

ウェーブレット分散を用いた 金融時系列の長期記憶性の分析

いなだまさかず
稲田将一

要 旨

本稿では、ある時系列が長期記憶性を有するかどうかを分析する手法を検討する。長期記憶性の有無を調べる方法の1つとして、スペクトル密度を用いる方法が知られているが、推定精度は高くないという問題がある。そこで、スペクトル密度に代えてウェーブレット分散を用いて長期記憶性の有無を調べる手法を示す。ウェーブレット分散を用いると、推定精度が向上する。実証分析として、株価（TOPIX）の日次収益率の長期時系列にウェーブレット分散を用いると、1970年代には長期記憶性が認められたものの、1980年代以降では長期記憶性が認められなかった。また、円/ドル・レートの日次変化率では、一貫して明確な長期記憶性は認められなかった。

キーワード：長期記憶性、ウェーブレット分散、スペクトル密度、
フラクショナル・インテグレーション過程

.....
本稿は、2006年3月に日本銀行で開催された「金融商品の価格付け手法とリスク管理技術の新潮流」をテーマとする研究報告会（FEテクニカル・ミーティング）への提出論文に加筆・修正を施したものである。同テクニカル・ミーティング参加者からは、貴重なコメントを多数頂戴した。記して感謝したい。ただし、本稿に示されている意見は、筆者個人に属し、日本銀行の公式見解を示すものではない。また、ありうべき誤りはすべて筆者個人に属する。

稲田将一 日本銀行金融研究所主査（E-mail: masakazu.inada@boj.or.jp）

1. はじめに

有名なブラック＝ショールズのオプション価格公式では、株価の収益率が、一定値のドリフトと一定値のボラティリティを伴ったブラウン運動で表現されるとの前提がおかれている。この前提は、それによって、計算が容易なオプション価格公式が導き出されることから、金融実務で多用されている。

しかし、ドリフトやボラティリティが一定値であるという前提は、株価収益率が過去の株価の履歴によらずに決められることを意味するが、現実の株価変動は、この前提に従っていないとの指摘が数多くなされている。例えば、ある日に、株価に大きな変動が生じると、しばらく大きな変動が続くことが指摘されている（ボラティリティ・クラスタリングの存在）。この指摘に基づけば、株価等の時系列過程は、過去の履歴に依存しない「無記憶過程」ではなく、過去の履歴に依存する「有記憶過程」であることになる。

以下、本稿では、ウェーブレット（wavelet）解析を用いることにより、株価収益率や為替レートに、十分遠い過去の影響を受ける「長期記憶性」が存在するかどうかを検証する^{1, 2}。

長期記憶性の有無を判断する際に留意しなければならないこととして、金融データの統計的性質が時間によって変化し得るという点が挙げられる。特に、超長期のデータを扱う際には、ある程度サンプルを区切って統計的性質の安定性を検討することも必要になる。例えば、平均値が大きくシフトするような階段状の異常値があると、時系列に長期記憶性があると誤って検定してしまうことがある³。そのため、異常値を含む時系列を分析対象とする場合には、それらの影響を除去して分析することを検討する必要がある。

本稿で示す長期記憶性の判定方法は、金融時系列のモデル化に当たって事前分析として使うことができる。例えば、オプション価格公式を導くことを目的として、株価収益率の時系列をモデル化するのであれば、本稿で示す長期記憶性の判定方法により当該時系列に長期記憶性が確認された際には、ブラック＝ショールズのオプション価格公式に適当な修正を施すことが考えられる⁴。また、長期記憶性がある時系列の将来変動を予測するためには、長期記憶性を的確に表現したモデルを使用することが望ましいであろう。

1 刈屋・勝浦 [1992] 等では、ボラティリティと異なり、株価収益率には長期記憶性はないと考えられている。なお、白石・高山 [1998] は、TOPIXと日経平均株価のボラティリティは長期記憶過程として表現することが妥当であると報告している。

2 経済データの長期記憶性を解説したものに、Baillie [1996] がある。

3 例えば、Gourieroux and Jasiak [2001] がある。

4 Elliott and Hoek [2003] は、連続時間の枠組みで、資産収益率に長期記憶性が認められる場合に、オプション価格公式に修正が必要となることを示した。

本稿の構成は以下のとおりである。2節では、長期記憶性の概念を説明し、長期記憶を表現するモデルを説明する。3節では、フーリエ解析とウェーブレット解析を用いて長期記憶性の有無を判定する手法を解説する。4節では、まず、人工的なデータで長期記憶性の有無を判定する。その際、ウェーブレット解析を用いた手法がフーリエ解析を用いた手法より恣意性が小さく、高精度で長期記憶性の有無を判定可能であることを示す。さらに、現実のデータとして株価（TOPIX）の日次収益率や円/ドル・レートの日次変化率を取り上げ、ウェーブレット解析を用いて長期記憶性の有無を調べる。その際には、推定期間をずらして長期記憶性がどのように変化してきたかを確認する。5節では、本稿のまとめを行う。補論では、本分析に必要なウェーブレット解析の基礎的事項を整理する。

2．長期記憶過程

本節では、長期記憶性という概念を説明し、長期記憶性を的確に表現するモデルとして、フラクショナル・ガウシアン・ノイズ過程とフラクショナル・インテグレーション過程を説明する。さらに、時系列が長期記憶性を保有しているか否かを検証する方法を述べる⁵。なお、以下では、数学的な厳密さよりもわかりやすさに焦点を当てて説明を行う。

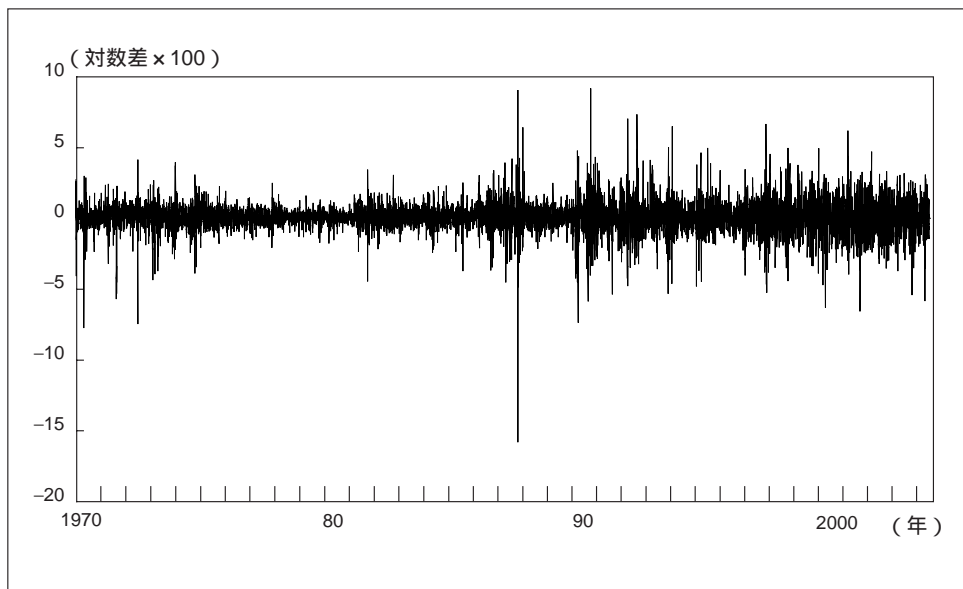
(1) 株価収益率の単位根検定

時系列分析の多くは、単位根検定によって、データが定常であるか否かをチェックするところから始まる。通常、定常過程をI(0)過程、1階階差をとった系列が定常になるような非定常過程をI(1)過程とそれぞれ呼ぶ。単位根検定では、統計的検定によりデータが定常か否か（I(0)過程、I(1)過程のいずれに従うか）の判断を行う⁶。具体的に、現実のデータの例として株価（TOPIX）収益率（図表1）を取り上げ、単位根検定として広く用いられているADFテスト（Augmented Dickey-Fuller test）やPPテスト（Phillips-Perron test）を行った結果が図表2である。これらの単位根検定は、時系列がI(1)過程に従っているという帰無仮説を検定する手法である。帰無仮説が棄却された場合には、時系列は定常過程（I(0)過程）に従っていると判断される。

5 長期記憶性に関する詳しい説明は、Beran [1994]、矢島 [2003] 等を参照されたい。

6 非定常データのうち、データの1階階差をとることによって、定常過程（I(0)過程）になるものがI(1)過程である。厳密には、非定常データには、より高階の階差をとることにより初めて定常過程となるものも存在し、 n 階差をとることによって定常過程になるものを、I(n)過程と呼ぶ。ただし、 $n > 1$ のI(n)過程は、ここでは議論の対象としない。

図表1 株価収益率の推移



図表2 株価収益率の単位根検定

ADFテスト		PPテスト		KPSSテスト
定数項あり	定数項なし	定数項あり	定数項なし	定数項あり
-35.973**	-35.916**	-84.174**	-84.136**	0.493*

備考：**1%有意水準、*5%有意水準。ADF検定のラグ次数は、AICにより判定。

図表2左・中列の結果によると、ADFテストとPPテストのいずれの場合も、1%有意水準で帰無仮説を棄却している。したがって、これからは、株価収益率はI(0)過程と判断される。ADFテストやPPテストの頑健性をチェックするため、Maddala and Kim [1998] で提唱されているように、帰無仮説と対立仮説を入れ替えたKPSSテスト (Kwiatkowski *et al.* [1992]) も同時に行った。その結果 (図表2右)、KPSSテストでも5%有意水準で帰無仮説を棄却しており、株価収益率がI(0)過程であるというADFテストやPPテストの結果とは異なる結果を示している⁷。

このような矛盾した結果が得られる理由として、各テストの検定力の問題が考えられるが、それ以外では、株価収益率はI(0)過程やI(1)過程のいずれにも属さず、両者の中間的性質を備えた系列と捉えることもできる。つまり、株価収益率をはじめ、現実のデータは、I(0)過程やI(1)過程に明確に区分することはできないと考えることも可能である。以下では、こうした問題意識のもとで、金融時系列の性質を探る。

⁷ 株価収益率がI(0)過程に従っているのであれば、KPSSテストの帰無仮説は棄却されない。

(2) 長期記憶性と短期記憶性

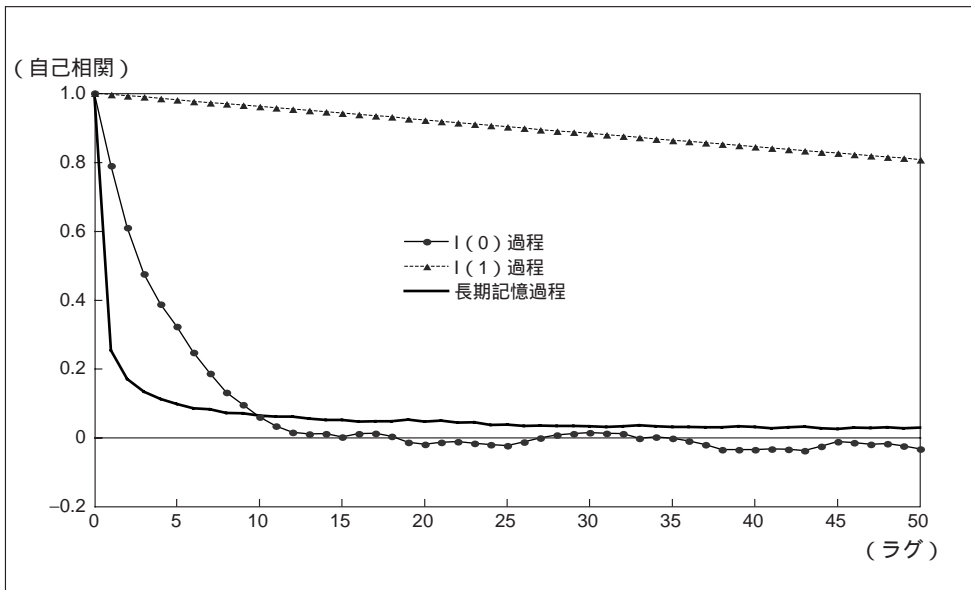
I(0)過程とI(1)過程は、標本自己相関関数でみると、次のような特徴を持つ。まず、I(0)過程は、急激（指数関数的）に自己相関がゼロに収束する。一方、I(1)過程には、長いラグであっても、非常に大きな自己相関が存在する（図表3）⁸。

I(0)過程は、遠い過去の影響をほとんど受けないことから、「短期記憶過程」と呼ぶことができる。これに対し、I(1)過程ほどの大きな自己相関は続かないが、I(0)過程ほどには急激に減衰せず、長いラグをとってもある程度自己相関が残っている過程を「長期記憶過程」と呼ぶ。したがって、長期記憶過程では、次に示すように自己相関関数 $\rho(h)$ の無限和が発散する。

$$\sum_{h=0}^{\infty} |\rho(h)| = \infty \quad (1)$$

I(1)過程も(1)式を満たすことは容易にわかるが、本稿で長期記憶過程という場合には、平均値や分散が一定であり、自己相関関数は時間差のみの関数となっているという意味で「定常な長期記憶過程」を指すものとする。I(1)過程は、分散が増大する過程で満たされないため、対象としない。(1)式と異なり、 $\rho(h)$ の無

図表3 I(0)過程、I(1)過程と長期記憶過程の標本自己相関関数の例



8 図表3では、 ϵ_t を標準正規分布に従う独立な乱数として、I(0)過程は $x_t = 0.8x_{t-1} + \epsilon_t$ 、I(1)過程は $x_t = x_{t-1} + \epsilon_t$ に従う系列 $\{x_t\}$ をそれぞれ人工的に生成し、それらの標本自己相関関数を計算した。また、長期記憶過程は、後述の(16)式で $d = 0.2$ とした $\{x_t\}$ を人工的に生成した。

限和が発散しないものが短期記憶過程である。このように自己相関関数の無限和の性質が異なるのは、短期記憶過程の $\rho(h)$ が急激に減衰する一方、長期記憶過程の $\rho(h)$ は緩やかに減衰することによるものである。以下では、短期記憶過程と長期記憶過程の統計的性質を議論し、(1)式が成立する $\rho(h)$ の条件を調べる。

有限で一定の期待値 μ_x と有限で一定の分散 σ_x^2 を持つ時系列 x_1, x_2, \dots, x_n を考える。それが上記のような定常性を満たすことを仮定すると、平均値 \bar{x} の分散は、 $n \rightarrow \infty$ でゼロに収束する必要がある。具体的に平均値 \bar{x} の分散を求めると、この時系列には自己相関があるため、以下のようになる。

$$\text{var}[\bar{x}] = \frac{\sigma_x^2}{n^2} \sum_{s=1}^n \sum_{t=1}^n \rho(s, t) . \quad (2)$$

ここで、 $\rho(s, t)$ は x_s と x_t の自己相関である。 $\{x_t\}$ が定常過程であれば、自己相関は、 $\rho(s, t) = \rho(|t - s|) = \rho(h)$ として、時間差 $h = |t - s|$ の関数として表現される。すなわち、(2)式は次のように書き直すことができる。

$$\begin{aligned} \text{var}[\bar{x}] &= \frac{\sigma_x^2}{n^2} \sum_{s=1}^n \sum_{t=1}^n \rho(|t - s|) = \frac{\sigma_x^2}{n} \left[\rho(0) + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \rho(h) \right] \\ &= \frac{\sigma_x^2}{n} [1 + \delta_n(\rho)] . \end{aligned} \quad (3)$$

ただし、 $\delta_n(\rho)$ は次のようにおいた。

$$\delta_n(\rho) = 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \rho(h) . \quad (4)$$

$n \rightarrow \infty$ で $\delta_n(\rho)$ がある一定値に収束するとき、 $\text{var}[\bar{x}]$ は n^{-1} のオーダーでゼロに収束する。このような時系列が短期記憶過程である。これまでの議論を拡張し、自己相関の強いデータに適用し得るように、 $n \rightarrow \infty$ における $\text{var}[\bar{x}]$ を、次のように記述する。

$$\text{var}[\bar{x}] \approx \sigma_x^2 c(\rho) n^{-\alpha} . \quad (5)$$

$c(\rho)$ は定数であり、(2)式との比較から以下のように表せる。

$$c(\rho) = \lim_{n \rightarrow \infty} n^{\alpha-2} \sum_{s \neq t} \rho(s, t) . \quad (6)$$

定常性を満たすように(5)式で示される $\text{var}[\bar{x}]$ が $n \rightarrow \infty$ でゼロに収束する時系列を考えると、 α は $0 < \alpha \leq 1$ の値をとることがわかる⁹。 $\alpha = 1$ であれば前述のように短

9 $\alpha > 1$ の場合は、短期記憶過程よりも自己相関が弱い時系列を意味することになるが、この場合には矛盾が生じる。なぜならば、最も自己相関が弱い時系列(1以上のラグでは全て無相関)を考えても、 $\text{var}[\bar{x}] = \sigma_x^2/n$ となり、短期記憶過程($\alpha = 1$)として記述されるためである。

期記憶過程である。自己相関が強ければ $0 < \alpha < 1$ になり、これが長期記憶過程である。つまり、長期記憶過程は、 $\text{var}[\bar{x}]$ の収束速度が n^{-1} のオーダーよりも遅い時系列である。

(6)式より、 $c(\rho)$ が一定値に収束するときには、 $\Sigma\rho(s, t)$ が $n^{2-\alpha}$ のオーダーに等しいことがわかる。 $\Sigma\rho(h)$ のオーダーは $n^{-1} \Sigma\rho(s, t)$ のオーダーに等しいことから、長期記憶過程の場合、 $\Sigma_{h=-(n-1)}^{n-1} \rho(h)$ は $0 < \alpha < 1$ として $n^{1-\alpha}$ のオーダーである。これを考慮すると、適当な定数 c_ρ を用いて、長期記憶過程の自己相関関数は次式のように、べき関数で減衰することがわかる。

$$\rho(h) = c_\rho |h|^{-\alpha}, \quad 0 < \alpha < 1. \quad (7)$$

(7)式のように緩やかに減衰する自己相関関数が(1)式を満たすことは、容易に確認される。

以下では、(7)式のように自己相関関数が緩やかに減衰する時系列を表現するモデルとして、フラクショナル・ガウシアン・ノイズ過程とフラクショナル・インテグレーション過程を説明する。

(3) フラクショナル・ガウシアン・ノイズ過程

はじめに、フラクショナル・ブラウン運動を説明する。フラクショナル・ブラウン運動の概念は、ブラウン運動のスケール不変性を拡張したものである。簡単化のため、離散時間で、原点を出発したブラウン運動に従う確率過程 $\{Y_t\}$ の変化を考える。時点 t までの各時点で生じた変化を X_s とすると、時点 t での確率過程 $\{Y_t\}$ の値は、以下のように表される。

$$Y_t = \sum_{s=1}^t X_s. \quad (8)$$

さて、任意の2時点 a と b ($a < b$) 間の、増分 $Y_b - Y_a$ に注目する。 $Y_b - Y_a$ の分布は正規分布で表現されるので、 $Y_b - Y_a$ の平均 m と分散 σ^2 は、それぞれ以下ようになる。

$$m = E[Y_b - Y_a] = 0, \quad (9)$$

$$\sigma^2 = E[(Y_b - Y_a)^2] = A^2(b-a). \quad (10)$$

なお、 A は定数であり、(10)式は次のように表現することもできる。

$$\sigma \propto (b-a)^{1/2}. \quad (11)$$

すなわち、 $Y_b - Y_a$ の標準偏差 σ と、時間間隔 $b-a$ の間にはスケール不変の関係がある。このスケール不変の関係を、一般的な形に拡張したものが、次の関係式である。

$$E[(Y_b - Y_a)^2] = A^2 |b - a|^{2H}. \quad (12)$$

上記(12)式の H は、ハースト指数と呼ばれる。ハースト指数の値域は、 $0 < H < 1$ である¹⁰。 Y_t の増分が(12)式で表現されるとき、 $\{Y_t\}$ をフラクショナル・ブラウン運動と呼ぶ。なお、 $H = 1/2$ のときは、 $\{Y_t\}$ は通常のブラウン運動となる。

フラクショナル・ブラウン運動の階差をとり、 $X_t \equiv Y_t - Y_{t-1}$ によって定義される $\{X_t\}$ を、フラクショナル・ガウシアン・ノイズ (Fractional Gaussian Noise : 以下、FGN) 過程と呼ぶ。(9)式の定義により、 X_t の期待値はゼロである。また、 $Y_0 = 0$ と(12)式より、 $a > 0$ で $E[Y_a^2] = A^2 a^{2H}$ となることから、再び(12)式を用いて、

$$E[Y_a Y_b] = \frac{A^2}{2} (a^{2H} + b^{2H} - |b - a|^{2H}), \quad (13)$$

を得る。(12)式より、 X_t の分散は A^2 で与えられることから、時間差を $h \equiv b - a$ とくと、(13)式を用いて X_t の自己相関関数 $\rho(h)$ は以下のように計算される。

$$\begin{aligned} \rho(h) &= \frac{\text{Cov}(X_a, X_b)}{\sqrt{\text{var}(X_a)} \sqrt{\text{var}(X_b)}} = \frac{E[X_a X_b]}{A^2} = \frac{1}{A^2} E[(Y_a - Y_{a-1})(Y_b - Y_{b-1})] \\ &= \frac{1}{A^2} \{E[Y_a Y_b] - E[Y_{a-1} Y_b] - E[Y_a Y_{b-1}] + E[Y_{a-1} Y_{b-1}]\} \\ &= \frac{1}{2} \{(h+1)^{2H} - 2h^{2H} + (h-1)^{2H}\}. \end{aligned} \quad (14)$$

自己相関関数 $\rho(h)$ は、時間差 h のみに依存しており、 X_t は定常性を満たすことが確認される。(14)式は、 $\rho(h)$ が h の関数 h^{2H} の2階階差で表現されることを表しており、 $h \rightarrow \infty$ では

$$\rho(h) \propto h^{2H-2}, \quad (15)$$

となる。(7)式との比較から、FGN過程は、 $1/2 < H < 1$ のとき長期記憶過程になることがわかる。

.....
¹⁰ $H=0$ であれば、時間間隔と無関係に $E[(Y_b - Y_a)^2]$ は一定値をとる。また、 $H < 0$ であれば時間間隔が大きくなるほど、 $E[(Y_b - Y_a)^2]$ は減衰することになるが、現実的にこのような振る舞いをする確率過程は考えにくい。また、(14)式より、 $H=1$ では $\rho(h) \equiv 1$ となる。さらに、(15)式より、 $H > 1$ では $h \rightarrow \infty$ で $\rho(h) = \infty$ となる。このため、ハースト指数は $0 < H < 1$ の値をとる。

(4) フラクショナル・インテグレーション過程

次に、自己相関関数が緩やかに減衰していくデータを表現する別のモデルとして、フラクショナル・インテグレーション過程を取り上げる。時系列 $\{x_t\} (t = 1, 2, \dots)$ をフラクショナル・インテグレーション過程として記述すると、実数 d を用いて以下のようになる。

$$(1-L)^d x_t = \epsilon_t \quad (16)$$

L はラグ・オペレータであり、 $L^h x_t = x_{t-h}$ となる。また、 ϵ_t はホワイト・ノイズである。なお、 $(1-L)^d$ は、以下の無限級数で記述される。

$$(1-L)^d = 1 - dL - \frac{d(1-d)}{2} L^2 - \frac{d(1-d)(2-d)}{3 \cdot 2} L^3 - \dots \quad (17)$$

以下では、フラクショナル・インテグレーション過程を $I(d)$ 過程と記述する。

ここで、 $I(d)$ 過程のパラメータ d のとりうる値はどのような範囲にあるかを考える。既述のように、本稿では、長期記憶過程として、定常、自己相関に持続性のある過程であると定義している。データが定常であるという条件から、 d のとり得る値は $-1/2 < d < 1/2$ となる (Granger and Joyeux [1980], Hosking [1981])。

さて、(16)式で記述された時系列の自己相関関数は、以下のように表せる。

$$\rho(h) = \frac{\Gamma(1-d)\Gamma(h+d)}{\Gamma(d)\Gamma(h+1-d)} \quad (18)$$

ここで、 $\Gamma(\cdot)$ はガンマ関数である。ガンマ関数では、 h が十分に大きいとき、以下の近似式が成り立つ (スターリングの公式)。

$$\frac{\Gamma(h+a)}{\Gamma(h+b)} \approx h^{a-b} \quad (19)$$

したがって、 $I(d)$ 過程の自己相関関数は、 $h \rightarrow \infty$ で、以下のように近似される。

$$\rho(h) \approx \frac{\Gamma(1-d)}{\Gamma(d)} |h|^{2d-1} \quad (20)$$

ここで、FGN過程で算出された(15)式と比較すると、 $I(d)$ 過程のパラメータ d とFGN過程のハースト指数 H の間には、以下の関係が成立することがわかる。

$$d = H - 1/2 \quad (21)$$

したがって、 d が $0 < d < 1/2$ の範囲にあるとき、(16)式で表現される $I(d)$ 過程 $\{x_t\}$ は長期記憶過程となる。また、 d がゼロに近ければ短期記憶過程に近く、 d が $1/2$ に近づくほど長期記憶の程度が大きいことになる。

3. 長期記憶パラメータの推定

前節では、長期記憶過程を適切に表現するモデルとして、FGN過程とI(d)過程をみてきた。FGN過程のハースト指数Hと、I(d)過程のパラメータdの間には、(21)式で示したように、 $d=H-1/2$ という関係が成立する。そこで、以下では、I(d)過程のみを用いて議論する。ある時系列がI(d)過程に従うとすると、推定されたパラメータdが $0 < d < 1/2$ の範囲にあれば、その時系列は長期記憶過程であることがわかる。このdを推定する方法は数多く提唱されているが、本稿では、スペクトル密度を用いた手法(フーリエ解析)とウェーブレット分散を用いた手法(ウェーブレット解析)を説明する¹¹。

(1) フーリエ解析を用いたパラメータの推定

I(d)過程の近似的な自己相関関数(20)式にフーリエ変換を施すことにより、スペクトル密度 $S(f)$ は以下ようになる。

$$S(f) = \sum_{h=-\infty}^{\infty} \rho(h) e^{-i2\pi fh} \approx c_d |f|^{-2d}. \quad (22)$$

ここで、 f は周波数を表し、 $-1/2 < f < 1/2$ の範囲にある。また、 c_d は定数である。(22)式は、データが長期記憶性を保有している場合($0 < d < 1/2$)、スペクトル密度が $f \rightarrow 0$ で発散することを意味している。つまり、スペクトル密度の原点付近での振る舞いを調べることで、時系列が長期記憶性を持つかどうかを判定することが可能である。dは、(22)式の両辺の対数をとった以下の式を用いて推定可能である。

$$\ln S(f) = \text{const.} + (-2d) \ln f. \quad (23)$$

しかし、現実のデータは有限長であるため、(22)式にあるような無限和をとることは不可能である。したがって、スペクトル密度の推定値として、次のペリオドグラム $\hat{I}(f_k)$ を用いる。

$$\hat{I}(f_k) = \frac{1}{N} \left(\sum_{t=1}^N (x_t - \bar{x}) e^{-i2\pi f_k t} \right)^2. \quad (24)$$

ただし、 N は $\{x_t\}$ の長さ、 \bar{x} は $\{x_t\}$ の平均、 f_k は $f_k = k/N$ (k は整数で $-1/2 < f_k < 1/2$)である。(24)式により $\hat{I}(f_k)$ を求めた後で、次式により最小2乗法等を用いて、 d を推定すればよい。

11 その他の長期記憶性の推定手法の詳細は、Beran [1994]、矢島 [2003] 等を参照されたい。

$$\ln \hat{I}(f_k) = \text{const.} + (-2d) \ln f_k . \quad (25)$$

ここで、 $k = 1, \dots, k_{\max}$ である。Geweke and Porter-Hudak [1983] では、 $I(d)$ 過程のスペクトル密度を近似せずに導出し、 d の推定に以下の式を用いている。

$$\ln \hat{I}(f_k) = \text{const.} - d \ln(4 \sin^2(\pi f_k)) . \quad (26)$$

(26)式より推定された d は、GPH推定量 (GPH estimator) と呼ばれる。ペリオドグラムの原点付近では $\sin(\pi f_k) \approx \pi f_k$ が成り立つため、(25)式は、(26)式の近似とみなせる。

(25)式により d を推定するためには、 k_{\max} をどのように選ぶかという問題がある。ペリオドグラムは、スペクトル密度の漸近不偏推定量ではあるが、一致推定量ではないことが知られている。したがって、 k_{\max} の選び方により d の推定値が異なることも考えられる。Lardic, Mignon, and Murtin [2003] によると、GPH推定量を計算する際、通常は、 k_{\max} は $N^{0.5}$ 程度としている。しかし、Hurvich, Deo, and Brodsky [1998] では、 d の真値とGPH推定量の差に関する期待値が最小となる k_{\max} を求め、最適な k_{\max} は $N^{0.8}$ であると報告している。後述の図表8で示すように、ペリオドグラムの分布は、ばらつきが非常に大きく、サンプル・サイズ次第で d の推定結果が大きく異なる。そのため、ペリオドグラムを用いた長期記憶パラメータの推定は、 k_{\max} の選び方という恣意性が入り込む手法であるといえる。

(2) ウェーブレット解析を利用したパラメータ推定

ここでは、ウェーブレット分散を用いた長期記憶パラメータの推定手法を説明する¹²。

Percival and Walden [2000] にならひ、レベル j に対応するスケール τ_j を、次のように定義する。

$$\tau_j \equiv 2^{j-1} . \quad (27)$$

サポート長 L のフィルタを用いたDWT (Discrete Wavelet Transform) によって計算される、レベル j のウェーブレット係数 $\{w_{j,t}\}$ の調整DWT分散 $\hat{\sigma}^2(w_j)$ は、以下で与えられる。

12 ウェーブレット分散をはじめ、ウェーブレット解析の基礎的概念は、補論を参照されたい。

$$2^j \hat{\sigma}^2(w_j) = \frac{\sum_{t=L_j}^{N/2^j} w_{j,t}^2}{N/2^j - L_j + 1} \quad (28)$$

ここで、 N は原データのサイズ（データ数）、 L_j は $(L-2)(1-2^{-j})+1$ 以上の最小の整数として定義される。

(28)式の分散の計算方法の概要を解説する。ここでは、通常の分散の計算と同様、ウェーブレット係数 $\{w_{j,t}\}$ の2乗平均を考えている。DWTでは、分解レベルが上がるとウェーブレット係数が半減するため、(28)式の左辺の 2^j という調整が必要となる。また、サポート長が2よりも大きいウェーブレット・フィルタでは、係数の端で始点と終点を接続したデータの影響を受けるため、 L_j という変数を導入して、端の係数を分散の計算から除去している（詳細は補論を参照）。

以下では、調整DWT分散がスペクトル密度 $S(f)$ とどのような関係にあるかをみる。

まず、フーリエ解析の議論から、時系列 $\{x_t\}$ の分散 σ_x^2 とスペクトル密度 $S(f)$ の間には、以下の関係がある。

$$\sigma_x^2 = \int_{-1/2}^{1/2} S(f) df = 2 \int_0^{1/2} S(f) df \quad (29)$$

(29)式は、全スペクトル密度の積分が原系列の分散に等しいことを意味している。別の解釈をすれば、原系列の分散は各周波数成分の強さに分解可能であるといえる。

次に、 $\{x_t\}$ が無数のレベルまでのウェーブレット分解が可能であれば、原系列の分散 σ_x^2 と調整DWT分散の関係は、次式で表せる。

$$\sigma_x^2 = \sum_{j=1}^{\infty} \hat{\sigma}^2(w_j) \quad (30)$$

(30)式は、調整DWT分散の和が原系列の分散に等しいことを意味している。逆にいうと、原系列の分散は各レベルの調整DWT分散に分解可能であると解釈される¹³。スペクトル密度もウェーブレット分散も、原系列の分散を周波数やレベル（スケール）に応じて分解しているという点で、類似の概念である。

ここで、ウェーブレット係数というレベルの概念と、スペクトル密度という周波数の概念が、どのような対応関係にあるかを考察する。DWTを1回実行することによって、データは高周波成分と低周波成分に分割される。まず、レベル1のウェーブレット係数が持つ情報は、高周波成分の情報であり、そのスケール（実質的な

13 この関係が成り立つことは、エネルギー保存則から導かれる。エネルギー保存則は、補論(3)節を参照されたい。

データ間隔)は原系列の2倍である¹⁴。したがって、レベル1のウェーブレット係数は、周波数 f との対応でいえば $1/4(=1/2^2)$ から $1/2$ までの情報を含んでいる¹⁵。これから、レベル1のウェーブレット分散は、周波数が $1/4$ から $1/2$ までのスペクトル密度と同等の情報を含んでいる。同様に、レベル2のウェーブレット係数は、周波数 f が $1/2^3$ から $1/2^2$ までの情報を含んでいる。このように、あるレベルのウェーブレット分散は、そのレベルに対応するスペクトル密度と等しい情報を保有しているとみなせる。つまり、スペクトル密度を、対応する分解レベルに応じて集計したものが、ウェーブレット分散である。

このことから、調整DWT分散とスペクトル密度の間には、以下の近似式が成立する¹⁶。

$$\hat{\sigma}^2(w_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S(f) df . \quad (31)$$

$I(d)$ 過程のスペクトル密度である(22)式の $S(f)$ を(31)式に代入すれば、調整DWT分散は、(27)式で定義した τ_j を用いて以下で計算される。

$$\hat{\sigma}^2(w_j) \propto \tau_j^{2d-1} . \quad (32)$$

(32)式の両辺の対数をとることで、次式が得られる。

$$\ln \hat{\sigma}^2(w_j) = \text{const.} + (2d-1) \ln \tau_j . \quad (33)$$

ただし、 $j=1, \dots, j_{\max}$ である。(33)式を用いることで、ウェーブレット解析を用いた d の推定が可能となる。

調整DWT分散を利用した(33)式から d を推定する場合でも、 j_{\max} をどのように選ぶか、あるいはウェーブレット・フィルタとして何を用いるかという恣意性は依然として残る。しかし、 j_{\max} をどう選ぶかという問題は、データの制約から、フーリエ解析での k_{\max} の選び方ほど恣意性は大きくないと考えられる。なぜならば、長期記憶性は低周波領域で顕著に表れるため j_{\max} は大きく選ぶことが望ましいが、ウェーブレット分散を計算するのに十分なウェーブレット係数を確保するには j_{\max} は小さい方が望ましいというトレード・オフがあるからである。したがって、上記とを同時に満たす j_{\max} は、原系列の個数からある程度制限される。

14 残り半分の情報はレベル1のスケール係数が保有している。

15 レベル1では、周波数 f が0から $1/4$ の情報はスケール係数に含まれる。

16 ウェーブレット・フィルタが理想的な高域通過フィルタであれば、(31)式は、近似式ではなく等式として成り立つ。代表的なウェーブレット・フィルタの周波数特性は、稲田・鎌田 [2004]、Gençay [2002]、Percival and Walden [2000]等を参照されたい。

この点を少し詳しく説明する。例えば、原系列は8,192 (=2¹³) 個であるとする。これは、最大で13回のウェーブレット変換を施せることを意味している。したがって、理論上は13個の異なるレベルごとにウェーブレット係数が計算される。しかし、レベル13のウェーブレット係数の個数は1であり、レベル12、11のウェーブレット係数の個数はそれぞれ2、4である。分散を計算するためには、こうした少数のデータでは不足である。常識的には、分散を計算するには10を超えるデータ数が必要であると考えられる。以上から、この場合、8,192=16×2⁹より、 j_{\max} は9とするのが妥当であると思われる。分散を計算するためのデータ数が16では不足と考えれば、8,192=32×2⁸より、 j_{\max} を8とするのが妥当である。このように、ウェーブレット解析では、レベルの選択に当たり、恣意性はわずかに残っているが、フーリエ解析ほどの恣意性はないと考えられる。

4 . 金融時系列と長期記憶性

前節で説明したように、ある時系列が長期記憶過程に従っているかどうかを判定する方法として、スペクトル密度やウェーブレット分散を用いる方法がある。本節では、これら2つの方法のどちらが優れた長期記憶性の判定方法であるか、恣意性の問題や推定精度の面から考察し¹⁷、現実のデータ（株価収益率や円/ドル・レートの日次変化率）に長期記憶性が存在するかを調べる。

(1) 人工的に作成したデータに対する長期記憶性の判定

現実のデータを扱う前に、人工的に作成したデータを用いて、これまで説明してきた長期記憶性の判定方法が、どれだけ有効であるかを確認する。人工的に生成するデータは、ホワイト・ノイズ ((34)式)、 $d = 0.2$ のI(d)過程((35)式)の2種類である。

$$x_t = \epsilon_t, \quad (34)$$

$$(1-L)^d x_t = \epsilon_t. \quad (35)$$

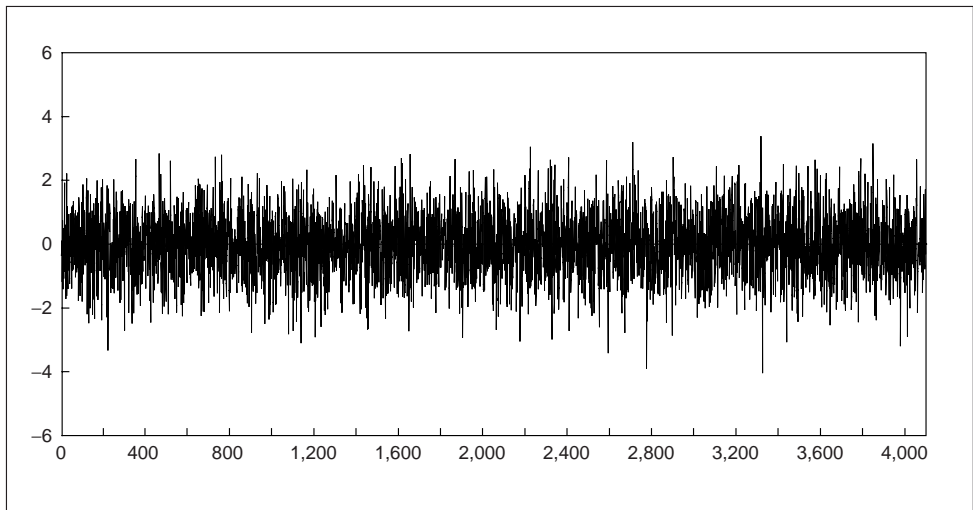
ここで、 ϵ_t は平均0かつ分散一定の標準正規乱数である。データは、4,096個を生成する。なお、長期記憶性のある時系列を表現する、より一般的なモデルとして、

17 I(d)過程のパラメータdを推定する方法として、ウェーブレット解析とフーリエ解析を比較した先行研究に、Jensen [1999]がある。この研究では、フーリエ解析を用いた手法としてGeweke and Porter-Hudak [1983]が提唱した手法を用いている((25)式ではなく(26)式を用いている)。同研究でも、本稿と同様に、ウェーブレット解析の方が高精度でdを推定できるとの結論を得ている。

ARFIMA (Autoregressive Fractional Integration Moving Average) 過程がある。本稿の直接の目的は、金融時系列に長期記憶性があるかどうかを確認することであって、より当てはまりのよいモデルを探索することではない。したがって、(35)式で与えられる $I(d)$ 過程のパラメータ d を推定することで、十分に長期記憶性の有無を判別することができる。本稿で示した(25)式や(33)式の手法であれば、AR過程やMA過程の効果の影響を受けることなく、長期記憶性の有無を判断可能である¹⁸。(34)式と(35)式に従うデータの動きを視覚的に比べるため、図表4にホワイト・ノイズ、図表5に $I(0.2)$ 過程の動きを示す。

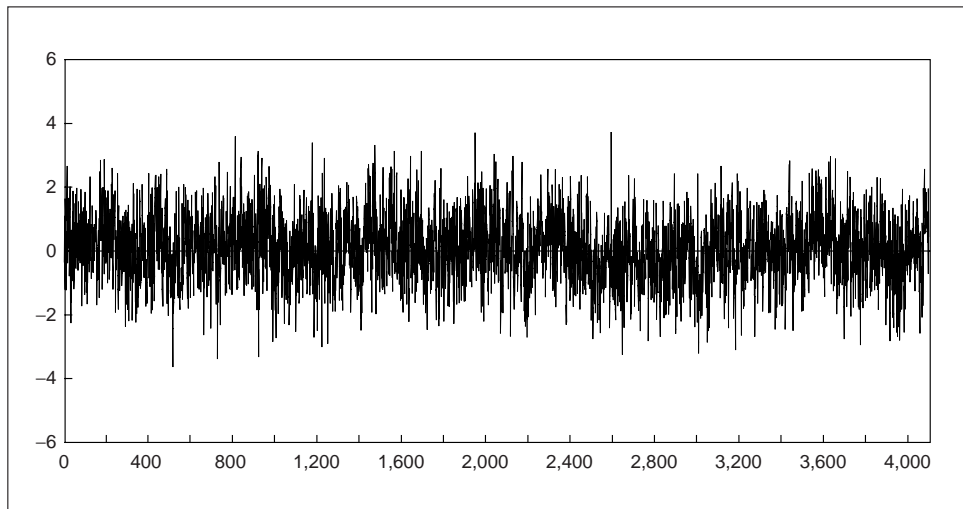
上記2つのデータを事前の情報なく観察しても、両者の統計的な違いがどこにあるのか区別することは難しい。視覚的には統計的性質をほとんど区別することができない2つのデータに、フーリエ解析をベースとする(25)式(ペリオドグラムを利用)、ウェーブレット解析をベースとする(33)式(調整DWT分散を利用)を用いて、それぞれ d を推定する。 d の推定には最小2乗法を用いる。先述のように、フーリエ解析では、推定に利用する原点付近のサンプル・サイズをどう設定するかという恣意性が存在する。また、ウェーブレット解析でも、数多く存在するウェーブレット・フィルタのうち、どのフィルタを用いるか、という恣意性がある。こうした恣意性が分析結果にどのような影響を与えるかを調べるため、異なるサンプル・サイズ、異なるウェーブレット・フィルタを用いて、 d を推定した。サンプル・サイズは、 $4,096^{0.5}=64$ 、 $4,096^{0.8}\approx 776$ となるため、ここでは、55、60、65、70、775の5通

図表4 ホワイト・ノイズ



18 AR過程やMA過程により生じる変動は、高周波領域のスペクトル密度に反映される。したがって、周波数領域の原点付近(低周波領域)のデータであれば、AR過程やMA過程の影響を受けずに、長期記憶パラメータ d を推定することが可能である。このことは、長期記憶性の有無を検定するに当たり、モデルの特定化が重要ではないことを示している。

図表5 I(0.2) 過程



りとした。また、ウェーブレットの種類は、最も扱いやすいハール・ウェーブレットに加え、サポート長が4と12のドビッシェー・ウェーブレットを採用した¹⁹。各設定での d の推定結果を以下に示す（図表6、7）。

ウェーブレット解析では、レベル7までの調整DWT分散を用いた。これは、 $4,096=16 \times 2^8$ である一方で、 $D(4)$ ウェーブレットや $D(12)$ ウェーブレットのように、サポート長が2より長いウェーブレットでは、レベル8のウェーブレット分散の計算のためのデータ数が16を下回るためである。レベル7であれば、 $D(12)$ ウェーブレットでも、ウェーブレット分散の計算に十分なデータ数が確保される。

フーリエ解析を用いた推定結果（図表6）によると、ホワイト・ノイズのときは、サンプル・サイズによらず d の推定値はゼロに近い値をとる。また、 $d=0$ の帰無仮説は棄却されない。したがって、ホワイト・ノイズを長期記憶過程と誤って判断する可能性は極めて低いと考えられる。一方、 $I(0.2)$ 過程では、 $d=0$ の帰無仮説は棄却されるが、 d の値はサンプル・サイズで異なるうえ、真値（0.2）から大きく乖離している。真値に最も近いサンプル・サイズ775の場合でも、真値からは25%乖離している。このように、フーリエ解析による $I(d)$ 過程のパラメータ推定では、原点付近のサンプル・サイズによって d の推定値が異なるうえ、真の値を得ることが困難である。

次に、ウェーブレット解析による推定方法（図表7）では、ハール・ウェーブレット、 $D(4)$ ウェーブレット、および $D(12)$ ウェーブレットの全てで d の値が、ほぼ正確に推定された。ホワイト・ノイズの推定では、 d は0に近い値となり、 $d=0$ の帰無仮説も棄却されない。よって、フーリエ解析と同様に、ホワイト・ノイズを長期記

19 サポート長が L のドビッシェー・ウェーブレットを $D(L)$ と表記する。

図表6 フーリエ解析による人工データのパラメータ推定

I(0) 過程		I(0.2) 過程	
サンプル・サイズ	d	サンプル・サイズ	d
55	-0.039 (0.090)	55	0.374** (0.077)
60	-0.048 (0.084)	60	0.415** (0.080)
65	-0.039 (0.079)	65	0.370** (0.077)
70	-0.060 (0.074)	70	0.360** (0.074)
775	-0.025 (0.023)	775	0.250** (0.023)

備考：括弧内は標準誤差。**1%有意水準。

図表7 ウェーブレット解析による人工データのパラメータ推定

I(0) 過程		I(0.2) 過程	
	d		d
ハール	-0.015 (0.015)	ハール	0.216** (0.021)
D(4)	-0.016 (0.009)	D(4)	0.207** (0.016)
D(12)	0.010 (0.010)	D(12)	0.171** (0.020)

備考：括弧内は標準誤差。**1%有意水準。

憶過程であると判断する可能性は極めて低いと考えられる。また、I(0.2)過程の推定では、 d の値は0.2に近く（誤差は高々15%）、 $d = 0$ の帰無仮説は、1%有意水準で棄却されている²⁰。つまり、ウェーブレット解析を用いた方法では、ウェーブレットの種類によらず、ホワイト・ノイズと長期記憶過程の違いを区別したうえで、 d の値をほぼ正確に求められることになる。

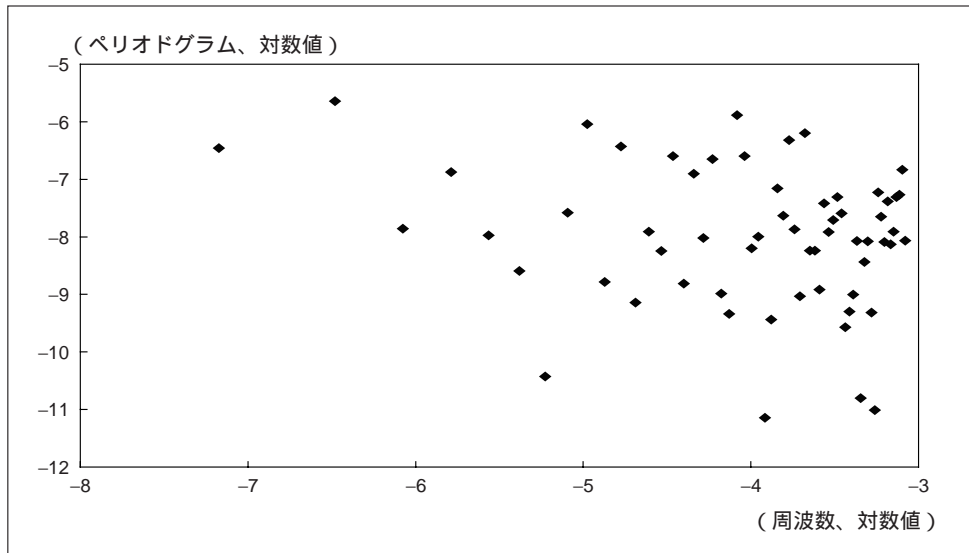
先述のように、フーリエ解析を用いた方法もウェーブレット解析を用いた方法も、本質的には同等の概念である。では、どのような理由で、ウェーブレット解析によ

20 ハール・ウェーブレットの場合には、レベル8でもウェーブレット分散を計算するのに十分なデータが得られる。そこで、ハール・ウェーブレットを用いてレベル8までの調整DWT分散を計算し、 d の値を推定した。その結果、ホワイト・ノイズでは、-0.035という値が得られ、 $d = 0$ の帰無仮説は棄却されなかった。また、I(0.2)過程では、0.206という値が得られ、 $d = 0$ の帰無仮説は1%有意水準で棄却された。ウェーブレット解析では、分解するレベルの程度を変えてもほぼ同様の結果を得たことから、レベルの選択に関する恣意性は、フーリエ解析のサンプル・サイズの恣意性ほど大きな問題ではないといえる。

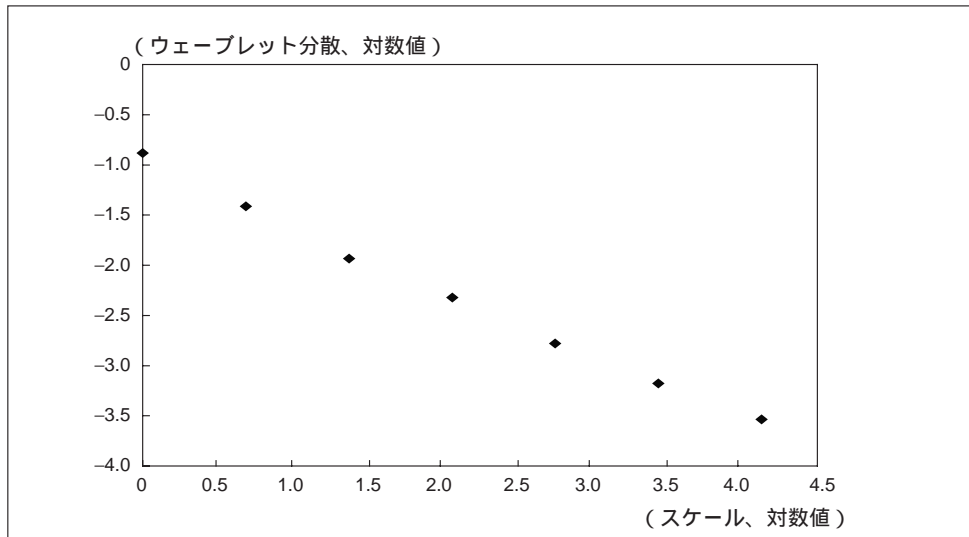
る推定精度の方が高いのであろうか。その理由はペリオドグラムとウェーブレット分散のの違いにある。まず、先述のように、スペクトル密度の推定値であるペリオドグラムは、漸近的不偏性はあるも一致性を持たない。そのため、ペリオドグラムの対数プロットは、図表8のようにばらつき、サンプル・サイズ次第で、 d の推定値は大きく異なる。

一方、ウェーブレット分散の対数プロットは、図表9のように、視覚的に、ほとんど直線とみなすことができる。これは、ウェーブレット分散が、(31)式で表現

図表8 (0.2) 過程のペリオドグラム



図表9 (0.2) 過程のウェーブレット分散



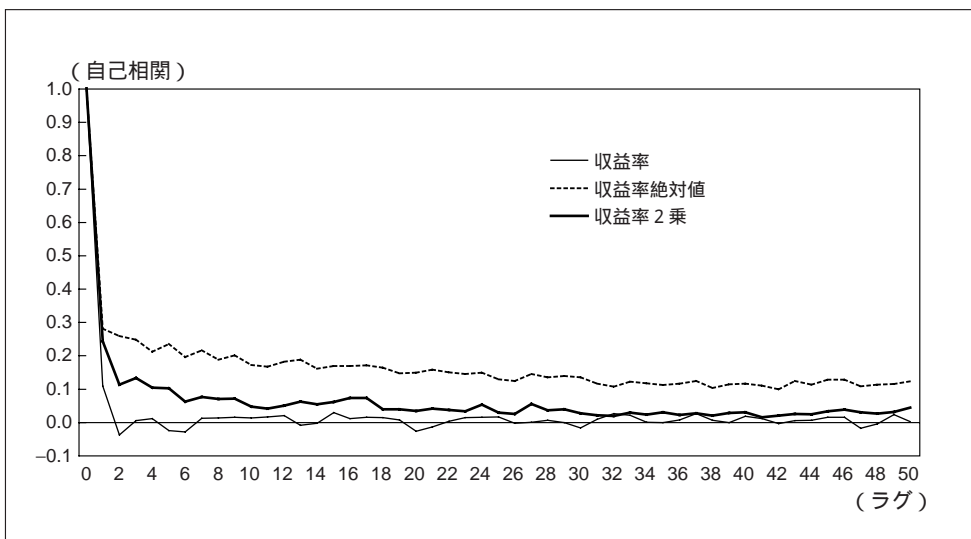
されるように、スペクトル密度の集計値で与えられるためである。つまり、ペリオドグラムを均すことで、図表8のようなばらつきを抑えることが可能になるのである。

(2) 金融時系列の長期記憶性

これまでみてきたように、ウェーブレット解析（調整DWT分散）を用いることで、 $I(d)$ 過程のパラメータ d をほぼ正確に求められることがわかった。そこで、現実のデータとして、株価収益率（前掲図表1）が長期記憶過程に従うかどうかを、ウェーブレット解析を用いて調べる。繰り返しになるが、ある時系列が長期記憶性を保有しているかどうかを視覚的に調べる方法は自己相関関数をみることである。株価収益率の自己相関関数をみると、非常に早くゼロに収束しており、株価収益率は、長期記憶過程には従わず、 $I(0)$ 過程とみなすことが妥当であるように思える（図表10）²¹。しかし、2節でみたように、KPSSテストは、株価収益率が $I(0)$ 過程に従うという帰無仮説を棄却している。

そこで、株価収益率を $I(d)$ 過程に当てはめ、ハール・ウェーブレットを利用して d を推定した。ハール・ウェーブレットを用いた理由は、以下のとおりである。DWTを実行する際、サポート長が短いほど計算時間は短縮されるほか、サポート長が短いほど原データのなかで分析に使えないデータの比率が小さくなる。また、

図表10 株価収益率の自己相関関数



21 むしろ、収益率絶対値や収益率2乗の自己相関関数は減衰が緩やかであることから、ボラティリティに長期記憶性が含まれていると考えられる。

人工的なデータを生成し、複数のウェーブレットを用いてI(d)過程のdを推定した結果、ほぼ同一の結果が得られた。したがって、ウェーブレットの選択はdの推定にほとんど影響を与えないと考え、サポート長が2であるハール・ウェーブレットを選択した。

現実の金融時系列として、非常に長期のデータを得ることが可能である、TOPIX収益率(サンプル期間:1970年6月4日~2005年7月29日、サンプル・サイズ:8,704 $\leq 17 \times 2^9$)を取り上げ、長期記憶パラメータdを(33)式により推定すると、 $d = 0.040$ という値が得られた。この標準誤差は0.020であり、この結果に基づけば、株価収益率に有意な長期記憶性を認めることはできないことになる。しかし、ここでの長期記憶パラメータdの計算では、金融時系列の性質が30年近い長期間不変であることを仮定している。しかし、約30年の超長期のデータのなかには、構造変化や時系列の本質的な変動とは異なる大きなノイズも含まれており、この仮定が実際に成立しているかは疑わしい。

こうしたことから、dを推定する際、構造変化やノイズの影響を軽減するため、サンプル期間を短い期間に区切ってdの変化を調べる。

しかし、(33)式を用いてパラメータdを推定するには、できるだけ多くのサンプルを用いる必要がある。つまり、ある程度高いレベルまでウェーブレット分散を計算しなければならない。その一方で、サンプル数が少ないと高レベルのウェーブレット分散を計算することができない。このように、サンプル・サイズの選択には、構造変化やノイズの影響回避という問題と、なるべく多くのウェーブレット分散を計算する必要性という、トレード・オフがある。

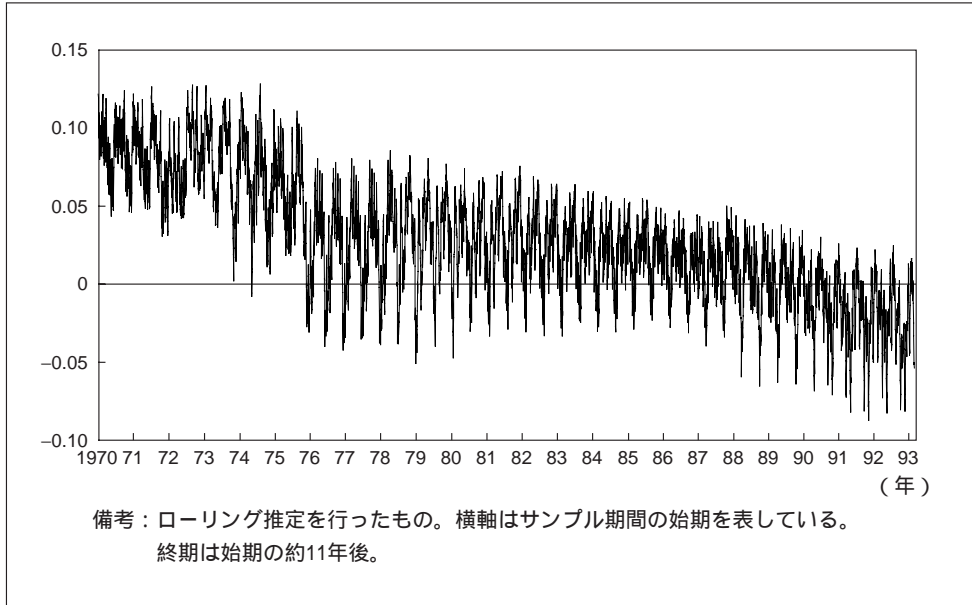
そこで、このトレード・オフに対処するために、サンプル・サイズを2,944 ($= 23 \times 2^7$)としてローリング推定を行うことにした。このサンプル・サイズであれば、レベル7までのウェーブレット分散を計算することができるため、(33)式を用いたパラメータdの推定が可能である。また、サンプルの対象となる期間も11年程度であるため、構造変化がどこかの時点で生じていた場合、長期記憶パラメータdの違いとして捉えることが可能となる²²。図表11が推定の結果である。

図表11をみると、長期記憶パラメータdは趨勢的に低下していることがわかる。また、短期的にも大きな変動が観察される。このように、サンプル期間を少し変えただけでdの値が大きく変化するという現象は、以下で述べるような異常値が存在することで発生する。

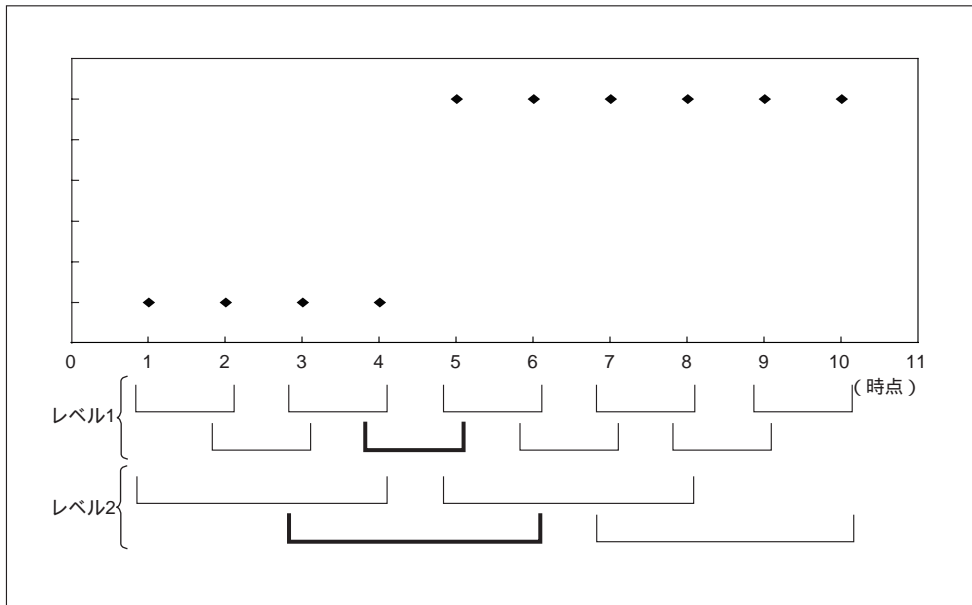
株価収益率の推移(前掲図表1)をみると、まれに大きな変動が観察される。こうした大きな変動を、本稿では異常値と呼ぶ。まず、異常値の代表例として、図表12のような階段状の異常値を考える。階段状の異常値は、大局的にみれば短期の変動

22 もっとも、構造変化は、ある特定の時点を境に急激な変化が生じるだけでなく、徐々に変化していくこともある。後者の場合、明確に構造変化の発生時点を特定することはできないが、大雑把に、1980年代と1990年代を比較し、時系列の性質に変化があるかどうかの判別はできる。

図表11 株価収益率の長期記憶パラメータ(1)



図表12 階段状の異常値の検出



というより、長期に亘る変動といえる。したがって、階段状の異常値は、ウェーブレット係数の高レベル（低周波成分）で検出されることが望ましい。しかし、サンプル期間の設定によっては、レベル1やレベル2の低レベルのウェーブレット係数で、階段状の異常性が検出されることもある。以下、この点を説明する。図表12には、簡単化のため10個の時点のみを掲げている。

補論(1)節で説明しているハール・ウェーブレットの形状を勘案すると、レベル1のウェーブレット係数は、連続する2つの原データにおける差を表現している。つまり、レベル1のウェーブレット係数列は、隣り合う2つの原データの組から構成されている。係数列は、始点と終点（サンプル期間）の選び方によって、必ず、次の2つの群のどちらかに分類される。

$$\dots, \{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7, x_8\}, \{x_9, x_{10}\}, \dots \quad (36)$$

$$\dots, \{x_2, x_3\}, \{x_4, x_5\}, \{x_6, x_7\}, \{x_8, x_9\}, \dots \quad (37)$$

図表12の階段状のデータ（ x_4 と x_5 の間に段差が発生）では、(36)式の組合せとなるようサンプル期間を選択した場合、レベル1で異常性を検出することができない（全ての組合せの差分がゼロになる）。しかし、(37)式の組合せとなるようサンプル期間を選択すると、レベル1で異常性を検出することができてしまう。そして、レベル1で異常性を検出した後は、より高いレベルで異常性を捉えることはできない。

次に、レベル1で異常値を検出することができなかった(36)式の組合せで、レベル2のウェーブレット係数を計算するときの組合せは、次の2つである。

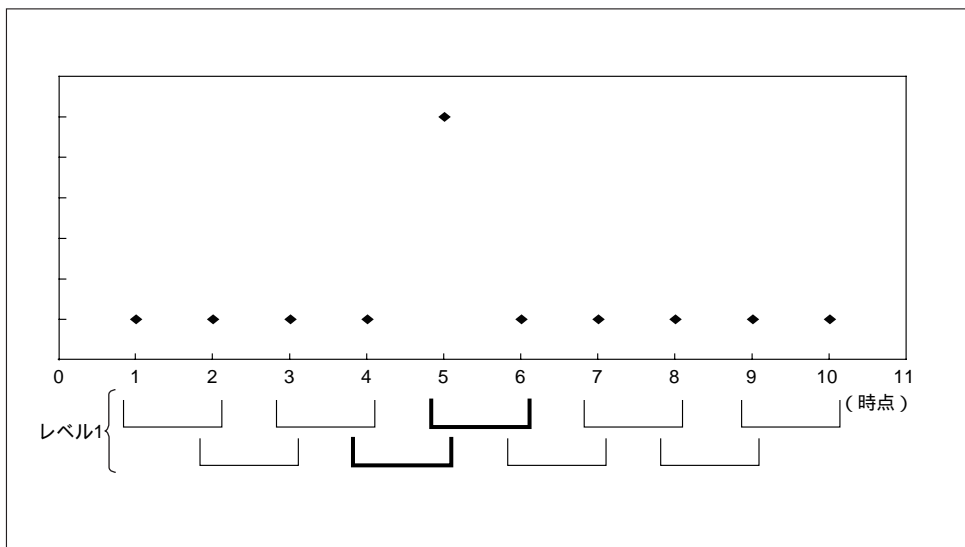
$$\dots, \{x_1, x_2, x_3, x_4\}, \{x_5, x_6, x_7, x_8\}, \dots \quad (38)$$

$$\dots, \{x_3, x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}\}, \dots \quad (39)$$

この場合も、(38)式の組合せでは異常性を検出することができないが、(39)式の組合せでは異常性を検出することができる。このように、階段状の異常値が存在するとき、サンプル期間のわずかな違いが、異常性を検出するレベルに大きな影響を与える。ウェーブレット係数が大きな値をとれば、そこから計算されるウェーブレット分散も大きな値となるため、サンプル期間のわずかな違いは、ウェーブレット分散の大きな差になり得るのである。

階段状の異常値以外の代表的な異常値として、スパイク状の異常値を考える（図表13）。スパイク状の異常値は、階段状の異常値とは逆に短期的変動であり、ウェーブレット係数の低レベル（高周波成分）で検出されることが望ましい。そして、スパイク状の異常値は、ウェーブレット係数の低レベルで異常性を確実に捕捉される。以下では、図表13の簡単な例を用いてこの点を確認する。

図表13 スパイク状の異常値の検出



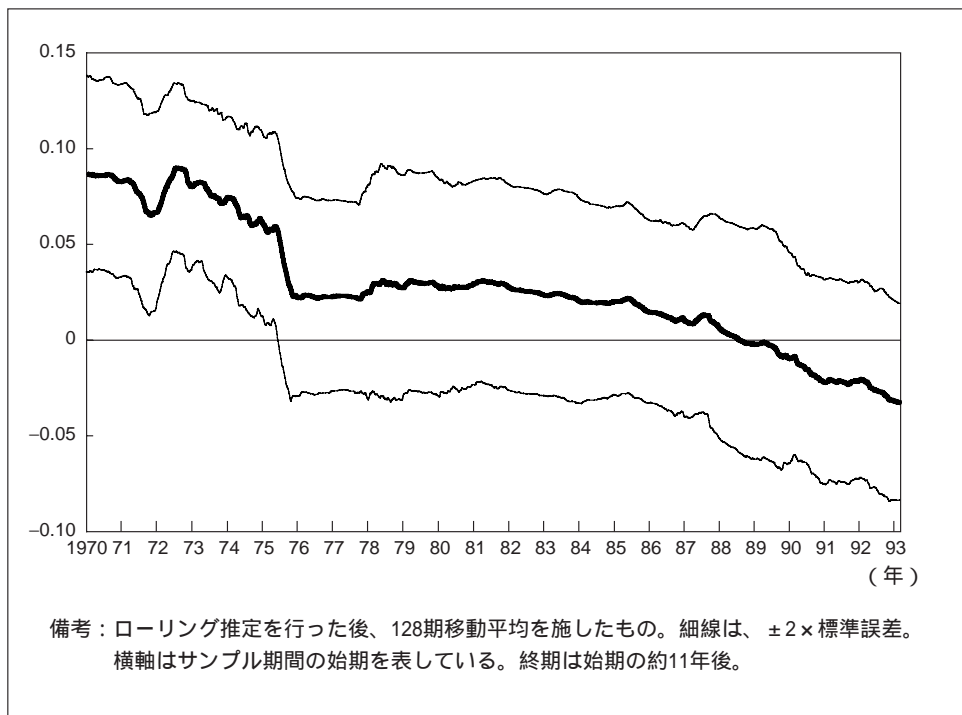
先に示したとおり、レベル1のウェーブレット係数は、2つの隣り合うデータの差として計算され、その組合せは2通りである。図表13のように、 x_5 で大きな値を示す系列を考える。どのような組合せを選んでも、 x_5 を含む差分のみが大きな値を示し、その他の差分はゼロになることが容易に確認される。すなわち、(36)式の組合せでは $\{x_5, x_6\}$ の差分のみが、(37)式の組合せでは $\{x_4, x_5\}$ の差分のみが大きな値を示す。このように、スパイク状の異常値は、サンプル期間をどのように選んでもレベル1で確実に捉えられる。また、このようなスパイク状の異常値であれば、サンプル期間に関係なく、ウェーブレット分散は一定値をとる。

このように、スパイク状の異常値が存在しても、サンプル期間のずれはウェーブレット分散の計算に影響を与えないようにみえる。しかし、スパイク状の異常値が連続して存在する場合には、状況が異なる。具体的に、 $x_5 = \alpha$ 、 $x_6 = -\alpha$ 、その他がゼロであるケースを考える。このとき、(36)式の組合せと(37)式の組合せから計算されるウェーブレット係数は異なり、ウェーブレット分散の値も違うことが容易に確認される。したがって、スパイク状の異常値であっても、サンプル期間のずれがウェーブレット分散の値に影響を与え得る。

ここで、前掲図表11に戻る。局所的には、サンプル期間をわずかにずらただけで、長期記憶パラメータ d は大きく変動している。これは、株価収益率の長期記憶性が短期間で大きく変化すると考えるよりもむしろ、これまで議論したように、異常値の存在によって局所的に大きく変動していると解釈することができる。そこで、128 (=2⁷) 期移動平均をかけて、長期記憶パラメータの趨勢部分を抽出した(図表14)。

図表14をみると、1975年から1976年にかけて大きな段差が発生していることがわ

図表14 株価収益率の長期記憶パラメータ(2)

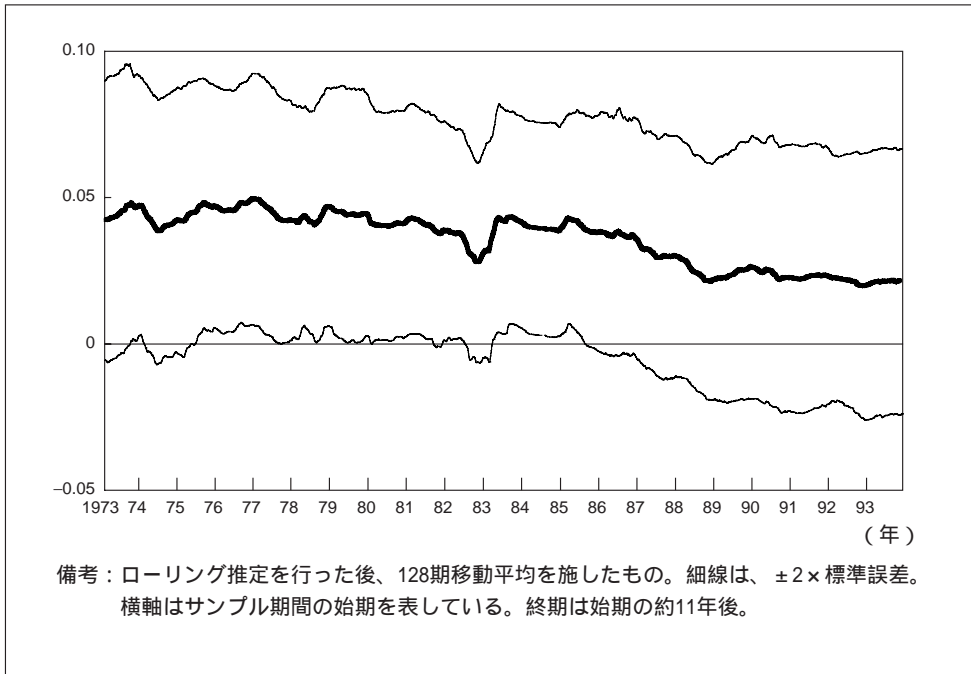


かる。これは、サンプル始期である1975年から、サンプル終期に当たる1987年のどこかで構造変化が生じていたことを示している。構造変化が生じた具体的な時点²³を知ることはできないが、1970年代前半を始期とする11年間の株価収益率には有意な長期記憶性が存在していたが、1980年代以降の株価収益率には有意な長期記憶性が存在しないと解釈することができる。つまり、近年の株価収益率と30年前の株価収益率とでは、統計的な性質が異なっていると、みなすことができる。前述のように複数の単位根検定で、株価収益率がI(0)過程ともI(1)過程とも明確な区別がつかなかった理由は、1970年代における有意な長期記憶性であると考えられる。

次に、長期のデータを取得可能である円/ドル・レートの日次変化率を対象に、ローリング推定を行った後で128期移動平均をとり、長期記憶パラメータの推移を求めた。その結果が図表15である。円/ドル・レートの日次変化率では、株価収益率と異なり、全期間を通じて長期記憶パラメータに大きな変化はなく、ほぼ一定の値をとっている。また、長期記憶パラメータの水準をみる限り、全期間を通じて明確な長期記憶性を認めることはできない。

23 もっとも、構造変化は特定の時点で急激に生じるだけでなく、徐々に時間をかけて進行することもある。後者のような状況では、そもそも、構造変化発生の詳細時点を知ることは困難になる。

図表15 円/ドル・レートの日次変化率の長期記憶パラメータ



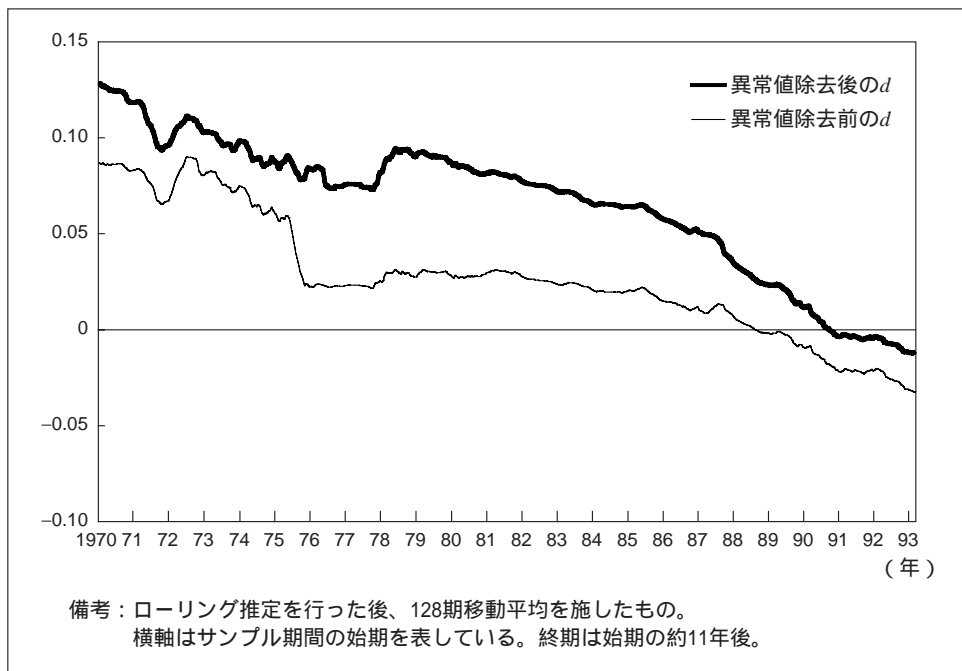
(3) 金融時系列に含まれる異常値

異常値の除去作業には恣意性が入り込むとはいえ、異常値が含まれていると、異常値以外の本質的な変動の性質を知ることができない。そこで、異常値を各レベルの標準偏差の3.3倍として除去することにした²⁴。異常値を除去した後の d の推定値 d_1 は、異常値除去前の推定値 d_0 と異なることが予想される。したがって、異常値除去前後の d を比較すれば、異常値の大半がスパイク状の異常値か、あるいは階段状の異常値かを判定することができる。つまり、 $d_1 > d_0$ であれば、異常値は低レベル部分に存在していたことを意味するため、スパイク状の異常値が大半であることになる。 $d_1 < d_0$ であれば、異常値は高レベル部分に存在していたことを意味するため、階段状の異常値が大半であることになる。株価収益率から異常値を除去したうえで、長期記憶パラメータをローリング推定した結果が図表16である。

株価収益率の異常値除去後の長期記憶パラメータ d の推定値は、異常値除去前に比べ大きくなった。先行研究によると、時系列中に階段状の異常値が混在していると、長期記憶性があると判断する傾向が強くなる（Gourieroux and Jasiak [2001]）。しかし、図表16をみると、全期間を通じて、異常値除去後の d の値が大きくなっている。こうしたことから、株価収益率に混在している異常値は階段状のものではなく、スパイク状の異常値が主であったと考えられる。

24 ウェーブレット係数の分布が正規分布に従うのであれば、全体の約0.1%が異常値として除去される。

図表16 異常値除去後の長期記憶パラメータ（株価収益率）



次に、検出した異常値が、いつの時点で発生したのかを検討する。DWTでは、例えばレベル1で検出した異常値は、2営業日のいずれかでの異常値であり、時点も特定される。しかし、高レベルになるほど分解の解像度は粗くなり、例えばレベル5で検出した異常値は、32営業日のうちのどこかでレベル・シフトが発生したことまでしかわからない。

こうした問題点を解決するのが、MODWT (Maximal Overlap Discrete Wavelet Transform) である。DWTでは高レベルになるほどウェーブレット係数の個数は減少し、係数同士の間隔は広がるため、高レベルのウェーブレット係数から詳細な時間情報を得ることは不可能になる。しかし、MODWTでは、レベルによらず、係数の間隔は基のデータと等しい。つまり、DWTによって、高レベルで異常値を検出した場合に、異常値発生時点を知りたければ、MODWTを用いればよいことになる。

最初からMODWTで異常値を特定することも考えられるが、それには問題がある。MODWTのウェーブレット係数には、非常に高い系列相関が存在するため、ある1日に大きな変動が生じると、その時点だけでなく前後の期間でも相対的に大きなウェーブレット係数が得られてしまう。これによって、本来そうではない時点でも、異常値が発生したと誤認する可能性が高くなる。

MODWTを用いることにより、株価収益率（サンプル期間:1970年6月4日～2005年7月29日）のデータから検出された詳細な異常値発生時点を一覧にしたものが、図表17である。異常値がいつ発生したかを知ることができれば、具体的なイベントと関連付けて、異常値発生の原因を探ることができる。ただし、先述のように、サ

図表17 異常値発生時点

レベル1	1971年 8月20日	レベル2	1972年 6月23日
	1972年 6月23日		1973年12月25日
	1973年 2月 2日		1985年 8月 1日
	1974年10月 9日		1986年10月23日
	1974年10月28日		1987年 7月23日
	1987年 4月13日		1987年10月19日
	1987年10月19日		1987年10月22日
	1987年10月21日		1988年 1月 6日
	1987年10月29日		1990年 4月 4日
	1987年11月11日		1990年10月 3日
	1988年 1月 6日		1990年10月 5日
	1990年 2月26日		1991年 8月21日
	1990年 4月 2日		1992年 4月 6日
	1990年 8月 7日		1992年 4月17日
	1990年 8月13日		1992年 8月21日
	1990年 8月15日		1992年 9月 1日
	1990年 8月23日		1993年11月29日
	1990年10月 1日		1997年11月17日
	1990年11月 9日		1998年10月 7日
	1991年 1月16日		1999年10月18日
	1991年 8月19日		2000年 2月22日
	1992年 4月 9日		2000年12月20日
	1992年 8月24日		2001年 9月18日
	1992年 8月26日		2002年10月11日
	1992年11月17日		2003年10月22日
	1993年12月 8日		2004年 5月10日
	1994年 1月24日		
	1994年 1月28日	レベル3	1971年 8月18日
	1995年 2月27日		1972年 6月26日
	1995年 3月28日		1987年 5月 6日
	1995年 4月 3日		1987年10月20日
	1997年10月29日		1987年11月17日
	1997年11月17日		1990年 3月28日
	1997年11月19日		1990年 8月29日
	1997年11月21日		1990年10月 3日
	1998年 9月 4日		1992年 8月24日
	1998年 9月24日		1997年 1月 9日
	1998年10月 6日		1997年11月18日
	1998年10月12日		1997年12月24日
	1999年 3月18日		1999年 9月30日
	1999年 8月31日		2000年 3月13日
	2000年 1月11日		2000年 4月18日
	2000年 4月14日		2001年 3月14日
	2000年 5月11日		2002年 1月10日
	2001年 3月19日		
	2001年 3月26日	レベル4	1971年 8月18日
	2001年 3月28日		1973年 2月20日
	2001年 9月12日		1973年12月26日
	2001年 9月14日		1987年10月30日
	2003年 3月24日		1992年 8月24日
	2004年 5月 7日		1992年 8月28日
	2004年 5月17日		1998年10月20日
	2005年 4月18日		
		レベル5	1971年 8月19日
			1990年10月24日
			1992年 4月30日
			2001年10月 5日

サンプル期間をずらすだけで異常値を検出するレベルが異なる。この意味で、図表17の結果は、あくまでも1つの目安という位置付けとなる。

5 . むすび

本稿では、主に時系列が長期記憶性を保有しているかどうかを、ウェーブレットを用いて判定する方法を説明した。ウェーブレット分散を用いた長期記憶性の判断は、比較的恣意性に影響されることなく、高い精度で行える。しかし、異常値ともいえるような大きな変動があると、サンプル期間がわずかに異なるだけで、推定結果に大きな違いが生じ得ることには注意が必要である。

時系列として、株価（TOPIX）の日次収益率と円／ドル・レートの日次変化率を用いて、長期記憶性を分析したところ、1970年代の株価収益率に長期記憶性があることがわかったが、最近のデータでは、どちらにも長期記憶性は認められなかった。したがって、現在は、冒頭で述べたように、株価や円／ドル・レートを原資産とするオプションの価格を計算するためや、資産価格の将来変動の予測のために長期記憶性を考慮する必要性は小さいといえる。

また、長期記憶性が存在するという事は、過去の株価の推移を用いて、将来の株価を予測できることになるという意味で、市場は効率的ではないと解釈することができる。このような解釈を前提とすれば、本邦株式市場は、時日の経過とともに効率的になってきた一方で、円／ドル・レート市場は、本邦株式市場とは異なり、以前から効率的であったといえることができる。

最後に、本稿では扱わなかったが、今後検討に値すると思われる点を挙げて本稿を締め括る。まず、円／ドル・レートと本邦株価を用いた分析で得られた、市場の違いによる効率性の違いが、どのような背景（取引制度や市場参加者等の違い）に基づくものであるのかを検討することは意義があることであると思われる。次に、最近、株価等のティック・データの入手が容易になってきたこともあり、リアライズド・ボラティリティ（Realized Volatility）による研究が多数行われている。ボラティリティは、変化率よりも強い長期記憶性を有していると考えられていることから、ウェーブレット解析をリアライズド・ボラティリティに適用することで、その長期記憶性を分析することは興味深い研究テーマであると思われる。

補論．ウェーブレット解析の基礎的概念

ウェーブレット変換は、線形フィルタリングの一種で、フーリエ変換との類似点が幾つかある。フーリエ変換を実行すると、時間に関する情報は失われる代わりに、周波数に関する詳細な情報が得られる。これは、フーリエ変換が、原系列を三角関数という恒久的な波によりフィルタリングを行うからである。それに対し、ウェーブレット変換は、恒久的な波の代わりに局所的な波を用いる。局所的な波の位置と大きさをみることによって、不確定性原理の許す範囲内で、時間と周波数の情報を同時に得ることができる。

ウェーブレット変換では、フィルタリングに用いる局所的な波が無数に存在するため、変換によって得られる結果は一意ではない。本稿では、ハール・ウェーブレット、ドビッシェー・ウェーブレットにより、分析を行っている。以下では、直交ウェーブレット変換である離散ウェーブレット変換 (Discrete Wavelet Transform : 以下、DWT) と、DWTの弱点を補うMODWT (Maximal Overlap Discrete Wavelet Transform) によるウェーブレット変換の概要を説明する²⁵。

(1) DWT

DWTの実行には、ウェーブレット・フィルタとスケーリング・フィルタという2つのフィルタを必要とする。ウェーブレット・フィルタ h は、以下の(A-1)~(A-3)式を満たす。

$$\sum_{i=1}^L h_i = 0, \quad (\text{A-1})$$

$$\sum_{i=1}^L h_i^2 = 1, \quad (\text{A-2})$$

$$\sum_{i=1}^L h_i h_{i+2n} = 0. \quad (n \text{ は } 0 \text{ 以外の整数}) \quad (\text{A-3})$$

ここで、 L はフィルタの長さを表し、サポートと呼ばれる。(A-1)式は、ウェーブレット・フィルタが全体として「山」と「谷」から構成されていることを意味しており、高周波成分を検出する高域通過フィルタ (high-pass filter) であることを示している。また、(A-2)式は規格化条件、(A-3)式は直交条件である。

スケーリング・フィルタ g は、ウェーブレット・フィルタと独立ではなく、次の直交鏡像変換によって関係付けられる。

25 ウェーブレット解析を経済分析に応用した解説として、稲田・鎌田 [2004]、田中 [2003]、Gençay [2002]、Percival and Walden [2000]、Schleicher [2002] 等がある。

$$g_i = (-1)^i h_{L-i+1} \Leftrightarrow h_i = (-1)^{i-1} g_{L-i+1} . \quad (\text{A-4})$$

スケーリング・フィルタは、以下の(A-5)~(A-7)式を満たす。

$$\sum_{i=1}^L g_i = \sqrt{2} , \quad (\text{A-5})$$

$$\sum_{i=1}^L g_i^2 = 1 , \quad (\text{A-6})$$

$$\sum_{i=1}^L g_i g_{i+2n} = 0 . \quad (n \text{は} 0 \text{以外の整数}) \quad (\text{A-7})$$

(A-5)式は、スケーリング・フィルタが全体として1つの「山」として構成されていることを意味しており、低周波成分を検出する低域通過フィルタ (low-pass filter) であることを示している。また、(A-6)式と(A-7)式は、それぞれ、規格化条件と直交条件である。さらに、ウェーブレット・フィルタとスケーリング・フィルタ間の性質として、偶数倍のシフトで直交関係がある。

$$\sum_{i=1}^L g_i h_{i+2n} = 0 . \quad (n \text{は整数}) \quad (\text{A-8})$$

なお、本稿の分析で主に使用するハール・ウェーブレットでは、ウェーブレット・フィルタとスケーリング・フィルタは、以下のように表せる。

$$h = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right), \quad g = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) . \quad (\text{A-9})$$

ハール・ウェーブレット以外の例として、サポート長が4のドビッシェー・ウェーブレットでは、以下ようになる²⁶。

$$h = \left(\frac{1-\sqrt{3}}{4\sqrt{2}}, \frac{-3+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{-1-\sqrt{3}}{4\sqrt{2}} \right), \quad (\text{A-10})$$

$$g = \left(\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}} \right) .$$

このようなフィルタを用いて、長さ N の原系列 x にDWTを施す。DWTを実行すると、ウェーブレット係数とスケーリング係数が得られる。レベル j のウェーブレット

26 本稿では、サポート長が L のドビッシェー・ウェーブレットを、 $D(L)$ と表記している。

ト係数を w_j 、同スケール係数を v_j とする。レベル0のスケール係数は、原系列に等しいので、 $v_0 \equiv x$ と定義する。ここでレベル $j-1$ のスケール係数 v_{j-1} にウェーブレット・フィルタ h を適用すると、レベル j のウェーブレット係数 w_j が得られる。また、レベル $j-1$ のスケール係数 v_{j-1} にスケールリング・フィルタ g を適用すると、レベル j のスケールリング係数が得られる。

DWTによるフィルタリングは、データをトレンドとトレンドからの乖離に分割する作業である。最初のフィルタリングにより得られたトレンド成分を、次のフィルタリングで、より大きなスケールのトレンドとトレンドからの乖離に分割する。この一連の作業を繰り返していくのがDWTである。これらのフィルタリングを数式で表したものが、(A-11)式、(A-12)式である²⁷。

$$w_{j,t} = \sum_{k=1}^L h_k v_{j-1, \{(2t-k) \bmod (N/2^{j-1})\} + 1} \quad (A-11)$$

$$v_{j,t} = \sum_{k=1}^L g_k v_{j-1, \{(2t-k) \bmod (N/2^{j-1})\} + 1} \quad (A-12)$$

ここで、以下の3点に注意する必要がある。第1に、これらフィルタリングの操作は、畳込みの計算であって内積の計算ではない。第2に、フィルタのサポート長が2より大きい場合、データの端では、始点と終点を接続したデータを使うため、その影響を受ける²⁸。第3に、フィルタを通過すると、ウェーブレット係数とスケールリング係数の個数は半減する(ダウン・サンプリング)。したがって、レベル j までDWTを繰り返すためには、データは 2^j の倍数個だけ必要となる。

なお、ウェーブレット係数とスケールリング係数を合わせると、フィルタリングを行う前と等しい個数のアウトプットが得られており、全体の情報は失われていない。

(2) MODWT

先述のとおり、DWTを実行するためにはサンプル・サイズ(データの個数)が 2^j の倍数でなくてはならないという制約がある。この制約は、特にサンプル・サイズが小さいときに深刻な制約となり得る。また、高レベルのウェーブレット係数の間隔は、原系列と比べて非常に大きくなる。これは、ウェーブレット係数は、高レベルになるほど、位置に関する情報が曖昧になることを意味している。この

27 $A \bmod B$ は、 A の B を法とする演算子であって $0 \sim B-1$ の整数をとる。

28 (A-11)式、(A-12)式は、畳込みは基のデータを2つずらしながらフィルタと掛け合わせることを表現している。したがって、フィルタのサポート長が2のとき、畳込みはデータの過不足なく行えることがわかる。逆に、サポート長が2より大きいと基のデータが足りなくなり、始点と終点を接続したデータを用いないと畳込みが行えないことを意味している。

DWTの弱点を補う手法として、MODWTがある。MODWTに用いるフィルタ \tilde{h} と \tilde{g} は、DWTで用いたフィルタ h と g を使って、次のように表せる。

$$\tilde{h} = \frac{h}{\sqrt{2}}, \quad \tilde{g} = \frac{g}{\sqrt{2}}. \quad (\text{A-13})$$

レベル j のウェーブレット係数 \tilde{w}_j とスケーリング係数 \tilde{v}_j は、それぞれ、以下のように計算される。なお、 $\tilde{v}_0 \equiv x$ である。

$$\tilde{w}_{j,t} = \sum_{k=1}^L \tilde{h}_k \tilde{v}_{j-1, \{(t-2^{j-1}(k-1)-1) \bmod N\}+1}. \quad (\text{A-14})$$

$$\tilde{v}_{j,t} = \sum_{k=1}^L \tilde{g}_k \tilde{v}_{j-1, \{(t-2^{j-1}(k-1)-1) \bmod N\}+1}. \quad (\text{A-15})$$

MODWTは、DWTと違い、ダウン・サンプリングされないことに注意が必要である。このため、ウェーブレット係数の個数はどのレベルでも、基のデータと同数である。一方で、DWTによるウェーブレット係数が基の情報を保持していたことと比べると、MODWTによるウェーブレット係数は、余分な情報を含んでいることがわかる。そのため、MODWTは、非効率なフィルタリングであり、直交条件も満たさない。MODWTの非効率さは、サンプル・サイズの影響を受けない利点の代償である。

(3) ウェーブレット分散

ある時系列 $\{a_t\}$ の分散 σ_a^2 は、次のように計算される。

$$\sigma_a^2 = E[(a_t - E[a_t])^2]. \quad (\text{A-16})$$

ここから類推すると、十分に大きなサポートを持つウェーブレットを用いれば、ウェーブレット係数の平均値はゼロとみなせる²⁹ため、ウェーブレット係数 $\{w_{j,t}\}$ や $\{\tilde{w}_{j,t}\}$ の分散(ウェーブレット分散)も、単純に $E[w_{j,t}^2]$ や $E[\tilde{w}_{j,t}^2]$ となると思われるが、実際にはそうではない。以下では、この点を説明する。

ウェーブレット変換の重要な性質の1つとして、エネルギー保存則がある。ここで、エネルギーとは、ある系列の2乗和をいう。DWTとMODWTは、ともにエネルギー保存則を満たす。つまり、レベル j までのウェーブレット変換を実行したとき、以下が成立する。

29 $I(d)$ 過程で、ウェーブレット・フィルタのサポート長が $L > 2d$ であれば、ウェーブレット係数の期待値はゼロである(Percival and Walden [2000])。本稿では、高々 $d=1/2$ の場合を扱っており、最もサポート長の短いハール・ウェーブレットであっても、ウェーブレット係数の期待値はゼロである。

$$\|x_t\|^2 = \sum_{j=1}^J \|w_{j,t}\|^2 + \|v_{J,t}\|^2, \quad (\text{A-17})$$

$$\|x_t\|^2 = \sum_{j=1}^J \|\tilde{w}_{j,t}\|^2 + \|\tilde{v}_{J,t}\|^2. \quad (\text{A-18})$$

さらに、無限回のウェーブレット変換が可能であると仮定すれば、(A-17)式や(A-18)式はそれぞれ以下になる。

$$\|x_t\|^2 = \sum_{j=1}^{\infty} \|w_{j,t}\|^2, \quad (\text{A-19})$$

$$\|x_t\|^2 = \sum_{j=1}^{\infty} \|\tilde{w}_{j,t}\|^2. \quad (\text{A-20})$$

このように、原系列のエネルギーは、ウェーブレット係数のエネルギーに分解可能である。エネルギー保存則によって、原系列の変動が主としてどのレベル(スケール)によってもたらされているかの情報を得ることが可能になる。

MODWTを実行すると、分解レベルが異なってもウェーブレット係数の個数は等しいため、分解レベルの異なるウェーブレット分散同士の大小を比較することができる。これは、MODWTでは、 $E[\tilde{w}_{j,t}^2]$ の値に意味があることを示唆している。その一方で、DWTを実行すると、分解レベルが上がるにつれ、得られるウェーブレット係数は半減する。このため、分解レベルの異なるDWTのウェーブレット分散を単純に比較することができず、分散に $E[w_{j,t}^2]$ の値をそのまま用いてはならないことになる。結局、DWTのウェーブレット分散は、 $E[w_{j,t}^2]$ ではなく、 $E[w_{j,t}^2]/2^j$ という調整が必要になる。これによって、分解レベルの異なるウェーブレット分散の比較が可能となる。

前述のように、サポート長 L が2より大きなウェーブレット・フィルタを用いると、ウェーブレット係数の端では、始点と終点を接続したデータの影響を受ける。そこで、真のデータから算出される部分のみで、ウェーブレット係数の2乗平均をとる。この場合のMODWTの具体形は、 $\tilde{L}_j \equiv (2^j - 1)(L - 1) + 1$ を用いて次のように記述される。

$$\hat{\sigma}^2(\tilde{w}_j) = \frac{\sum_{t=\tilde{L}_j}^N \tilde{w}_{j,t}^2}{N - \tilde{L}_j + 1}. \quad (\text{A-21})$$

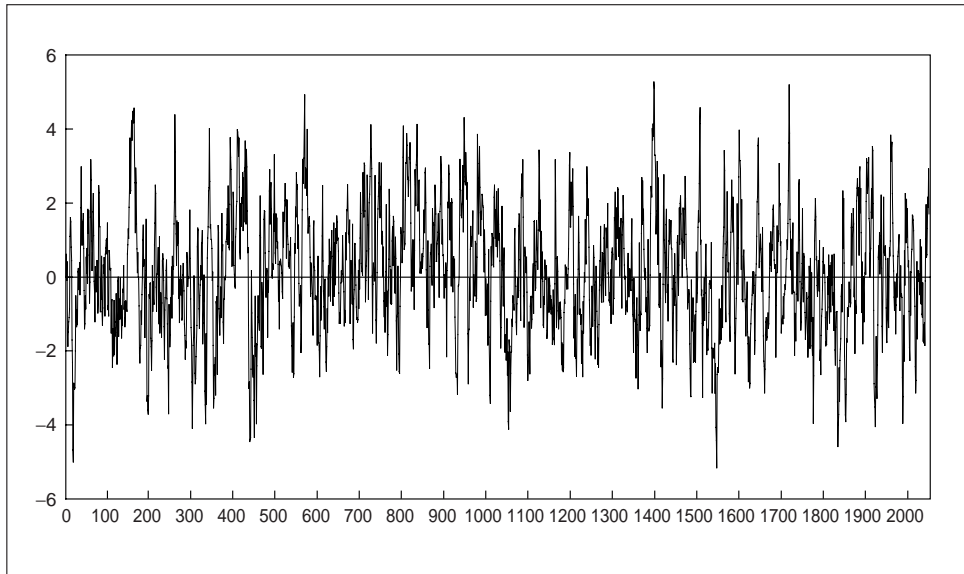
DWTの場合は、 $(L - 2)(1 - 2^{-j}) + 1$ 以上で最小の整数として定義される L_j を用いて、次のように表現される。

$$2^j \hat{\sigma}^2(w_j) = \frac{\sum_{t=L_j}^{N/2^j} w_{j,t}^2}{N/2^j - L_j + 1}. \quad (\text{A-22})$$

本稿では、(A-21)式で計算される $\hat{\sigma}^2(\tilde{w}_j)$ をMODWT分散と呼び、(A-22)式で計算される $\hat{\sigma}^2(w_j)$ を調整DWT分散と呼んでいる。

図表A-1のデータのDWT分散 ($E[w_{j,t}^2]$)、調整DWT分散、およびMODWT分散を図表A-2に示す。図表A-2から、調整DWT分散とMODWT分散は、近い値をとることがわかる。

図表A-1 ある人工データ



図表A-2 ウェーブレット分散（ハール）の比較

レベル	DWT分散	調整DWT分散	MODWT分散
1	1.005	0.523	0.491
2	1.176	0.294	0.302
3	1.222	0.153	0.156
4	1.257	0.079	0.073
5	1.042	0.033	0.036
6	1.097	0.017	0.018
7	0.928	0.007	0.010

参考文献

- 稲田将一・鎌田康一郎、「ウェーブレットによる経済分析」、『金融研究』第23巻第1号、日本銀行金融研究所、2004年、1～62頁
- 刈屋武昭・勝浦正樹、「株価・為替レート時系列変動の長期依存性の検証」、『一橋論叢』第108巻第6号、一橋大学一橋学会、1992年、939～946頁
- 白石典義・高山俊則、「株価収益率ボラティリティの長期依存性とロングメモリーモデル」、森棟公夫・刈屋武昭編『ジャフィー・ジャーナル、リスク管理と金融・証券投資戦略』、東洋経済新報社、1998年、123～150頁
- 田中勝人、「ウェーブレット解析」、刈屋武昭・矢島美寛・田中勝人・竹内啓著『経済時系列の統計：その数理的基礎』、岩波書店、2003年、281～311頁
- 矢島美寛、「長期記憶をもつ時系列モデル」、刈屋武昭・矢島美寛・田中勝人・竹内啓著『経済時系列の統計：その数理的基礎』、岩波書店、2003年、103～202頁
- Baillie, R. T., “Long Memory Processes and Fractional Integration in Econometrics,” *Journal of Econometrics*, 73 (1), 1996, pp. 5-59.
- Beran, J., *Statistics for Long-Memory Processes*, Chapman and Hall, 1994.
- Elliott, R. J., and J. V. der Hoek, “A General Fractional White Noise Theory and Applications to Finance,” *Mathematical Finance*, 13 (2), 2003, pp. 301-330.
- Gençay, R., F. Selçuk, and B. Whitcher, *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, 2002.
- Geweke, J., and S. Porter-Hudak, “The Estimation and Application of Long-Memory Time Series Models,” *Journal of Time Series Analysis*, 4 (4), 1983, pp. 221-238.
- Gourieroux, C., and J. Jasiak, “Memory and Infrequent Breaks,” *Economics Letter*, 70 (1), 2001, pp. 29-41.
- Granger, C. W. J., and R. Joyeux, “An Introduction to Long Memory Time Series Models and Fractional Differencing,” *Journal of Time Series Analysis*, 1 (1), 1980, pp. 15-29.
- Hosking, J. R. M., “Fractional differencing,” *Biometrika*, 68 (1), 1981, pp. 165-176.
- Hurvich, C. M., R. Deo, and J. Brodsky, “The Mean Squared Error of Geweke and Poter-Hudak’s Estimator of the Memory Parameter of a Long-Memory Time Series,” *Journal of Time Series Analysis*, 19 (1), 1998, pp. 19-46.
- Jensen, M. J., “Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator of the Long-Memory Parameter,” *Journal of Forecasting*, 18 (1), 1999, pp. 17-32.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root,” *Journal of Econometrics*, 54 (1-3), 1992, pp. 159-178.
- Lardic, S., V. Mignon, and F. Murtin, “Frequency-Domain Estimation of Fractionally Integrated Process: Impact of Short-Term Components on the Bandwidth Choice,” Working Paper 2003-08, University of Paris X.

- Maddala, G. S., and In-Moo Kim, *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 1998.
- Percival, D. B., and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.
- Schleicher, C., "An Introduction to Wavelets for Economists," Working Paper 2002-3, Bank of Canada, 2002.