What Do Sub-second Price Data Tell Us about the Arrowhead Stock Market?

Mieko Tanaka-Yamawaki
Dept. Math. Sciences, School of Interdisciplinary Math.
Meiji University, 1-21-4, Nakano, Nakano-ku,
Tokyo, 164-8525, Japan
e-mail: mieko@meiji.ac.jp

Masanori Yamanaka
Dept. Physics, School of Science and Engineering
Nihon University, 14-8-1, Kanda-Surugadai, Chiyoda-ku
Tokyo 101-8308, Japan
e-mail: yamanaka@phys.cst.nihon-u.ac.jp

***Abstract*—In this paper, we study ultrafast stock time series of the newly developed arrowhead trading system in Tokyo Market, in order to investigate the statistical nature of the stock time series under sub-second time scales. We also compare the current result to the past study on longer time scale up to a few minutes. It is shown that the empirical distributions obtained in this study follow the scaling law of the Lévy stable distribution of index α ranging from 1.4 to 2.0.**

Keywords-stock time series; arrowhead market, statistical distribution, scaling phenomena.

I. INTRODUCTION

The science of price fluctuation was initiated by a French Mathematician Luis Bachelier, who recognized the nature of price fluctuation as the random walk (Brownian motion) in 1900 [1], which was five years earlier than Albert Einstein's formulation of random walk in physics. This tradition is still carried over in the basic theory of financial technology to evaluate the derivative prices, such as Black-Sholes-Merton (BSM) formula [2]-[5].

However, it is well-known that the BSM formula often fails to describe the real world. While the important parameter σ (volatility) is assumed to be a certain constant in the above formula, there is no reliable way to compute its value theoretically.

Two empirical ways are often used to obtain the value of σ: One is the 'historical volatility', or the realized volatility', to compute the average values of the standard deviation over the historical price data over a fixed length, such as 2 weeks. Another is the 'implied volatility' to obtain σ by inversely solving the BSM formula from for the actual price time series of the option prices. However, the obtained values σ are not a constant but varies as a function of K (the target price of each option) of the same option for different terms T. This is known as the 'smile curve' because the σ-K plot shapes concave and resembles the 'smile' mark. Considering the importance of the derivation of the BSM formula in financial engineering, it is essential to solve the problem of σ.

Another problem of the BSM formula is the basic assumption of Gaussian nature of price fluctuation, which is incompatible with the observed 'fat-tail', or 'narrow-neck' nature of the actual statistics of the price fluctuation. Moreover, it is widely accepted that the price fluctuation has the scale invariant property, which is incompatible with the Gaussian distribution, since Gaussian distribution is bounded by the scale of variance, or standard deviation, σ. In order to remedy this situation, a scale-invariant distribution called Lévy stable distribution is proposed and the index α=1.4 was discussed widely [6][7]. Although the infinite variance in Lévy stable distribution is not mathematically compatible to framework of option pricing theory, actual price fluctuations behave more like Lévy stable distribution than Gaussian. Thus, it is a highly challenging problem to clarify the statistical nature of price fluctuation in various range of time resolutions.

In this work, we investigate the new world of arrowhead stock market [8], not only to determine the shape of the statistical distribution, but also to examine various results obtained so far, such as cross correlation between different stocks using random matrix theory-oriented principal component analyses and related techniques [9]-[11], for the stock market of normal speed to the results on this newly developed arrowhead market.

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 flows. In Section II, we summarize the formulation of price dynamics. In Section III, we show the result of our former analysis [12] 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 IV, we analyze newly obtained full arrowhead stock price data of the years 2015-2016 [12] 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 ($1.4<\alpha<2.1$). Finally, Section V is devoted to the conclusion.

II. FORMULATION OF PRICE DYNAMICS

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

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

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 [6] then followed by Mantegna and Stanley [7] that the probability distribution of asset returns follow Lévy stable distribution, defined as

 ISBN: 978-1-61208-620-0

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

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

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

The first parameter $\boldsymbol{\alpha}$ characterizes the distribution and is called Lévy index, taking the range of $\mathbf{1 \le \alpha \le 2}$, and the second parameter $\boldsymbol{\beta}$ is proportional to the time interval Δt, as follows.

$$\boldsymbol{\beta = \gamma \Delta t} \quad (4)$$

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 $\boldsymbol{\beta}$ to be multiplied by n, without changing the Lévy index α. However, this model of price movements naturally assumes a limitation on the maximum number of steps, 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 $\boldsymbol{\Delta t}$.

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

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 $\boldsymbol{\Delta t}$ =1-100 min [7].

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

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

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

III. 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) [13] 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).

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, as shown in Fig.1 [8]. 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).

As shown in Fig.2, histograms of various values of the scale parameter Δt in Fig. 1 overlap on a single distribution by rescaling acording to Eq.(12) if the parameter α is chosen to the value α=1.4.

The scale invariance of the statistical distribution of price increments can be checked using two other methods. One way is to use Eq. (11) for checking the straightness of the log-log plot of P(0) vs. Δt and also to obtain the Lévy index α from the slope ($-1/\alpha$) of the plot. By means of the least square fit, the best fit line turns out

By using this data set, we investigated the statistical distribution of the price increment

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

as shown in Fig.3. The Lévy index α obtained in this result is α=1/0.709=1.41, which is consistent to in Fig. 2. [7]

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 [14] 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

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

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, as shown in Table 1 [12].

IV. Full arrowhead price data

Recently, full arrowhead price data became available via the web page of JPX [8]. Compared to the Data-A, the data sizes are incredibly large. They are compared in Table 2. 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.

We began our analysis from the most active stock, code number 8306. We first pick up the stock prices every 100 millisecond interval to make a time series of the stock prices from October, 2015 to December 2016. Based on this data file, we draw the empirical probability distributions for various values of Δt. For the sake of simplicity, we focus on the graphs for Δt= 8, 16, 32, 64, 128, 256, 512,1024, 2048, 4096, and 8192 (×100ms). Those eleven histograms are simultaneously shown in Fig.4. 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.5. All the eleven histograms corresponding to Δt= 8, 16, 32, 64, 128, 256, 512,1024, 2048, 4096, and 8192 (×100ms) can be scaled to a single curve by choosing c=1.5 and the corresponding index is around 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 $1.4 < \alpha < 2$, as shown in Table III. The uncertainty of P(0) comes from the nature of the price data, since the observed number of unchanged price contains numerous counts of the ‘absence of trade’ on top of the ‘trade with the same price’. It is hard to distinguish those two by the data. However, it is possible to estimate the true value of P(0) in such a way for the probability P(Z) to satisfy a smoothness by removing excess P(0) from the data.

V. 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 [12], however, the distribution turned out to be the same as the result of one-minute resolution data in [7]. 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 in [12].

References

[1] L. J. B. Bachelier, Theory of Speculation, 1900.

[2] F. Black and M. Scholes, Journal of Political Economy, Vol. 81, pp. 637-654, 1973.

[3] R. Merton, The Review of Economics and Statistics, Vol. 51 pp. 247–257, 1969.

[4] R. C. Merton, The Bell Journal of Economics and Management Science, Vol. 4, pp. 141–183, 1973.

[5] R. C. Merton, “Option Pricing When Underlying Stock Returns Are Discontinuous”, Journal of Financial Economics, Vol. 3, pp. 125-144, 1976.

[6] B. B. Mandelbrot, “The variation of Certain Speculative Prices”, J. Buisiness, Vol. 36, pp.394-419, 1963.

[7] R. N. Mantegna and H.E. Stanley, Nature Vol.376, pp.46-49, 1995.

[8] www.jpx.co.jp/english/corporate/news-releases/0060/20150924-01.html; [retrieved March 2018]

[9] S. Nagata and K. Inui, “Does high-speed trading enhances market efficiency?”, Journal of Trading, Vol.3, pp.75-81, 2014

[10] M. Tanaka-Yamawaki et al.;“Trend-extraction of Stock Prices in the American Market by Means of RMT-PCA”; Intelligent Decision Technologies,SIST10, pp.637-646, 2011.

[11] X.Yang, Y.Mikamori, and M. Tanaka-Yamawaki, Procedia Computer Science, Vol.22, pp.1172-1181, 2013.

[12] M. Tanaka-Yamawaki, Procedia Computer Science, Vol. 112, pp. 1439–1447, 2017.

[13] www.tse.marketimpactview.com [retrieved March 2018]
[14] finance.google.com/finance [retrieved March 2018]

TABLE I. THE K-L DIVERGENCE OF THE BEST FIT INDICES OF LÉVY STABLE DISTRIBUTION AND DATA-B

data (1 min return)	α	β	K-L DIV.
Ave. of 440 returns	1.40	5.4×10^{-6}	0.039
9503	1.55	10.0×10^{-6}	0.286
7201	1.65	3.9×10^{-6}	0.423
6502	1.55	8.8×10^{-6}	0.156

TABLE II. THE DATA SIZES OF ACTIVE COMPANIES ARE COMPARED TO DATA-A

Stock code	Data-A	Nov. 2016	Dec. 2016
# of companies	100	3806	3820
8306	0	36,330,455	31,418,450
7203	640,800	13,430,448	10,297,008
9984	0	9,984,780	14,009,724

TABLE III. THE SCALE FACTOR AND THE VALUES OF LÉVY INDEX

$c = (\Delta t)^{1/\alpha}$	1.4	1.5	1.6
α	2.06	1.71	1.47

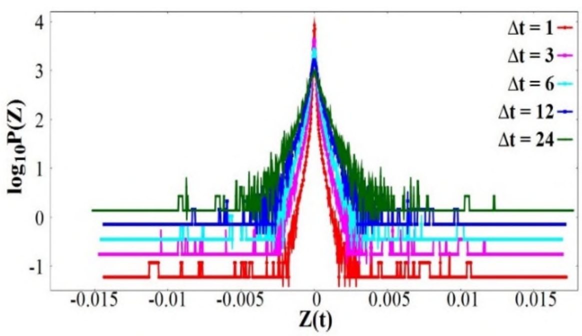

Fig.1 The histograms of statsitical distribution of Z for Δt=1 (5 sec), 3 (15 sec), 6 (30 sec), 12 (1 min), and 24 (2 min).

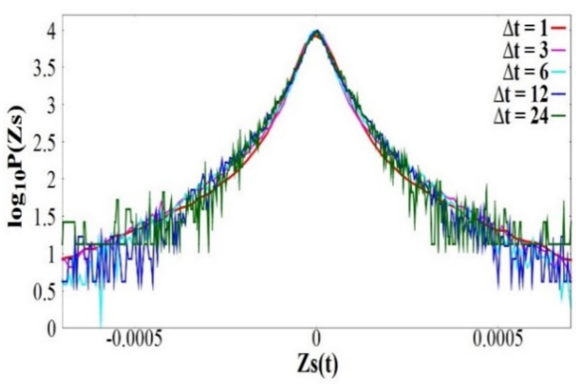

Fig.2 The histograms in Fig.1 are well scaled by Zs vs. $\log_{10}$P(Zs) for Δt=1(equivalent to 5s), 3(15s), 6(30s), 12(1 min), and 24(2 min) for the case of $\alpha = 1.4$.

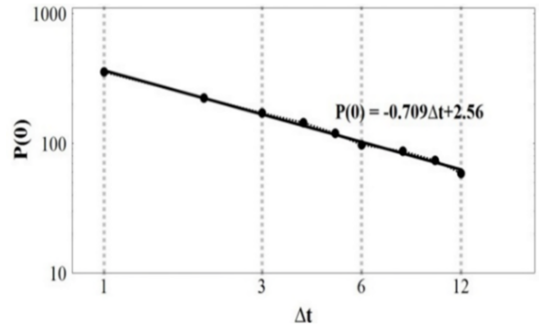

Fig. 3 The least square fit derives α =1.41, consistent to the result from Fig. 2.

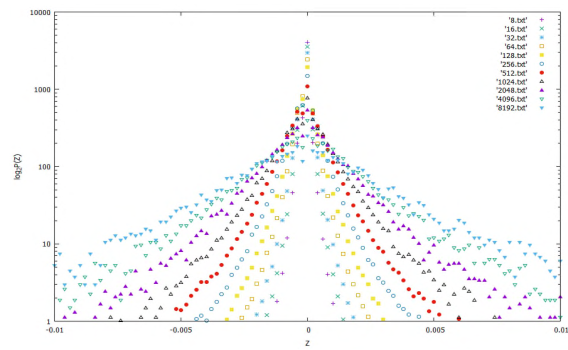

Fig. 4 The histograms of 100ms returns of stock code 8306 are compared for various levels of coarse graining, 8.txt, 32.txt,.., 8192.txt, corresponding to the time scales, Δt =8-8192 (unit 100ms).

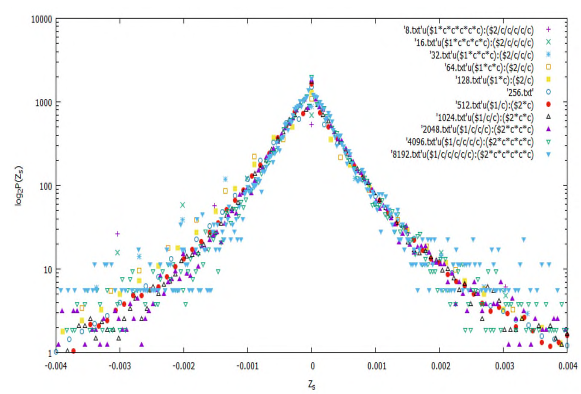

Fig. 5 The six histograms in Fig. 2 of different time resolutions, Δt =8, 32, 128, 512, 2048, 8192 (unit 100ms) can be rescaled to overlap on a single curve by properly choosing the scaling factor $c = (\Delta t)^{1/\alpha}$. This figure shows the case of c=1.5 which derives the index α = 1.7.