 390 (2011) 837–846

Contents lists available at ScienceDirect

Physica A

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

Cross-correlations in volume space: Differences between buy and sell volumes

Sun Young Lee *, Dong Il Hwang, Min Jae Kim, In Gyu Koh, Soo Yong Kim

Department of Physics, Korea Advanced Institute of Science and Technology, Republic of Korea

ARTICLE INFO

Article history:
Received 27 May 2010
Received in revised form 14 October 2010
Available online 26 November 2010

Keywords:
Cross-correlations
Volumes
Pulling effect
Minimal spanning tree
Epps effect
Frequency scale
Synchronization

ABSTRACT

We study the cross-correlations of buy and sell volumes on the Korean stock market in high frequency. We observe that the pulling effects of volumes are as small as that of returns. The properties of the correlations of buy and sell volumes differ. They are explained by the degree of synchronization of stock volumes. Further, the pulling effects on the minimal spanning tree are studied. In minimal spanning trees with directed links, the large pulling effects are clustered at the center, not uniformly distributed. The Epps effect of buy and sell volumes are observed. The reversal of the cross-correlations of buy and sell volumes is also detected.

1. Introduction

Cross-correlation is the general method for exploring the relation between random variables (or their own elements). In physics, it is used in order to analyze various variables even from quantum dot to climates [1,2]. It is also frequently used in finance to understand not only structural relations in stock returns but also the information-flow between the stocks [3–7]. The correlation concept in finance has been introduced by physicists. They applied additional techniques such as minimal spanning tree (MST), principal component analysis (PCA), and random matrix theory (RMT) to the correlation [8–13]. They also developed a new method like detrended cross-correlation analysis which is similar to detrended fluctuation analysis (DFA) [14,15].

Cross-correlation is divided into time-dependent and equal-time ones. Most of the previous studies of time-dependent cross-correlations were done in a daily scale. However, due to improved communication technologies, the speed of information-flow in market is quite fast. In electronic stock markets such as the Korean stock market, NASDAQ and NYSE, the information is reflected quickly in stocks. These trends have necessitated studies in much shorter scale than a day. Kullmann et al. and Toth et al. studied the time-dependent cross-correlations empirically in high frequency [12,13]. They show that a time shift existed in the cross-correlations of returns which is known as the pulling effect or the lead–lag effect. However, this effect is small and weak and it is gradually becoming smaller and smaller as time goes by. It shows that the information-flow between stocks is fast enough to support the efficient market hypothesis.

The Epps effect is a well-known effect for equal-time cross-correlations [16]. The effect represents that the cross-correlations of stock returns depend on the data sampling frequency. The cross-correlations are smaller according to the decrease of their time window. It is generally explained by the lead–lag effect between stock returns and asynchronicity of

* Corresponding author.
E-mail address: Archangel@kaist.ac.kr (S.Y. Lee).

0378-4371/$ – see front matter Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.11.012

stock trades. Toth et al. [17] introduced another source of the effect. A cross-correlation for any sampling time depends on the decay of lagged auto-correlations of shorter sampling time. Their result empirically fits the Epps effect well.

Precedent studies on the cross-correlations, both time-dependent and equal-time, were done by using log returns of stocks [5,12,13,16,17]. In the finance market, both stock returns and stock volumes are two of the most important aspects [18–20]. Prices (or returns) move generally in response to the aggregate supply and demand of quotes. The supply and demand are shown by trade and quotes (TAQ) volume data. Therefore, studying stock volumes is necessary and important to predict the returns of stocks.

In this paper, we study both time-dependent and equal-time cross-correlations by using buy and sell trading volumes of stocks in high frequency. We first check the existence of pulling effect in volume space and compare the pulling effects for both returns and volumes of stocks. We also observe the difference of pulling effects in buy and sell volumes. Buy volumes of stocks are more synchronized than sell volumes in high frequency. Through the directed MST of buy and sell volumes, we suggest that the different reaction times of stocks about the macroeconomic events influence the pulling effect. The equal-time cross-correlation of buy and sell volumes are reversed according to their sampling time scale.

This paper is organized as follows. In the next section we introduce the construction method of buy and sell volumes from the TAQ database. We describe the methodology of our computations. In Section 3 we analyze the time-dependent cross-correlations in order to study the pulling effects. Then we observe the characteristics of the MST with directed links. In Section 4 we calculate the Epps effect of buy and sell volumes separately, and compare them. The paper ends with a brief discussion in Section 5.

2. Methodology

2.1. Data

We analyzed the TAQ database for 204 days from Jan. 2 2006–Nov. 31 2006. We selected 199 stocks that were included in Korea Stock Price Index 200 (KOSPI 200) during the period. The index represents the Korean stock market and it is the standard of pricing for derivatives. Most of these stocks are actively traded for the purposes of hedge and speculation.

Unlike returns, volumes have the U-shape pattern in intraday [21]. This means that the shape of the trading volume graph in intraday resembles the capital letter U. This characteristic of volumes could distort the time-dependent cross-correlations. In order to remove this intraday trend, we constructed the volume data from 9:30 AM–14:30 PM, which is the intraday period that lies between 30 min after the market opens and 30 min before it closes. We cross checked the correlations from our method against those calculated by using the detrended data [22]. Their detailed results are given in Appendix A.

In electronic stock markets like the Korean stock market, ordered quotes are displayed in a limit order book, and these combinations determine the buy and sell volumes. If the bid quote's price is ordered higher than the inside asking price of the limit order book, then the deal is achieved. We defined the trade volume, generated by the quote, as the buy volume. The sell volume is the trade volume that is traded by an ask quote whose price is lower than the inside bidding price of the limit order book.

To make sure that our results are consistent, we divided the data into several sub-data and analyzed each sub-data. First, we divided the data into 4 sub-period data which consists of about 50 days each. Secondly, we divided the data into 2 sub-list data by capitalized orders. Our time-dependent cross-correlations were calculated in all sub-data and they revealed consistent behaviors. In Section 3, the time-dependent cross-correlations of the entire data is displayed.

2.2. Time-dependent cross-correlation

In Korea, stocks are traded in the national exchange during a limited time, not for 24 h. The time series data of stocks has different behaviors at the boundaries of the date due to the overnight effect. This effect has an impact on the time-dependent cross-correlations. In order to remove this effect, we calculated the time-dependent cross-correlations throughout the day independently, and analyzed their average over the period.

In this paper, we used the buy and sell volumes as random variables that were generally defined in the function of correlation. For example, we defined $\mathbb{B}$ as the buy volumes traded in a time window Δt. The definition of the time-dependent cross-correlations is

$$C_{\Delta t}^{A,B}(\tau) = \frac{\langle \mathbb{B}_{\Delta t}^{A}(t+\tau)\mathbb{B}_{\Delta t}^{B}(t)\rangle - \langle \mathbb{B}_{\Delta t}^{A}(t+\tau)\rangle\langle \mathbb{B}_{\Delta t}^{B}(t)\rangle}{\sigma_A \sigma_B} \tag{1}$$

where the notation $\langle \cdot \rangle$ means average and $\sigma^2 = \langle (\mathbb{B}_{\Delta t}(t) - \langle \mathbb{B}_{\Delta t}(t)\rangle)^2\rangle$. The equal-time cross-correlation is $C_{\Delta t}^{A,B}(0)$.

We followed the method and time window Δt used in Refs. [12,13]. It is helpful to compare our pulling effects with their pulling effects. They suggested an overlapping method to calculate the correlation instead of using distinct time data in high frequency. Because trades in a short time window are rare, it is not easy to find the maximum value of correlations that can fluctuate widely. Also, if the time shift which has a maximum cross-correlation is smaller than the time window Δt, we can detect it by using the overlapping method.