論　文

株価ティックデータの統計特性について
（その1）

譚　　康　融

1．まえがき

最近，株式・為替などの金融マーケットのミクロ構造に関する研究，特に短時間において，収益率の分布はどう変わるか，市場裁定はとうていあるかどうかに関する研究が注目されている。そのため，従来，金融工学でよく用いられていた株価・為替レットの日次や月データなどよりも，ティック単位で市場の詳しい取引情報（例えば，Ask, Bid, Transaction Price, Volume）を記録したティックデータを用いた研究が必要になる。幸い，IT（Information Technology）技術の発展により，従来の汎用大型コンピュータや，並列UNIXサーバーなどしかできない情報処理の仕事は，いま，パソコンでも，この膨大なティックデータをも処理できるようになった。しかしながら，計算スピードや，システムの安定性などを考えると，まだUNIXサーバーあるいは，伝統の大型汎用機のほうが優れていると言えよう。

また，近年，数多くの統計パッケージが開発・リリースされ，煩わしいプログラミングをしなくても，より複雑な統計処理ができることや，優れたグラフィックス機能を使用できることなど，使いやすくなっている一方，各パッケージの精度の違い，バグの影響で，場合によって，ずいぶん違う結果がそれぞれのパッケー

ジから得られることもあり得る。さらに，ある意味では，パッケージの内部の仕組みはブラックボックスになっているため，間違った処理がされても，ユーザーサイドはそのことを察知しずらいという面もあるが，できれば，ユーザー自らプログラムを組んだほうが確実であろう。それはコストとパーフォマンスとのトレードオフである。

最近，株式リターンの分布は，パーレット分布，あるいは，Levy 分布であるといった見解が数多く発表されているが[1][2]，しかしながら，実際の金融マーケットのデータを分析すると，例えば，2次モーメントは存在しており，いくつか実情と理論と合致していないところが実証されている。また，分布のテールについては，冪乗現象が観測されている。本論文は，これらの現象と議論を踏まえて，A社の株価ティックデータセットを用いて，その統計特性を調べ，ティックデータの場合はどのような確率現象および確率構造が現れるかについて研究する。本文は以下のように構成されている。2においては，データのクリーンや，それから解析用に使われるデータセットの作成について述べ，主にティックデータセットの仕組みについての説明，分析に使われる各データセットの作成を整理する。3においては，それぞれ作成したデータセットを用いて，ヒストグラム，収益率の分布，および標準偏差とサンプリングインターバルの関係を調べる。また，本稿でまとめた結果は，中間報告で研究全体の一部にすぎないため，その1（Part One）としてまとめている。むすびにおいては，今後の研究課題について述べる。

2　データセットについて

本研究に用いられたA社の株価ティックデータセットは，1990年1月から1991年1月までの取引が秒単位で記録されているものである。詳細内容は，取引のDATE＋TIME（12桁），ASK PRICE（6桁），BID PRICE（6桁），TRANS-

ACTION PRICE（6桁），およびTRANSACTION VOLUME（6桁）となっている。具体的なイメージは表1に示されている。

データセットには単に垂れ流しで刻々の取引情報を記録しているため，分析用にするまでは，クリーンの作業は不可欠ものである。本研究はクリーン作業をおこなった後，以下の分析用データセットを作成した。また，他の研究にも示されたように[3]，夕方や，深夜あるいは早朝の取引状況は，通常時間帯の取引状況とかなり異なっており，安定した価格データのプロセスを得るため，データセットを作成する際は，9:30AMから4:00PMまでの取引データのみを抽出して利用し，他の時間帯のデータを作成したデータセットから取り除いた。また，取引のない時刻における株価は，両端の株価に基づいて，補間値を用いている。なお，全部データを使った研究結果は，次の研究に公表する予定である。

表1　データセットのフォーマット

Date	Time	Ask	Bid	Trans Pri	Volume
900101	090310	105.3	105.2	105.2	011000
900101	090310	105.3	105.2	105.2	008000
			………		
900101	090311	105.3	105.2	105.3	020000
900101	090311	105.3	105.2	105.3	020000
			………		
910101	160000	111.2	111.2	111.2	015000

表2　データセットの種類

データセット名	時間間隔（sec）	データ数
DS1	60	4900
DS2	120	4900
DS3	300	4900
DS4	1200	5500
DS5	3000	5000
…	…	…

3　分析の結果

具体的な説明をおこなうまえに，まず，各期の収益率 r_i は式(1)，すなわち，

$$r_{\Delta t} = logp(t + \Delta t) - logp(t) \tag{1}$$

と定義するが，株価が Δt において，ごく少ししか動かない場合は，

$$r_{\Delta t} \approx \frac{p(t+\Delta t) - p(t)}{p(t)} \tag{2}$$

となる。収益率の時系列データの一部は図1に示されている。

また，各分布の比較をしやすくするため，収益率を以下のように正規化している。すなわち，

$$sr_{\Delta t} \approx \frac{r_{\Delta t} - \langle r_{\Delta t} \rangle}{v} \tag{3}$$

$$v^2 = \langle r^2_{\Delta t} \rangle - \langle r_{\Delta t} \rangle^2 \tag{4}$$

となる。なお，$r_{\Delta t}$ を計算する際，Transaction Price を用いた。

3.1　ヒストグラムについて

図2には，全体のティックデータを用いて，収益率 $r_{\Delta t}$ を計算した時系列データを，平均0，標準偏差1に正規化し，それをヒストグラムにしたものが示されている。図から分かるように，全体の分布はほぼ対称しているが，正規分布よりもはるかに尖がっており，プラスとマイナスのテールが観測されている。また，時間間隔 Δt が大きくなるにつれて，収益率の分布は次第に正規分布に近づくことが，図3から明らかである。

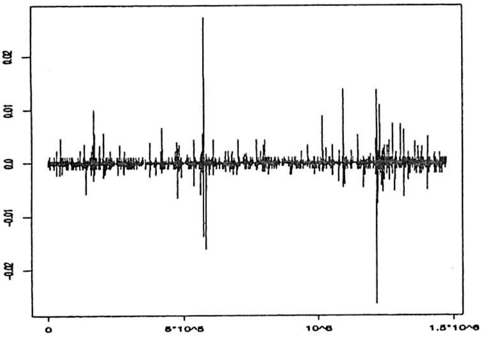


図１：収益率の時系列

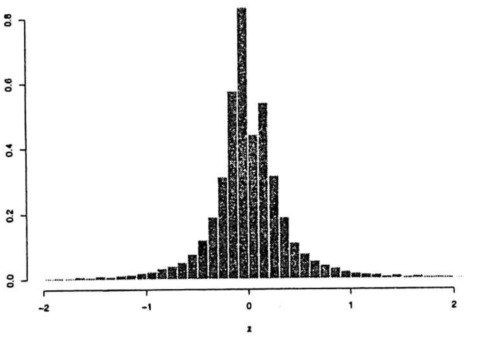


図２：ヒストグラム

3.2　収益率の分布について

次は，収益率がある値よりも大きいときの確率 $P(>\Delta x,\ x>0)$ の分布を考える。図４にはこの累積確率分布が示されている。図から分かるように指数 $x^{-\alpha}$ 級のテールが観測されている。各 $sr_{\Delta t}$ のリジュンにおいて，指数 α のフィットバリューは以下のようになっている。

（ケース1：5minutes の場合（Positive））

$$\alpha = \begin{cases} 2.169 & \text{if } 0.5 \leq sr_{\Delta t} \leq 3,\ R^2 = 0.971 \\ 3.602 & \text{if } 3.0 \leq sr_{\Delta t},\ R^2 = 0.974 \end{cases}$$

（ケース2：10minutes の場合（Positive））

$$\alpha = \begin{cases} 2.30 & \text{if } 0.5 \leq sr_{\Delta t} \leq 3,\ R^2 = 0.902 \\ 6.54 & \text{if } 3.0 \leq sr_{\Delta t},\ R^2 = 0.961 \end{cases}$$

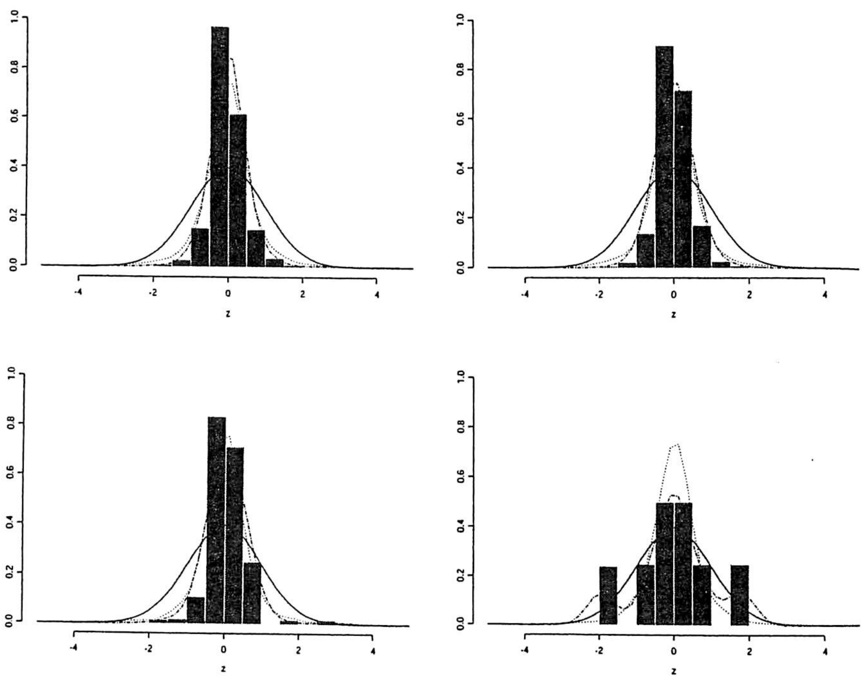
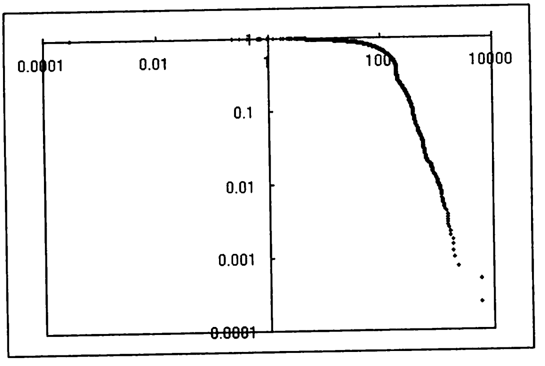


図3： 5, 20, 200, 200000 minutes のヒストグラム

実線：N (0, 1)
破線：推定した密度関数



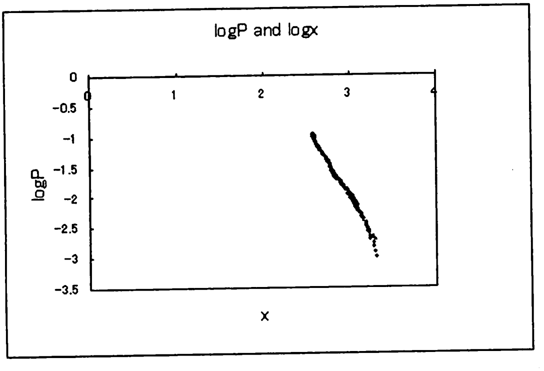


図4：$P>(\Delta x)$，（$\Delta x>0$）の累積確率分布

（ケース3：5minutes の場合（Negative））

$$\alpha = \begin{cases} 1.67 & \text{if } 0.5 \leq sr_{\Delta t} \leq 1.0,\ R^2 = 0.905 \\ 2.29 & \text{if } 1.0 \leq sr_{\Delta t} \leq 3.0,\ R^2 = 0.989 \end{cases}$$

（ケース4：10minutes の場合（Negative））

$$\alpha = \begin{cases} 2.32 & \text{if } 0.5 \leq sr_{\Delta t} \leq 1.0,\ R^2 = 0.970 \\ 2.64 & \text{if } 1.0 \leq sr_{\Delta t} \leq 3.0,\ R^2 = 0.971 \end{cases}$$

以上の計測から，Δt の変化につれて，テールの指数値も変化しているのが分かる。

3.3　Δt と $\sigma_{\Delta t}$ との関係

さらに，作成した各データセットを用いて，それぞれの標準偏差 $\sigma_{\Delta t}$ 計算して，Δt と $\sigma_{\Delta t}$ との関係を調べ，次の関係式を発見した。すなわち，

$$\sigma_{\Delta t} \propto (\Delta t)^k \tag{5}$$

実際の計算においては，三つのケースに分けておこなった。

（ケース１）

まず，時間間隔 1sec，60sec，120sec，および 300sec とそれぞれのデータセットの標準偏差を用いて，回帰分析をおこない，以下の式を得られた。

$$log(\sigma_{\Delta t}) = 0.511 log(\Delta t) - 3.9989 \tag{6}$$

（ケース2）

次に時間間隔 300sec，1200sec，3000sec，および 12000sec とそれぞれの標準偏差を用いて，回帰すると，

$$log(\sigma_{\Delta t}) = 0.5614 log(\Delta t) - 4.0014 \tag{7}$$

が得られた。

（ケース3）

上述した二つのグループデータをあわせて，回帰分析をおこない，

$$log(\sigma_{\Delta t}) = 0.569 log(\Delta t) - 4.0585 \tag{8}$$

が得られた。その関係は，図５からも確認できよう。いずれの回帰式において，各式の R^2 が 0.90 以上となっていることが分かる。また，いうまでもなく，指数 K の値は 0.5 である場合は，ランダムウォークになる。

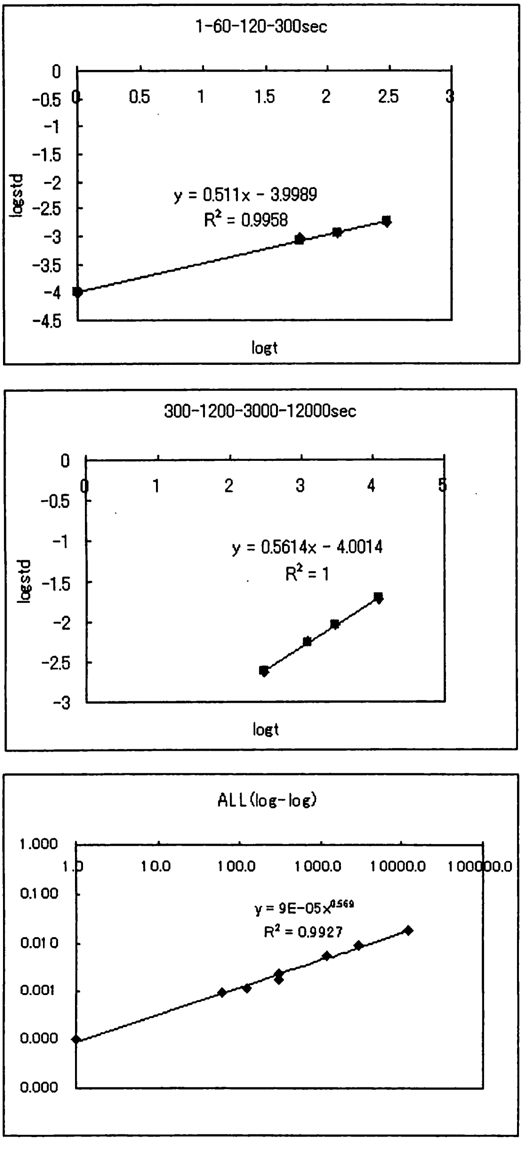


図５：標準偏差 $\sigma_{\Delta t}$ と時間間隔 Δt との関係

4　むすび

本研究は，主要業務時間帯（朝九時半から午後四時まで）のティックデータを用いて，その株価変動の統計特性を調べた。時間間隔を大きく取れば大きく取るほど，収益率の分布は正規分布に近づくことが分かった，また，分布のテールについて，実際の指数の値をフィットしてみた。特に時間スケールを大きく取れば大きく取るほど，テール部の指数の値が大きくなるという傾向が明らかになった。さらに時間スケールと収益率の標準偏差との関係についても調べ，回帰分析により，べき乗の関係をつかんだ。これにより，短時間内において，収益率の変化は，ランダムウォークに近いということは分かった。また，この論文は，全体研究の中間報告であり，全データセットを用いた分析の結果は，今後の研究で公表する予定である。さらに，株価だけではなく，為替や，先物などのティックデータも用いて，マーケットのミクロストラクチャーを研究する予定である。

〔参考文献〕

[1] Pareto, V., Manuel d'Economie Politique, Paris, 1927.
[2] Levy, P., Theorie de l'Addition des Variables Aleatoires, Gauthier-Villars, Paris, 1937.
[3] Stoll, H., and Whaley, R.(1990), "Stock market structure and Volatility", Review of Financial Studies, 3, 37-71.
[4] R. N. Mantegna and H. E. Stanley, Nature (London) 376, 46, 1995.
[5] I. Chang and D. Stauffer, Eur. Phys. J. B6, 543, 1998.
[6] A. Pagaan, J., Empirical Finance 3, 15,1996.
[7] R. N. Mantegna and H. E. Stanley, "An introduction to econophysics: correlations and complexity in finance", Cambridge University Press, 2000.
[8] A Car, E, and Lequeux P., "Trading rule profits and the underlying time series properties" First International conference on High Financial Frequency Data in Finance, Zurich, Switzerland, 1995.
[9] Lux, T., Marchesi, M., Nature (London) 297, 498(1999).