

消費費日間の階層・空間構造

荒 木 孝 治
橋 本 紀 子

要 約

近年、多くの物理学者が、従来物理学で用いられてきた手法や方法論を金融市場の分析に適用し、経済学的に見て意味のある情報を抽出することに成功している。

Mantegna (1998, 1999) は、市場で取り引きされる株式銘柄を階層的に分類する方法を提案し、これにより、従来見いだすことができなかった経済的に有意な分類結果を示した。Mantegna の方法は、操作としては、株式銘柄の価格変動の時系列データのみを用い、すべての銘柄の組み合わせ(対)の相関行列を求めるだけのシンプルなものである。援用する手法も、物理や統計学の世界では広く知られたクラスター分析(統計学)や最小木(グラフ理論)といった方法である。しかし、この方法を、異分野である金融データ、株価の変動分析に適用することにより、従来にない視点からの新しい結果が得られたのである。

本稿は、Mantegna の方法を、日本の家計消費に関する「国民経済計算」および「家計調査」の費目データに適用し、その分析結果を提示した。また、社会学や生態学の分野で発展してきたネットワーク分析の手法を援用して、Mantegna の方法を補完・拡張することを試みた。

キーワード：相関係数行列、最小木(MST)、デンドログラム、ネットワーク分析、家計調査、国民経済計算

分類番号：02-33、02-41

1 はじめに

近年、物理学の手法を社会科学に適用するという新しい学問分野が開拓され、経済物理学(econophysics)と呼ばれている。精力的に研究成果が報告されており、たとえば株式市場の分析において斬新な切り口が提供されている¹⁾。

経済物理学の代表的な研究者の1人である Mantegna は、株価の時系列データを用いて、市場で取引される株式銘柄間の位置関係を見出す新しい方法を提案した。たとえば、Mantegna (1998, 1999) は Dow Jones Industrial Average (DJIA) を構成する株式銘柄および Standard and Poor's 500 (S&P500) 指数を構成する株式銘柄に対してこの方法を

適用し、銘柄間の階層的関係を示した。また、Bonanno *et al.* (2000) は、Morgan Stanley Capital International (MSCI) 指数に適用し、分析結果を報告している。

Mantegna の方法は次の手順で行われる。まず、時系列データの変動に関する情報を抽出するため、2つの時系列の変動が互いに関連している強さを相関係数の形で表現する。考察対象となる株式銘柄に関連して、全期間、すべての可能な組み合わせの相関を求めると、相関行列が得られる。この方法により、同時に複数の時系列データを扱うことが可能になる。相関行列には分析対象となる期間を通じた株式全体の変動に関する情報が凝縮されている。

次に、相関係数に一定の変換を施し、距離行列を定義する。この距離に対して統計学のクラスタ分析およびグラフ理論の手法である最小木を適用すれば、株価指数を構成する株式銘柄の位置的関係、階層・空間構造を知ることができる。

Mantegna が用いた手法は新しく開発されたものではなく、統計学やグラフ理論の分野では広く知られた、ありふれたものである。しかし、それらを株式データという異分野で扱われてきたデータに適用することにより、従来経済学の分野では知ることのできなかった株式銘柄の階層構造を、株価の時系列データのみから知ることができるのである。Mantegna の手法の意義は、従来からある手法を異なる分野の新しい対象に適用し、従来見出すことのできなかった結果を得た点にある。

本稿は、Mantegna の方法を、日本の家計消費データに適用する。用いるデータは「国民経済計算」および「家計調査」（全世帯、勤労者世帯）の3種類の費目別支出額データであり、それぞれの費目間に内在する空間的、階層的な構造を明らかにしていく。

2 Mantegna の方法

あるポートフォリオに含まれる n 個の株式銘柄を考える。なお、この節での説明には、日次データを想定する。第 t 日における銘柄 i の株価（たとえば終値）を $P_i(t)$ とし、前日 ($t-1$) との対数価格の差を

$$Y_i = \ln(P_i(t)) - \ln(P_i(t-1)) = \ln(P_i(t)/P_i(t-1)), \quad t=2, \dots, T$$

とする (Y_i については、 t を省略している)。 Y_i はいわゆるリターンに相当する量である。次に、この値を、各銘柄ごとに Y_i の平均と標準偏差を用いて標準化する。それを改めて Y_i と表記する。

二つの変量の関係の強さの尺度として最もよく用いられる統計量として Pearson の相関係数がある。二つの銘柄がそれぞれ変化する仕方の関連の強さの尺度として、この相関係数を用いることにする。

$$r_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}}$$

ここで、 $\langle \dots \rangle$ は時間に関する平均を求めることを示しており、調査期間全体の中での取引日に対応するすべての値を用いて平均を計算する。たとえば、 $x = \{x_1, x_2, \dots, x_T\}$ のとき、

$$\langle x \rangle = \sum_{i=1}^T x_i / T$$

となる。

相関係数 r_{ij} は、二つの株価が時間の経過につれて変動する変化のパターンの類似性を、厳密には線形関係の強さを計測している。相関係数は -1 から 1 までの値をとり、 1 に近いとき二つの株式の変化の仕方は正の相関が強い（一方が上昇/下落するときは、他方も上昇/下落する）、 -1 に近いときは負の相関が強い（一方が上昇/下落するときは、他方は下落/上昇する）、 0 に近いときには互いの動きにはあまり関連性がないと判断される。

n 個の銘柄を考える場合、 n 行 n 列の相関（係数）行列 C 、

$$C = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nn} \end{bmatrix}$$

を構成することができる。 C の要素は n^2 個あるが、 C が対称行列 ($r_{ij} = r_{ji}$) であるため、また、対角要素 r_{ii} がすべて 1 であることから、この行列は $n(n-1)/2$ 個の要素のみによって特徴づけられている。

これらの要素中には株価変動に関する情報が含まれていることが期待され、この情報を何らかの手法を用いて抽出できれば非常に有用である。Mantegna は情報抽出のため最小木やクラスター分析の適用を考えたが、これらの手法を利用するには、相関行列 C を用いて銘柄間の距離を定める必要がある。次項でこの点に関する Mantegna の工夫について解説する。

2.1 距離

相関行列に基づいて、統計学の手法であるクラスター分析を行い、グラフ理論のツールである最小木 (MST: Minimal Spanning Tree) を求めれば、株価変動に関する有用な情報が入手できると考えられる。これらの手法は、分析対象間の距離（非類似度ともいう）に基づく手法なので、距離の定義を定め、その値を求めなければならない。

相関係数そのものを距離として利用することはできない。なぜなら、相関係数は、ユークリッド空間において距離が満足すべき 3 つの条件

$$(i) \ d(i, j) = 0 \Leftrightarrow i = j \quad (ii) \ d(i, j) = d(j, i) \quad (iii) \ d(i, j) \leq d(i, k) + d(k, j)$$

の (iii) を満たさないからである。このため Mantegna (1998) は、相関係数を変換し、距離

を定義する式として $d(i, j) = \sqrt{2(1-r_{ij})}$ を採用した²⁾。すべての相関係数を用いてこの距離 $d(i, j)$ を求めると、 n 行 n 列の距離行列 $D = [d(i, j)]$ が得られる。

2.2 クラスタ分析 — デンドログラム

距離行列 D を用いてクラスタ分析を行う。クラスタ分析は距離(非類似度)を基準に、分析対象を、同一の属性を持つ集団(クラスタ)として分類する手法である。

クラスタリングを進める方法には、全体をひとつのクラスタとして考え、順次これを細分していく分割的方法と、逆に、各個体がそれぞれひとつのクラスタを構成していると考え、これらの個体のうち距離が近いものを順次統合していく併合的方法がある。ここでは併合的方法を採用する。

デンドログラム(樹形図)は、クラスタ分析の手法の一つであり、距離を縦軸にとり、分類対象とする個体を横軸に配置して、統合/分割のプロセスがわかるように樹木の形に表したものである。統合/分割の際にクラスタ間の距離を考える必要があるが、その取り方にはさまざまある。最短距離法はその代表的なもので、各クラスタに含まれる個体の組み合わせ間の距離のうち最短のものをクラスタ間の距離とする³⁾。

2.3 最小木(MST)

MSTは、いくつかの点(ノード)から構成される集合とそれぞれの点の間の距離が定められているとき、関連が強いノードを選択して結合し、ノード間の空間的配置を提示する手法である。複雑に配置されたノードに関して、ノード間の関係を調べるために考え出された手法であり、複雑な現象の内部に潜んでいる構造をスケルトンとして抽出することが可能となる。具体的には、 n 個の分析対象および対象間の距離が定義されているとき、距離の合計が最小になるように、すべての対象を $n-1$ 本の線で結ぶことにより得られるツリー状のグラフが MST である⁴⁾。

Mantegna が行った株価の分析では、各銘柄を一つのノードと考え、銘柄間の距離に基づいて MST を求めることにより、株式間に内在する構造すなわち銘柄の階層的な位置関係を距離行列から得ることができる。

3 家計消費データの分析

本節では、日本の家計部門の消費支出行動を表したデータである「国民経済計算」(経済企画庁、2000年1月の省庁再編後は内閣府経済社会総合研究所)および「家計調査」(総務庁統計局、省庁再編後は総務省統計局)の費目別データに Mantegna の手法を適用し、消費支出

額の時系列データのみを用いて相関係数を計算し、これらにクラスター分析や MST の手法を援用することにより、費目間の空間構造を示す。

3.1 データ

データとして、次の3つのデータを用いる。

「家計調査」のデータより、下記の10費目を支出対象とする、全国の全世帯および勤労者世帯⁵⁾の支出データの2種を用いる。いずれも1970年1月から1998年12月までの月次データ、348期を利用する⁶⁾。

- | | | |
|--------------|-----------|----------|
| 1. 食料 | 2. 住居 | 3. 光熱・水道 |
| 4. 家具・家事用品 | 5. 被服及び履物 | 6. 保健医療 |
| 7. 交通・通信 | 8. 教育 | 9. 教養娯楽 |
| 10. その他の消費支出 | | |

3つめとして、「国民経済計算」の家計の目的別最終消費支出データを用いる。費目は次に示す8項目で、1970年第1四半期から98年第4四半期までの116期間の四半期データを利用する。

- | | |
|---------------|---------------|
| FDH : 食品飲料・煙草 | APH : 衣服・はきもの |
| RLH : 家賃・水道光熱 | FHH : 家具器具・雑費 |
| MHH : 医療・保健 | TCH : 交通・通信 |
| REH : 娯楽・教育文化 | MSH : その他 |

「家計調査」と「国民経済計算」は、日本の家計消費の動きを表す代表的な調査である。両者にはマイクロベースあるいはマクロベースというとらえ方の違いがあるが、本稿では消費費目への支出配分を通して各費目の空間関係を分析していくため、両者ともを利用していく。

両者の費目内容は、項目名が同じ場合には概ねその内容は似通っているものの、たとえば煙草代は家計調査ではその他に、国民経済計算では食品に入っているなど、若干の相違が見られる。また、住居費の扱いに大きな違いがある。国民経済計算では帰属家賃を住居費に含めているが、家計調査では住居費は“宅地および住宅に関連する支出。ただし、宅地・住宅の購入費は実質支出以外のなかの財産購入として扱い、住居費支出には含めない”とされている。すなわち、住居費に含まれるのは家賃地代、設備修繕・維持の費用のみであり、持家の帰属家賃は含まれていない。日本では持家率は60%を越えていることを考えると、この住居費の扱いは不十分な点が多いと考えられている。これらの点に留意しながら、本稿では、家計調査および国民経済計算における費目の空間構造について分析し、得られる結果に違いがないか、あるいは、両調査の整合性、費目別分析を行っていく上での妥当性について比較・

検討していく。

家計調査データについては、全国を対象とする全世帯データ、勤労者世帯データの2種の結果を比較・検討していく。注5で述べたように全世帯は勤労者世帯をも含んでいるため間接的な検証になるが、勤労者世帯と個人営業世帯やその他の世帯の行動結果に大きな違いがあれば、両データの結果に差が生じると考えられるからである。

次節以降の分析では、家計調査と国民経済計算、家計調査における全世帯と勤労者世帯という2組の比較に配慮しながら、消費費目の空間構造について明らかにしていく。

3.2 予備的分析

費目 i ($i=1, \dots, 10$ または $i=1, \dots, 8$) の第 t 期 ($t=1, \dots, 348$ または $t=1, \dots, 116$) における消費支出額を $P_i(t)$ とする。

Mantegna 他は株式データの分析において前日比の対数変換値である

$$Y_i = \ln(P_i(t)/P_i(t-1))$$

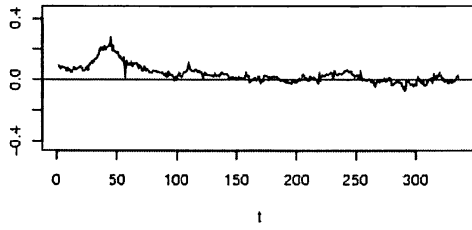
を用いたが、本稿では、対前年比の対数値を求める。すなわち、家計調査データについては前年同月の支出額との比の対数変換値 $Y_i = \ln(P_i(t)/P_i(t-12))$ を、国民経済計算については前年同期に対する $Y_i = \ln(P_i(t)/P_i(t-4))$ を求め、標準化したものを改めて Y_i とする。これは、費目によっては特定の月・期に消費額が大きく変動するものがあり、その値がヒストグラムや散布図において外れ値や層別要因として表れることに対処するためである⁷⁾。

まず、時系列的に Y_i がどう変化しているかを検討するために線グラフを見る。家計調査の各費目の動きを図1に、国民経済計算の結果を図2に示す⁸⁾。なお、家計調査において、対象世帯の違いによりデータの動きに大きな違いは見られなかったため、ここでは全世帯のみの結果を示している。(以下、費目名は、その費目中影響の大きな項目名のみを表示する。)

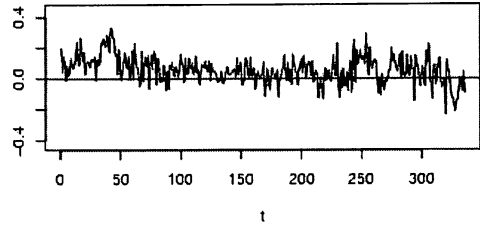
いずれの線グラフを見ても、 Y_i がマイナスの値を取る場合はあまり多くない。

家計調査では1. 食料、9. 娯楽、10. その他の変動が小さいこと、逆に、2. 住居、4. 家具、6. 保険医療、7. 交通通信、8. 教育の変動が大きいことが観察される。特に、4. 家具では、期間初期には大きなプラスの変動、後半には一時的ではあるが大きな落ち込みが見られる。7. 交通通信では、初期には大きな変動がみられるものの、その後変動幅は次第に小さくなっている。多くの費目では、度合いこそ違え、観察期間前半と比べて後半に変動の幅が小さくなっている。この傾向が見られないのは、2. 住居、4. 家具のみである。

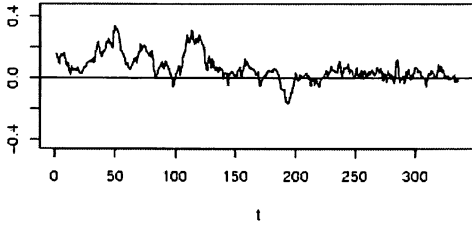
国民経済計算では、ほとんどの費目で、期間後半に変動の幅が小さくなっている。この傾向に反したのは、FHH(家具)のみであり、家計調査でも見られた一時的ではあるが大きな落ち込みが観察されている。さらに注目すべきは、RLH(家賃・光熱)の動きである。RLH



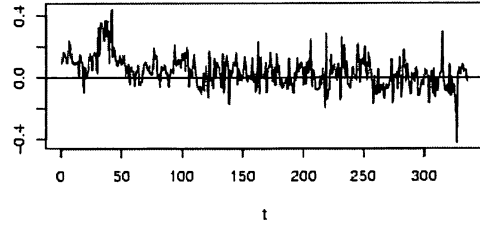
1. 食料



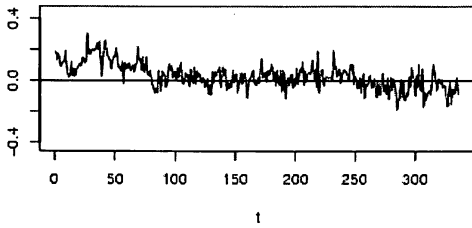
2. 住居



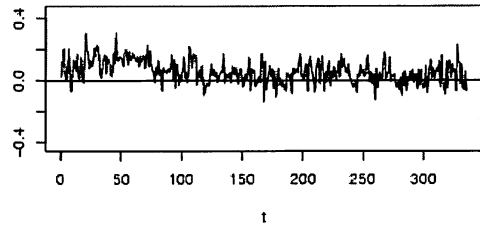
3. 光熱



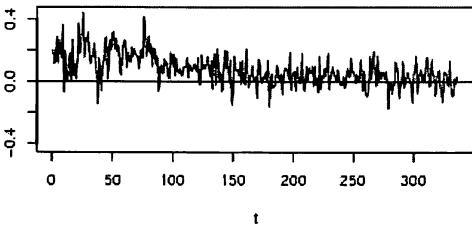
4. 家具



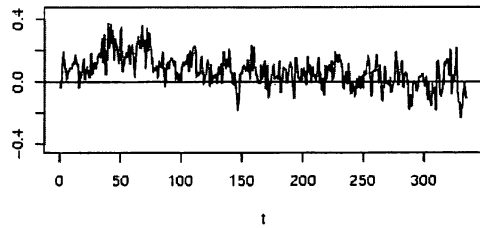
5. 被服



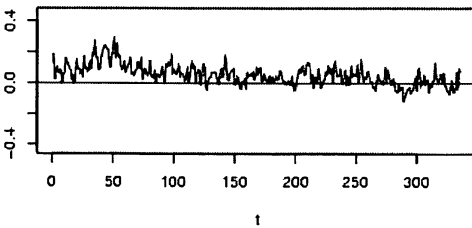
6. 保健医療



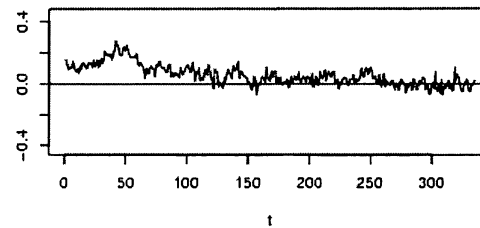
7. 交通・通信



8. 教育



9. 娯楽



10. その他

図1 家計調査費目 (全世帯) の線グラフ

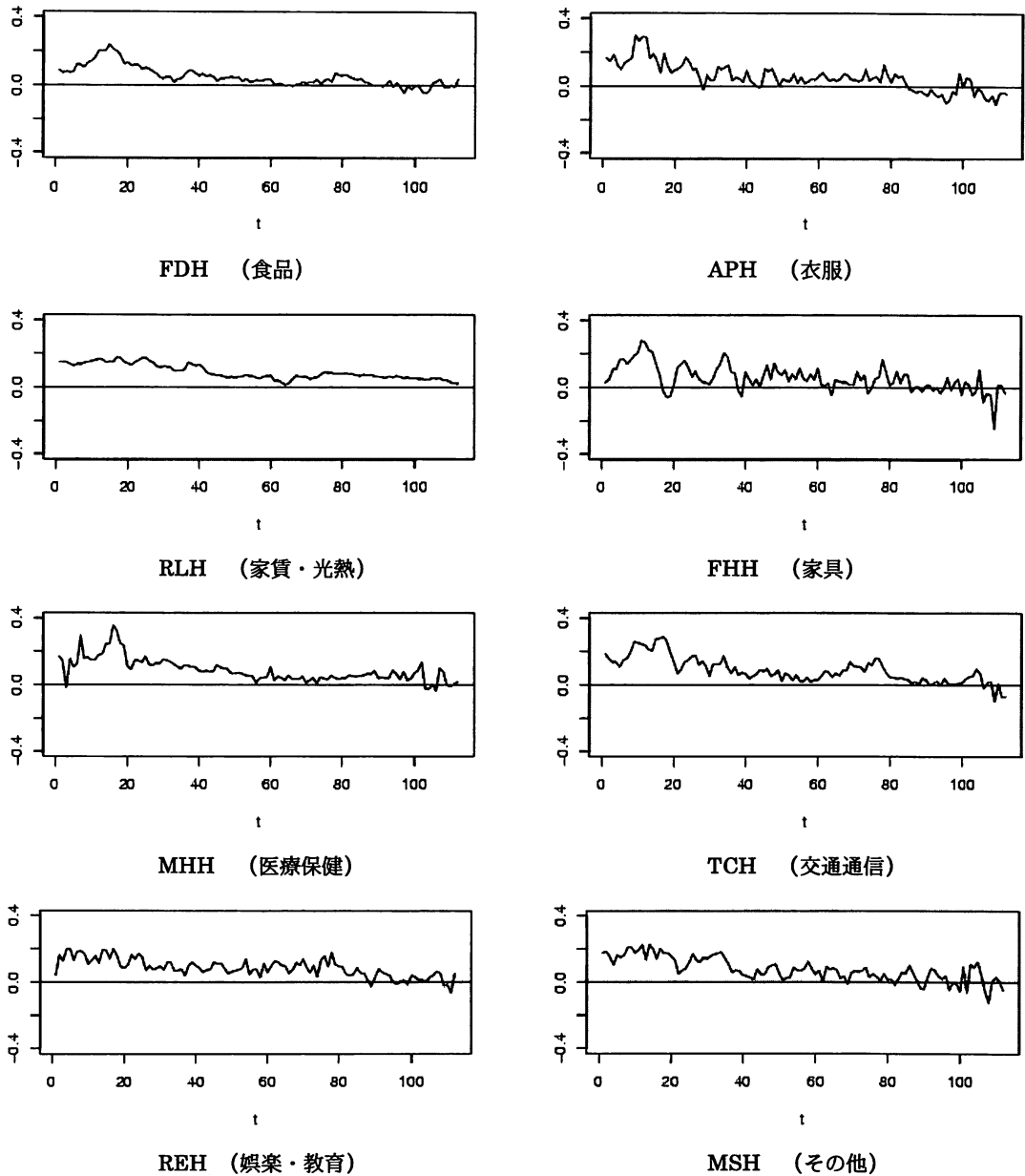


図2 国民経済計算費目の線グラフ

は、費目としては家計調査の2. 住居と3. 光熱を合わせ、さらに帰属家賃を加えたものに対応する。家計調査の2. 住居と3. 光熱ともが期間前半には大きな変動を見せていたこと、また2. 住居については後半も大きく変動しているにもかかわらず、期間を通じて非常に安定した動きを見せている。この点より、両調査における住居関連費目の動きはかなり異なることが予想される。また、国民経済計算では、家計調査における住居と光熱費という傾向の異なるものをあわせていることに注意を払う必要がある。この点に関しては、ヒストグラム

を用いて、後ほど詳しく検討する。

次に、全期間のデータに基づいて Pearson の相関係数 r_{ij} を求め、相関行列 C を定める。国民経済計算については全世帯と勤労者世帯に対して各10行10列、家計調査については8行8列の相関行列が求まる。費目間のペアの組み合わせは、家計調査で ${}_{10}C_2=45$ 個、国民経済計算で ${}_8C_2=28$ 個ある。

ここで、Pearson の相関係数を求めることに意味があるかどうか、たとえば正規性の仮定が満足されているかどうか、また外れ値と見られるデータがないか等をチェックするために、予備的な分析を行っておく。

まず、費目 i に対する Y_i のヒストグラムを作成し、図3に示す。家計調査 (全世帯) では若干の歪みがみられるものの、いずれの費目でもほぼ左右対称で一山型のヒストグラムが得

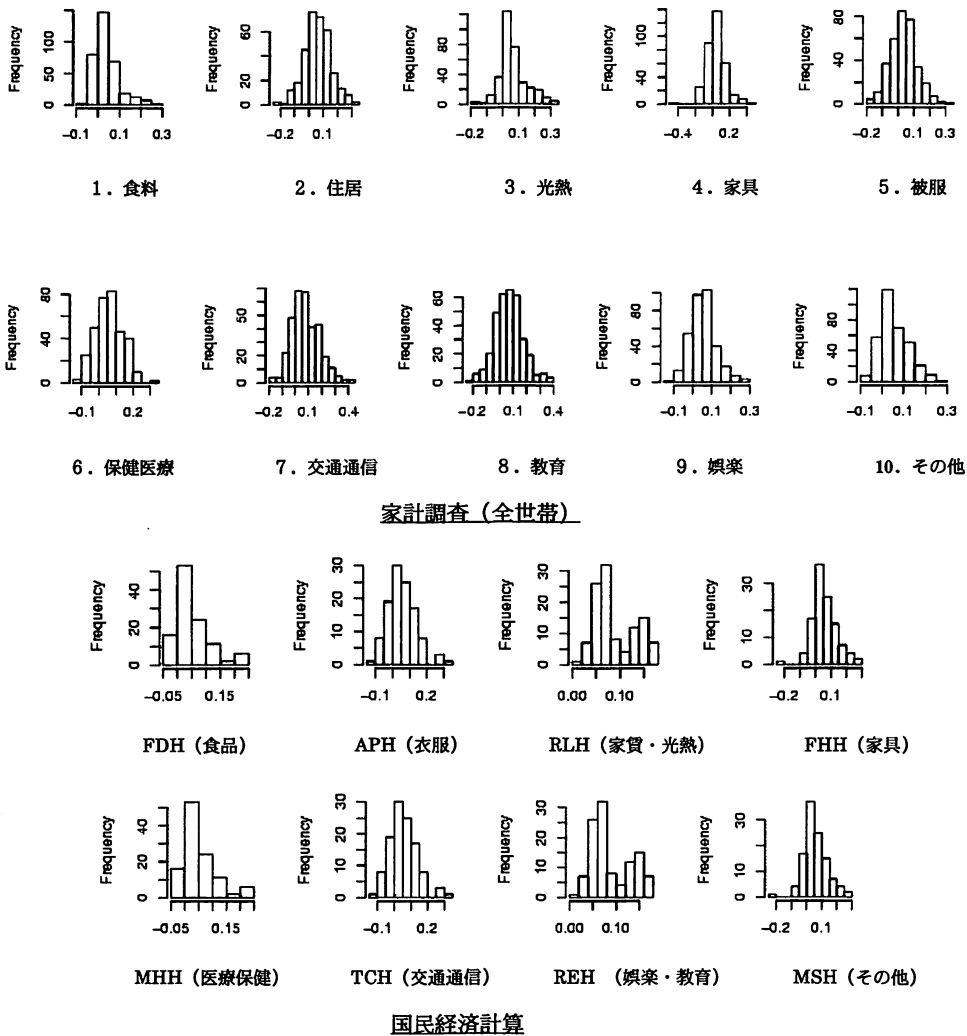
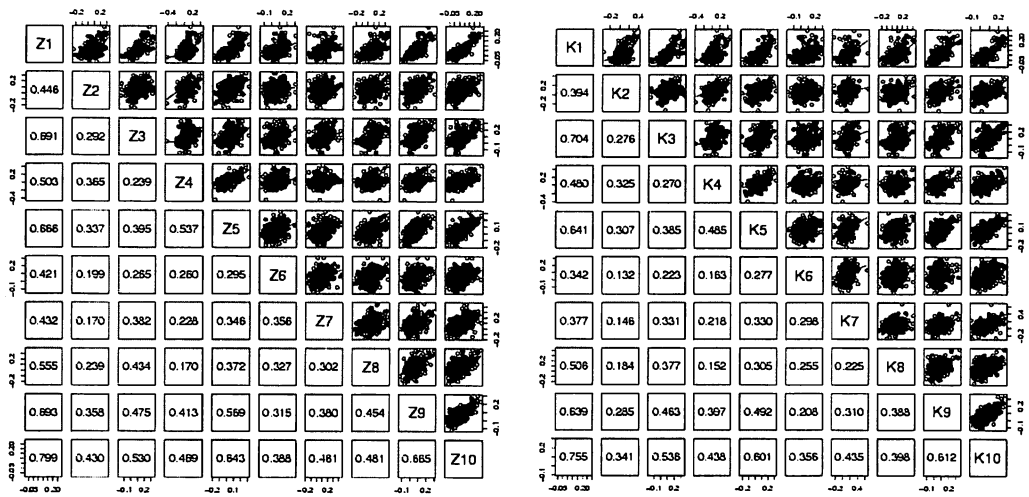


図3 費目別増加率のヒストグラム

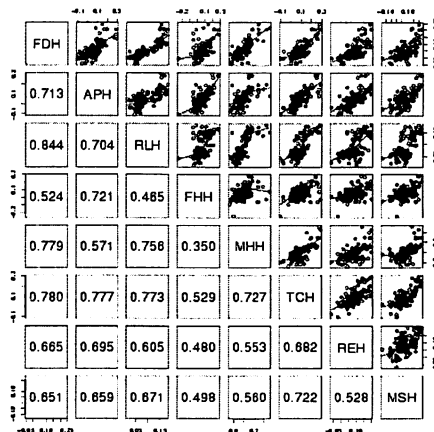
られている。なお、勤労者世帯データを用いても結果はほぼ同じであるので、スペースの関係から図は省略する。これに対し、国民経済計算ではRLH(家賃・光熱)とREH(娯楽・教育)の2費目で二山形となっており、層別の可能性が示唆される。このことは家計調査では異なる動きを持つ二つの費目、RLHについては住居と光熱、REHについては教育と娯楽を合計して費目が構成されていることによる可能性が考えられる。

次に、各費目間の散布図を行列形式にして、図4に示す。図4では、対角部分に費目番号(Zは全世帯、Kは勤労者世帯を示す)および略号、左下三角部分に相関係数、右上三角部分に散布図を示している。散布図および相関係数から、特に強い相関は見られない。若干の



家計調査(全世帯)

家計調査(勤労者世帯)



国民経済計算

図4 相関・散布図行列

外れ値が見られるものの、層別の可能性や強い非線形の関係を示しているものも特に見られない。

先の増加率のヒストグラムでも観察されたように、家計調査の2種のデータ、全世帯と勤労者世帯データの間ではほとんど違いは観察されないが、家計調査と国民経済計算の結果には大きな違いが見られる。両調査の費目分類が異なることの影響と考えられる。

次に、相関係数のヒストグラムおよび正規カーネルを用いた推定密度を図5に示す。

図より、ヒストグラムおよび推定密度は、家計調査、国民経済計算ともほぼ左右対称であるものの、家計調査ではプラス側に、国民経済計算ではマイナス側に若干歪んでいる。相関係数はすべて正の値である。国民経済計算では、過半数が0.6以上の強い相関を示しているが、

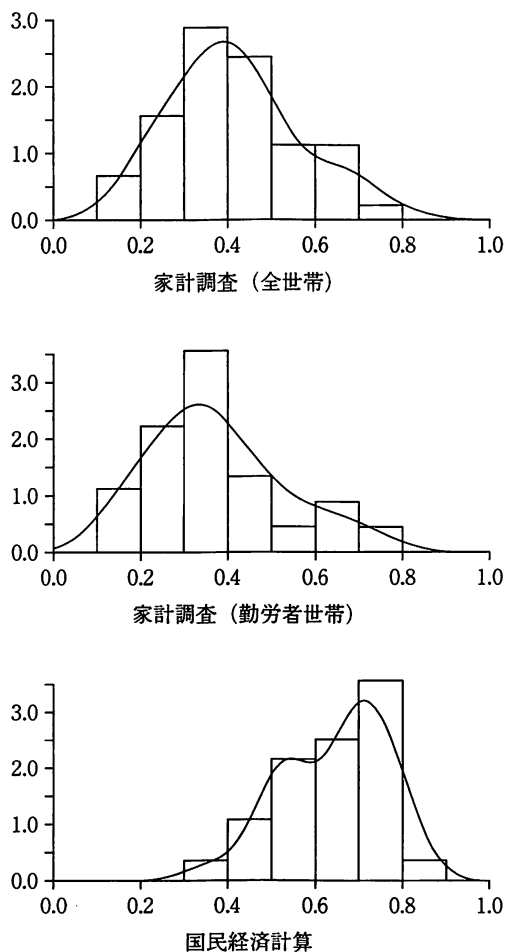


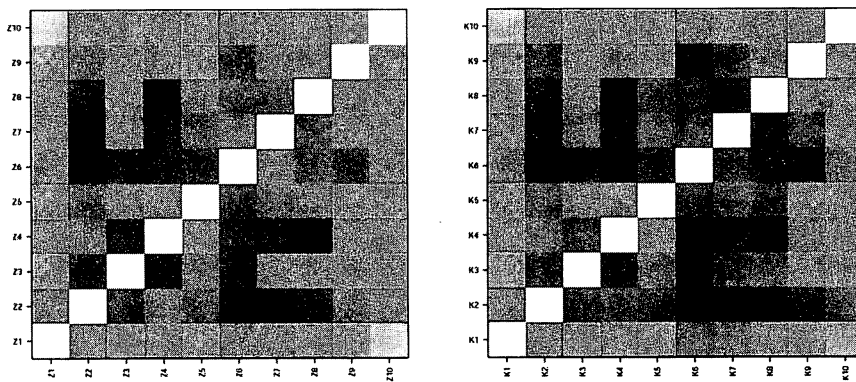
図5 調査別対数増加率のヒストグラム

家計調査の場合は、小さな値の相関係数が多く見られる。

家計調査の2種類のデータによる結果を比較すると、上記の2つの傾向（プラス側への歪み、国民経済計算に比較して相関係数の値が小さいこと）は、勤労者世帯でより強く観察される。全世帯は、勤労者世帯および個人営業世帯、その他の世帯を含んでいる。このことより、個人営業世帯あるいはその他の世帯のみのデータが利用可能ならば、その属性は勤労者世帯と異なる可能性があるかと推測できる。

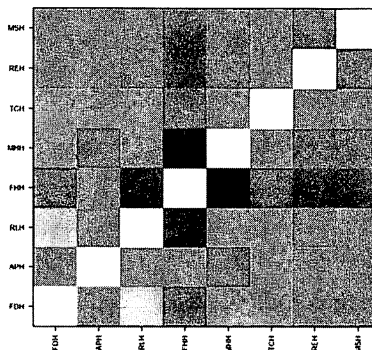
しかしながら、家計調査の結果は全世帯の場合であれ、国民経済計算を用いた結果とはその性格が大きく異なる。家計調査における全世帯は注5で述べたように日本で消費活動を行う全ての家計を対象とはしていない点には留意すべきであるが、以上の結果より、両調査の費目分類が異なることによる効果が大いことがヒストグラムより読みとることができる。

相関の強さを色の違いで表示すると、相関行列は図6のようになる。各グリッドは相関行列の要素に対応し、相関係数が1に近いほど白く、0に近いほど黒く表現されている。これ



家計調査 (全世帯)

家計調査 (勤労者世帯)



国民経済計算

図6 相関係数のイメージ図

は、バイオインフォマティクスでよく用いられる関係分析の手法である。これより家計調査の二つのデータの相関行列が似ているのがわかる。ここでも、家計調査の2種のデータは、対象世帯が異なるにもかかわらずそのパターンにほとんど違いがないことが観察される。一方、国民経済計算のパターンはこれらと明確に異なっている。

3.3 デンドログラム

Mantegna (1998) に従い、距離 $d(i, j) = \sqrt{2(1-r_{ij})}$ に基づいて距離（非類似度）行列 $D = [d(i, j)]$ を定め、デンドログラムを求める。最短距離法によるデンドログラムを図7に示す。

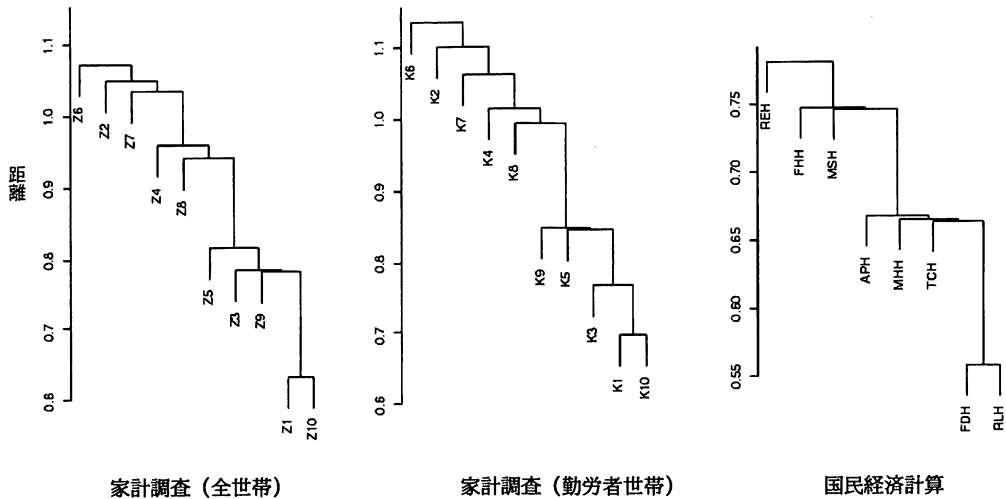


図7 デンドログラム

図7より、最も距離が短い費目は、家計調査では全世帯、勤労者世帯データとも1. 食料と10. その他である。これに対し、国民経済計算データではFDH（食品）とRLH（家賃・光熱）の距離が最も近い。これらがクラスタリングの最初に統合され、一つのクラスターを構成する。順次距離の離れた費目まで考慮していくことにより、他の費目がこのクラスターに結合されていく。

家計調査を用いた場合、全世帯を対象としても勤労者世帯を対象としてもほぼ同じ木構造が得られている。1. 食料と10. その他のクラスターに次の段階で近い費目は3. 光熱、5. 被服、9. 娯楽であり、続いて、8. 教育、4. 家具、7. 交通通信、2. 住居、6. 保健医療の変化のパターンが似ていることがわかる。全世帯と勤労者世帯で異なったのは、9. 娯楽の距離が、全世帯では他の8品目中1. 食料と10. その他に一番近かったのに対し、勤

労者世帯では、3. 光熱、5. 被服に次いだ位置にあることのみである。

一方、国民経済計算では、FDH(食品)とRLH(家賃・光熱)に続いて、TCH(交通通信)、MHH(医療保険)、APH(衣服)の3費目が、さらにMSH(その他)、FHH(家具)、最後にREH(娯楽・教育)がつけ加わる形で費目が結合されていく。

家計調査、国民経済計算のいずれのデータでも食品は類似の中心となる費目であるが、国民経済計算ではRLH(家賃・光熱)がこれともっとも関連する費目となっているのに対し、家計調査では3. 光熱の位置は第2グループに属するものの、2. 住居の位置は10品目中9番目と食品から離れた費目であるとの結果が得られている。被服や家具の位置は両データで大きな違いは見られないが、国民経済計算で第2グループに属するTCH(交通通信)やMHH(医療保険)が家計調査ではもっとも距離が離れている(それぞれ7、6が対応)であり、逆に、家計調査では重要な費目と考えられる8. 教育や9. 娯楽が、国民経済計算(REH)では大きく離れて観察されるなど、調査により、費目の位置関係についてはかなり大きな違いが見られる。

デンドログラム全体としては結合が1本の鎖状(chain)になっており、明確なクラスターを得ていない。もっとも関連する2つの費目から見た他の費目の位置関係、いずれの費目が近しいかについては判明するものの、消費費目全体のグループ分けはできていない。これは、最小距離法が持つ欠点であり、次に求めるMSTはこの欠点を補う意味も持つ。

3.4 最小木

クラスター分析の最短距離法は、構成の仕方から推測できるようにMSTと同値なものである⁹⁾。MSTは最終的な空間配置の姿であり、デンドログラムはそれに至るプロセスを示している。これらは異なる情報を持ち、そのため両者を作成して考察する必要がある¹⁰⁾。

家計調査データおよび国民経済計算データに対して得たMSTを図8に示す¹¹⁾。図の各楕円(ノード)は費目を示し、費目番号または略号をラベルとして記している。直線で結ばれている費目が、全体として、もっとも関連の強い費目と考えることができる。直線につけたラベルが費目間の距離の大きさである。

図8より、全世帯、勤労者世帯の両者において1. 食品を中心とするクラスターが構成されている。距離の大きさは別として、両者で位置付けが異なるのは6. 保健医療の位置である。全世帯では1. 食品と直接関連づけられているが、勤労者世帯では10. その他を通じて間接的に結びついている。

一方、国民経済計算では明確なクラスターは構成されていないが、強いて言えばFDH(食品)とTCH(交通通信)を中心とする2つのグループが形成されているといえる。FDH(食

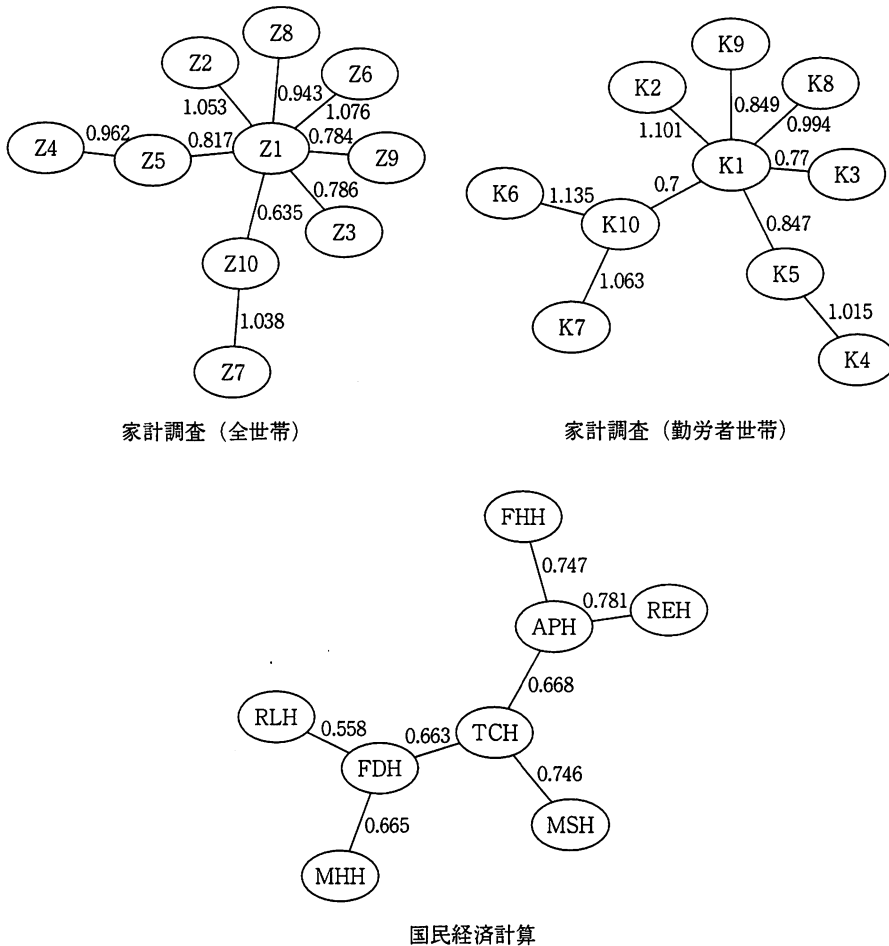


図8 費目間距離に基づくMST

品)を中心とするのはRLH(家賃・光熱)とMHH(医療保健)という生活の根幹に関わる、生活必需度の高い費目であり、TCH(交通通信)に関連するのは、MSH(その他)およびAPH(衣服)、REH(娯楽・教育)、FHH(家具)といった生活に余裕が出てきた後その支出が増えると思われるより必需度の低い費目である。

相関行列の推定には誤差が伴うため、ランダムな系列であってもノード間に見かけ上の距離が発生し、MSTを求めることが可能となる。そのため図8に示すMSTが意味のあるものなのか、あるいは単なるデータのばらつきや推定の誤差による偶然的産物なのかを客観的に判断する必要がある。Mantegnaはこの方法を与えていない。次節では、ネットワーク理論を用いて、これを客観的に判断するための方法を提示する。

4 ネットワーク分析

本節では、ネットワーク理論を援用して、第3節で求めた MST の分析を進める。図8において他のノードを多く集めているもの、すなわち、連結の中心となるものが見られるが、これらに注目し、これらが偶然の結果ではないことを客観的に判定する。本稿ではそのための方法を二つ考える。一つは、中心となるノードの「中心性」が他のノードのそれと有意に異なるかどうか検証する方法であり¹²⁾、もう一つはランダムに生成したネットワークの中心性の分布と図8に示した MST の中心性の分布とを比較する方法である。

一般にグラフ G は、ノードとノード間を結合する線の集まりである。これらの関係は接続行列を用いて記述できる。グラフの最も簡単な表現方法として、接続行列 R の要素 x_{ij} を、

$$x_{ij} = \begin{cases} 1 & \text{ノード } a_i \text{ と } a_j \text{ が結合されているとき} \\ 0 & \text{結合されていないとき} \end{cases}$$

と定めることができる。

中心性の指標の代表的なものに次数 (degree) がある。これは、ネットワークにおいて接続数の最も多いノードがグラフの中心であるという考え方に基づく指標である。 a_i ノードの次数 $d(a_i)$ を

$$d(a_i) = \sum_{j=1}^n x_{ij}$$

と定義する¹³⁾。この次数を用いて、第3節で得た MST の有効性を検証する。

MST より各費目の次数を求め、そのヒストグラムを図9に示す。ヒストグラムより次数1のノードが圧倒的に多いが、家計調査では大きな次数のノードが各一つ(食品)存在することがわかる。これら大きな次数の銘柄が他の銘柄と有意に異なるかどうかをチェックするため、箱ひげ図を作成し、図10に示す。

次数の箱ひげ図より、全世帯と勤労者世帯では上部のひげを越える点が1点ある。この点に対応する費目が食品である。一般に箱ひげ図で、ひげを超える点は外れ値と判断する。よって、これらの銘柄はランダムな変動のために偶然大きな次数を持ったのではなく、他の次数とは統計的に異なると考えることができる。すなわち、家計調査では、食品は他の費目と異なり、特別な位置を占めている。これに対して国民経済計算では、特別な位置を占める費目があることを次数から示すことはできない。

上記の方法は次数の分布が大きく歪んでいる場合に有効でない可能性がある。そのためもう一つの方法も試みる。ランダムに発生させたデータの相関行列より MST を構成し、その最大次数を求める。家計調査に対しては、長さ348(期数に対応)の乱数を10組(費目数に対応)

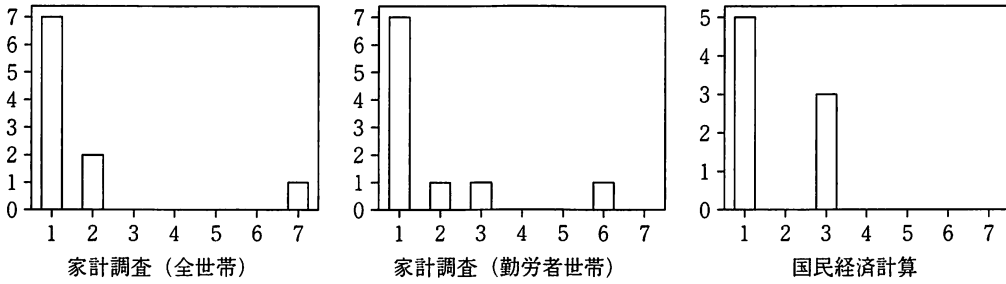


図9 次数のヒストグラム

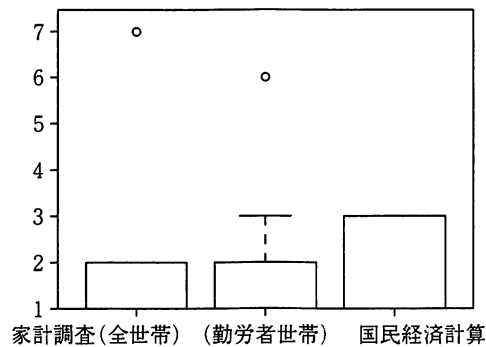


図10 次数の箱ひげ図

発生させ、その相関行列を求める。これより MST を構成し、最大次数を求める。これを5000回繰り返し、5000個の MST の最大次数のヒストグラムを作成する。乱数として正規乱数、対数正規乱数、一様乱数を用いた場合を図11に示す。

シミュレーションより、正規乱数では最大次数が6となったのが20回、7は1回、8以上は0回であった¹⁴⁾。対数正規乱数では最大次数が6となったのが24回、7以上は0回、一様乱数では、最大次数が6のものは36回、7は1回、8以上は0回であった。いずれも、

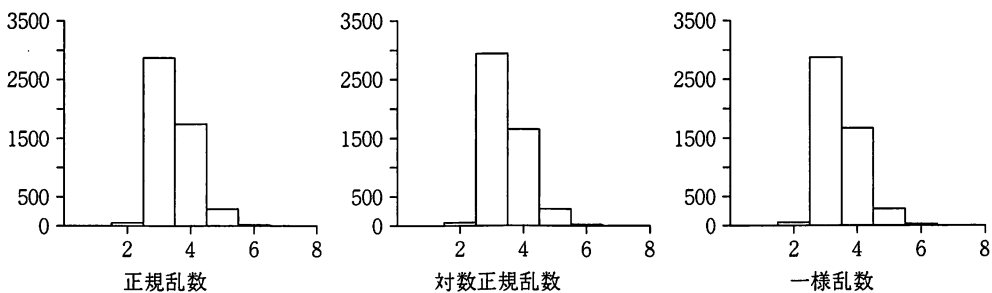


図11 次数のヒストグラム (シミュレーション結果)

ランダムに作成した MST の最大次数が 6 以上となる割合は 1%以下であり、最大次数が 6 以上となっている家計調査の MST はランダムなものとは異なると判断できる。

国民経済計算に対応する同様のシミュレーションを行ったが、これに関しては図 9 から予想されることではあるが、有意な結果が得られなかった。

5 終わりに

本稿では、Mantegna が提唱した方法を、家計調査、国民経済計算という 2 つの消費データに適用し、その費目間の空間構造について検討した。

株式データと比較するならばデータ数が少ないこと、今回利用したのが大費目であることなどの限界はあったが、消費費目間の階層構造について一定の結果が得られた。

家計調査については、全世帯を対象とするデータ、勤労者世帯のみを対象とするデータの 2 つを利用した。得られた結果はほぼ同じ傾向を示したものの、若干の違いが見られた。この点は今後より詳細なデータを用いて検討していく必要があるだろう。

家計調査では、費目が 1. 食料を中心に空間的に位置しているという明確な結論が得られたが、国民経済計算を用いた場合、途中のプロセスで含意のある結果が散見はされたものの、最終的には有意と考えられる結論は導き出せなかった。この最大の理由は、予備的考察で検討したヒストグラム(図 3)で指摘したように、RLH(家賃・水道)および REH(娯楽・教育)に層別の必要があること、言い換えるならば、現在の国民経済計算費目は、異種のもの混合になっているからであると考えられる。このことより、費目間の配分問題等個別消費の問題を考えていく上では、国民経済計算の利用は、十二分に注意して行う必要があるであろう。

なお、上記の国民経済計算にデータ上の問題点があったため、家計調査の住居費の動きが妥当と考えられるか、住居費目のために他の品目の支出行動の把握に問題が生じていないか等の問題については、必ずしも十二分な検討ができなかった。この点は、記して、今後の課題としたい。

注

- 1) 経済物理学を概観するには、Mantegna and Stanley (2000)、Bouchaud and Potters (2000) 参照。
- 2) この変換の意味については、Mantegna and Stanley (2000) 参照。
- 3) 以上、宮本 (1999) 参照。
- 4) MST の詳細については、たとえば Deo (1974) 参照。
- 5) 家計調査によると、家計調査の調査対象は、農林漁家世帯や単身世帯、あるいは外国人世帯を除く世帯であり、その全体を「全世帯」と呼んでいる。全世帯は、勤労者世帯と勤労者以外の世帯(個人営

業世帯やその他の世帯、無職世帯も含まれる)に二分される。勤労者世帯は、“世帯主が会社、官公庁、学校、工場、商店などに勤めている世帯”である。ただし、世帯主が会社団体の役員である世帯は、その他の世帯に分類される。

- 6) データの期間は、国民経済計算が新 SNA に移行した1970年以降とした。また、計算の時点で、年単位で見て国民経済計算の確報が得られていた1998年までとした。
 なお、本文中で述べた調査対象者の問題を解決するために、総務庁は、2000年より「家計調査」に農林漁家世帯を含め結果を公表している。また、1995年から、単身世帯の消費を把握するため「単身世帯収支調査」を開始しており、今後、2人以上世帯を対象とした「家計調査」と「単身世帯収支調査」の結果を合わせ、「家計総世帯集計結果」として四半期ごとに公表していくことを発表している。本稿で用いたのは、これらの変更が行われる以前の期間のデータである。
- 7) 先に Araki and Hashimoto (2002) で、家計調査の食品12費目について同様の分析を行ったが、その際、12費目ともで各年末に(11月から12月にかけて)支出額が大きく増加、翌年年頭には減少する動きが観察された。また、食品以外でも、教育支出は学期始めに増加するなど、費目独自の季節パターンがいくつか観察された。この問題を解決するため、前期比ではなく対前年比の対数値を用いて分析を行った。
- 8) 以下、特に断らない限り、計算および図の作成は GNU プロジェクトの一つであるデータ分析ソフトウェア R およびそのライブラリを利用して作成している。詳しくは The Comprehensive R Archive Network (<http://cran.r-project.org/>) 参照。
- 9) 宮本 (1999) 参照。
- 10) Legendre and Legendre (1984) 参照。
- 11) グラフの作成には、オープンソース・ソフトウェアの GraphViz (<http://www.research.att.com/sw/tools/graphviz/>) を用いた。
- 12) 荒木 (2001) は日経平均データでこの方法を試みた。
- 13) 中心性については Wasserman and Faust (1994) 参照。
- 14) 最大次数はノード数を n とするとき、 $n-1$ となる。

参考文献

- [1] 荒木孝治 (2001). 「株式市場における空間構造」, 『関西大学商学論集』第46巻第5号, 1-18.
- [2] Araki, T. and N. Hashimoto (2002). Hierarchical structure in food consumption. *Kansai University Review of Economics* 4 (forthcoming).
- [3] Bonanno, G., N. Vandewalle and R. N. Mantegna (2000). Taxonomy of stock market indices. *Physical Review E* 62, R7615-R7618.
- [4] Bouchaud, J-P, and M. Potters (2000). *Theory of financial risks - From statistical physics to risk management*. Cambridge University Press.
- [5] Deo, N. (1974). *Graph theory with applications to engineering and computer science*. Prentice Hall.
- [6] Legendre, P. and L. Legendre (1984). *Numerical ecology*. Elsevier.
- [7] Mantegna, R. N. (1998). Hierarchical structure in financial markets. *European Physical Journal B* 11, 193-197.
- [8] Mantegna, R. N. (1999). Information and hierarchical structure in financial markets. *Computer Physics Communications* 121-122, 153-156.
- [9] Mantegna, R. N., and H. E. Stanley (2000). *An introduction to econophysics - Correlation and*

complexity in finance. Cambridge University Press.

- [10] 宮本定明 (1999). 『クラスター分析入門 ファジィクラスタリングの理論と応用』 森北出版.
- [11] Wasserman, S. and K. Faust (1994). *Social network analysis: Methods and applications*. Cambridge University Press.