Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 126 (2018) 1029–1036

Procedia Computer Science

www.elsevier.com/locate/procedia

22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Statistical distribution of the sub-second price fluctuation in the latest arrowhead market

Mieko Tanaka-Yamawaki[a*] Masanori Yamanaka[b], Yumihiko S. Ikura[a]

[a]*Department of Mathematical Sciences Based on Modeling and Analysis, School of Interdisciplinary Mathematical Sciences, Meiji University, 1-21-4 Nakano, Nakano-ku, Tokyo, 164-8525, Japan*

[b]*Department of Physics, School of Science and Engineering, Nihon University, 14-8-1, Kanda-Surugadai, Chiyoda-ku Tokyo 101-8308, Japan*

Abstract

We study the statistical nature of stock price fluctuation in the ultrafast stock trading system called the "arrowhead market". The first stage of this "arrowhead market" was started in the beginning of January 2010 at Tokyo Security Exchange market in order to reach the speed of transactions at the level of one millisecond. Then it was upgraded to the second stage to reach the speed of half a millisecond in September 2015, after the merge of Tokyo market and Osaka market to become Japan Exchange Group, Inc. Due to the immense size of the data, only one stock from October 2015 to December 2016 is treated in this article. The empirical distributions of stock prices converted to stock returns every 0.1 second follow the scaling law of the Lévy stable distribution of index $\alpha = 1.7 \pm 0.3$, as a result of scaling study of the coarse graining in the range of 0.8 seconds to 13 minutes. This result clarified that the distribution is consistent to the Lévy stable distribution, the same as our past study on the first stage of the arrowhead market, as a result of coarse graining in the range of 5 seconds to 2 minutes. Although preliminary, this result may imply that the efficient market is realized in the "arrowhead market".

© 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of KES International.

Keywords: *arrowhead stock maket; statistical distribution; Lévy stable distribution; index α = 1.7;scaling phenomena*

*Corresponding author. Tel. +81-3-5343-8337

E-mail address: mieko@meiji.ac.jp

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of KES International.
10.1016/j.procs.2018.08.039

1. Introduction

It is known that the essential nature of price fluctuation is the random walk (Brownian motion) [1]. However, the details of the price motion tell us a lot about the state of the market. In particular, a notable analysis on short-time price change showed a possibility of deviation from the pure random walk [2]. We have been studying the various properties of price increments in recent years. In particular, the predictability by using the evolutional computation [3-4], trend extraction based on the RMT-oriented principal component analysis [5], possible applications of the randomness level of the price time series [6], and so on.

Meanwhile, the stock exchange system met huge reformation, and it changed from buying and selling by a person into the mechanical dealings using internet. As a result, the trading speed as well as traded amount increased incredibly. In Japan, Tokyo Security Exchange Market (TSE) and Osaka Security Exchange Market (OSE) merged to establish Japan Exchange Group, Inc. in 2013. A few years before the merge, TSE started a new trading system called "Arrowhead system" that makes the trading speed as short as a millisecond from January, 2010. This system was further renovated in the middle of September, 2015 to the 2nd stage arrowhead system that makes the trading speed to half a millisecond [7]. This renovation attracted a lot of attention among the financial technology. In particular, the discussion on the possibility of checking the efficient market hypothesis in this ultrafast trading system [8].

We also became interested in the nature of this ultrafast price motions in this new trading system. As a first step, we collected 5 second sample price time series of the arrowhead market from April to December in 2013, by downloading the price time series every day throughout the period, from the homepage of the TMIV (Tokyo Security Exchange Market Impact View) [9]. This internet site was prepared for traders to view the real-time motion of the market. For this reason, 100 stocks are selected to include various sizes and business types.

Our result using this TMIV data was reported in KES2017. Based on extensive numerical analysis, we concluded that the statistical distribution of the average of 5 second returns obeys a scale-invariant distribution called Lévy stable distribution of index α=1.4. This result was supported by another analysis using 1-minute price time series downloaded from the google site [10]. We have reported those result in KES2017 [11].

In this paper, we report the statistical distributions of price fluctuation obtained from the sub-second range to a few minutes, in order to show their scale invariant property. The rest of the paper is structured as follows. In Section 2, we review the reason of stable distribution in the price dynamics. In Section 3, we summarize the result of our former analysis on the TMIV data [9] using five-second sampled prices of 100 companies of Tokyo market in 2013, in which the average stock prices per 5 second are well described by Lévy stable distribution of index α=1.4, based on the fact that the distribution follows the scale invariance for a wide range of time scale Δt=1 to 12. In Section 4, we analyze newly obtained full arrowhead stock price data of the years 2015-2016 [8] to show that the scale invariance seems to hold in the range of 0.8 second to one hour, although the estimated range of index α is rather broad (α = 1.7 $\pm$ 0.3). Finally, Section 5 is devoted for the conclusion.

2. Reasons of Stable Distribution

We are interested in the statistical distribution of the price increment, which is often called as log-return

$$\boldsymbol{Z(t) = logX(t + \Delta t) - log\, X(t)}$$
(1)

of the asset price $X(t)$ at time t and the same price $X(t+\Delta t)$ at $t+\Delta t$, to clarify whether the statistical distribution of the price returns is not purely Gaussian but has fat-tails and narrow necks. Several decades ago, it was pointed out by Mandelbrot then followed by Mantegna and Stanley [2] that the probability distribution of asset returns follow Lévy stable distribution, defined as

$$f_{\alpha,\beta}(Z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}}\, dk \qquad (2)$$

which is the Fourier transform of the kernel $F(k)$ given by

$$F_{\alpha,\beta}(k) = e^{-\beta|k|^{\alpha}} \qquad (3)$$

The first parameter α characterizes the distribution and is called Lévy index, taking the range of $1 \leq \alpha \leq 2$, and the second parameter β is proportional to the time interval Δt, as follows.

$$\beta = \gamma \Delta t$$
(4)

Equation (4) can be understood as follows. The stable distribution holds the same index α under convolution of two stochastic variables following the same stable distributions: *i.e.*, $z=x+y$ follows Lévy stable distribution of index α if both x and y follow Lévy stable distribution of the same index α. This means that the distribution of asset returns at 5 seconds ($\Delta t=1$) follows the same distribution as the same asset returns at 10 seconds ($\Delta t=2$).

$$f_{\Delta t=2}(z) = \int_0^z f_{\Delta t=1}(x) f_{\Delta t=1}(z-x)dx \quad (5)$$

In the Fourier space, a convolution is reduced to a product of the Fourier kernels.

$$F_{\Delta t=2}(k) = (F_{\Delta t=1}(k))^2 \quad (6)$$

which can be generated to the case of n steps to have

$$F_{\Delta t=n}(k) = (F_{\Delta t=1}(k))^n \quad (7)$$

A series of n steps yields β to be multiplied by n, without changing the Lévy index α. As long as the statistical distribution of the stochastic variable z(t) in Eq. (1) has the form of stable distribution in Eq.(3), the distribution of one step of price change follows the same distribution of n-steps with the same index α but β multiplied by n.

Note that (2) can be integrated for two special cases, α=1 and α=2, first of which is the Lorentz distribution,

$$P_{\alpha=1,\beta}(Z) = \frac{\beta}{\pi} \frac{1}{\beta^2+Z^2} \quad (8)$$

and the second is the normal (Gaussian) distribution.

$$P_{\alpha=2,\beta}(Z) = \frac{1}{2\sqrt{\pi\beta}} exp\left(-\frac{Z^2}{4\beta}\right) \quad (9)$$

For general values of α, the distribution is computed by numerically integrating Eq. (2).

The scale invariant property of Lévy stable distribution is derived from Eq. (2),

$$P_{\alpha,\beta}\left(Z/(\Delta t)^{1/\alpha}\right) = (\Delta t)^{1/\alpha} \; P_{\alpha,\beta\Delta t}(Z) \quad (10)$$

Setting Z=0 in Eq. (10), Lévy index α is estimated by comparing the height of the distribution $P_{\Delta t}(0)$ for various values of Δt.

$$log\;(P_{\alpha,\beta\Delta t}(0)) = -\frac{1}{\alpha} log\;(\Delta t) + log(P_{\alpha,\beta}(0)) \quad (11)$$

The above scenario was applied on American stock index S&P500, per 1 minute for 1984-1985, and per 15 seconds for 1986-1989, which was well-fitted to Lévy stable distribution around the center of the distribution, and the scale invariant property was proved in the range of Δt =1-100 min [2].

The scale invariant property of Lévy stable distribution is derived from Eq. (2),

$$P_{\alpha,\;\beta}(Z_s) = (\Delta t)^{1/\alpha} \; P_{\alpha,\;\beta\Delta t} \quad (12)$$

$$Z_s = Z/(\Delta t)^{1/\alpha} \quad (13)$$

3. Preliminary result by 5 second sampled data

Although the arrowhead trading system was introduced in Tokyo Security Exchange (TSE) on January 4, 2010, it was hard for us to access to the full numerical data due to its huge size. Tokyo Market Impact View (TMIV) [9] offered us an opportunity to download sampled prices of 100 selected stocks per 5 seconds for a limited time from

April to December, 2013 (Data-A).

Table 1. The data sizes of active stocks are compared to Data-A.

Stock code	Data-A(2013 100 stocks)	2016 Feb.	2016 Nov.	2016 Dec.
7203	640,800	20,259,440	13,430,448	10,297,008
8306	0	41,367,900	36,330,455	31,418,450
9984	0	25,953,408	9,984,780	14,009,724

We investigated the statistical property of the price increment of TMIV, and obtained the empirical probability distribution of the average of the 100 stock prices for various time intervals Δt=1, corresponding to the interval of 5 seconds, 3, 6, 12, 24, corresponding to the interval of 2 minutes. If the statistical distribution if the price increments $Z(t)$ is indeed the scale-invariant distribution, those histograms of five different values of Δt should overlap each other after the scaling transformations of Eq. (12) and Eq. (13). We have found that the histograms of various values of the scale parameter Δt overlap on a single distribution by rescaling acording to Eq.(12) if the paraneter α is chosen to the value α=1.4.

This result was cross checked by using Eq.(11) by fitting the slope of log P (0) vs. $log\ (\Delta t)$ as the inverse of Lévy index α.

$$log\ \mathrm{P}(0) = -\ 0.709\ log(\Delta t)+2.56 \tag{14}$$

The result α=1/0.709=1.41 is consistent to the result obtained above.

So far, we have seen that our analyses on Data-A (5 seconds resolution of TSE arrowhead market) gave us a consistent result. However, a question remains. The price increments looked like purely random in early 20th century. However, it was shown that the price changes are governed by the scale invariance under high resolution analyses. Also, it became clear that the probability distribution of the price changes is featured by the fat-tails and a narrow neck. We have to clarify to what level of resolution this phenomenon goes on. We need to determine whether or not the scale invariant property is valid under the arrowhead market in which the assets are traded under ultra-high resolution shorter than a millisecond.

Before getting into the arrowhead market, we attempted to check our results to another independent data of 1 minute resolution, downloaded from Google Finance site [10] for the duration of June 16, 2015 to November 4, 2015. We call this data Data-B. However, the time resolution (frequency) of this Data set is not as small as the previous Data-A and the scaling method is not suitable to analyze this data. We need a different method for Data-B. Since we cannot compare the distributions of different time resolutions, we adopt another method to search for the best value of α to minimize Kulbuck-Leibler divergence (K-L divergence) between two probability distributions $p(x)$ and $q(x)$ defined by

$$D(p||q) = \sum_{x \in X} p(x) log\ \frac{p(x)}{q(x)} \tag{15}$$

We compute $D(p||q)$ in Eq. (15) by setting $p(x)$ and $q(x)$, as the probability distribution of the 1- minute price increments (return) and the corresponding Lévy stable distribution for various values of α and β. The best fitted result for those parameters is consistent with the case of Data-A [9].

4. Full arrowhead price data

Recently, full arrowhead price data became available via the web page of JPX [7]. Compared to the Data-A, the data sizes are incredibly large. They are compared in Table I. The most active stock in Nov. 2016 has over 36 million data points in one month, and the sum of Nov. and Dec. 2016 has comparable size to that of total 100 companies in Data-A.

Moreover, the times of trades are utterly irregular in the case of arrowhead data, while Data-A has exactly 3600 points each day. Typical directory sizes for January-December 2012 are illustrated in Fig.1, and the typical sizes of top-ranked active stocks are shown in Table II, taking as example from February 2016.

201201 contains 2439 files
1301
1305
..
9997
201202 contains 2442 files
1301
1305

Fig. 1. Typical directory size of arrowhead data.

We began our analysis from the most active stock, code number 8306, and obtained empirical probability distribution by using two different ways of data processing.

(1) The data sampled into 100 millisecond intervals

We first pick up the stock prices every 100 millisecond interval to make a time series of the stock prices of fixed time interval from October, 2015 to December 2016. Since the original data do not necessarily have a price at every 100 milliseconds, the nearest price before that point is filled for that position. Thus a long queue of the same price is created if no trade occurs for a long time.
An example of the created fixed-time interval data is illustrated in Fig.2. This example shows a case where the price was 572.9 at the 13-hour-42-minute-54-second, and so on.

Fixed time interval of 0.1 second	
13 42 54. 0	572.9
13 42 54.1	572.9
13 42 54.2	573.0
13 42 54.3	573.0
13 42 54.4	573.0
13 42 54.5	573.0

Fig.2 The original arrowhead data is formed into time series of a fixed time interval of 100 milliseconds.

Based on this data file, we draw the empirical probability distributions for various values of Δt. We focus on the graphs for Δt= 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, and 8192 (×100ms). Those eleven graphs are simultaneously shown in Fig.3.

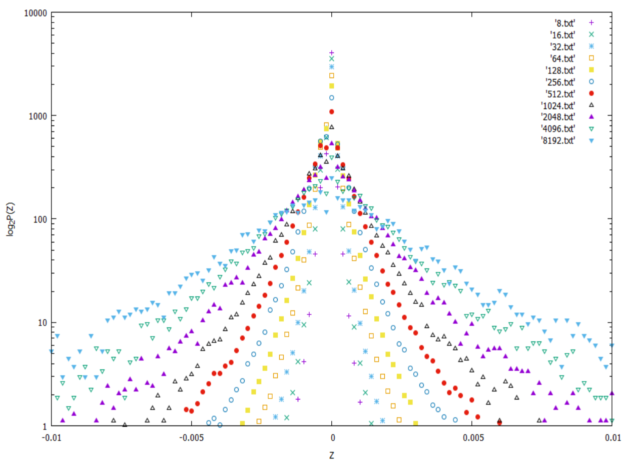

Fig.3 The histograms of 100ms returns of stock code 8306 are compared for various levels of coarse graining, 8.txt, 32.txt,.., 8192.txt, corresponding to the time scales, Δt =8-8192 (unit 100ms).

The graph for Δt= 8 has the smallest width on the horizontal axis Z and the tallest height on the vertical axis $\log_2 P(Z)$, and the graph for Δt= 16 is slightly smaller width in Z and shorter height in the vertical axis. Those histograms of regularly increasing time scales seem to obey some regularity. If they obey a scale-invariant distribution such as Lévy stable distribution, we should be able to identify the scaling factor $c = (\Delta t)^{1/\alpha}$. For example, the graphs for Δt= 8 should overlap the graph for Δt= 16 by multiply Z by the factor $c=2^{1/\alpha}$ and divide the vertical axis by the same factor c. Applying the same rule on all the eleven histograms, they should be able to overlap on a single distribution if the factor c is properly chosen. This is done by choosing c=1.5 as shown in Fig.4. All the eleven histograms corresponding to Δt= 8-8192 (×100ms) can be scaled to a single curve by choosing c=1.5 and the corresponding index is around α=1.7.

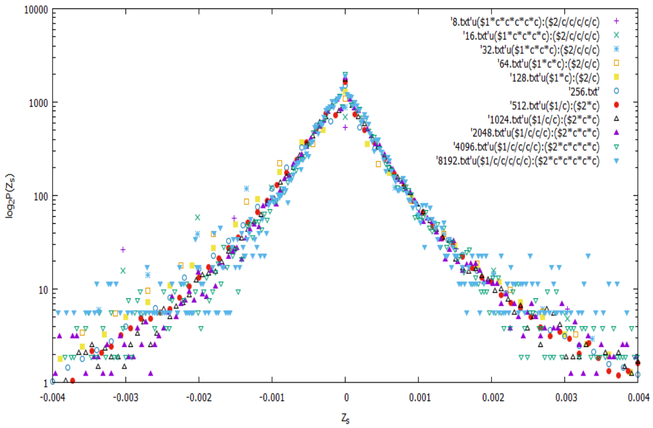

Fig.4 The histograms in Fig.3 of different time resolutions, Δt =8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192 (unit 100ms) can be rescaled to overlap on a single curve by properly choosing the scaling factor $c = (\Delta t)^{1/\alpha}$ =1.5 which derives the index α =1.7.

Unfortunately, the resolution of this estimate is not high and the accuracy of the factor c varies in the range of $1.4 < c < 1.6$ according to the estimation of P(0). This uncertainty of c implies the uncertainty of the index α, in the range of $1.4 < \alpha < 2$, or $\alpha = 1.7 \pm 0.3$ as shown in Table 2. The uncertainty of P(0) comes from the excess zeros created in the process of making the price data per 100 milliseconds.

Table 2. The scale factor and the values of Levy index.

$c = (2)^{-1/\alpha}$	1.4	1.6
α	2.06	1/47

(2) All the prices are used disregarding time stamps

In order to remove excess zeros, we adopted another way of setting the arrowhead data. Instead of forming the price data by a fixed time interval, all the price data by disregarding the time stamps are used, as illustrated in Fig. 5. In doing so, we can use the entire set of arrowhead price data, in exchange of sacrificing the information of actual time stamps. Although the correspondence to the actual time interval is lost, a big advantage in the second way of data processing is the lack of excess zero problem. Moreover, processing time is shorter due to the simplicity of the process. By drawing histograms for various levels of coarse graining parameter dt=8,…,2048, we recognize a similar scaling property observed in the first method of analysis shown in Fig.3, as shown in Fig.6. The histograms in Fig.6 are tested for the possibility of scaling. It turns out that the scaling is achieved at the value of the parameter α =2. Although this value stays within the range of error bar of the result in the first way of data processing, α =2 implies Gaussian distribution, which conflicts the fat-tail distribution widely accepted nowadays.

yr. mo. day	hour. min. sec.	price
20160229	145950355000	487.700000
20160229	145950422000	487.600000
20160229	145950516400	487.600000
20160229	145950517500	487.600000
20160229	145950685100	487.700000

For Δt=4, log-returns are computed as follows.

Z(1)=log(487.6)-log(487.7)
Z(2)=log(487.6)-log(487.6)
Z(3)=log(487.6)-log(487.6)
Z(4)=log(487.7)-log(487.6)

Fig.5 The second method of arrowhead data processing of taking all the prices without time stamps illustrated.

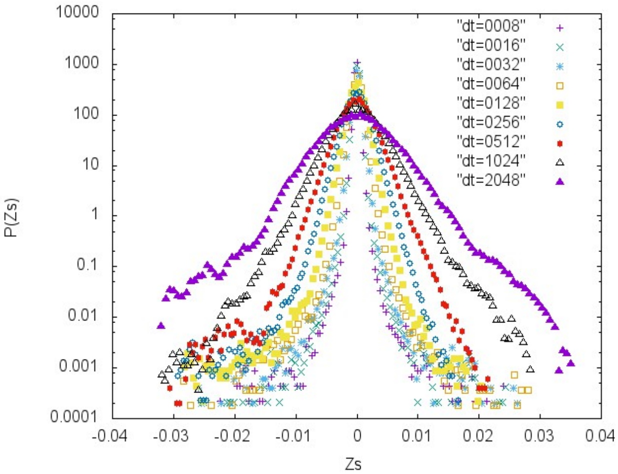

Fig.6 The histograms of 100ms returns of stock code 8306 are compared for various levels of coarse graining, dt=8, dt=16,.., dt=2048, for the arrowhead price data processed in the second way(no time information).

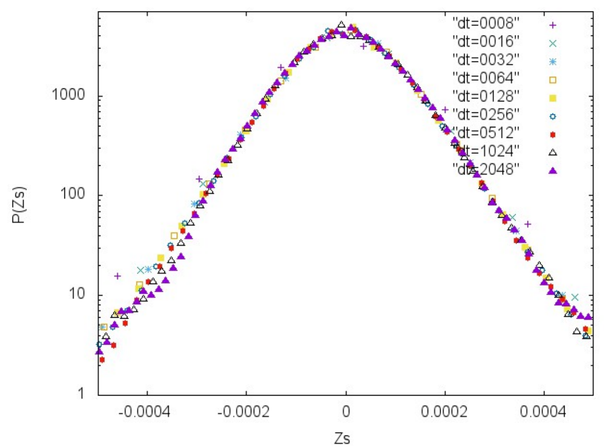

Fig.7 The 9 curves in Fig.6 of different levels of coarse graining, dt=8, 16, 32, 64, 128, 256, 512, 1024, 2048 can be rescaled to overlap on a single curve by properly choosing the scaling factor $c = (\Delta t)^{1/\alpha}$. This figure shows the case of index $\alpha = 2$.

5. Conclusion

We focused in this work to discover possible new elements to characterize the price changes under ultrafast market transactions of sub-millisecond intervals in the arrowhead market, operated in Tokyo market from 2010. In particular, we investigated the shape of the statistical distribution of the price increments. Especially, we obtained the probability distribution of the asset returns and examined the central part of the distribution utilizing its scale-invariant property. In our previous work using 5 second resolution data [11], however, the distribution turned out to be the same as the result of one-minute resolution data in [2]. In this paper we show, using the new data of 100ms resolution, that the same kind of scale-invariant statistical distribution holds for the sub-second motion of price changes, although the index to characterize the scale invariance comes out to be α=1.7. Considering various uncertainties, this value is roughly consistent to our previous result.

References

[1] Bachelier, L. J. B.; "Theory of Speculation", 1900; Bachelier, L. J. B., Davis, M. and Etheridge, A.; "Louis Bachelier's theory of speculation : the origins of modern finance", Princeton University Press, 2006.

[2] Mantegna, R. N. and Stanley, H.E.(1995); "Scaling behaviour in the dynamics of an economic index", Nature Vol.376, pp.46-49.

[3] Tanaka-Yamawaki, M. and Motoyama, T.; "Predicting the Tick-wise Price Fluctuations by Means of Evolutionary Computation", IEEE-CEC2004, pp.955-958, 2004; Tanaka-Yamawaki, M. and Tokuoka, S.; Physica A 383, pp.125–133, 2007.

[4] Tanaka-Yamawaki, M.; "Cross Correlation of Intra-day Stock Prices in Comparison to Random Matrix Theory", Intelligent Information Management(http://www.scrp.org), Vol.3, 65-70, 2011.

[5] Tanaka-Yamawaki, M.; Kido, T. and Itoi, R.; "Trend Extraction of Stock Prices in the American Market by Means of RMT-PCA", Intelligent Decision Technologies (SIST10,Springer), pp.637-646, 2011.

[6] Yang, X., Mikamori, Y. and Tanaka-Yamawaki, M.; "Predicting the security levels of stock investment by using the RMT-test", KES2013: Procedia Computer Science, Vol/22, pp.1172-1181, 2013.

[7] www.jpx.co.jp/english/corporate/news-releases/0060/20150924 -01.html.

[8] Nagata, S. and Inui, K.; " Does high-speed trading enhances market efficiency? Empirical Analysis on "Arrowhead" of Tokyo Stock Exchange", Journal of Trading, Vol.3 (4), pp.75-81, 2014.

[9] www.tse.marketimpactview.com/.

[10] finance.google.com/finance/.

[11] Tanaka-Yamawaki, M.; "Statistical Distribution of the Arrowhead Price Fluctuation", KES2017: Procedia Computer Science, Vol. 112, pp. 1439–1447, 2017.