

株価指数データと混合正規モデル

三 浦 良 造

1 はじめに

株価の変動を数学モデルで記述しようとする試みは古くからある。そのアプローチの仕方は大きく分けて2通りある。ひとつは株価の対数の差を確率変数とみなしてそれが確率的に独立で同一分布に従うと仮定しその分布形を同定しようとするものである。もうひとつは対数価格の差の独立性を仮定しないで過去に依存した時系列モデルをあてはめようとするものである。後者の方が現実描写にすぐれていると思われる。近年の研究では対数価格差の条件付分布が詳しく研究されている。しかし本稿ではまず前者のアプローチをとる。その理由は2つある。まず後者のアプローチの方が最終的にはすぐれて現実描写的であってもそれが前者のアプローチのどういう欠点を改良したものかが未だよく見えないからである。もうひとつの理由は株式を原証券とするオプション価格との関連である。対数価格差を正規分布又は混合正規分布で表わせばオプション価格は明示的に表わされる。現実の価格変動が例えば独立同一分布性の仮定と分布が混合正規であるという仮定からどのようにはずれているかを摘出すれば今後の参考になるだろう。本稿はそういう意味で株価変動モデル構築へむけての準備作業のひとつである。

対数価格差が従う分布のあてはめとしては過去に正規分布、混合正規分布、学生t分布そして安定分布が試みられている。混合正規分布については最近アルゴリズムも改良されてきた。本稿では混合正規分布のあてはめをまず2節で行う。これによってどの時期にどういう正規分布がどのように混

合されるかが独立同一分布性の仮定のもとで見える。現実の株価変動が2節のアプローチでは説明できない部分を明確にするために独立同一性を検定する統計量を計算する。これは対数価格差の線形および非線形な依存関係に敏感に反応するもので、名称は確定していないが现阶段では Correlation Integral 統計量又は BDS 統計量と呼ばれているものである。

2 混合正規モデルと母数推定

いくつかの正規母集団があり一回毎の観測はこれらのうちの1つの母集団から観測されるときこの観測値は混合正規分布に従うといわれる。各回の観測でどの母集団から観測するかは確率的であるとし、各回のこの母集団の選択は他の回の母集団選択には影響を与えない、つまり確率的に独立であるとする。さらに各回の母集団からの観測は確率的に独立であるとする。

データの分布形（ヒストグラム）が一山型（unimodal）でなかったり、分布の片側又は両側の裾が長いようなときに混合正規分布による分布型の同定がよく行われる。本稿では株価（指数値）の対数値の差 $x_t = \log S_{t+1} - \log S_t$, $t=1, 2, \dots, T$ の分布が正規型よりも裾が長いために混合正規分布によってデータの分布に対する一応の理解を得ることにした。混合正規モデルについての説明は Kon (1984) [8] と三浦 (1988) [10], さらに詳しくは McLachlan and Basford (1988) [9] にあるのでここでは詳しく行わない。

観測が m 個の正規母集団のうちの1つから得られるとして、各母集団の密度関数を

$$f(x; \mu_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right\}$$

$$j=1, 2, \dots, k$$

と書く。観測が母集団 $j(j=1, 2, \dots, k)$ から得られる確率を λ_j と書くと同回の観測の密度関数、あるいは混合正規分布の密度関数は

$$f_0(x; \lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_k, \sigma_1, \dots, \sigma_k)$$

$$= \sum_{j=1}^k \lambda_j f(x; \mu_j, \sigma_j)$$

である。ただし $\sum_{j=1}^k \lambda_j = 1, \lambda_j > 0, j=1, 2, \dots, k$ である。

いま確率変数 X_1, X_2, \dots, X_T が互いに独立で同一の混合正規分布に従うとする。観測値 (確率変数の実現値) x_1, x_2, \dots, x_T が得られるときの尤度関数は

$$L(\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_k, \sigma_1, \dots, \sigma_k; x_1, x_2, \dots, x_T) \\ = \prod_{i=1}^T \left\{ \sum_{j=1}^k \lambda_j f(x_i; \mu_j, \sigma_j) \right\}$$

である。観測値 x_1, \dots, x_T にもとづく母数 λ, μ, σ の標準的な推定は最尤推定でありそれはこの尤度関数を最大にする母数の値として得られる。それは対数尤度を最大にする母数の値と同じだから

$$\theta = (\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_k, \sigma_1, \dots, \sigma_k) \\ \varphi = (x_1, \dots, x_T)$$

とにおいて

$$l(\theta; x) = \sum_{i=1}^T \left[\log \left\{ \sum_{j=1}^k \lambda_j f(x_i; \mu_j, \sigma_j) \right\} \right] \\ \equiv \sum_{i=1}^T [\log f_0(x_i; \theta)]$$

を最大にする θ の値である。それはまた

$$\frac{\partial}{\partial \theta} l(\theta; x) = \sum_{i=1}^T \left[\frac{\partial}{\partial \theta} \log f_0(x_i; \theta) \right] \\ = \sum_{i=1}^T \frac{\frac{\partial}{\partial \theta} f_0(x_i; \theta)}{f_0(x_i; \theta)} \\ = \sum_{i=1}^T \sum_{j=1}^k \frac{f(x_i; \mu_j, \sigma_j)}{f_0(x_i; \theta)} \cdot \frac{\frac{\partial}{\partial \theta} [\lambda_j \cdot f(x_i; \mu_j, \sigma_j)]}{f(x_i; \mu_j, \sigma_j)} \\ = 0 \text{ (ゼロベクトル)}$$

をみます。これは書き換えれば

$$\sum_{i=1}^T \frac{f(x_i, \mu_j, \sigma_j) - f(x_i; \mu_k, \sigma_k)}{f_0(x_i; \theta)} = 0, \quad j=1, 2, \dots, k-1$$

$$\sum_{i=1}^T \tau_{ij} \frac{\partial}{\partial(\mu_j, \sigma_j)} \log f(x_i: \mu_j, \sigma_j) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad j=1, 2, \dots, k$$

である。ただし

$$\tau_{ij} = \frac{\lambda_j f(x_i: \mu_j, \sigma_j)}{f_0(x_i: \theta)}$$

である。

このような最尤推定値を求める手順と考え方は McLachlan and Basford [9] によく説明されている。そこに引用されている Hathaway (1986) [2] [3] のアルゴリズムを本稿では用いた。それは E-M アルゴリズムと呼ばれる計算手順 (と考え方) の応用の一例であり、 θ の関数

$$\sum_{i=1}^T \sum_{j=1}^k \tau_{ij} \log \lambda_j + \sum_{i=1}^T \sum_{j=1}^k \tau_{ij} \log f(x_i: \mu_j, \sigma_j)$$

を作り最大化する。Kon [8] の引用にあるように Kiefer (1978) [6] はこの最尤推定を定める方程式には根があることを本稿の設定より少し一般的な形 (回帰の場合) について証明している。そこでは初期値のとり方についての提案も行っている。

$(\lambda_j, \mu_j, \sigma_j), j=1, 2, \dots, k$ の推定値を得たあとで、各観測値 $x_i, i=1, 2, \dots, T$ がどの母集団から来たかを定める必要があるが、Kon [8], McLachlan [9] (11 ページ) に従って

$$\hat{\lambda}_j f(x_i: \hat{\mu}_j, \hat{\sigma}_j)$$

が最大である母集団をえらぶこととした。

以上のようなモデルと計算手順に従って Topix の日次データを 2 つ又は 3 つの正規母集団に分けることを試みた。1985 年 4 月 30 日から 1990 年 12 月 6 日までの 1500 日分のデータを用いた。

図 6 の一部分では Topix の値 S_t の変化を t を横軸にして折れ線グラフで示している。図 1 は S_t の対数値の差 $x_t = \log S_{t+1} - \log S_t$ を t を横軸にして折れ線グラフで示した。オプション価格理論でよく仮定されるように S_t が対数正規過程に従うならば x_t は単一の正規母集団からの観測値なのだがしばし

図 1

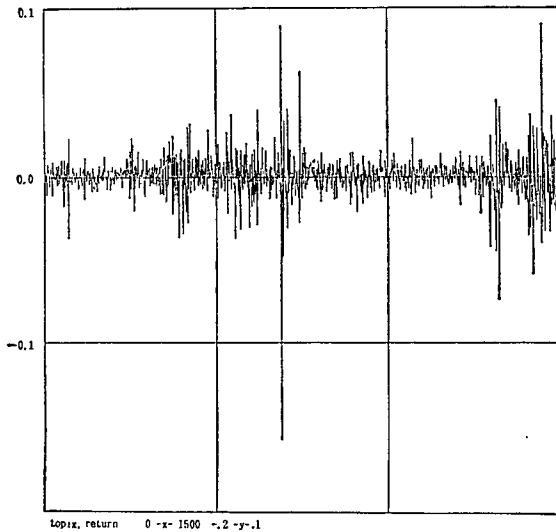


図 2

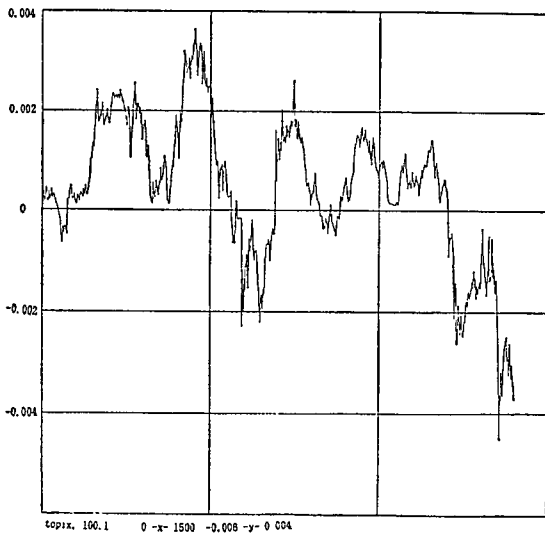


図 3

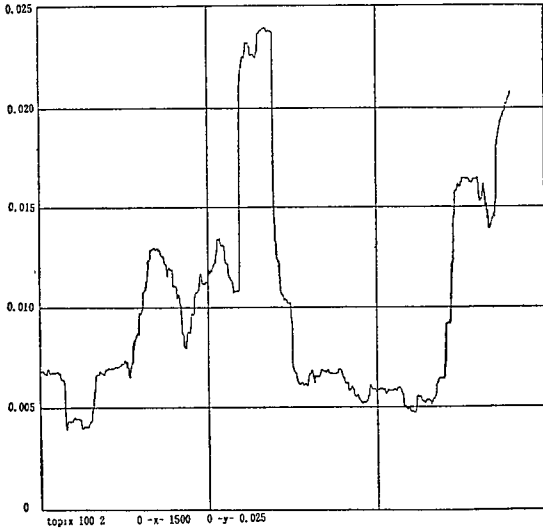


図 4

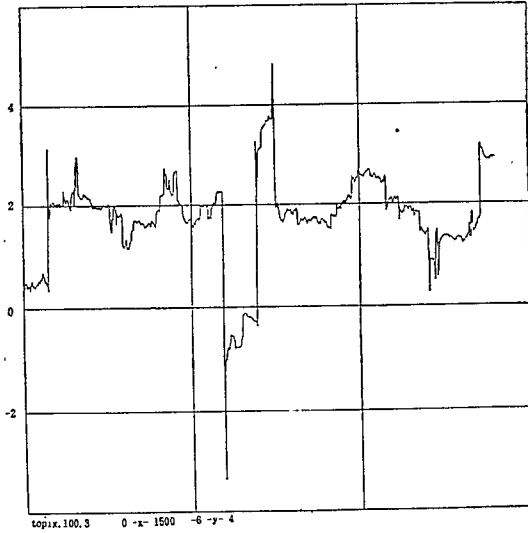
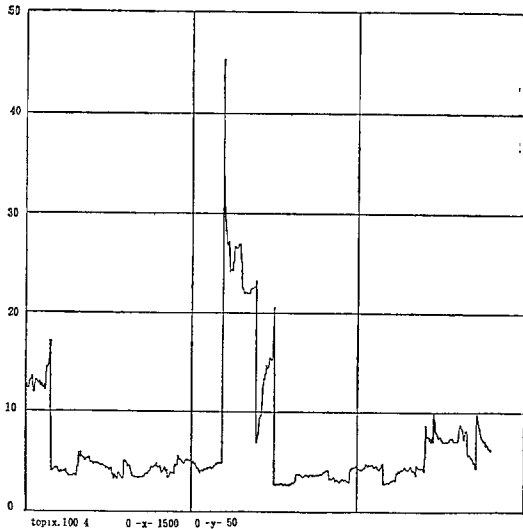


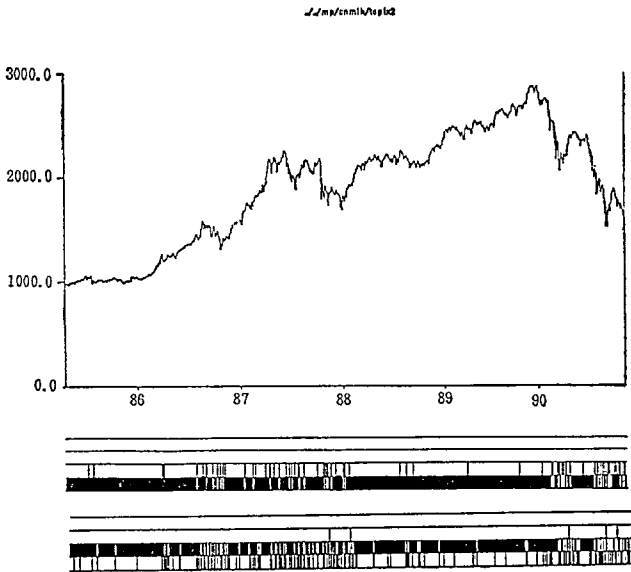
図 5



ば指摘されているように (例えば刈屋他 (1989) [5] を参照せよ.) 実際の x_t は正規分布に従ってはいないようである. 図 1 を一見して分かるように時期によってパラツキの大きさも異なるし時々飛び離れた値もみられる. 図 2 と図 3 ではそれぞれ 100 日間の標本平均と標本標準偏差を 1 日ずつずらして移動平均, 移動標準偏差として表示した. きちんとした検定の手順を待つ必要がない位によく変化している. 参考のために図 4 と図 5 ではそれぞれゆがみ又はねじれ (Skewness) と尖り (Kurtosis) の様子を計る 3 次と 4 次のモーメントを前の 2 つの図と同様の手順で 1 日ずつずらしてグラフで示した. このような統計量を用いて正規性の検定を行うこともできるが新味はないのでここでは行わない.

図 6 は 1500 日分の x_t をひとまとめにしてひとつの混合正規分布に当てはめた結果を示している. 上段の折れ線グラフで Topix の S_t を示した. 中段は混合正規分布で $k=2$ としたときに各日の x_t が 2 つのうちのどちらの母集団に属するかを示している. 黒い縦線が所属を示している. 下段は同様に $k=3$ と

図6



したときの x_t の所属を示している。 $k=2$ と $k=3$ とではどちらが尤もらしいかというきちんとした判定は本稿では行わない。ここでは S_t の変化を x_t を通してつぎの2つの視点からみることとし後のモデル構築のための準備としたいからである。1つは $X_t, t=1, 2, \dots, T$ が独立同一分布に従うと仮定するモデル（つまりここでは混合正規モデル）のもとでデータの特徴を見ることである。2つは $X_t, t=1, 2, \dots, T$ の互いの依存性（確率的従属性）がどの時期に強いかを見ることである。

表1には $k=2$ と $k=3$ の場合の母数の推定値を示した。

図1で予想できたように x_t がかなりばらついている時期には2つの母集団からの x_t が混在している。よくまとまっている時期は1つの母集団からの x_t が続いている。飛び離れた x_t の値は $k=3$ の場合に最も分散の大きな母集団に所属するように判定されていることも見える。表1から見られる特徴は

表 1

		λ	μ	σ
$k=2$	1	0.852	0.00081	0.00664
	2	0.148	-0.00222	0.0258
		λ	μ	σ
$k=3$	1	0.68	0.00090	0.0054
	2	0.31	-0.00065	0.01547
	3	0.01	-0.00445	0.06231

(Topix の特徴であると主張する論理的根拠は未だないのだが) $k=2$, $k=3$ のどちらの場合においても μ の値の小さい母集団の方が大きな σ の値をもっていることである。これは Kogure and Takeuchi (1991) [7] の指摘する x_t の条件付分散が $|x_t|$ だけでなく x_t の正負の符号にも依存するという特徴と合わせて今後のモデル作りの参考とすべきであろう。ここで1つ注意しておくことがある。各 x_t が3つの正規母集団のうちのどれに属するかを判定するために $\hat{\lambda}_j f(x_t; \hat{\mu}_j, \hat{\sigma}_j)$ が最も大きい j を選んだのだが、上の場合第3の母集団には6個の x_t が属すると判定された。 $\lambda_3=0.01$ だから1500個のうちの1パーセント、つまり15個程度が期待されるがそれより非常に少ない個数しか得られなかった。この判定方法の弱点であろうか。

$k=2$ とした場合にはデータの85%が分散の大きな(そして平均の小さい)母集団に属し残る15%が分散が小さい(そして平均の大きい)母集団に属するという結果であった。しかし $k=3$ とすると飛び離れた x_t を含めてデータの1%が分散が最大である母集団にまとめられ、その影響でデータの31%が中位の分散をもつ母集団に属し、データの68%が低位の分散をもつ母集団に属するという結果になった。 $k=3$ の結果の方が飛び離れた x_t をよく説明しているので分かりやすい。

$k=3$ の場合について1500日間における時期毎の差異をみると1986年の後半と1987年のほぼ全体と1990年のほぼ全体において低位と中位の分散をもつ母集団がよく混合され、さらに高位の分散をもつ母集団も混合されている。1985年と1988年と1989年においては低位の分散をもつ母集団のデータが大

半を占めている。このようにしてみると前者の時期はボラティリティが相対的に大きい時期で後者はボラティリティが小さい時期としてその構造と内容を伴って理解される。

1986年から1990年までの5年間について各年毎に混合正規分布の当てはめも行った。その結果は次節で示す。

ある時期において x_t の確率的独立性が強く認められればその時期におけるオプション価格は正規性の確認のもとでブラック・ショールズのオプション価格式を用いてよい。そこでの分布の正規性が認められなくて混合正規分布が尤もらしいならば、ブラック・ショールズ式を用いるのはやめて混合正規分布を基礎にしたオプション価格式を用いればよい。この価格式については前稿三浦[10]に説明した。そのときの母数の推定のためには本節でとりあげた推定方式が有力であろう。

各時期における x_t の確率的独立性をチェックするために種々の検定統計量が開発されている。とくに近年は x_t の線形依存性だけでなく非線形な依存性も検出する統計量が提案されている。次節で用いる BDS 統計量はそのうちの有力なひとつである。次節ではこの統計量を用いて Topix 日次データの x_t の互いの依存性の様子を見ることにする。

3 BDS 統計量

BDS 統計量は Correlation integral (日本語では相関積分とでも呼ばばよいか。) と称される統計量をもとにしている。correlation integral はもともと物理学の研究者がカオス風の決定論的システム (chaotic deterministic system) と確率的システム (stochastic system) を区別するために考案したものである (Grassberger and Procaccia (1983) [1] 参照)。それを3人の研究者 Brock, Dechert, and Scheinkman が観測値の独立同一分布性を検定する統計量として応用したので Hsieh (1989) [4] にならって本稿でも BDS 統計量と呼ぶことにする。この統計量については簡単な性質しか知られておらず現在数理統計学の分野で研究の対象となっている。(例えば Wolf (1990) [11] がある。)

いずれ統計量の性質がもっとよく知られるうちに例えば相関積分統計量とでも
 いうような内容を表わす呼称が定まるであろう。

BDS 統計量を定義する。 X_1, X_2, \dots, X_T を独立な確率変数としそれぞれ同一の分布 F に従うとする。連続する N 個の確率変数 $X_t, X_{t+1}, \dots, X_{t+N-1}$ からなる N -次元ベクトルを \underline{X}_t^N と書く。つまり $\underline{X}_t^N = (X_t, X_{t+1}, \dots, X_{t+N-1})'$ である。つぎに \underline{X}_t^N と \underline{X}_s^N の近さを測るものとして

$$I_l(\underline{X}_t^N, \underline{X}_s^N) = \begin{cases} 1 & l \geq \max_{1 \leq j \leq N-1} |X_{t+j} - X_{s+j}| \text{ のとき} \\ 0 & l < \max_{1 \leq j \leq N-1} |X_{t+j} - X_{s+j}| \text{ のとき} \end{cases}$$

を定義しておく。これらを用いて correlation integral は

$$C_N(l, T) = \frac{2}{T_N(T_N-1)} \sum_{t < s} I_l(\underline{X}_t^N, \underline{X}_s^N)$$

と定義される。ここで $T_N = T - N + 1$ である。

$X_t, 1 \leq t \leq T$ が独立で同一分布に従うという仮定のもとで $C_N(l, T)$ の期待値と分散の推定を構成し、それを用いて標準化された Correlation integral が $T \rightarrow \infty$ において漸近正規分布に従うことが証明されていれば一応の検定ができる。Hsieh [4] は BDS 3 氏が証明が示したとしてその結果をつぎのように紹介している。

任意に固定された N と l に対して、 $T \rightarrow \infty$ のとき $C_N(l, T)$ は $C(l)^N$ に確率 1 で収束する。ただし

$$\begin{aligned} C(l) &= \lim_{T \rightarrow \infty} \frac{2}{T(T-1)} \sum_{t < s} I_l(\underline{X}_t^1, \underline{X}_s^1) \\ &\equiv \lim_{T \rightarrow \infty} C_1(l, T) \end{aligned}$$

である。さらに $\sqrt{T}[C_N(l, T) - \{C_1(l, T)\}^N]$ が $T \rightarrow \infty$ のとき平均ゼロで分散

$$\sigma_N^2(l) = 4 \left[K^N + 2 \sum_{j=1}^{N-1} K^{N-j} C^{2j} + (N-1)^2 C^{2N} - N^2 K C^{2N-2} \right]$$

の正規分布に従う。ただし

$$C \equiv C(l) = \int [F(x+l) - F(x-l)] dF(x) \equiv P_r \{ |X_t - X_s| < l \}, \quad t < s$$

$$\begin{aligned}
 K &\equiv K(l) = \int [F(x+l) - F(x-l)]^2 dF(x) \\
 &\equiv P_r \{ |X_t - X_s| < l \text{ かつ } |X_s - X_r| < l, t < s < r \}
 \end{aligned}$$

である。 $C_1(l, T)$ は $C(l)$ の推定量であり $K(l)$ は

$$K(l, T) = \frac{6}{T_N(T_N-1)(T_N-2)} \sum_{t < s < r} I_l(X_t, X_s) I_l(X_s, X_r)$$

によって推定される。こうして BDS 統計量は

$$\text{BDS}(N, l, T) = \sqrt{T} \{C_N(l, T) - [C_1(l, T)]^N\} / \sigma_N(l, T)$$

と定義される。これは $T \rightarrow \infty$ のとき標準正規分布に従う。ここで $\sigma_N^2(l, T)$ は上の $\sigma_N(l)^2$ の定義式のなかで K を $K(l, T)$ で C を $C_1(l, T)$ でおきかえたものである。

Topix 日次データ $x_t, t=1, 2, \dots, 1500$ に対して $T=100, l=0.005, N=5$ とし て BDS 統計量を計算し、連続する 100 日分の x_t 全体を 1 日ずつずらして いわば移動 BDS 統計値をグラフにしたものが図 7A である。それぞれの BDS 統計値の算出に用いた平均の推定値 ($[C_1(l, T)]^N$) と標準偏差の推定値 ($\sigma_N(l, T)$) をそれぞれ図 7B と図 7C にグラフとして示した。

Hsieh [4] のシミュレーションによる研究では BDS 統計量は X_t の独立同一分布性を帰無仮説とするとき単に X_t の線形だけでなく非線形な確率的従属性を対立仮説として設定しても有効な検定統計量であると認められている。図 7A をみれば粗っぽくみても BDS 統計値が 2 を越える時期つまり x_t の独立同一性が棄却される時期が 6 回ある。しかし一定して -0.5 から 1 の間にある時期もある。BDS 統計値が大きいのは X_t^N の集積が独立同一分布性のもとで想定されるより以上に頻繁に生じることを意味する。 X_t^N の集積とは N 個の連続する X_t がよく似た変動パターンを示すことである。これが頻繁に生じる時期は X_t の何らかの従属性が強いのでさらにふみこんだモデル化が必要である。ただ BDS 統計値は l と N の値の選択に依存して微妙に変化するので注意を要する。つぎに図 7B をみると correlation integral の平均値 (の推定値) も時期に依り異なることが分る。BDS 統計値が 1 よりも小さい時期にこの平均

図 7A

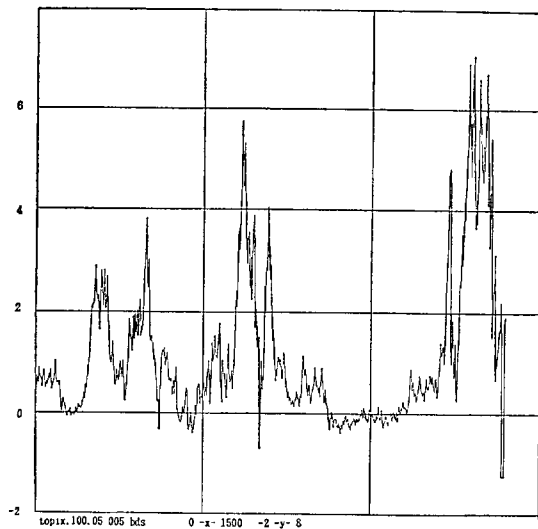


図 7B

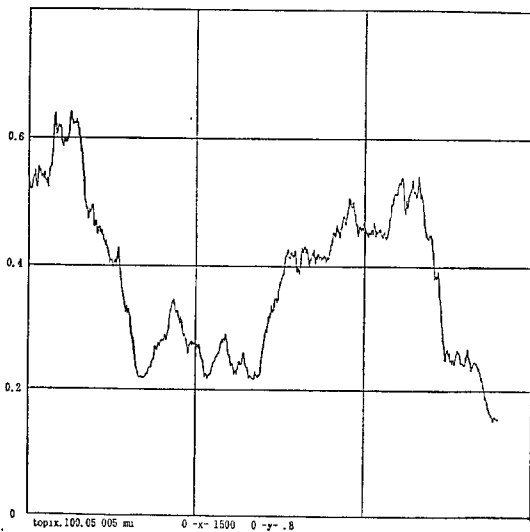
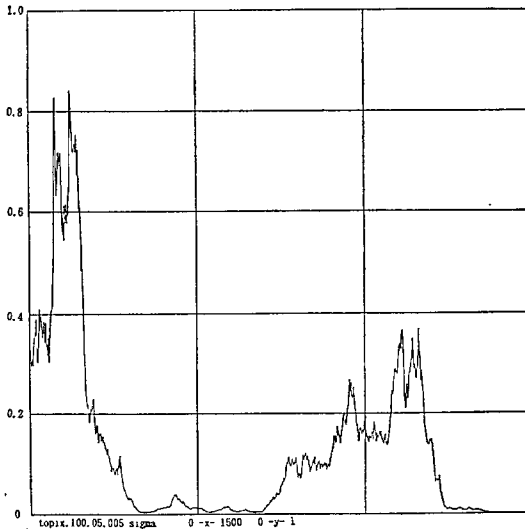


図7C



が大きいように見えるが初期つまり1985年から1986年にかけてはそうでもない。この平均値が理論的にどういう値をとるかは x_t の独立同一分布性のもとではよく分っているようだがその他の場合は未だ知られておらず、やっと昨年、自己回帰とか移動平均モデルの場合について研究成果が報告された状況である (Wolf, [11] 参照)。この平均値そしてつぎにみる標準偏差値が大幅に変化するということは独立同一分布性の仮定のもとでは F の形が変わること、又は独立同一分布性の条件のうち独立性と同一分布性のどちらかあるいは両者が満たされないことに対応する。図7Cをみると初期を除いては σ_N の値はBDS統計値が大きいときにゼロに近く(小さく)BDS統計値が小さいときには大きいようである。従って1986年のある時点以前と以降については図7Aから図7Cでみる限り x_t は異なる構造をもっているとみてよい。

BDS統計値は時刻が連続する N 個の X_t の(ベクトルとしての)集積を計測し、混合正規分布によるあてはめは時刻を無視して X_t の値の集団を X_t の

独立同一分布性の仮定のもとで2つ又は3つに分けている。この2種類の趣旨の異なる分析結果を並べて検討することはいくらかの意義があるだろう。時刻が連続している X_t がすべてひとつの母集団に属しながらベクトルとしての集積をみせる場合は S_t は比較的小きざみにしかし単調に上昇又は下降するだろうし、これが X_t が2つ又は3つの母集団に交互に属しながら生じる場合は S_t の上昇又は下降のきざみの様子が粗くなったり小さくなったりするだろう。そういう視点も含めてつぎの図をみる。1986年から1990年の5年間に対して各1年間それぞれについて個別に混合正規分布の当てはめを行った。表2と図8Aから図8Eにその結果を示した。比較がしやすいように図7AのBDS統計値のグラフと図6の S_t のグラフを時期を合わせて並べておいた。

まず $k=2$ の場合をみる。1988年はデータの99%が1つの正規母集団に属する。この年の S_t のグラフをみると年初に上昇してその後は昇降をくり返している。 X_t が独立同一分布に従っているという条件にふさわしいふるまいをみせている。BDS統計値をみると年初に値が非常に大きくその後は小さい値で安定している。値の大きさが S_t の上昇の時期に対応している。

X_t が、独立で同一の正規分布に従うとき、平均 μ がかなり大きくて σ が小さければ S_t は(平均的には)順調に上昇する。ちなみに X_t が負となる確率は

$$P\{X_t < 0\} = P\left\{\frac{X_t - \mu}{\sigma} < -\frac{\mu}{\sigma}\right\} = \Phi\left(-\frac{\mu}{\sigma}\right)$$

である。ただし $\Phi(\cdot)$ は標準正規分布関数である。例えば $\mu=0.001, \sigma=0.006$ とすれば

$$P\{X_t < 0\} = \Phi\left(-\frac{1}{6}\right) > 0.43$$

である。 X_t が負となる確率がこのように大きいと X_t の確率的独立性の仮定のもとでは S_t の単調な上昇、例えば1988年の年初のような上昇は稀にしか生じないだろう。このような上昇は X_t の確率的独立性が失われているとみて自然である。それに対応するかのよう X_t の従属性を示すBDS統計値はこ

表 2

$k=2$		λ	μ	σ
1986年	1	0.74	0.0022	0.0062
	2	0.26	-0.00080	0.015
1987年	1	0.911	0.0010	0.0098
	2	0.089	-0.0061	0.045
1988年	1	0.99	0.00097	0.0066
	2	0.01	0.0054	0.027
1989年	1	0.82	0.00062	0.0048
	2	0.18	0.0017	0.0082
1990年	1	0.69	-0.0026	0.011
	2	0.31	-0.0019	0.027
$k=3$				
1986年	1	0.58599	0.00220	0.00545
	2	0.40742	0.00090	0.01208
	3	0.00659	-0.03579	0.00106
1987年	1	0.68162	0.00104	0.00788
	2	0.30526	-0.00026	0.01893
	3	0.01311	-0.02229	0.09576
1988年	1	0.84975	0.00184	0.005999
	2	0.14658	-0.00392	0.00763
	3	0.00368	0.06359	0.00001
1989年	1	0.84541	0.00092	0.00471
	2	0.12309	0.00170	0.00919
	3	0.03150	-0.00665	0.00087
1990年	1	0.77722	-0.00289	0.01466
	2	0.11863	0.00115	0.03477
	3	0.10416	0.00052	0.00141

の時期には高い。1988年についてはこのように理解される。その特徴はまとめると X_t の独立同一分布性（しかも単一の正規分布に従う）の仮定のもとでは稀にしか生じないはずだが実は S_t の単調な上昇又は下降が生じており、しかも高い BDS 統計値が対応していることである。この特徴をもつ時期を他にさがすと1989年の終り頃がそのようである。この年は $\mu/\sigma=6/48$ であり X_t が負となる確率は1988年のそれより大きい。それにもかかわらず1年間の大半は S_t が順調に上昇し、年末近くになって上昇の度合が強くなる。この上昇の

図 8A

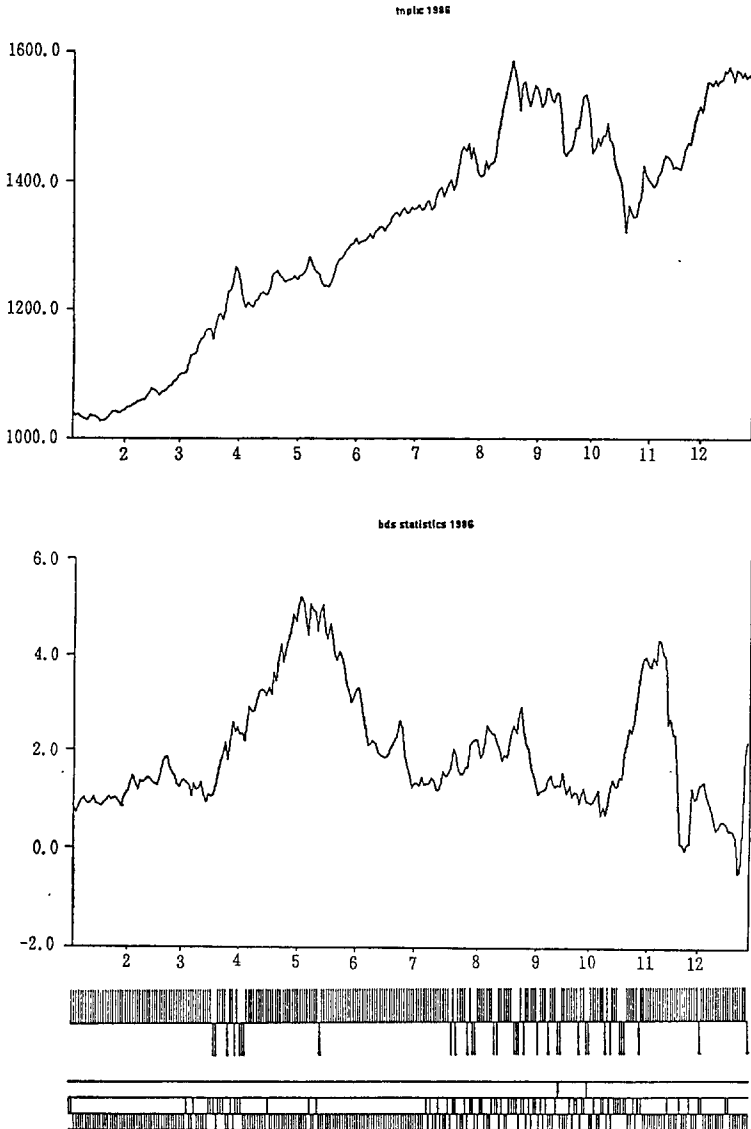


図 8B

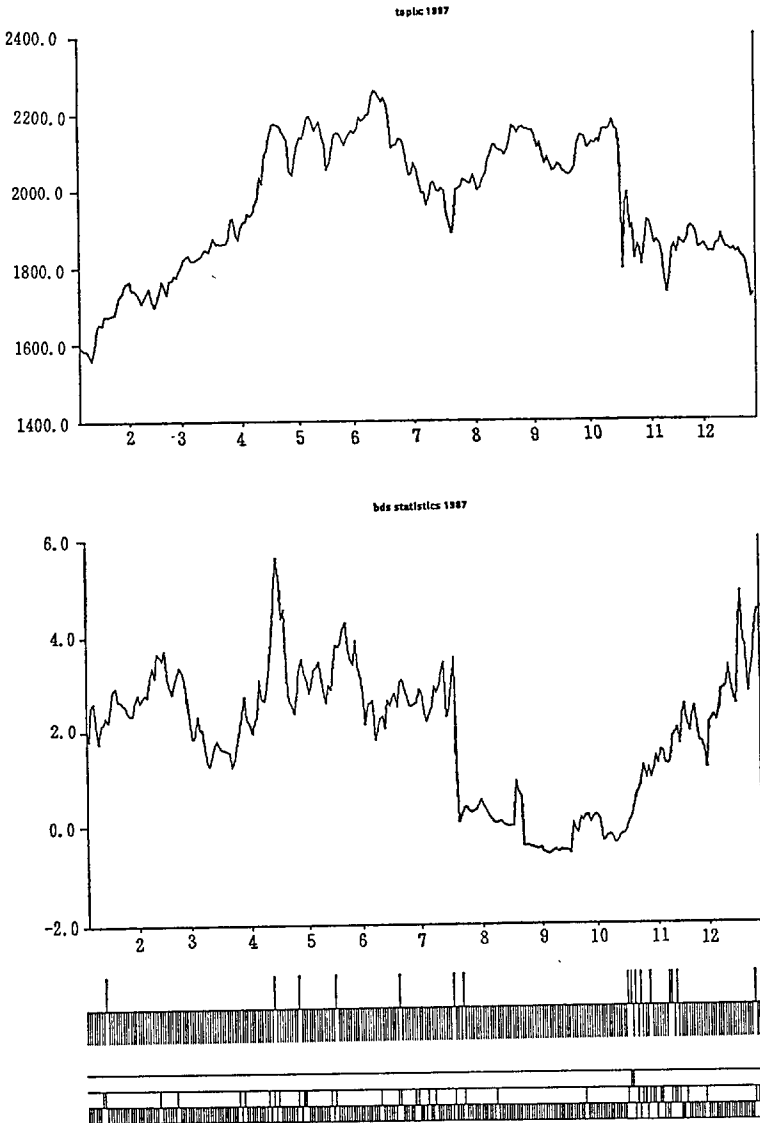


図 8C

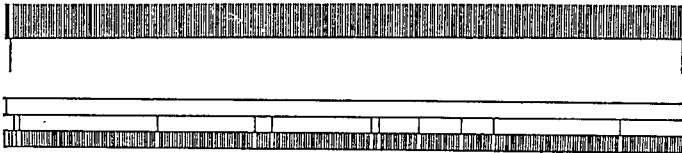
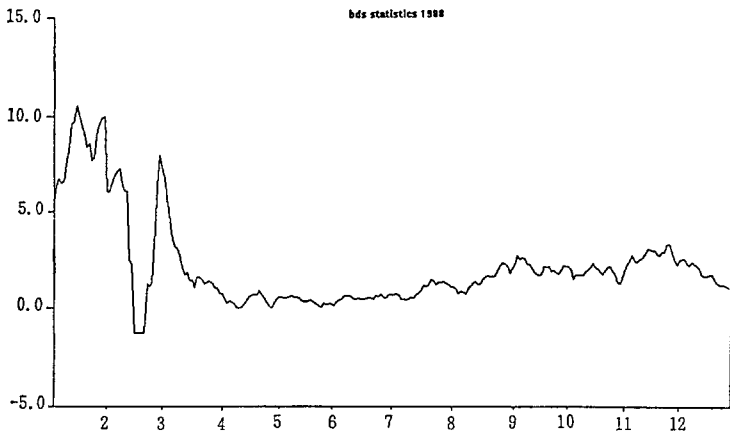
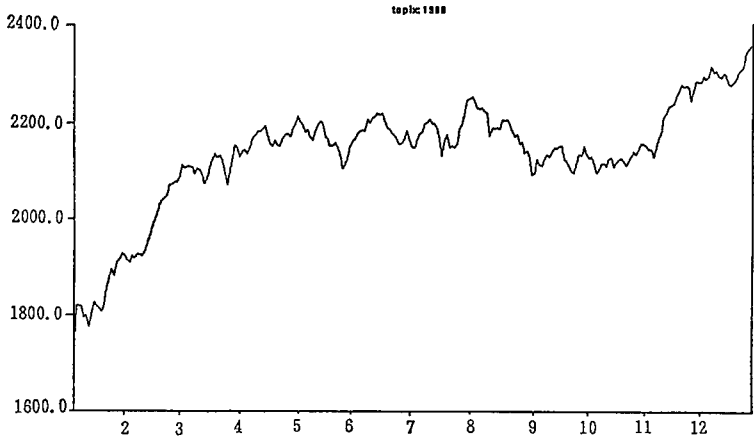


図 8D

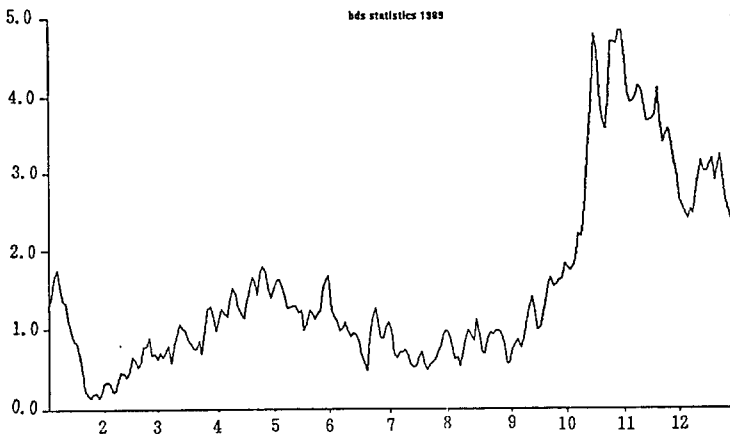
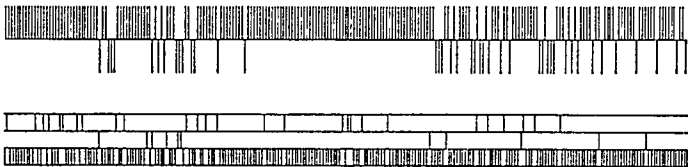
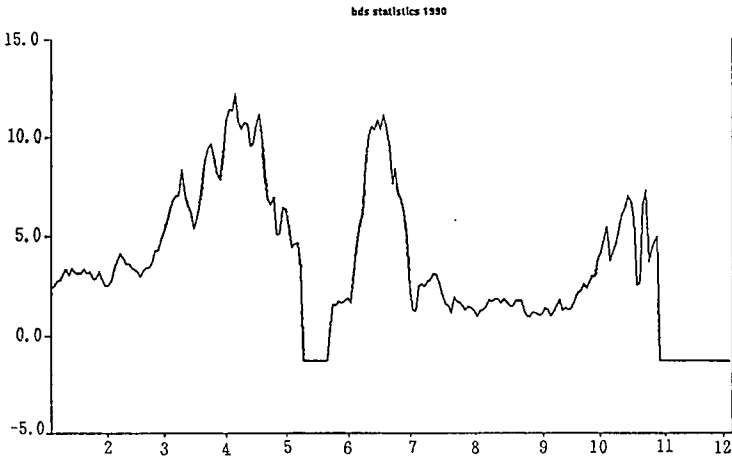
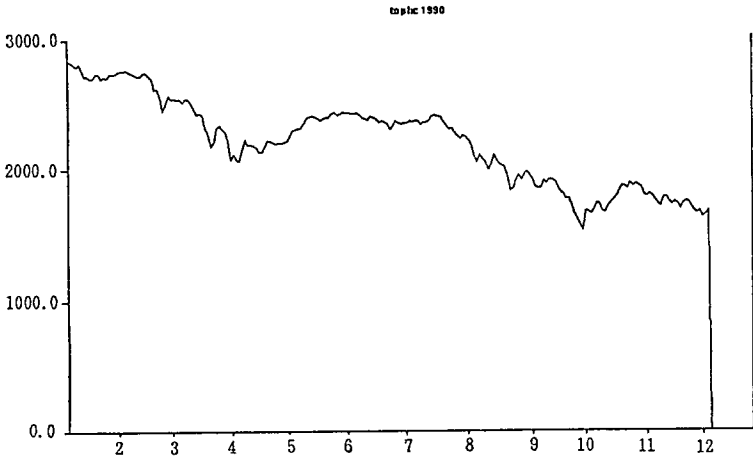


図 8E



強さが BDS 統計値の高さに対応している。

1986年には S_t が急上昇する時期は2つあるが同時期の BDS 統計値は高い。また1990年6月頃の単調な下降期にも BDS 統計値は高い。しかもこの時期のデータはほぼ単一の正規母集団に属する。他の上昇又は下降期は2つの母集団が混合しているため X_t の従属性(非独立性)とばらつき(σ)の大きさを同時に考慮する必要がある。

S_t の単調な上昇又は下降の傾向を BDS 統計値に対応させたが、しかしすでにふれたように単に BDS 統計値だけに依存させるのはいけない。例えば1990年の場合は $\mu/\sigma=26/110$ とか $\mu/\sigma=19/270$ であるから X_t が負となる確率は60%近くもある。従って X_t が独立であっても平均的に S_t は下降傾向にある。この平均的下降傾向を考慮したうえでさらに詳しく分析する必要がある。

つぎに $k=3$ の場合をみよう。1986年と1987年については正規母集団の混合が生じる時期は $k=2$ のときと同様である。1988年と1989年については S_t の変化が谷になるところの x_t がひとつの母集団に分類されたようである。1990年については平均が最もゼロに近くて分散が最も小さい集団がとりだされた。この年は $k=2$ の場合と比べて分類の様子が大幅に変わった。

$k=3$ の場合は各年で特徴的な x_t を第3の母集団としてとりだしているが λ_3 の値(母集団比率)が小さい年ではその位置づけを明確にするのが困難である。全体についていえることは $k=2$ の場合と比べて BDS 統計値との(観察による)対応づけがうまくいかない。つぎに表2で示す母数推定値について述べておく。前節でみた各正規母集団の(μ, σ)の大小関係についてであるが μ が小さい程 σ が大きいという関係は(λ_3 が微小であるところは除いて) S_t が全体に上昇期であるときには成立している。1990年にはそれが成立していないが逆に μ が大きいときに σ も大きいという傾向が成立しようにみえるが第3の母集団の母数がそれをさまたげている。しかしこの傾向の観察は S_t の下降期の市場の特徴であるという可能性をもっているので今後の課題としたい。

X_t の依存関係の存在を簡単にみておくために X_t の自己相関、つまり $(X_t,$

図 9A

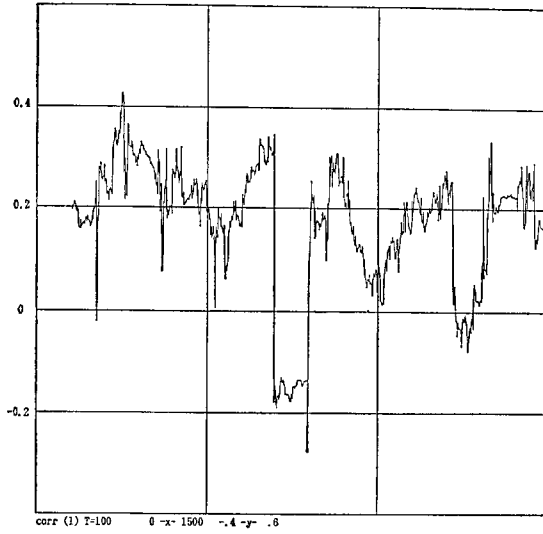


図 9B

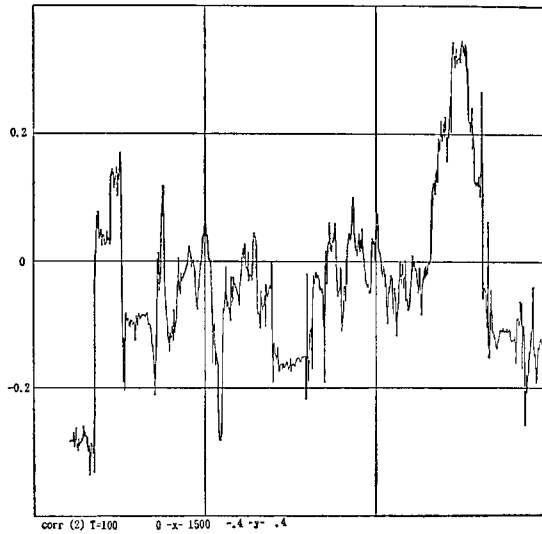


図 9C

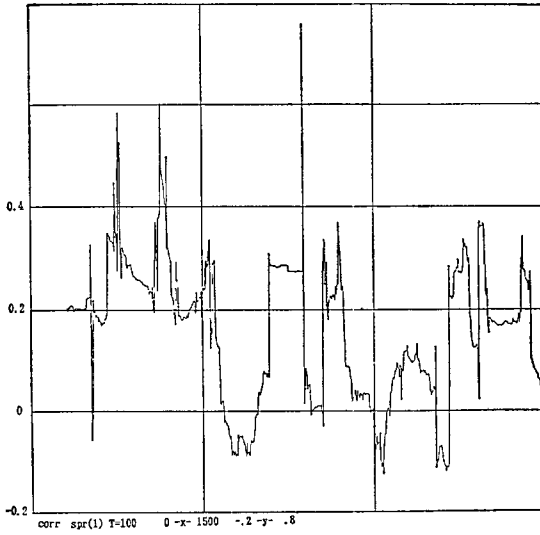
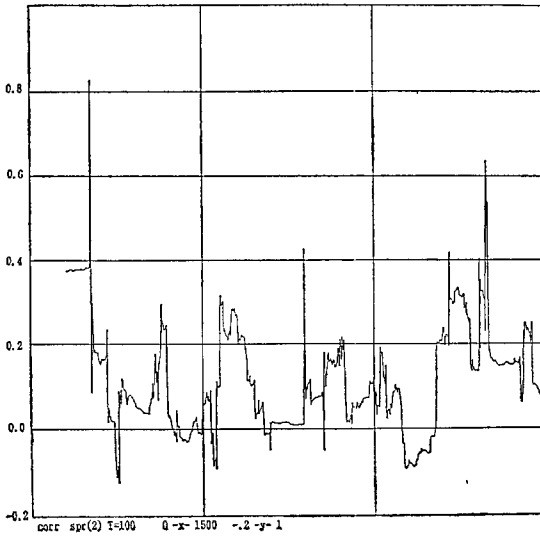


図 9D



X_{t+i}), $i=1, 2$ の 100 日移動相関をそれぞれ図 9A と図 9B に示した。さらに (X_t^2, X_{t+i}^2) , $i=1, 2$ の 100 日移動相関をそれぞれ図 9C と図 9D に示した。 X_t^2 の自己相関の方が X_t のそれよりもわずかに高いようにみえる。これは線形の依存性もあるがもっと強い非線形の依存性の存在を示している。この依存性の存在と BDS 統計値との対応については今後の課題としたい。

4 まとめと今後の課題

混合正規分布のあてはめによってつぎのことが分った。まず $k=2$ でも $k=3$ でも正規分布の混合が各時期均等にあるのではなく時期によって混合に濃淡がある。まだ検討の余地があるが混合される正規母集団の母数値にある傾向が見られる。指数値の上昇期には分散値が大きい程平均値が小さい。下降期はその傾向がくずれてはぼ逆向きになる。

BDS 統計値を併置してつぎのことが観察された。正規分布の母数値からは想定しにくい程の指数値の単調な上昇又は下降が生じている時期の BDS 統計値が高い。これは対数価格差のあいだの依存関係の存在を示していると思われる。

今後の課題のひとつは正規分布の混合がよく生じているときに各時点の正規母集団の選択が過去に依存したもののか否かを調べることである。もうひとつは対数価格差の線形依存関係と非線形依存関係を表現することである。このとき市場の環境を示す他の変量を用いてこれらの依存関係の背景が理解されることが望ましい。

注) 本稿のデータ分析のためにデータの提供と計算の実行について(株)日興リサーチセンター投資工学研究所の御協力を得ました。ここに記して感謝します。もし計算結果に誤りがあれば、それは著者の責任であることをお断りしておきます。

本稿の研究は日本証券奨学財団研究調査助成金の援助をうけました。

参考文献

- [1] Grassberger, P. and Procaccia, I (1983) "Characterization of strange attractors," Physics Review Letters, Vol. 50. pp. 316—319.

- [2] Hathaway, R. J. (1986 a), "Another Interpretation of the EM Algorithm for mixture Distributions." *Statistics and Probability Letters*. Vol. 4. pp. 53—56.
- [3] Hathaway, R. J. (1986 b). "A Constrained EM Algorithm for Univariate Normal Mixtures." *Journal of Statistical Computing and Simulations*. Vol. 23. pp. 211—230.
- [4] Hsieh, D. A. (1989). "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates." *Journal of Business*. Vol 62, No. 3, pp. 339—368.
- [5] 刈屋武昭他編著 (1989年)「日本の株価変動」東洋経済新報社
- [6] Kiefer, N. (1978). "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model." *Econometrica* 46. pp. 427—434.
- [7] Kogure, A. and Takeuchi, Y. (1991). "The asymmetric property of the conditional variances in the Japanese stock market: some empirical evidence against the ARCH and Taylor models.' mimeographed.
- [8] Kon, S. J. (1984). "Models of Stock Returns—A Comparison," *The Journal of Finance*, Vol. 39, No. 1. pp. 147—165.
- [9] McLachlan, G. J. and Basford, K. E. (1988) 「Mixture Models. Inference and Applications To Clustering.」 Marcel Dekker, Inc. New York and Basel.
- [10] 三浦良造 (1989) "混合型株価変動モデルとオプション価格" 一橋論叢 第102巻 第5号, pp. 620—644 (48)—(72))
- [11] Wolf, R. C. L. (1990). "A note on the behavior of the correlation integral in the presence of a time series." *Biometrika*, Vol. 77, No. 4, pp. 639—697.

(一橋大学教授)