Time-Space Scaling of Financial Time Series

Yoshiaki Kumagai[1]

Institute for Economic and Industrial Studies, Keio University, 2-15-45, Mita, Minato-ku, Tokyo 108-8345, Japan

Summary. We propose a new method to describe scaling behavior of time series. We introduce an extension of extreme values: maximum and minimum. By using these extreme values determined by time-space scaling with a spread (or width), functions of this spread are defined. One is the number of these extreme values, and the other is the total variation among these extreme values. These functions are independent of time scale. In high frequency data, observations can occur at varying time intervals. In particular, on fractal analysis, interpolation influences the results. Using these functions, we can analyze non-equidistant data without interpolation. Moreover the problem of choosing the appropriate time scale to use for analyzing market data is avoided. In other words, 'time' is defined by fluctuations here. Lastly, these functions are related to a viewpoint of investor whose transaction costs coincide with the spread.

Key Words: Non-equidistant data, Transaction costs, Tick-by-tick, Fractal, Scaling

1 Introduction

High frequency data is not equidistant in physical time by nature. For instance, transactions in tick-by-tick data do not occur in equal intervals throughout the day. We can convert them into the data of equal intervals by interpolation. But the choice of the interpolation method influences the nature of the time series. In particular, on fractal analysis, the effect of interpolation is significant to the results. Further, between transactions, prices are not existing, not unobserved. Thus, interpolation is not proper for tick-by-tick data. In this paper, we present a new method to analyze non-equidistant data of financial time series. In fractal analysis of one-dimensional time series, scaling is applied on time, along the horizontal axis. We introduce a new type of scaling of fluctuations, scaling together with the vertical axis.

Moreover, even if data is equidistant in physical time, there remains the problem of choosing the appropriate time scale to use for analyzing market data. There are at most three candidates for the proper time scale: the physical time, the trading time, and the number of transactions. For instance, Clark [1] assumed that events important to the pricing occur at a random, not uniform rate through time. In short, 'time' which evolve economic activities is not passing uniformly. Since it is difficult to select from these three candidates, this paper proposes a method to avoid this difficulty. Using local maxima and minima determined by time-space scaling with spread C, we can construct some functions independent of time scale. Using any of these time scales, the function is the same. In this paper, 'time' can be defined in terms of the number of fluctuations.

Lastly, this new method has a close relation to transaction costs and viewpoints of investors. We observe the price series from viewpoints of investors. The investors pay some costs at each transaction. The transaction costs include bid-ask spread and sale commission, for example. Therefore, if fluctuations of price are smaller than these costs, they can not make a profit. For this reason, the investors observe the price series within a tolerance of these costs. They ignore fluctuations smaller than transaction costs.

This paper is organized as follows. Section 2 defines extreme values determined by time-space scaling with a spread C. Further the number of extreme values and the total variation are formulated as functions of C. Section 3 applies this method to financial time series. The properties of the functions are described using tick-by-tick data.

2 Scaling of fluctuations

First, we present an extension of extreme values of function. In Kumagai [2], *local maximum determined by time-space scaling with* C was defined as follows. Suppose g is a real function defined on R^1. We say that g has a *local maximum determined by time-space scaling with* C at a point t, if there exist $d_1, d_2 > 0$ such that $g(t-d_1) < g(t) - C, g(t+d_2) < g(t) - C$, and in $[t-d_1, t+d_2]$, g attains its maximum at t. Figure 1 shows an example of time series and its extreme values. As shown in Figure 1, by observing the graph through a slit (denoted by two parallel dashed lines), we can detect these maxima easily. The spread (or width) of this slit is C, and the length is not limited. The local maximum determined by time-space scaling with 0 means the ordinary local maximum. The width C is the accuracy of measurement of extreme values. That is, we determine the local maximum within an accuracy of C.

Local minimum determined by time-space scaling with C is defined likewise. Suppose g is a real function defined on R^1. We say that g has a *local minimum determined by time-space scaling with* C at a point t, if there exist $d_1, d_2 > 0$ such that $g(t-d_1) > g(t) + C, g(t+d_2) > g(t) + C$, and in $[t-d_1, t+d_2]$, g attains its minimum at t. Figure 1 shows an example.

Now we define two functions using these extreme values. Throughout this section, suppose that the time series $g(t)$ is defined in a given interval $[0, S]$. Let C denote a vertical spread (or width). In this interval$[0, S]$, we can detect extreme values determined by time-space scaling with C. Suppose $g(t)$ attains its extreme values at $t_1^C < t_2^C < t_3^C < \cdots$, ordered increasingly. With $g(t)$ fixed, the number of these maxima and minima in $[0, S]$ is a function of C. Let

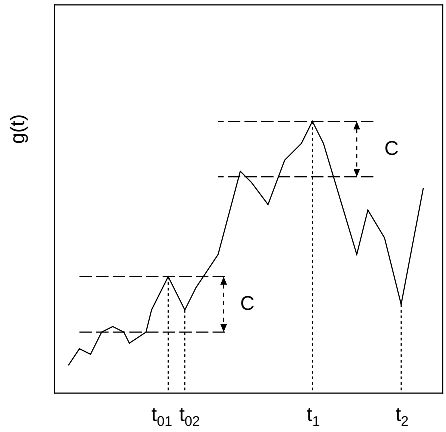

Fig. 1. Example of extreme values determined by time-space scaling with spread C. The zigzag line denote a function $g(t)$. The function g attains its local maximum determined by time-space scaling with C at t_1, but g does not attain its maximum at t_{01}. Using a pair of parallels whose distance is C (denoted by *dashed lines*), we can detect these extreme values. On the other hand, g attain its local minimum with C at t_2, not at t_{02}.

$m(C)$ denote this number of the extreme values. Figure 2 shows an example of the function m calculated from empirical data [1] . The data are chosen from three categories: currency, commodity futures and stock [2] .

The property of this function is the following. By definition, $m(C)$ is monotone decreasing in the wider sense. Further m is a step function continuous from the right, not from the left. If C_L is larger than the range: $C_L \geq \max g(t) - \min g(t)$, $m(C_L) = 0$. Besides, $m(0)$ is the number of the ordinary extreme values in the interval. Let C_1 be one of the points at which m is discontinuous. Then there are some pairs of the extreme values: maximum and minimum whose difference is C_1. Let k denote the number of these pairs. In other words, the number of fluctuations whose size is equal to C_1 is k. Assuming that a positive value δ is sufficiently small, we obtain

$$\begin{aligned} m(C_1) - m(C_1 - \delta) &= 2k, \\ m(C_1 + \delta) - m(C_1) &= 0. \end{aligned} \tag{1}$$

[1] All tick-by-tick data used in this study are supplied from Cyber Trading/ Risk Laboratory of Keio University.

[2] Using a method similar to this paper, Kumagai [3] reported the analysis on daily closing price of commodity futures markets: gold futures in Tokyo Commodity Exchange and soybean futures in Tokyo Grain Exchange.

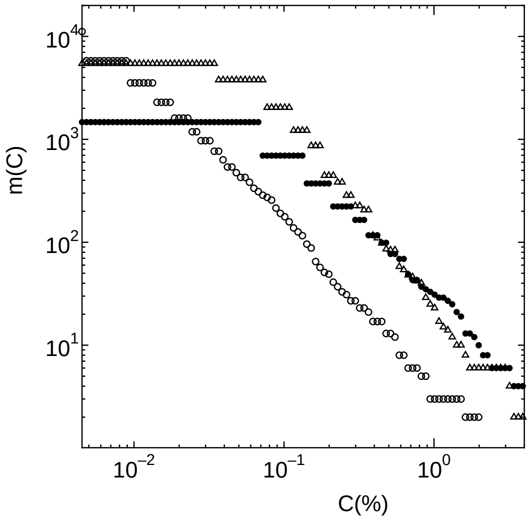

Fig. 2. Double logarithmic plot of the number of extreme values: $m(C)$ with C. The sample period is 5 trading days from May 8, 2000 to May 12, 2000, for three contracts: USD/JPY exchange rate (○), WTI (1st. month) in New York Mercantile Exchange (△), Hitachi in Tokyo Stock Exchange (●)

Next we introduce a kind of total variation of time series. Let $R(C)$ denote the total sum of absolute variation between neighboring extreme values:

$$R(C) \equiv \sum_{i=1}^{m(C)-1} |g(t_{i+1}) - g(t_i)|. \tag{2}$$

Figure 3 shows an example. The data used are the same as Fig. 2.

By definition, $R(C)$ is monotone decreasing in the wider sense. Furthermore R is a step function continuous from the right as m. If C_L is larger than the range: $C_L \geq \max g(t) - \min g(t)$, $R(C_L) = 0$. If $C = 0$, R is the total sum of absolute variations : $R(0) = \lim_{n\to\infty} \sum_{k=1}^{n} |g(kS/n) - g((k-1)S/n)|$. Let C_1 be one of the points at which m is discontinuous as before. Further let k denote the number of the pairs of the extreme values: maximum and minimum whose difference is C_1. This means that k is the number of fluctuations whose size is C_1. Assuming that a positive value δ is sufficiently small, we obtain

$$R(C_1) - R(C_1 - \delta) = C_1(m(C_1) - m(C_1 - \delta)) = 2kC_1,$$
$$R(C_1 + \delta) - R(C_1) = C_1(m(C_1 - \delta) - m(C_1)) = 0. \tag{3}$$

The first equality of each of Equations (3) holds everywhere. Thus, in general, the relation between the total variation and the number of extreme values

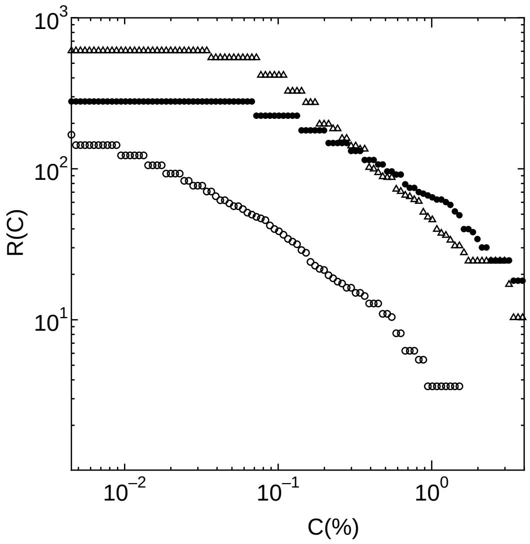

Fig. 3. Double logarithmic plot of the total variations: $R(C)$ with C. The sample period is 5 trading days from May 8, 2000 to May 12, 2000, for three contracts: USD/JPY exchange rate ($\circ$), WTI (1st. month) in New York Mercantile Exchange ($\triangle$), Hitachi in Tokyo Stock Exchange ($\bullet$)

with C is

$$R(C) - R(C - \delta) = C(m(C) - m(C - \delta)). \tag{4}$$

Let $v : [0, S] \to [0, S]$ be a monotone increasing continuous function. To show the dependence of the function clearly, we designate m by $m(C, g(t))$ and R by $R(C, g(t))$ for a while. Suppose $g(t)$ attains its extreme values at $t_1^C < t_2^C < t_3^C < \cdots$, ordered increasingly. Then, $g(v(t))$ attain its extreme values at $v(t_1^C) < v(t_2^C) < v(t_3^C) < \cdots$. Thus, the number of the extreme values m does not change by this transformation of time scale: $m(C, g(t)) = m(C, g(v(t)))$. Further, the total variation R is also independent of time scale: $R(C, g(t)) = R(C, g(v(t)))$. In other words, $m(C)$ and $R(C)$ are invariants for time scale stretching [3] . This property is significant to investigate non-equidistant data like tick-by-tick. Furthermore, there is no absolute time scale in financial time series. Thus, in the next section, we apply this method to financial time series.

[3] The function $m(C)$ is also an invariant for space scale stretching. The transformation using a monotone continuous function can not change m.

3 Application to financial time series

Above defined functions are related to the ex-post optimal trading. Let $g(t)$ denote a price at time point t. Here we classify the costs for investor into two categories: one is in proportion to the number of transactions, and the other is in proportion to the period of taking position. Let C denote the former: the costs at every transaction per unit. These costs include bid-ask spread, tax, and sale commission. On the other hand, the costs in proportion to the period of taking position are not included. For example, opportunity costs of margin requirement are not considered here.

We assume that C is constant throughout the interval $[0, S]$. As previous section, local maxima and minima can be determined by time-space scaling with C. By using these extreme values, the functions $m(C)$ and $R(C)$ are obtained. Throughout this section, we assume that the return of safety assets is equal to zero. Hence, if the investors buy at local minima and sell at local maxima, their profit is maximized. Because of the uncertainty of the price, none of them actually can trade at these extreme values. Therefore, this strategy is optimal in ex-post meaning: maximizing the profit when the complete expectation is obtained. In short, the investors observe the price series with the required accuracy in consideration of the transaction costs. In this strategy, the number of transactions in the interval $[0, S]$ is $m(C)$. Besides, the total variation $R(C)$ corresponds to the sum of the return and the transactions costs per unit. Thus, subtracting the transaction costs from the total variation, the return per unit is defined as follows:

$$\Pi(C) \equiv R(C) - Cm(C). \tag{5}$$

Since this profit is obtained by the ex-post optimal trading, it can be said as maximum profit. Figure 4 shows an example. The data used are the same as Fig. 2 and 3.

By definition, Π is non-negative monotone decreasing. As m and R, Π is continuous from the right. Furthermore, even at the discontinuous points of m and R, by Equation (4), Π is also continuous from the left. Let C_1 denote one of these points,

$$\lim_{\delta \to 0} \Pi(C_1 - \delta) = \Pi(C_1). \tag{6}$$

Therefore $\Pi(C)$ is continuous everywhere.

Next, we describe the relation between the return Π and the number of extreme values, i.e. the number of transactions m [4] . The left-hand derivative of Π: $\Pi'_{-}(C)$ is

$$\Pi'_{-}(C) \equiv \lim_{\delta \to 0} \frac{\Pi(C) - \Pi(C - \delta)}{\delta} = -m(C - \delta). \tag{7}$$

[4] Kumagai [3] mentioned this briefly.

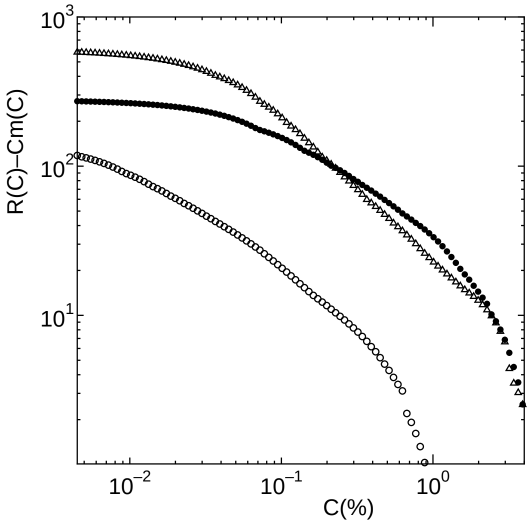

Fig. 4. Double logarithmic plot of the profit per unit: $\Pi(C) \equiv R(C) - Cm(C)$ with C. The sample period is 5 trading days from May 8, 2000 to May 12, 2000, for three contracts: USD/JPY exchange rate (○), WTI (1st. month) in New York Mercantile Exchange (△), Hitachi in Tokyo Stock Exchange (●)

On the other hand, the right-hand derivative $\Pi'_+(C)$ is

$$\Pi'_+(C) \equiv \lim_{\delta \to 0} \frac{\Pi(C+\delta) - \Pi(C)}{\delta} = -m(C). \tag{8}$$

Hence Π is not differentiable at the discontinuous points of m (and R) .

As previously described, m and R are independent of time scale. Hence, Π is also independent of time scale. These functions m, R, and Π are invariants of time scale stretching. This property is important from the following reasons. Firstly, transactions of high frequency data do not occur equidistantly in physical time. Moreover, in the financial market, there is no appropriate choice of time scale. Since there are merits and demerits in any candidates for time scale, it is difficult to select the best. The ordinary fractal analysis depends on time scale. Therefore, this independence to time scale is important property in analysis on financial time series.

Figure 5 shows a simple example of time series and its nest structure. Figure 6 shows the time series and nest structure of foreign exchange rates. Figure 7 magnifies a part of Fig. 6 . In these figures, the solid line shows the time series, and the nested rectangles of gray and white describe the nest structure of the fluctuations. With transaction costs C fixed, we divide the whole interval at maxima or minima determined by time-space scaling with C. An increasing interval from maximum to minimum is denoted by white. A decreasing interval from maximum to minimum is denoted by gray. Smaller

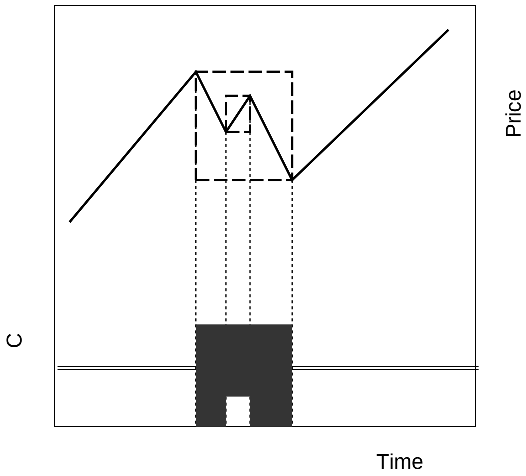

Fig. 5. An example of nest structure of fluctuations. The zigzag line (*solid lines*) shows a time series $g(t)$ for the right-side scale. The two nested rectangles (*dashed lines*) represent the pairs of the extreme values: maximum and minimum. Shifting down these rectangles (along *dotted lines*), we put their bases on the horizontal axis. The outer gray rectangle represents the decreasing period, and the inner white rectangle represents the increasing period

fluctuations are observed by higher resolution. One rectangle corresponds to a pair of the extreme values: maximum and minimum. In each rectangle, there are some smaller rectangles. Further, in these rectangles, there are some much smaller ones. These figures show the nest structure of fluctuations. The line of level C crosses these rectangles at the extreme values determined by time-space scaling with C. These intersections divide the line of level C into intervals painted by white or black. The intervals shown by white line denote increasing periods, and the intervals shown by solid lines denote decreasing periods. This shows how an investor with transaction costs C observes the price time series. In these figures, the function $m(C)$ corresponds to the number of intersections of the rectangles and the horizontal line of level C. Further, $R(C)$ corresponds to the sum of the length of vertical laterals of the rectangles crossed by this line of level C. In Fig. 5, $m(C)$ and $R(C)$ are invariants for time-space scale stretching in the dashed rectangles whose height is smaller than C. This means that they do not vary with fluctuation or interpolation in the rectangles. In other words, 'time' passes by when a fluctuation occurs, not when the trading time passes on.

4 Conclusion

In this paper, we proposed a new method to analyze non-equidistant data. Our method has three remarkable features as follows. Firstly, using this

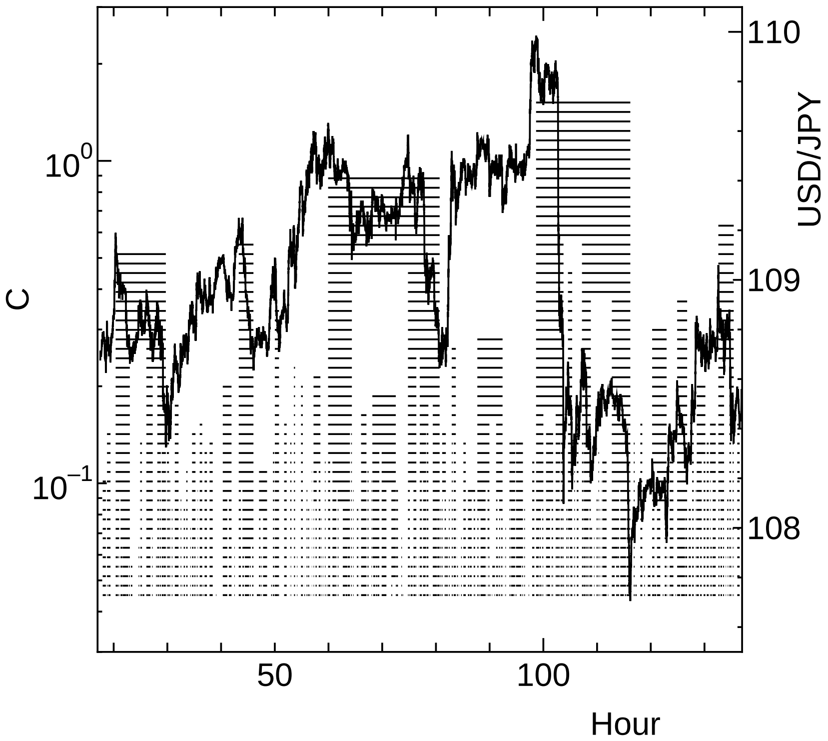

Fig. 6. Time series and nest structure. The prices are plotted in solid line along right-hand scale. Nests of the rectangles represent nests of fluctuations. A gray (striped) or white rectangle represents a pair of extreme values: maximum and minimum. The altitude of the rectangle, which is measured by left-hand logarithmic scale, represents the transaction costs with which these extreme values are determined. The data set is tick-by-tick currency exchange rates of USD/JPY from May 8, 2000 to May 12, 2000. The number of data is 36040 The origin of the time scale is 0:00 May 8, 2000 (GMT)

method, we can analyze non-equidistant data without interpolation. Secondly, the result is independent of time scale. Lastly, the variables and the functions are related to a viewpoint of investor with transaction costs. Financial market is not temporally homogeneous. This method captures a new type of scaling behavior for non-equidistant data. We plan to improve this method, and to study high frequency data of financial time series.

Acknowledgements

We would like to thank Hideki Takayasu for organizing the symposium. We are grateful to Souichiro Moridaira and Cyber Trading/ Risk Laboratory of

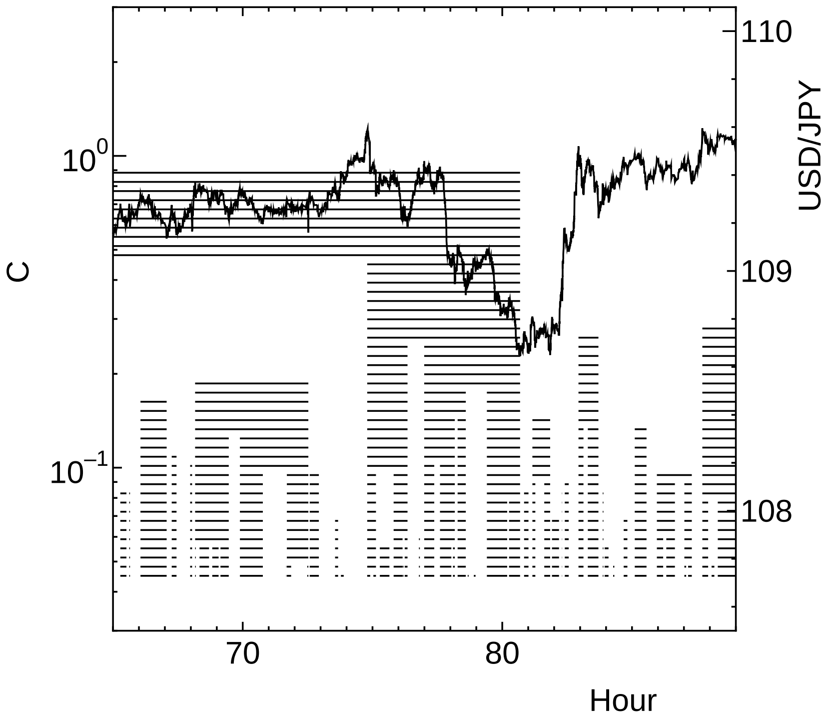

Fig. 7. Magnification of a part of Fig. 6. The prices are plotted in solid line along right-hand scale. Nests of the rectangles represent nests of fluctuations. A gray (striped) or white rectangle represents a pair of extreme values: maximum and minimum. The altitude of the rectangle, which is measured by left-hand logarithmic scale, represents the transaction costs with which these extreme values are determined. The data set is tick-by-tick currency exchange rates of USD/JPY in May 10, 2000. The origin of the time scale is 0:00 May 8, 2000 (GMT)

Keio University for supplying the data. We also acknowledge the continuing guidance and encouragement of Gyoichi Iwata and Yukitami Tsuji.

References

1. Clark P K (1973) Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica 41: 135-155
2. Kumagai Y (1999) Speculative Market Model with Transaction Costs and Endogenous Horizons . Mita Business Rev (in Japanese) Mita Syougaku Kenkyuu 42, 4: 71-92
3. Kumagai Y (2000) Analysis of Trading Volume and Liquidity by Speculative Market Model with Transaction Costs and Endogenous Horizons (in Japanese)