Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 1439–1447

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Statistical Distribution of the Arrowhead Price Fluctuation

Mieko Tanaka-Yamawaki*

Department of Mathematical Sciences Based on Modeling and Analysis,
School of Interdisciplinary Mathematics, Meiji University, 1-21-4, Nakano, Tokyo, 164-8525, Japan

Abstract

The arrowhead system is a new trading system that allows transactions in a millisecond, launched in January 2010 at Tokyo Stock Exchange, Inc. (TSE) and was upgraded in September, 2015 to half a millisecond. In order to investigate the nature of the price fluctuation in the arrowhead market, we have analyzed price time series of 100 stocks sampled per 5 seconds for the duration of 9 months in 2013 consists of 640, 800 data points taken from TSE. The result shows the central part of the statistical distribution of the average returns of 100 selected stocks can be identified as the Lévy's stable distribution of index α =1.4, the same as the previous result reported in 1995 for the data of S&P500 index (1983-1989), and also for the Nikkei Index price in mid 1990's. However, due to the lack of data points, our result does not mean to constrain the shape of the tails of the distribution thus there is no difficulty concerning to the divergent variances of Lévy's stable distribution.

Peer-review under responsibility of KES International

Keywords: Price Fluctuation, , TSE Arrowhead Market, Lévy's Stable Distribution, Statistical Distribution Function

1. Introduction

How the price moves is one of the most fascinating questions to everyone. Although the price changes are intrinsically random [1], a great amount of effort is being devoted to predict the future performance of financial markets to avoid various kind of risks [2, 4-7]. Financial Technology (FT) has been developed as a fruit of such effort. The most remarkable outcomes of FT would be the derivation of the Black-Sholes formula (BSM formula) [3] for the European style option pricings in which the price (C) of the call option can be written as

* Corresponding author.
E-mail address: mieko.tanakay@gmail.com

Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.106

$$C(S_t, t) = N(d_1)S_t - N(d_2)Ke^{-r(T-t)} \tag{1}$$

and the price(P) of the put option

$$\begin{aligned} P(S_t, t) &= Ke^{-r(T-t)} - S_t + C(S_t, t) \\ &= N(-d_2)Ke^{-r(T-t)} - N(-d_1)S_t \end{aligned} \tag{2}$$

where

$$N(x) = \int_{-\infty}^{x} e^{\frac{-z^2}{2}} dz \tag{3}$$

and the arguments are

$$d_1 = \frac{log\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{4}$$

and

$$d_2 = d_1 - \sigma\sqrt{T-t} \tag{5}$$

respectively.

In short, the option prices can be calculated with high accuracy under the assumption that the probability distribution of the asset (e.g., stock) returns is Gaussian, and the volatility of returns are known.

However, it is well-known that the BSM formula 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. Usually, two empirical ways are 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 above formula from Eqs.(1-4) 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 of the time interval before the target date of the [8-10]. This is known as the 'smile curve' because the σ-K plot shapes concave and resembles the 'smile' mark.

Another problem of the above formula is on the assumption of the Gaussian nature of price fluctuation. If the distribution is not Gaussian, the use of function $N(x)$ in Eq.(3) for the option price determination in Eqs.(1)and (2) are no longer justified and the entire framework of option pricing theory might be altered. This fact prompts us examine the real-world data of price fluctuation. Plenty of works in the field of Econophysics have been devoted for this question. Based on the exhaustive data analysis of European and American stock prices, the distribution function is not Gaussian but has

(1) Fat tail
(2) Narrow neck

More specifically, the distribution is identified as the Lévy's stable distribution of index α =1.4 for the American stock index S&P500 with the time resolution per one minute [11].

However, this result has been heavily criticized due to the difficulty of infinitely large variance in the Lévy stable distribution and the authors later denied their result.

In this paper, we attempt to add some new knowledge toward this question by analyzing the stock price times series in 2013 with the time resolution of 5 seconds, much higher than the data used in Ref. [11].

Nomenclature

S_t	spot price of the underlying asset
T-t	time to maturity (time interval from the current time t to the target T)
K	target price at T

r	interest rate
σ	volatility of returns of the underlying asset

2. Arrowhead Market

Supported by the rapid progress in computer hardware and software, Tokyo Security Exchange (TSE) introduced a new system to process starting from January 4, 2010, which aims drastic improvement in the following three points [15]:

(1) reliability
(2) usability
(3) transaction speed to make a millisecond security exchanges possible

This new system was further improved five years later in September 2015 to make the transaction to half a millisecond resolution. Unfortunately, we are not able to access to the latest dataset. In this paper, we report our result of analyzing the arrowhead market data from April to December 2013, in order to help understanding the statistical nature of stock price fluctuation.

3. Scale Invariance of the Lévy Stable Distribution

As is well known, the price increments, often represented by the log-returns [13],

$$Z(t) = logX(t + \Delta t) - log\, X(t) \tag{6}$$

of the asset price *X(t)* at time t and a later time at $t+\Delta t$ are 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 that the probability distribution of asset returns follow Lévy stable distribution, defined as

$$f_{\alpha,\beta}(Z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikZ-\beta|k|^{\alpha}}\, dk \tag{7}$$

by assuming that the return in Eq.(6) as the stochastic variable in Eq.(7). Which is the Fourier transform of the kernel *F(k)* given by

$$F_{\alpha,\beta}(k) = e^{-\beta|k|^{\alpha}} \tag{8}$$

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

$$\beta = \gamma\Delta t \tag{9}$$

Eq. (9) 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 (Δt=1) follows the same distribution as the same asset returns at 10 seconds (Δt=2).

$$f_{\Delta t=2}(z) = \int_0^z f_{\Delta t=1}(x) f_{\Delta t=1}(z-x)dx \tag{10}$$

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

$$F_{\Delta t=2}(k) = (F_{\Delta t=1}(k))^2 \tag{11}$$

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

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

A series of n steps yields β 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 E.q (7) can be integrated for two special cases, α=1 and α=2, first of which is the Lorentz distribution,

$$P_{\alpha=1,\beta}(Z) = \frac{\beta}{\pi}\frac{1}{\beta^2 + Z^2} \tag{13}$$

and the second is the normal (Gauss) distribution.

$$P_{\alpha=2,\beta}(Z) = \frac{1}{2\sqrt{\pi\beta}}exp(-\frac{Z^2}{4\beta}) \tag{14}$$

For general values of α, the distribution is computed by numerically integrating Eq. (7).

The scale invariant property of Lévy stable distribution is derived from Eq. (7),

$$P_{\alpha,\beta}\left(Z/(\Delta t)^{1/\alpha}\right) = (\Delta t)^{1/\alpha} P_{\alpha,\beta\Delta t}(Z) \tag{15}$$

Setting Z=0 in Eq. (15), Lévy index α is estimated by comparing the height of the distribution $P_{\Delta t}(0)$ for various values of Δt.

$$log\ (P_{\alpha,\beta\Delta t}(0)) = -\frac{1}{\alpha}log\ (\Delta t) + log(P_{\alpha,\beta}(0)) \tag{16}$$

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 min to 100min.[11]

4. Non Gaussian Nature of Price Return Fluctuation

The science of price fluctuation was started by Luis Bachelier[1], who described the price change as a process of random walk, in 1900 as a doctoral thesis submitted to Ecole Normale Superiore. This work set the direction of financial technology toward the stochastic treatment of price fluctuation.

The first evidence against Gaussian distribution was pointed out by Mandelbrot [2] who examined the price changes of cotton prices in 1950's. Later, Mantegna and Stanley [11], Bouchaud and Potters [12] pointed out that the price returns of S&P500, the stock index of 500 common stocks in American market can well be fitted by Lévy stable distribution of index $\alpha = 1.4$., followed by many related works.

5. Scaling Analysis of the Arrowhead Transaction Prices Per Five Second: Case of Data-A

The aim of this work is to investigate whether or not the above scenario still fits the recent Japanese market, the ultra-fast trading system called "the arrowhead market". This system was introduced in Jan. 2010 to make the trading interval as short as 1 millisecond, then upgraded to 0.5 millisecond after September 2015. The best available data that we have obtained are the prices of 100 selected stocks per 5 second intervals offered by TMIV (TOSHO Market Impact View), from April 1, 2013 to December 25, 2013. We have collected by downloading the daily collection of such prices per 5 seconds on the web page of the Tokyo Market Impact View open for a limited term from April 1, 2013 to December 25, 2013. The size of the data is 640800 points for each stock price. We call this set as Data-A.

The original price time series of stock returns averaged over 100 stocks are processed to draw a histogram of return Z defined in Eq. (6) for the horizontal axis and the relative frequency of appearance as the probability distribution in log scale for the vertical axis. The increment of Z is set by setting the width of each partition to be (max-min)/1200. The size of the partition is selected to be 1200 so that every partition has at least one data point. This histogram corresponds to 5 second sampled data is labelled by Δt=1. Next, a coarse graining by skipping two points out of three consecutive points is applied to have a histogram corresponds to 15 second sampled data labelled by Δt=3. Likewise, histograms corresponding to 30 second sampled data labelled as Δt=6, and 60 second sampled data labelled by Δt=12 are constructed. We plot all the histograms from Δt=1 to Δt=12 in Fig.1.

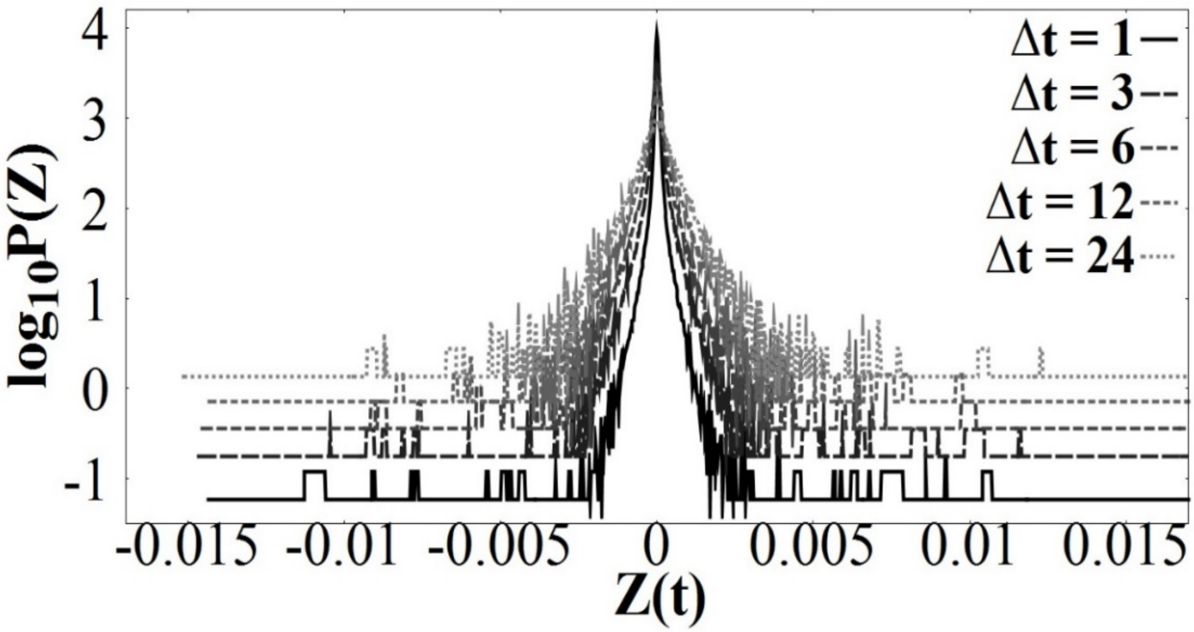

Fig. 1. The histograms of frequency of appearance of various Z for Δt=1(5 seconds),3,12,60,120 (2 minutes) are plotted in half-logarithmic scale.

Note that the number of data points becomes 53400 for Δt=12 and the accuracy of the histogram falls considerably. Table 1 summarizes the relation between the data sizes and the corresponding time intervals.

Table 1. Numbers of data points for each time scale

Δt	Actual time interval[s]	Number of data points
1	5	640800
3	15	213600
6	30	106800
12	60	53400

If the distribution of Z is indeed described by Lévy's stable distribution, all the histograms of different time resolutions Δt=1 to Δt=12 in Fig.1 should overlap exactly on the same curve of Eq. (7), by applying the scaling transformation of both axis as in Eq. (15) as long as both parameters α and β are correctly chosen.

This is achieved by rescaling both the return Z by the scaled Zs [13]

$$Z_s = Z/(\Delta t)^{1/\alpha} \tag{17}$$

in the horizontal axis, and the probability P(Z) by the scaled P(Zs) as follows.

$$P_{\alpha,\beta}(Z_s) = (\Delta t)^{1/\alpha} P_{\alpha,\beta\Delta t}(Z) \tag{18}$$

We search for the best-fit value of α in the range of $1 \leq \alpha \leq 2$ and found the best value to be α=1.4. In Fig. 2, the result for α=1.8 in the upper figure and for α=1.4 in the lower figure are shown.

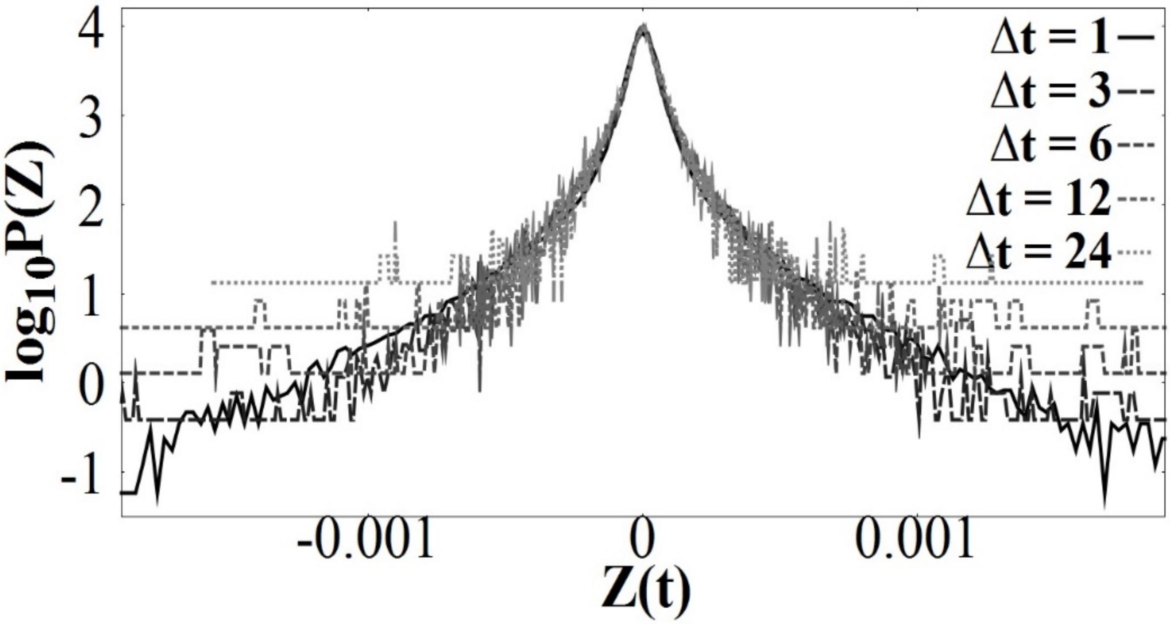

Fig. 2. Scaled histograms for Zs vs. log P(Zs) for *Δt*=1,3,12,60,120.for the case of $\alpha = 1.4$.

An alternative, independent way of quantification of parameter α is achieved by means of least square fittings of Eq. (16), from which the inverse of the parameter α is obtained as the slope of a linear equation between *P(0)* and Δt in logarithmic scale. The result by fitting the 10 points from Δ*t*=1 to $\Delta t = 24$ derives α=1.38. However, α=1.41 is derived if the point of $\Delta t = 24$ is excluded for the reason of low statistics. This result is shown in Fig.2. [14]

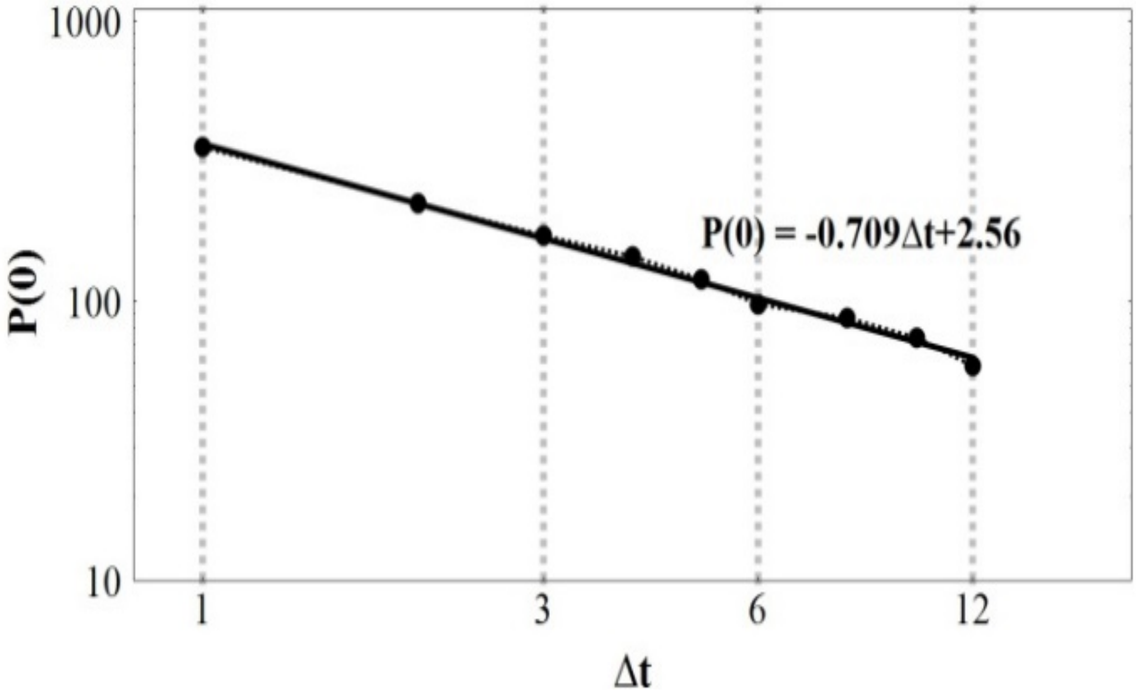

Fig.3 The least square fit of Eq. (16) derives α =1.41.

6. Statistical Distribution of the Recent Arrowhead Prices per 1 minutes to minimize the K-L Divergence

In spite of the fundamental change in the transaction system of TSE, the arrowhead data sampled by 5 seconds in 9 months of 2013 exhibited similar behaviour for at least in the central part of the distribution, being consistent to the former result Lévy's stable distribution on the American data in 1984-1989. While Arrowhead Data-A that we used in this analysis offered us to analyse the range of Δt from 5 seconds to 2 minutes, the result is not essentially different from the result of Ref.[3] that used the range of Δt = 1 min. to 1000 min.

For the sake of comparison, another set of price change per 1 min. are downloaded containing 29,386 data points per stock for 440 stocks of TSE from Google Finance site from June 16, 2015 to November 4, 2015. We call this data as 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 analyse 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

$$D(p||q) = \sum_{x \in X} p(x) log \frac{p(x)}{q(x)} \tag{19}$$

We compute D(p||q) in Eq. (19) by setting p(x) and q(x), as the probability distribution of the 1- minute price increments (return) and the corresponding Lévy's stable distribution for various values of α and β. The best fitted result for those parameters is consistent to the case of Data-A, as shown in Table 2.

Table 2 The K-L divergence between Lévy's stable distribution and the distribution of 1-minute returns in Data-B (average returns 440 stocks in TSE 2015).

Stock return data (1 minute interval)	α	β	K-L divergence
Average returns of 440 stocks	1.40	5.4×10^{-6}	0.039
Kansai Electric	1.55	10.0×10^{-6}	0.286
Nissan Motors	1.65	3.9×10^{-6}	0.423
Toshiba	1.55	8.8×10^{-6}	0.156

7. Discussion and Future Perspectives

In this work, we aimed to examine the nature of the arrowhead financial market in Tokyo Security Exchange Market, where, a new ultrafast trading system of sub-millisecond transaction time has been operated from 2010. Especially, we focused on the probability distribution of the asset returns and examined the central part of the distribution utilizing its scale-invariant property. The result turned out to be the same as in Ref.[11]. Based on this, we have to conclude that, even at the 5 second resolution level, average asset returns of small sizes corresponding to the central part of the probability distribution behaves in the same way as the 1 minute resolution level. Two questions remain. The first question is the reason why this occurs for the average returns and not for individual asset returns. There is an evidence in the computer simulation to show that the combinations of different indices of Lévy's stable distribution in the range of $1<\alpha<2$ indeed derive α=1.4 similar to ours [16]. The second is on the fat-tail behavior of the returns. Intuitively, convolutions of too many steps will eventually cause a breaking down of the Lévy's random walk model. Future works using a wide range of price data should be able to answer those questions.

Acknowledgments

The author is grateful to Mr. Takeru Fuii and Mr. Katsutoshi Yoshii for helping data analysis, and to the members of Knowledge Engineering Laboratory in the Graduate School of Engineering, Tottori University, Drs. Ippei Shimada and Masanori Yamanaka of Nihon University, and the members of econo-physics group in the research collaboration of the Institute of Statistical Mathematics for encouraging discussions.

References

1. Bachelier, L.J.B. (1900) *Theorie de la Speculation;* th`ese. Annales Scientifiques de l'Ecole Normale Sup´erieure, III-17:21–86, 1900. Reprinted by Jacques Gabay, Paris(1995), English Translation: Cootner (ed.), (1964) Random Character of Stock Market Prices, Massachusetts Institute of Technology, pages 17-78 or S. Haberman and T.A. Sibett (1995) (eds.), History of Actuarial Science, VII, 15-78. London.
2. Mandelbrot, B. B. (1963) *The Variation of Certain Speculative Prices,* The Journal of Business, Vol. 36, No. 4 (Oct., 1963), pp. 394-419.
3. Black, F., Scholes, M. (1973) *The Pricing of Options and Corporate Liabilities,* Journal of Political Economy, Vol. 81, pp.637–654.
4. S.R. Pliska (1997) *Introduction to mathematical finance*. Blackwell, Oxford.
5. Wilmott, P. (1998) *Derivatives, The Theory and Practice of Financial Engineering*. Wiley.
6. Geman, H., Madan, D.B., Pliska, T.Vorst, editors (2002) *Bachelier Congress 2000,* Mathematical Finance, Springer Finance. Springer-Verlag.
7 . Geman,H. D.Madan, M.Yor (2001) *Stochastic Volatility, Jumps and Hidden Time Changes,* Finance and Stochastics, Vol. 6, Issue 1, pp.63-90
8. Derman, E.,Kani, I. (1994) *Riding on a Smile,* RISK, Vol. 7 (issue 2 Feb.1994), pp. 139-145
9. Derman, E. and Kani, I. (1996) *The Volatility Smile and Its Implied Volatility*. Goldman Sach
10. Dupire, B(1997) *Pricing and Hedging with Smile*. in Mathematics of Derivative Securities, Cambridge
11. Mantegna, R.N., H.E. Stanley (1995) *Scaling behaviour in the dynamics of an economic index*, Nature Vol.376, pp.46-49.
12. Bouchaud, J. P., Potters, M. .(1999) *Theorie des Risques Financieres*, Alea-Saclay
13. Mantegna, R.N., Stanley, H.E.(2000) *An Introduction to Econophysics:Correlations and Complexity in Finance*,
14. Tanaka-Yamawaki, M. (2016) *Statistical Property of Prices in the Arrowhead Market-Lévy's Stable Distribution Or Else?-*, talk at APEC-SSS2016 (Univ. Tokyo, August 24-26, 2016)
15. http://www.jpx.co.jp/corporate/news-releases/0060/20150924-01.html

16. Umeno, K. (private communication in March 2017)