Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 544–551

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Market efficiency is truly enhanced in sub-second trading market

Mieko Tanaka-Yamawaki[a]*, Masanori Yamanaka[b]

[a]*Dept. Mathematical Sciences based on Modeling and Analysis, Meiji University, 1-21-4, Nakano, Nakano-ku, Tokyo, 164-8525 Japan*
[b]*Dept. Physics, College of Science and Engineering, Nihon University, 14-8-1, Kanda-Surugadai, Chiyoda-ku Tokyo, 101-8308 Japan*

Abstract

This article deals with the statistical nature of the price fluctuation. Although the price movements are roughly described as the Brownian motion, a certain discrepancy exists between the theoretically anticipated pure Gaussian random walk and the observed heavy-tailed distributions. In order to settle down this problem, the authors have surveyed massive database supplied by the newly developed stock trading system called "arrowhead trading system" in Tokyo Security Exchange Market (TSE), and have reached a conclusion supported by various numerical results, that the fundamental process of arrowhead price movements can be identified as the Gaussian random walk, or reaching toward Gaussian random walk corresponding to the index of Lévy stable distribution, $\alpha = 2.0$, to realize the efficient market as the time intervals going smaller from a few minutes to a few seconds. Also, the compound motion made by multiple consecutive motions that corresponds to the tail part of the distribution or the average price like various indices such as TSE Index in Japan or S&P500 in the U.S.A. tends to show the scale-free property corresponding to α = 1.4 - 1.7.

Keywords: Stock price fluctuation; arrowhead market; efficient market; statistical distribution; Gaussian distribution

1. Introduction

Since the work of Luis Bachelier [1], the nature of price fluctuation is known to be the random walk (Brownian

* Corresponding author. Tel.: +81-3-5343-8337; fax: +81-3-5343-8337.
E-mail address: mieko.tanakay@gmail.com, mieko@meiji.ac.jp

10.1016/j.procs.2019.09.209

motion) and the theory of financial technology is established on this base to evaluate, e.g., derivative prices such as Black-Sholes-Merton (BSM) formula [2,3,4]. However, the BSM formula has several kinds of difficulty, and often fails to describe the real world. For example, the parameter σ (volatility) is assumed to be a certain constant in the above formula because there is no reliable way to compute its value theoretically. A commonly used substitute for the volatility is the 'historical volatility', or the realized volatility', that is the average values of the standard deviation over the historical price data over a fixed length, such as 2 weeks. Another substitute 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.

From physics point of view, on the other hand, price fluctuation provides us a fruitful playground to test the relationship between the scaling and critical phenomena [6,7]. Stimulated by the discovery of scale-free property in price time series [8], a new field of interdisciplinary science on the price motions and its related phenomena have been practiced under the name of "econo-physics" [9,10]. There the universality of scale-free phenomena in a wide range of sciences have been much emphasized. Among them, the price dynamics was depicted as a typical example of having the scale-free property over the widest range of time resolution [11-13].

This direction of thought. however, often caused frustration between the 'scale-free' hypothesis in physics and the 'random walk' hypothesis in financial engineering. We attempted to solve this problem by means of direct numerical study of large-sized financial data provided by the recently developed ultra-high speed transaction at half a millisecond in Tokyo market, called 'arrowhead market'.

The rest of the paper is structured as follows. In Section 2, the formulation of price dynamics is summarized. Then the latest result of numerical analyses based on newly obtained full arrowhead stock price data are described in Section 3 and 4. Based on the detailed analysis of the most actively traded stock price of code number 8306 for four and a half years of August 2014- December 2018 to show that the fundamental price move behaves as the Gaussian random walk as expected by financial engineering. Then we proceed to show that the scale-free mechanism holds for the price motions of multiple steps as well as for the average prices like financial index, based on the scaling analysis of the stock index named TSE index for 14 years of April 2005- December 2018. Finally, Section 5 is devoted for the conclusion.

2. Statistical distribution of the price increments

The main variable of our interest is the price increment

$$z = X(t + \Delta t) - X(t) \tag{1}$$

of the asset price $X(t)$ time t and $X(t + \Delta t)$ at $t + \Delta t$. The main task is to clarify whether the statistical distribution of the price increments is Gaussian, or not Gaussian having fat-tails and narrow necks. Several decades ago, it was pointed out by Mandelbrot then followed by Mantegna and Stanley that the probability distribution of asset returns follow Lévy stable distribution, defined as

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

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

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

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.

$$\beta = \gamma \Delta t \tag{4}$$

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

$$P_{\alpha=1,\beta}(Z) = \frac{\beta}{\pi}\frac{1}{\beta^2+Z^2} \tag{5}$$

and the second is the normal (Gaussian) distribution**.**

$$P_{\alpha=2,\beta}(Z) = \frac{1}{2\sqrt{\pi\beta}} exp\left(-\frac{Z^2}{4\beta}\right) \tag{6}$$

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),

$$P_{\alpha,\beta}\left(Z/(\Delta t)^{1/\alpha}\right) = (\Delta t)^{1/\alpha}\ P_{\alpha,\beta\Delta t}(Z) \tag{7}$$

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.

$$log(\,P_{\alpha,\beta\Delta t}(0)) = -\frac{1}{\alpha} log(\Delta t) + log(\,P_{\alpha,\beta}(0)) \tag{8}$$

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 =1-1000 min. [9].

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

$$P_{\alpha,\ \beta}(\,Z_s) = (\Delta t)^{1/\alpha}\ P_{\alpha,\ \beta\Delta t}(Z) \tag{9}$$

$$Z_s = Z/(\Delta t)^{1/\alpha} \tag{10}$$

Except the case of Gauss distribution α =2.0, when the volatility parameter σ (equivalent to the standard distribution) determines the typical scale of the distribution, Lévy stable distributions of index $\boldsymbol{\alpha}$ < 2.0 has no parameter to determine the scale, and they are the scale-free distribution, or the scale-invariant distribution.

3. Fundamental price dynamics seems to be Gaussian

3.1. Millisecond data became available by the 'arrowhead market' in Tokyo market

The methodology stated above was originally proposed in Ref. [9, 10] to show that the price fluctuation has scale-invariant property, originally suggested in Ref. [8] to explain risk in the speculative market. In both cases, Lévy stable distribution of index α < 2.0 was observed**.**

Amazingly enough, the scale-invariance was observed for the extensive range of time resolution from a few minutes to s few days, for various financial time series [12-14]. However, those results are based on the index time series, which are 'average' prices of various different stock prices, or 'sum' of prices of the same stock taken at different times. It is not quite evident whether the fundamental process of price fluctuation obeys Gauss distribution or non-Gaussian 'scale-invariant' stable distribution, unless any clear evidence of non-Gaussian distribution comes out of fundamental analysis.

The arrowhead market, that started in Tokyo market from the beginning of January 2010, is expected to give us a good chance to observe the fundamental process of price fluctuation. Having this hope in mind, the authors focus on analysing the empirical statistical distribution pf price fluctuation.

At first, it was highly difficult to access the full trade data. Only an extremely limited data of every 5 second interval became available on the webpage of TMIV (Tokyo Market Impact View) for only 100 selected stocks, from April to Dec. 2013, which provides us, 720 points per hour, times 5 hours, times 179 days. The resulting data size is only 640,800 data points per each stock. The number of data points is not sufficient to draw a statistical distribution by a single stock time series only. Still the average of the 100 stocks makes us possible to draw histograms. The conclusion at that time was that the price fluctuation of this 100 stocks follow the Lévy stable distribution of index α=1.4. This result was checked by using three different ways of analyses: scaling fit, time resolution dependence of P(0), and the Kullbuck-Leibler Divergence [14].

The second preliminary analysis of the same line of thought was performed by using the full arrowhead data of half a millisecond resolution for 15 months from October ,2015 to the end of 2016 [15]. However, the data are highly sparse and not periodic. For example, the most active stock (8306) in the most actively traded month (December, 2016) had only 31,418,450 transactions, which is merely 0.3 percent of the total transaction chances of 1,000,000,000 (2000 times per second * 60 minutes *60 hours * 5 hours * 28 days). That means roughly 99.7 per cent out of all the possible transactions at every 0.0005 second are practically vacant. Therefore, the authors had to give up the maximum resolution, and decided to sample the original time series into rougher series per 0.1 second interval. Based on such sampled data, the authors observed scale-invariant property quite well and identified the index α of Lévy distribution to be α = 1.7 $\pm$0.3. However, this range of index is too wide by simultaneously allowing the Gaussian (α =2.0) and the old-time Lévy index α=1.4 obtained for 1-minute time series of the S&P500 index thirty years ago as well as the average time series of 100 selected stocks per 5 seconds in our first analysis [14]. On the other hand, it is widely accepted in the real-world that the assumption of financial technology stands over the belief of Gaussian fluctuation of fundamental moves of prices.

In order to narrow down the error bar in the estimation of index α, more data are necessary. It turns out a complete set of transaction prices, volumes and the transaction time to the accuracy of half of a millisecond for all the stocks traded in Tokyo Security Exchange market (TSE, hereafter) for seven years from 2012 to 2018, delivered by the JPX cloud service (http://www.jpx.co.jp/) became available recently. This is a great opportunity to observe and analyse the fastest transaction data available at this point, and to explore the world of sub-millisecond market system and observe how the current status of security exchange market.

3.2. Single stock price time series may exhibit the elementary price dynamics

It is expected that the single stock time series at the minimum time interval contains the basic information on the elementary price dynamics. At this moment, however, it is difficult to extract the price changes at the elementary level.
our previous experience on analysing various market indices and average stock prices. At first, when we did not have sufficient data length to analyse, we used index time series in order to obtain some sort of smooth distributions. Those 'mixed' time series tend to have smaller index values. Thus we expected that even if the fundamental motion of price fluctuation is Gaussian, there is a possibility of observing non-Gaussian distribution for mixed (averaged) time series.

In order to identify the index α based on the scaling property of Eq. (9), we pick up the most actively traded stock 8306. Although the total data set, from the beginning of 2010 when the 'Arrowhead system' has started to the end of 2018, is available, only the period of August, 2014 - September, 2018 can be consistently used for the analysis of the stock coded by 8306. The reason is due to the change of price resolution from 3 digits to 4 digits made by the system upgrade made during July 2014.

Considering various limitations coming from the sparseness of the traded record, the optimum choice of the minimum time resolution turns out to be Δt=3.75 seconds, mainly to suppress the excess-zero-problem. The resulting number of total data points for 8306 becomes 4,776,995. By using this set of data, the parameter α is estimated by two different methods.

<Method 1>: Scaling the empirical probability distributions of various time resolutions by properly choosing α

The first method is to use the scaling property in Eq.(7). This is illustrated by two graphs in Fig.1. In the left figure, eight graphs of empirical probability distributions for various time resolutions, corresponding to eight different time resolutions of Δt = 3.75, 7.5, 15, 30, 60, 120, 240, 480 seconds are simultaneously plotted. Those numbers are chosen to have Δt = 2 in Eq. (7). If the price time series of our concern obey a scale-invariant distribution such as Lévy stable distribution, we should be able to identify the scaling factor $c = (\Delta t)^{1/\alpha} = 2^{1/\alpha}$ so that all the eight lines to overlap on a single distribution if the factor c is properly chosen. All the eight histograms can be scaled to a single curve by choosing $c=\sqrt{2}$ and the corresponding index is around α=2.0, as shown in the right figure of Fig.1.

<Method 2>: Extract α from the slope of the log-log plot of P(0) and Δt

Another method can be used to obtain the value of index α from Eq. (8). Corresponding to the time resolution Δt = 3.75, 7.5, 15, 30, 60, 120, 240, 480, the peak P(0) of the probability distribution decreases according to Table 1, from which the index α as the inverse of the slope of the log-log plot shown in Fig.2 is α=1.982, consistent to α=2.0 obtained from the graphical approach above. The stability of the index α is supported by the straightness of the line in Fig.2.

Therefore, arrowhead price data for 8306 can be identified as the Gaussian distribution with high confidence level. Although the time resolution 3.75 seconds is still too large to be identified as the elementary process as a whole, it is expected that the large portion of fundamental price changes are contained in this range.

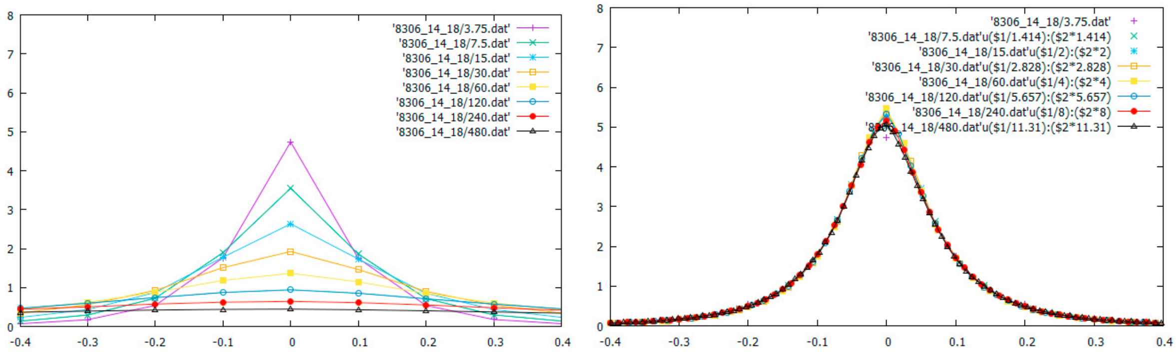

Fig.1 (left figure) Probability distribution of stock code 8306 (August 1, 2014 - September 30, 2018) as a function of stock increment z computed at various levels of time resolution Δt =1, 2, 4, 8, 16, 32, 64, 128 (unit 3.75s.) are simultaneously plotted. (right figure) The probability distributions in the left figure are scaled to overlap on a single curve by properly choosing the scaling factor to be c =$(\Delta t)^{1/\alpha} = 2^{0.5}$ implying $\alpha = 2.0$.

Table 1. Peak height P(0) of the probability distribution vs. the time resolution Δt.

Δt	3.75s.	7.5s.	15s.	30s.	60s.	120s.	480s.	960s.	1920s.	3840s.
P(0)	4792	3599	2683	1956	1384	958.9	673.4	461.5	301.1	214.2

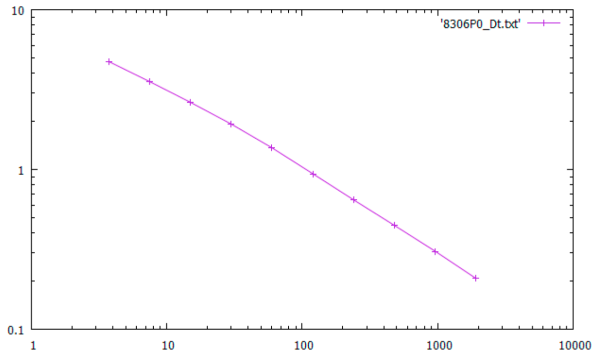

Fig.2 Plot of logP(0) vs log(Δt) in Table 1.

Note that in our previous analysis using log-returns defined in Eq.(1) between all the adjacent transactions of 8306, we disregarded time stamps and simply took the log-difference between two adjacent prices, such as

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

for all the recorded prices for i=1, …, n, where n represents the total number of data points. The statistical distribution of the resulting time series of Z(i) can be fitted to Gaussian distribution with index α=2.0, as shown in the last two figures of out last work (Fig. 5 and Fig. 6 in Ref. 15). In that analysis, most data points practically belong to sub-second region, although not at any fixed times, thus such data mostly consist of single motions. This fact can be regarded as another proof, from a different direction, for the statement that a single price moves according to Gaussian random walk.

The same conclusion has also been obtained in many other stocks. However, the density of non-vacant data points for other stocks are even lower than the case of 8306 and the figures are less clear than the case of 8306. We expect to have much denser transactions for many other stocks, where we should be able to obtain clear-cut figures in the near future.

Based on the above mentioned results, we conclude that the scaling analysis based on Eq.(9) suggests us that a single stock under sub-second (‘Arrowhead’) time resolution obeys Gaussian random walk whose probability distribution has the stable distribution of index α =2.0. Yet a serious question remains on the fat-tail property of the statistical distribution of price increments. It is clear that the both tails of Fig. 1 and Fig. 2 seem far from the Gaussian (normal) distribution. The only possible explanation is to admit that, although a single price change behaves like the Gaussian random walk, multiple steps of price fluctuation are not. Individual steps are not mutually independent but influenced by interaction. Suppose if they were indeed i.i.d., then the stable distribution holds the same index $\boldsymbol{\alpha}$ 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 Δt=1 follows the same distribution as the same asset returns at Δt=2.

$$f_{\Delta t=2}(z) = \int_0^z f_{\Delta t=1}(x) f_{\Delta t=1}(z-x)dx \tag{11}$$

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

$$F_{\Delta t=2}(k) = (F_{\Delta t=1}(k))^2 \tag{12}$$

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

$$F_{\Delta t=n}(k) = (F_{\Delta t=1}(k))^n \tag{13}$$

The fact that both tales are not Gaussian tells us the simple convolution story in Eqs. (11)-(13) fails in the price dynamics.

4 . Index time series have smaller α

Recall that scale-free Lévy stable distributions of index α=1.4 < 2.0 was derived for S&500 in Ref. [9]. The same result was also derived in our previous work in Ref. [14] for the average stock price increments over 100 selected stocks. Taking this coincidence seriously, we can draw another scenario that index time series are non-Gaussian due to a compound effect.

Just like the multiple steps in the region of both tails in the single stock time series, summing up over different stocks may also cause a kind of compound effect on the process connecting multiple elementary steps of price motions to make various indices. In order to confirm this scenario by the real data, we have analyzed various stock indices available in the JPX cloud service (http://www.jpx.co.jp/). Among them, we select an example of TSE-big-company index. The total number of data points are 3,834,275 from April 2005 to December 2018.

<Method 1> Scaling the empirical probability distributions of various time resolutions by properly choosing α

Parallel to the case of the single stock data for 8306 in the previous section, the first method to extract α from the data is to use the scaling property of Eq.(7). This is illustrated by two graphs in Fig.3. In the left figure, six graphs of empirical probability distributions of different levels of coarse graining, named as 15001.dat, 15002.dat, 15003.dat, 15004.dat, and 15005.dat, corresponding to the time resolution Δt = 15, 30, 60, 120, 240, 480 seconds are simultaneously plotted. Those numbers are chosen to have Δt = 2 in Eq. (7). If the price time series obey a scale-invariant distribution such as Lévy stable distribution, the value of α can be obtained by properly choosing the scale factor $c = (\Delta t)^{1/\alpha} = 2^{1/\alpha}$ so that all the six lines to overlap on a single distribution. This is done by choosing c = *1.503* and the corresponding index is around α = *1.7* by means of Eq. (9) taking Δt = 2, as shown in the right figure of Fig.3.

<Method 2>: Extract α from the slope of the log-log plot of P(0) and Δt

The second method to obtain the value of index α is to use Eq. (8). Corresponding to the time resolution Δt = 15, 30, 60, 120, 240, 480 second, the peak P(0) of the probability distribution decreases according to Table 1, from which the index α as the inverse of the slope of the log-log plot shown in Fig.2 is α=1.982*,* consistent to α=2.0 obtained from the graphical approach above. The stability of the index α is supported by the straightness of the line in Fig.2.

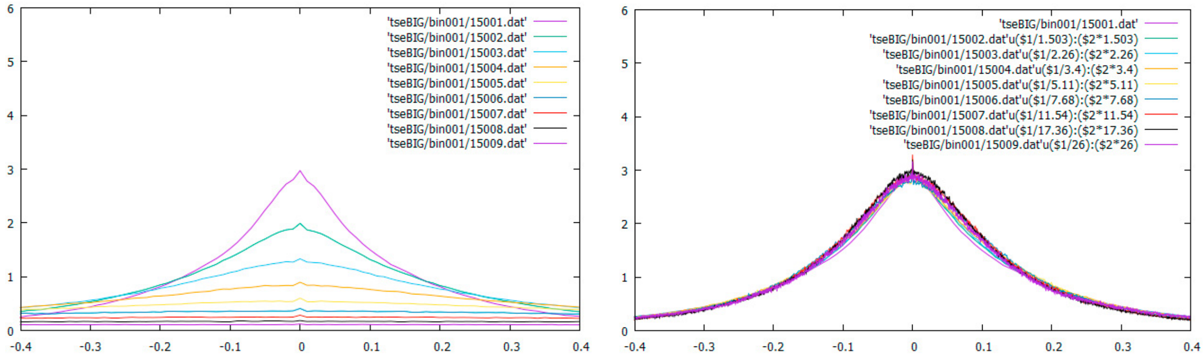

Fig.3 (left figure) Probability distribution of tse-BIG-company-index (April, 2005 - December, 2018) as a function of price increment z computed at nine levels of time resolution Δt =15s, 30s, …, 64min. corresponding to the lines named 15001.dat, 15002.dat, .., 15009 respectively are simultaneously plotted. (right figure) The probability distributions in the left figure are scaled to overlap on a single curve by properly choosing the scaling factor to be c =$(\Delta t)^{1/\alpha} = 1.503$ implying $\alpha = 1.7$.

Table 2 Peak height P(0) of the probability distribution vs. the time resolution Δt.

Δt	15sec.	30sec.	1min.	2min.	4min.	8min.	16min.	32min.	64min.
P(0)	2.973	1.994	1.334	0.8983	0.6030	0.4098	0.2854	0.1843	0.1218

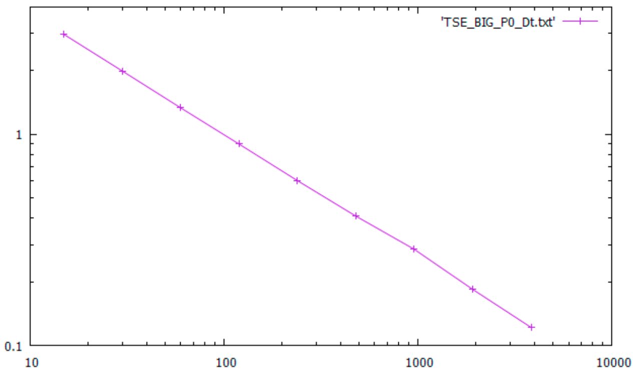

Fig.4 Plot of P(0) vs log(Δt) in Table 2.

5. Conclusion

In this paper, we focused to clarify whether the price fluctuation under sub-second transaction exhibit a sign of Gaussian, or non-Gaussian, by directly analyzing the full transaction data which became available recently. Our result shows that the price time series of a single stock (8306) for the duration of 50 months in the period of August, 2014-Septermber, 2018 behaves Gaussian, corresponding to the stable distribution of index α = 2.0, in the most part of the actual transactions. We have reached this conclusion based on the fact that the scaling analysis of distribution functions of various time resolution can be overlapped on a single function for the index α=2.0. However, the both tails of the distribution deviate from the pure Gauss function, although the probability that the tails are not significantly large. As a possible answer to explain this difficulty, we postulated that the central region and the tail region are made up from different mechanisms. Especially in the arrowhead market where the transaction speed is very fast, the central region of the distribution consists of fundamental motions of prices, thus corresponds to Gaussian random walk with index α = 2.0. On the other hand, the tail region of the distribution consists of compound motions made by multiple consecutive motions, where the mutual dependence of elementally moves cause the deviation from the pure convolution mechanism in Eq.(11)-(13). We also postulate that various indices made up of averaging many stocks deviate from Gaussian, due to similar mechanism as the case of the tail region of distribution for single stock prices. In order to confirm this postulate, we have analyzed various stock indices and show as an example the case of tse-BIG-stock-index that they indeed fit the curve of $\alpha = 1.7 < 2.0$.

References

[1] Bachelier, J. B. (1900) "Theorie de la Speculation" *Annales scientifiques de l'Ecole Normale Superieure, Ser*. 3.17: 21-86.

[2] Black, F. and Scholes, M. (1973) "The Pricing of Options and Corporate Liabilities" *Journal of Political Economy* **81**: 637-654.

[3] Merton, R.C. (1969) "Lifetime Portfolio Selection under Uncertainty: the Continuous-Time Case" *The Review of Economics and Statistics*, **51**: 247–257.

[4] Merton, R. C. (1973) "Theory of Rational Option Pricing" *The Bell Journal of Economics and Management Science*, **4**: 141–183.

[5] Merton, R. C. (1976) "Option Pricing When Underlying Stock Returns Are Discontinuous" *Journal of Financial Economics* **3**: 125-144.

[6] Stanley, H. E. (1971) "Introduction to Phase Transitions and Critical Phenomena" Oxford University Press.

[7] Domb, C., Green, M.S., and Lebowitz, J.L. (eds.) (1972–2001) "Phase Transitions and Critical Phenomena" **1**-**20**, Academic Press.

[8] Mandelbrot, B. B. (1963) "The variation of Certain Speculative Prices" *J. Buisiness* **36**: pp.394-419.

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

[10] Mantegna, R. N., and Stanley, H.E. (2007) "Introduction to Econophysics:Correlations and Complexity in Finance" paperbuck.

[11] Gopikrishnan, P., Meyer, M., Amaral, L.A.N., and Stanley, H.E. (1998) "Inverse Cubic Law for the Distribution of Stock Price Variations" *Eur. Phys. J. (B)* **3**: 139-140.

[12] Gopikrishnan, P., Plerou, V.,, Amaral, L.A.N., Meyer, M., and Stanley, H.E. (1999) "Scaling of the Distribution of Fluctuations of Financial Market Indices" *Physical Review E* **60**: 5305-5316.

[13] Preis, T., Sneider, J.J., and H.E. Stanley (2011) "Switching Process in Financial Markets" *PNAS* **108**: 7674-7678.

[14] Tanaka-Yamawaki, M. (2017) "Statistical Disribution of the Arrowhead Price Fluctuation" *Procedia Computer Science* **112**: 1439–1447.

[15] Tanaka-Yamawaki, M., Yamanaka, M., and Ikura, Y.S. (2018) "Statistical Disribution of the Sub-Second Price Fluctuation in the Latest Arrowhead Market" *Procedia Computer Science* **126**: 1029–1036.