ウェーブレット変換と多次元尺度法による時系列の類似性分析とその応用

矢加部, 正幸

https://doi.org/10.15017/3000259

出版情報：経済論究. 108, pp.113-125, 2000-11-28. 九州大学大学院経済学会
バージョン：

ウェーブレット変換と多次元尺度法による時系列の類似性分析とその応用

矢加部　正　幸

1　まえがき

対象をいくつかのカテゴリに分類する方法として，特徴量を比較することにより距離情報を求め，カテゴリの重心を用いて所属を推定することが行われる(1)。応用分野としては，手書文字の特徴ベクトルを用いた分類，ステレオ画像などカメラ入力を用いて対象物体を特定するなどの研究がある(1)。このような対象の明確な分類に対して，時系列の類似性解析に関しては，これまでGrangerによる共和分(Cointegration)がある程度で，研究例は極めて少ないのが現状である(2)(3)。しかしながら，時系列の類似性を用いる方法は視覚的に判断をする分野では重要であり，例えば，心電図，脳波を用いた診断などでは，経験的に蓄積された症例と時系列の対応関係が利用される(4)。また，経済分野においても例えば，株式投資の場合は，投資家は過去の株価データと現在の株価を時系列としてとらえた場合に類似性を視覚的にとらえ，時系列予測を行っている(5)。この場合，同様な変化をする時系列の類似性をとらえているが，計量的（メトリック）な情報を個別的に分析しているのではなく，非計量的（ノンメトリック）な比較を行っていると考えられる。

時系列の類似性分析では，このようなメトリック，ノンメトリックな方法論の違いに注意することのほかに，用いる特徴量が問題となる。時系列データそのものを用いた場合には，情報の集約ができないため，一定の解析精度を得るための計算量が増大する。そのため，スペクトルを特徴量として用いることが考えられるが，短期的な時間域での特徴が失われる可能性がある。このような目的には，短時間スペクトルを記述できるウェーブレット変換を用いることが適切であると考えられる。更に，ウェーブレット変換を用いる利点として，時系列がフラクタルであるどうかを簡単に検証できること，また，そのフラクタル次元をウェーブレット変換から推定できることがあげられる。

本論文では，時系列のウェーブレット変換を入力情報として類似性分析を行う方法を考察するが，この場合，ノンメトリックな方法の代表的な手法である多次元尺度法(MDS：Multidimensional Scaling)を用いる(6)(7)。多次元尺度法を用いることにより，時系列の類似性を空間的表現できる利点があるほかに，主成分分析などのメトリックな解析手法により失われる可能性のある現実との対応関係を保存することができる(6)。更に，このような多次元尺度法による時系列の類似性分析の有効性を確認するため，日本の証券市場における株価時系列の分析に適用すると同時に，このような分析結果を用いた投資の有効性について検証する。

以下，2．では時系列のウェーブレット変換とフラクタル次元を入力情報として用いる基礎的な方法を整理する。3．では多次元尺度法の代表的な手法であるKruskalの方法について整理するとともに，

時系列分析に適用する場合の拡張を行う。4．では，多次元尺度法による時系列の類似性分析を日本の証券市場における株価データに適用するとともに，このような類似性分析にもとづく投資方法の有効性を検証する。

2　時系列のウェーブレット変換とフラクタル次元

2.1　ウェーブレット変換

時系列の類似性を分析する場合に，共和分分析などでは直接的に観測されたデータを入力データとして用いている。しかし，このような直接的なデータ入力は計算量の増大をまねき適切ではない。そのため，時系列が含む情報を集約して用いること，短期的な変動を考慮した分析とするなど，時系列の変換データを用いることが望ましい。従って，本論文では時系列のウェーブレット変換（ウェーブレット係数）を入力として用いることにする(8)-(10)。

更に，株価などに経済時系列に頻繁に見られる自己相似性（フラクタル性）を入力として用いることを考慮する。具体的には，時系列をフラクタル時系列の集合と見なした場合に，フラクタル次元を推定し，これらを入力として用いる。フラクタル次元は，ウェーブレット係数から容易に推定することができる。

いま，与えられた時系列 $x(t)$ を，次のようにウェーブレット変換する。

$$x(t)=\sum_{n}\sum_{m}x_n^m\psi_n^m(t). \tag{1}$$

$$x_n^m=\int_{-\infty}^{\infty}x(t)\psi_n^m(t)dt. \tag{2}$$

$\psi_n^m(t)$ はウェーブレット 基本関数 $\psi(t)$ に対する次のスケール，シフト変換とにより構成される。

$$\psi_n^m(t)=2^{m/2}\psi(2^m t-n). \tag{3}$$

ここで，m，n は，スケール変換，シフト変換のインデックスである。$\psi(t)$ としてはDaubechiesのウェーブレットを用いる(10)。

以下の議論では，このウェーブレット係数 x_n^m を時系列の類似性分析に用いるが，同時に，時系列のフラクタル性の尺度も入力として用いる(11)-(13)。時系列のフラクタル性は，自己相似性ないしは長期的相関ともよばれており，自己回帰過程における短期的な相関の対極にあると考えられている。フラクタル性は，株価など経済時系列において確認されており，以下ではこの検証結果を分析の入力として用いる。

$x(t)$ がフラクタル性をもつ場合，その平均的なスペクトルは周波数のべき乗で低下する性質を持っている。ことから，ウェーブレット係数 x_n^m の満たすべき条件として，次の関係式が得られる。

$$var(x_n^m)=\sigma^2 2^{-\gamma m}. \tag{4}$$

この関係式は，ウェーブレット基本関数の性質などを用いると証明できる(8)(9)(11)(12)。

式(4)の両変の対数をとると，mについての線形の直線となるので，左辺により計算されるデータに対して回帰直線を当てはめ，この直線との2乗平方誤差 R_w の大きさによりフラクタル性を判定できる。

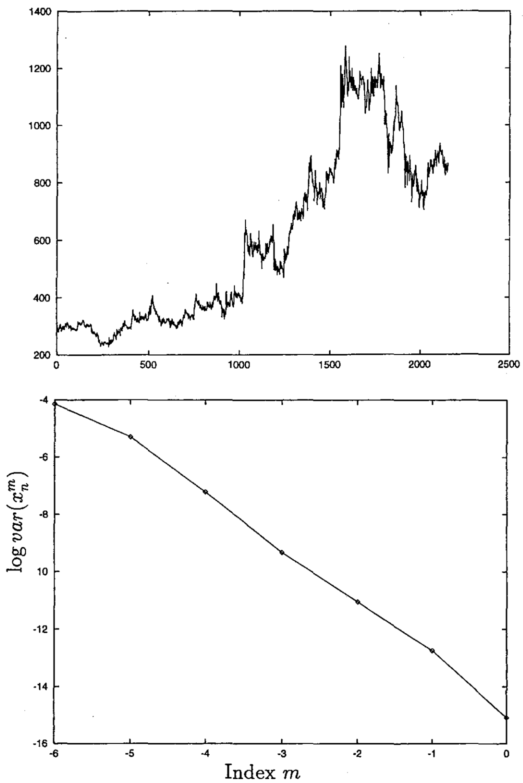

図１　株価とそのウェーブレット係数の分散 $\log var(x_n^m)$ の例

$$R_w = [\sum_m (\log(var(x_n^m)) - c_0 - c_1 m)^2]^{1/2} / (MX_r) \tag{5}$$

ただし，c_0，c_1 は回帰直線を当てはめた場合の係数であり，Mは添字mの取り得る個数である。具体的には，時系列のサンプリング間隔を１とした場合に，データの補間を考えないケースでは，ウェーブレット関数 $\psi_n^m(t)$ の最小の台は１となるので，$m=-4$，-3，-2，-1，0 などとなる。この場合は $M=5$ である。また，X_r は $\log(var(x_n^m))$ の最大値と最小値との差であり，この値により２乗平方誤差を正規化している。

フラクタル時系列について，式(5)の R_w を計算すると，理論的にはゼロになる。しかし，実際には計算誤差などが含まれるので，有限の小さな値となる。コンピュータシミュレーションにより発生させたfBm（fractional Brownian motion）について式(5)の R_w を計算すると，平均して約0.006となる。従って，この R_w の値がどの程度以下であればフラクタルと見なせるかを求める必要がある。ここでは，フラクタルとの対比で良く用いられ，明らかに指数関数的に相関が減少するARMAモデルについて R_w を計算し，このしきい値を推定する。

いくつかの次数の組合せ（AR次数p，MA次数q）をもつARMAモデルで発生させた時系列について平均的な R_w を計算することによりフラクタルであると判別する R_w のレベルを推定することができ

る。シミュレーションの詳細は省略するが，R_w のしきい値として0.03を採用し，この値より小さい値であればフラクタル性をもつと仮定する。

3　多次元尺度法の基礎

3.1　Kruskal法の原理

数値データにより特徴づけられた対象の類似性を検出する方法としては，通常，メトリック法が用いられる。しかし，メトリック法は対象の相関を厳密に解析するため，対象を表現する特徴量の定義が明確でない場合（大まかな傾向を検討する変量として用いている場合など）や，対象の観測データに雑音が多い場合には適当ではないことは知られている。本論文で変量として用いるウェーブレット変換係数も，このような性質を持っている。従って，ここでは特徴量をノンメトリックなデータとして処理する解析手法を用いることにする。

多次元尺度法（ノンメトリック法）は，入力データとして分類対象の非類似性を与え，この非類似性をできるだけ再現する対応関係を作成することにより，対象を n 次元空間の座標に配置する方法である。ノンメトリック法の中でもKruskal法は最も良く用いられる方法である(7)(8)。以下では，基本的な手法としてこの方法を用いるので，簡単に原理を整理しておく。

いま，対象 i と j の間の非近親性を δ_{ij} としておく。本論文では，ウェーブレット係数フラクタル次元などの特徴量 x_{ik} の2乗平方誤差を用いて，次により定義する。

$$\delta_{ij}=[\sum_{k=1}^{M}(x_{ik}-x_{jk})^2]^{1/2} \tag{6}$$

議論を簡単にするため，どの δ_{ij} も等しくないと仮定する。また，対象を配置する空間の次元 n は既知であるとする。このとき，n 個のなかの i 番目の対象の座標を

$$z_i=(z_{i1},\ z_{i2},\ \cdots,\ z_{in}) \tag{7}$$

とする。点 z_i と z_j との距離は，次で定義される。

$$d_{ij}=[\sum_{k=1}^{n}(z_{ik}-z_{jk})^2]^{1/2} \tag{8}$$

非近親性 δ_{ij} が大きいと d_{ij} も大きくなるような単調な関係になるように，できるだけ点の配置 z_i を求めることが多次元尺度法の原理である。一般には，非近親性 δ_{ij} を縦軸，距離 d_{ij} を横軸にとって描いた折れ線（区分線形の線）は単調増加にはならないので，距離 d_{ij} に替わって近似的な距離 $\hat{d}_{ij}$ をもとめて，この折れ線が単調増加となるようにする。折れ線がもとのデータにどれだけ対応しているかは，基本的には2つの距離 d_{ij}，$\hat{d}_{ij}$ の2乗誤差で定義される。これを適応度とよぶ。適応度が向上する方向に，点 z_i を配置する座標を修正していく。そのために最急降下法を用いる。これをアルゴリズムとして整理すると次のようになる。

ステップ1：初期設定

対象を配置する空間の次元 n の初期値を与えておく。n を1から順に増加させる場合には $n=1$ としておく。また，すべての点について配置する初期値を与えておく。乱数を発生させて $z_i=(z_{i1},\ z_{i2},$

…, z_{in}) ($i=1,\ 2,\ \cdots,\ N$) を構成するが，点は重なってはいけないこと，および，座標は次のような正規化を行っておくことに注意を要する。

$$\sum_{i=1}^{N} z_{ik}/N=0,\quad k=1,\ 2,\ \cdots,\ n \tag{9}$$

$$\sum_{i=1}^{N}\sum_{k=1}^{n} z_{ik}^2/n=1 \tag{10}$$

ステップ2：対象間の距離 d_{ij} の計算

このように与えた配置座標から，対象 i，j の距離を計算する。

ステップ3：ブロック評価の開始

連続するいくつかの距離をまとめてブロック b として，そのブロック内の平均距離が大小順になるようにブロック b_1, b_2, …, を作成する。ここで，ブロック b 内の平均距離は，ブロック内の距離 d_{ij}（個数B）を用いて，次で定義される。

$$d_b=\sum_{k=1}^{B} d_{ij}/B \tag{11}$$

このように，平均値が大小順に並ぶようにブロックを求める。隣接する3つのブロックを $b-$, b, $b+$ とするとき

$$d_b < d_{b+} \tag{12}$$

を満足するときこのブロックをup-satisfied（up-S）とよぶ。また，

$$d_{b-} < d_b \tag{13}$$

を満足するときこのブロックをdown-satisfied (dn-S) ただし，左端のブロックはup-S，右端のブロックはdn-Sとする。もし，d_{ij} が完全に大小の順序に並んでいれば各ブロックの両方を満足している。最初はすべての d_{ij} が1つのブロックをなしているとして開始する。初期段階では近似的な距離 $\hat{d}_{ij}$ は実際の距離 d_{ij} に等しい。

ステップ4：ブロックの更新

左端のブロック d_{i1j1} から開始して，次のブロック d_{i2j2} との関係がup-Sである ($d_{i1j1} < d_{i2j2}$) ならば，関係を満足するので，1つのブロック b_1 として表示する。従って，$d_{b1}=d_{i1j1}$ となる。これに対して ($d_{i1j1} < d_{i2j2}$) であるならば，1番目と2番目をまとめて第1ブロックとする。このブロックの距離は平均値

$$d_{b1}=(d_{i1j1}+d_{i2j2})/2 \tag{14}$$

を用いる。

次に，3番目の d_{i3j3} と比較する。もし $d_{b1} < d_{i3j3}$ であればup-Sである。従って，ブロックは b_1 左端となるので，up-S, dn-Sともに満足されるので次に進む。もし $db_1 > d_{i3j3}$ であるならば，up-Sを満たないので，第1ブロックに d_{i3j3} を加えて第1ブロックを作りなおす。このとき次のようになる。

$$d_{b1}=(d_{i1j1}+d_{i2j2}+d_{i3j3})/3 \tag{15}$$

ステップ5：操作の継続

以上のような操作を，各ブロックがup-S，dn-Sともに条件を満足するまで隣接する距離の計算，平均値の計算を続ける。もし，dn-Sを満足しない場合には，系列の上方からブロック化を逆に進め，up-S, dn-Sともに条件を満足するように平均距離を計算する。計算方法はステップ4と同じである。

ステップ6：適応度の計算

与えられた座標について，次のように定義される適合度（stress）を求める。

$$S=\sqrt{S^*/T^*} \tag{16}$$

$$S^*=\sum_{i<j}^{N}(d_{ij}-\hat{d}_{ij})^2,\quad T^*=\sum_{i<j}^{N}d_{ij}^2 \tag{17}$$

ステップ7：点の座標の更新

適合度が低下する方向に点の座標 z_{ks} は更新される。次の式を用いる。

$$\Delta z_{ks}=-\alpha S\sum_i\sum_j[(d_{ij}-\hat{d}_{ij})/S^*-d_{ij}/T^*]sig(z_{is}-z_{js}) \tag{18}$$

$$sig(a)=\begin{cases}1 & a>0\\ 0 & a=0\\ -1 & a<0\end{cases} \tag{19}$$

更新する値を用いて z_i を修正するが，解の近くでの収束を良好に保持するため，係数 α を適応的に変更するなどの方法が用いられる（詳細は省略する）。

ステップ8：収束の判定

適合度Sが目的とする数値以下に達したら計算を終了する。点の座標を配置する空間の次元 n を1つ増加させる。もし，この数値が目的とする最大次元を越えた場合には，全部の処理を終了する。そうでない場合には，ステップ2にもどって操作を繰り返す。

3.2　時系列分析へ適用する拡張

4　株価データへの適用

以下では，これまで述べたウェーブレット変換と多次元尺度法の応用例として，日本の証券市場における株価時系列の類似性を行う。用いる株価時系列は，CD-ROMとして刊行されている1983年から1994年まで，東京証券取引所で売買された銘柄に関するものであり，それぞれの株価データのサンプル数は約2000である(14)。

このような株価の類似性分析を行う場合に次のようなことに注意する必要がある。

(1)　平均株価との関連性

平均株価に近い位置にある銘柄は，市場平均に近い変動をしていると考えられる。このことは，市場より極端に低い利益となるケースもないが，大きな利益も期待できないグループである。

(2)　利益の大小

逆に，平均値より大きく離れている銘柄は，市場より大きな（小さな）利益となることが予想される。特に，大きな利益が期待できるケースでは，変動が大きいと予想され，いわゆるハイリスク・ハイリターンの状況にある。

(3)　銘柄のグループ分け

空間における位置が近い銘柄は，同じような株価変動を行うので，投資などにおいては同一グループとして取扱うことができる。一般的には，業種がこのようなグループに対応すると考えられるが，企業の財務構造などの依存する面ももっている。従って，業種で分類されるグループとは異なる所属となる銘柄が注目されることになるであろう。

4.1 次元数と時系列の長さ

実験に用いた株価を，以上のような3つの視点から分析した結果を表に示している。これらより，次のことが分かる。

株価の変動が類似していることを具体的に利用する分野としては，第1番目は投資における危険性の分散があげられる。同じ傾向の変動に従う株価を集中的に所有している場合には，株価上昇により大幅な利益が得られる場合もあるが，株価が下落する場合には，大きな危険性を伴うことになる。従って，傾向の異なる株価を組み合わせて購入することが行われる(株式ポートフォリオの理論)。株価の類似性分析の第2番目の利用分野としては，平均株価や為替レート変動との類似性を検証することがある。株式市場での取り引きでは平均株価を所有して得られる利益を基準とすることが多いこと，オプションとよばれる危険回避の手段は平均株価を媒介して行われることなど，平均株価との関連が重要である。また，輸出関連企業などの株価は為替レートとの関連が大きくなることが予想される。

多次元尺度法においては，対象を配置する最適な次元（ストレスが最小となる場合の次元）を求めることが問題となる。もし，次元が2あるいは3であるならば，株価を時系列として見た場合の類似性を視覚的にも判断することができる。また，株価をグループごとに分けた場合のそれぞれのグループの代表株価を選択することも容易となる。

このような最適次元は，株価のサンプル間隔にも依存していると考えられる。最初に，これらの関係をまとめておく。図2には，株価の多次元尺度構成における適合度と次元の関係の1例を示している。これより分かるように，最適な次元は，4前後の値となることが期待できる。

表1には，株価のサンプリング間隔を変えた場合に求められる最適な次元の平均値を示している。なお，分類に用いるデータ数を，それぞれの場合について同程度とするために，サンプリング間隔が短い場合には，比較対象とする株価時系列の長さも短くなる。従って，短期的な株価の類似性を分析していることになる。

表1では同時に多次元尺度法により分析する株価が少ない場合と，比較的多い場合とを考察している。結果より分かるように，株価の数が少ない場合の方が，対象を空間配置する最適な次元における適合度が小さくなっている。しかし，最適次元は同じ程度であり，分析方法を対象数に応じて細かく変更する必要はない。

表より分かるように，株価の類似性分析においては，次元を，ほぼ4にしておけば支障ないことが分かる。更に，株価のサンプリング間隔に関しては，余りにも短い間隔や，逆に余りにも長いサンプリング間隔では，それぞれの株価の特徴が失われる可能性がある。しかし，ここに示す結果からは，最低のサンプル数を100としておけば，類似性の判断，従って，空間に配置して相互を識別するための次元が特に高くなる現象は観測されない。

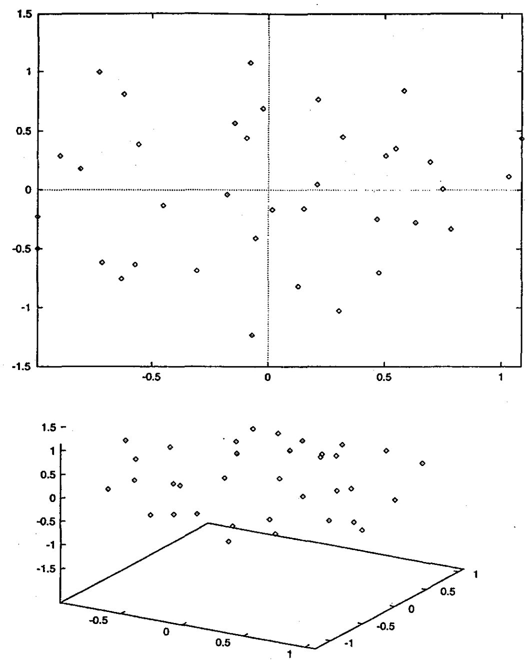

図2　多次元尺度法における株価配置の例（2，3次元）

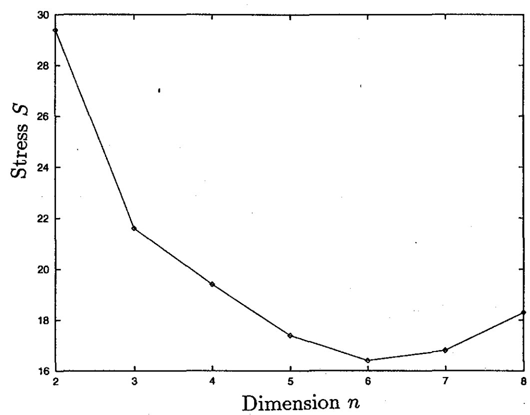

図3　適応度Sと次元nの関係の例

表1　12株式銘柄の多次元尺度の次元nと適応度%（M：データ長）

次元	M=100	M=500	M=1000	M=2000
n=2	29.8	29.4	31.0	33.0
n=3	22.9	21.6	22.0	24.8
n=4	19.2	19.4	20.3	20.7
n=5	15.8	17.4	16.7	14.9
n=6	16.1	16.4	16.8	15.1

表2　40株式銘柄の多次元尺度の次元nと適応度%（M：データ長）

次元	M=100	M=500	M=1000	M=2000
n=2	40.8	40.3	40.3	40.6
n=3	31.6	31.0	31.3	31.2
n=4	23.9	23.9	27.8	27.1
n=5	23.9	23.8	23.8	23.8
n=6	24.1	24.2	24.3	24.1

4.2　次元数とウェーブレット係数の個数

次に，解析に用いるウェーブレット係数の個数と次元数との関係を推定してみる。時系列のサンプリング周期を１とした場合に，時系列の総サンプル数をNとすると，計算可能であるウェーブレット係数のスケーリングインデックス（dilation index）の値は，それぞれの係数が少なくとも５個計算可能であると仮定すると，次のようになる。

いま，長期的な類似性と短期的な類似性について，ウェーブレット係数の個数と次元数との関係をもとめる場合を考える。すなわち，長期的な類似性を解析する場合には，インデックスの値の小さい範囲（周波数の低い成分）に注目し，短期的な類似性の場合にはインデックスの値が相対的に大きな成分に注目して，多次元尺度解析の距離情報の計算に用いることにする。このように，解析に用いるウェーブレット係数を制限しない場合には,長期短期の区別が明確ではない解析となる可能性がある。

このような２つのケースについて，適合度と次元の関係をまとめると，以下のようになる。まず，観測期間を短くした場合には，ウェーブレット係数の高周波数成分を分析に用いないでも，同様の多次元尺度構成が得られる。具体的には，用いるウェーブレット係数のスケーリングインデックスを半分にして，高周波数成分のみを用いても，適応度はほとんど変化しないか，場合によっては少しの改善が見られる(３%以内)。これは，短期的な株価の類似性分析では，高周波数成分のみに注目すれば良いことを意味している。観測期間を長くした場合に，低周波数成分のみに注目するために，ウェーブレット係数のスケーリングインデックスを半分にした場合，適応度はほとんど変化しないか，場合によっては少しの改善が見られる(２%以内)。このことにより，長期的な株価の類似性分析には低周波数成分のみに注目すれば良いことを意味している。

4.3　業種による類似性

株式市場では，景気変動などを考慮した売買決定が行われており，特に，業種ごとに影響する経済的な要因が存在することが予想される（例えば，輸出向け製品を多く製造している業界における為替レート変動の影響など)。従って，株価の類似性分析において，業種内部の株価の類似性が強くなることが予想される。

ここでは，このような業種における株価の類似性を，比較的簡単な方法による推定する。いま，それぞれの業種から，代表的な株価を選択し，この株価の多次元配置において，業種iにおいて，業種に属する株価の重心座標から半径r以内の範囲をさだめ，この業種以外の株価が含まれる確率p_iを計算して，その値が大きいほど業種の相互依存性が強いと判断する。解析する次元としては，いままで

表3 業種ごとの類似性の検証（半径 r に含まれる割合（%））

業種	金融保険	電気ガス	化学繊維	機械電機
値 p_i	1	12	16	1
業種	運輸通信	鉱業製鉄	食品	
値 p_i	1	1	10	

の議論で次元数4が適切であると考えられるので，この値に設定する。業種は7つとして，それぞれの業種に約40の株価が含まれる。ただし，これらを同時に解析することは困難であるので6ずつの株価を抽出して検証を行うことを繰り返した。グループ化する距離としては，ちょうど全部のサンプルがいずれかのグループに含まれる場合の値を用いている。

これらの結果を表にまとめている。結果より分かるように，業種ごとの集約は極めて良好であり本論文で提案する多次元尺度法による株価分類は有効であることを示している。業種ごとの差異としては，金融，運輸，鉱業製鉄，など，業種の特性（製造物の輸入輸出の関連，処理業務）などが比較的明確である業種では，業種内部での相互類似性が高くなっている（従って，数値 p_i は小さい）。これに対して，化学繊維や機械電機など，カバーする業務が比較的広い業種では，やや低くなる（p_i は高い）傾向にある。

4.4 株価の変動指標との比較分析

株式投資の伝統的な理論として，株価時系列の相関に相当する指標（β 値）を用いる方法がある。この理論はCAPM（Capital Asset Pricing Model）とよばれる[15]。この理論の詳細は省略するが，株式市場の情報の伝達が即時的であるなどの仮定をおいた場合に，ある銘柄 i の市場での平均利益 R_i は次の式で推定することができる。

$$R_i = R_F + \beta_i(R_M - R_F) \tag{20}$$

$$\beta_i = cov(x_i,\ x_M)/cov(x_M,\ x_M) \tag{21}$$

ここで，R_M，R_F は市場の平均利益，非危険資産（預金など）の平均利益であり，$cov(x,\ y)$ は x，y の共分散を意味する。

従って，市場の平均株価との連動が著しいほど利益は平均値に近く，β が大きい（小さい）ほど利益が大（小）となる。この β は1次元の量であるので，多次元尺度との比較は直接的ではない。しかし，次のような仮定のもとで，において得られた結果と比較すると，次のような結論が得られる。

β 値による類似性として，ここでは，単純に値の相対誤差をとり，これが小さいほど類似性が高いと判断する。β 値により求めたこのような類似性の判断と，多次元尺度法により求めた類似性の判断を，相互比較する方法として，類似していると判断された株価の中で，共通するものがどれだけ含まれているかを計算することにする。これにより，β 値による解析と多次元尺度解析にカバーする範囲が，ほぼ同じものであるか，あるいは大きく異なっているかの分析が可能となる。表4には，業種の β 値の平均を求め，これに近い β 値をもつ銘柄を類似しているとみなした場合に，業種以外の銘柄が含まれる確率 p_i を示している。これより分かるように，全体として，β 値による類似性分析では業種ごとのグループ化は明確ではない。これは，β という1つの値だけで分類している制限によるものと思われ

表4　業種ごとの類似性の検証（近いβ値を持つサンプルの割合（%））

業種	金融保険	電気ガス	化学繊維	機械電機
値p_i	42	16	27	24
業種	運輸通信	鉱業製鉄	食品	
値p_i	25	2	20	

る。従って，解析と多次元尺度解析にカバーする範囲は異なっていることが分かる。特に，業種を限定した場合には，かなり異なる範囲をそれぞれの類似グループとして指摘しており，同一の解釈を与えてはいない。

4.5　銘柄の組合せ問題への適用

株価の変動が類似していることを具体的に利用する分野として，投資における危険性の分散があげられる。同じ傾向の変動に従う株価を集中的に所有している場合には，株価上昇により大幅な利益が得られる場合もあるが，株価が下落する場合には，大きな危険性を伴うことになる。従って，傾向の異なる株価を組み合わせて購入することが行われる（株式ポートフォリオの理論）。

これまで述べてきた多次元尺度法による株価の類似性分析を投資理論に応用することを考える。株式投資の理論では，利益（リターン）と危険性（リスク）の相互関係が重視されており，利益の大きい株式には同時に大きなリスクが含まれていることが基本認識として存在する。特に，リスクを回避する方法として，株価変動パターンの異なる複数（少なくとも30以上）の銘柄を同時に保有することが提案されている。複数の銘柄をこのような目的で保有する方法をポートフォリオ選択とよんでいる。

以下では，多次元尺度解析の応用として，平均株価を中心として同じような距離にある株式をポートフォリオとして選択することにより，どの程度リスク回避が行えるか検証する。なお，比較の対象としては，従来から行われているβ値による投資理論を考える。すなわち，β値を尺度として用いた場合，β値が大きいほどハイリスク・ハイリターンであり，1は平均株価と同じ動きとなる。β値の場合には1次元の尺度しか用いることができない欠点があるが，期待される収益とβ値との推定式が得られている利点がある。

次のような前提条件を考慮する。シミュレーションの期間を1000日とする。株価の相対利益として，10日前の株価との比率を用いる。β値の計算には，この相対利益を用いる。このように相対利益値が判明している前提で，あらためて最初の日からT日間の相対利益を計算し，平均と分散を求める。ただし，このTを5日と固定すれば，β値による投資そのものであり，結果は自明であるので，やや長い期間（$T>30$）を設定する。

それぞれの類似性分析において，次のような投資方法を仮定する。

(1)　多次元尺度法による場合

平均株価の配置位置を中心として，距離dが大きな方からL個を選択する場合，距離dが小さい方からL個を選択する場合を考察する。この株式を保有しつづけたと仮定して，平均リターンR，リスク（分散）Vを計算する。

表5　本論文の手法による銘柄選択の結果（R，V）

	dの大きい順	dの小さい順
$L=5$	$R=14.7$，$V=0.56$	$R=6.8$，$V=0.002$
$L=10$	$R=10.2$，$V=0.28$	$R=5.8$，$V=0.02$
$L=20$	$R=7.8$，$V=0.14$	$R=5.3$，$V=0.002$
$L=30$	$R=6.9$，$V=0.09$	$R=5.2$，$V=0.002$

表6　従来の手法による銘柄選択の結果

	$\beta=1$に遠い順	$\beta=1$に近い順
$L=5$	$R=13.4$，$V=0.56$	$R=8.9$，$V=0.007$
$L=10$	$R=9.5$，$V=0.28$	$R=8.4$，$V=0.005$
$L=20$	$R=6.6$，$V=0.14$	$R=6.5$，$V=0.003$
$L=30$	$R=6.1$，$V=0.09$	$R=5.4$，$V=0.002$

(2)　β値を用いた場合

β値が1の場合が平均株価であるので，$\beta=1$値に近い方からL個の銘柄を選択した場合と，遠い方からL個選択した場合の2つのケースを考える。

表5，6にはこれらの結果を整理している。平均株価では，$R=4.0$，$V=0.009$である。これより分かるように，多次元尺度法を用いた投資選択においては，大きなリターンを得たい場合の選択が，β値による選択よりも適切に行われている(相対利益の平均値が大きい)。また，小リスクを目標とする選択においても，β値による方法よりは，リスク(分散)が小さくなっている。従って，本論文の方法により，距離dだけに注目して，目的とするリターン，リスクにあったポートフォリオを構成することが可能となっている。一方，β値による投資選択は，理論的には明快であるが，株価変動が必ずしも理論の仮定となっている条件を満足していないため，β値の大小によりリターン，リスクが分かれる結果とはなっていない。また，Lを大きくした場合には銘柄ごとのリスクやリターンが相殺され，2つの方法による相違は小さくなっているが，これは分散投資の方法をうらづけるものとなっている。

5　むすび

本論文では，時系列のウェーブレット変換を入力情報とする多次元尺度法により時系列の類似性分析を行う方法を考察し日本の証券市場における株価データを用いた評価を行った。まず，時系列データの情報をウェーブレット変換を用いて求め，対象の観測データに雑音が含まれる場合を想定し，特徴量をノンメトリックなデータとして処理する多次元尺度法を用いた。多次元尺度法としては，最も良く用いられるKruskal法を用い，対象間の非近親性距離の順序に，できるだけ近似する対象の空間配置を逐次計算により求めた。応用例として，日本の証券市場における株価分析をとりあげ，類似グループへの分割方法，およびこれを用いた安定的な組合せの選択方法を考察した。その結果，本論文の手法は，従来より，株式投資の分野で用いられているいわゆるβ値による手法よりも，類似性分析および，これを用いたリスク回避の方法においても優れていることが示される。

残された問題点としては，株価分類における適合度をさらに向上させる特徴量の選択，および実際の証券市場における検証があげられるが，今後，検討を進めていく予定である。

Reference

(1) 例えば，小川英光編：パターン認識・理解の新たな展開，コロナ社（1994）

(2) C.W.J. Granger: "Long memory relationships and the aggregation of dynamic models", J.Economet., 14, pp. 227-238 (1980)

(3) C.W.J. Granger: "Some properties of time series data and their use in econometric model specification", J. Economet., 16, pp.121-130 (1981)

(4) 例えば，K.P.Birma: "Rule based learning for more accurate ECG analysis", Trans., PAMI-4, 4, pp.366-380 (1982-07)

(5) 例えば，E.Gifford: Investor's Guide to Technical Analysis, Pitmas Publishing, London (1995)

(6) J.B.Kruskal: "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", Psychometrika, vol.29, pp.1-28, 1964.

(7) J.B.Kruskal: "Multidimensional scaling : A nemreical method", Psychometrika, vol.29, pp.115-129, 1964.

(8) G.W.Wornell and A.V.Oppenheim: "Estimation of fractal signals from noisy measurement using wavelets", *IEEE Trans. Signal Processing*, Vol.40, No.3, pp.611-623 (March 1992).

(9) G.W.Wornell; "Wavelet-based representation for the 1/f family of fractal processes", *Proc IEEE*, Vol.81, No.10, pp.1428-1450 (Oct. 1993).

(10) I.Daubechies: "Orthonormal bases of compactly supported wavelets", *Commun.Pure Appl.Math*, Vol.41, pp. 909-996 (Nov. 1988).

(11) 時永祥三, 森保洋, 宮崎明雄, 島津宣之："時系列のフラクタル性を用いた予測手法とその応用", 信学論(A), J79-A, 11, pp.1793-1800（1996-11）.

(12) 時永祥三，森保洋，宮崎明雄，島津宣之："スケール伸長変換およびウェーブレット変換によるパラメータ推定を用いたフラクタル時系列予測", 信学論(A), J79-A, 12, pp.2054-2062（1996-12）.

(13) Mandelbrot. B, "The Fractal Geometry of Nature," San Francisco, Freeman, 1982.

(14) 東洋経済新報社：株価CD-ROM，東洋経済新報社，1993

(15) W.F.Sharpe: Portfolio Theory and Capital Markets, McGraw-Hill (1970).