新しい経済指標作成の方法*

刈　屋　武　昭

§1 序

経済指標が用いられる領域は，物価指数，株価，福祉指標，景気指標，等，指標として表現する対象が一般に複雑で，必ずしも明確な概念として把握できにくい領域である。そこでは，経済理論は，物価理論のように重要な考え方を提供する場合もあるが，現象の複雑性と対象の把握(概念化)の困難性のゆえに十分力を発揮しているとはいいがたい。また物価理論でも，指標作成時における系列選択問題に対しては，その力を失う。一般に経済指標は，問題とする経済状況を把握することを主たる目的とするが，その時系列的変化を通じてしばしば予測のために用いられる。周知のように景気指標では，先行指標という形で最初から予測を目的として作られるものもある。

経済指標における予測は，「経済構造の変化は時として急激であるとしても，全体としては滑らかな変化をし，それが種々のデータ系列に統計的規則性として反映される」という考え方に基づいているといえよう。経済構造をブラックボックスとして扱う時系列分析も，広い意味でこのような統計的規則性を用いていると考えられる。その意味で経済構造がデータ発生に関して時系列論が述べる定常的メカニズムを保有していなくても，時系列論の応用は可能となろう。他方，経済理論では因果関係的発想をとるため，たとえば為替レート y_t の変動を予測するためには，その変動原因としての変数 $x_{1t},\cdots,x_{pt}$ をとり，

(1.1)　　$y_t=b_0+b_1x_{1t}+\cdots+b_px_{pt}+u_t$

として，y_t の変動をその原因の変動に求め，それによって予測可能性を確保する。これに対して，純粋な時系列モデルでは

(1.2)　　$y_t=c_0+c_1y_{t-1}+\cdots+c_py_{t-p}+u_t$

と表現する。また経済指標では，(1.1)の右辺を用いて

(1.3)　　$z_t\equiv a_0+a_1x_{1t}+\cdots+a_px_{pt}$

と表現する。もちろん為替レート y_t のように記述あるいは予測する対象が実数値の形で明確に定義されている場合，(1.3)のアプローチは一般にとられない。それは y_t の変動の情報を積極的に用いていないためである。しかし為替レートのように複雑で，かつ因果関係が十分はっきりしないような経済変数に対しては，(1.1)や(1.2)のアプローチでも必ずしも十分でない。こと予測に関しては特にそうである。いくつかの x_{it} は，原因でなく実は y_t の変動の結果と考えられる場合も多い。

本論文では

① 為替レートや株価等のような因果関係が必ずしもはっきりしない経済変数に対して，(1.1)，(1.2)のアプローチを補完するものとして(1.3)の視点によるアプローチを提案する。ただし，ここで考えている予測の対象となる変数は，為替レート，株価のような先物市場が成立し，今期の説明変数と将来期の予測の対象となる変数が相関(共変)関係をもつ場合である。

② 景気指標や福祉指標などで，前もって景気感や福祉の向上の感じを表わすものが，たとえばサーベイデータ等として利用可能なとき，そういった感じを表現する指標を，経済環境を表わすと考えられる個別系列を用いて作成する方法を示す。景気指標では，日銀の業況判断 DI(ディフュージョン・インデクス)が，その種のサ

＊　この指標作成法の開発にあたって，山一証券経済研究所の福山真弘，太田八十雄，山田貞二郎，原誠一，の各氏から景気指標との関連で多大な御教示を頂いた。ここに深く感謝する次第である。

ーベイデータの例である。またロジット・プロビット分析法との関連を述べる。

③　①, ②で用いる指標化の方法は，主成分分析をクラスター的分析と，指標化自体に2度用いるものであるが，クラスター的分析の視点は回帰分析での変数選択法として役に立つことを指摘する。

この指標作成の方法は，[5], [6] において応用された。その視点は②に基づくものである。過去3年間の間に指標のパフォーマンスはチェックされ，十分満足いくものであると判断されている。

§2　参照変数と指標作成の方法

この節では，為替レートの変動の記述予測等のように外から参照変数あるいは目的変数という形で表現対象が与えられている場合(外的基準がある場合，すなわち教師つきの場合)について，新しい指標作成の方法を示す。次節ではその参照変数が直接に得られないために，サーベイデータにその役割を求める場合を考察する。

一般に，経済指標作成の方法は，次のステップを踏んで行われる。

Ⅰ. 作成しようとする指標(表現対象)にレレヴァントと考えられるデータ系列を収集する。

Ⅱ. 収集された系列から指標作成のための系列を選択する。

Ⅲ. 選択された系列を指標化する。

Ⅳ. 作成された指標のパフォーマンスをチェックする。

上記の指標作成のプロセスで，基本的に重要であり，またむづかしいステップはⅡである。実際，多くの指標作成のプロセスで恣意性が最も多く介入するのがこの段階である。この困難性は表現対象が複雑で明確に定義できない場合はもとより，為替レートのように表現対象がはっきりしている場合でも起る。後者の場合は回帰分析の変数選択問題に似ている。実際§4で示すように，以下のⅡのステップを処理する方法は，回帰分析の変数選択法としても用いることができる。他方，表現対象が明確な形で与えられない場合として，たとえば景気指標では，一般に経済的重要性，転換のタイミング，データの速報性等，いくつかの基準を設け，その基準に照して専門家が判断して系列の選択が行われる。この論文では，すでに述べたように，参照変数という形で外から指標の表現対象ともいうべき変数が与えられていることを仮定するため，Ⅱの系列の選択が客観的になされる。もちろん参照変数の存在を想定することは，景気等の概念を単純化しすぎるため，その方法の応用を制約するものであるが，次節でみるようにサーベイデータを時系列的に得ることによって，景気指標，福祉指標，生計費指数等多くの場合に適用可能となる。また通産省の鉄鋼業景気指標 [4] では，最初から景気を「売上高営業利益率の変動」と定義しているため，これがそのまま参照変数となる。したがってこの場合，売上高営業利益率にレレヴァントな多くのデータ系列の集合から，以下に述べる方法によって客観的に系列を選択し，指標化が可能となる。

まず y_t を表現対象としての参照変数とし，x_{it} は y_t を指標化するため収集された p 種のデータ系列とする $(i=1, \cdots, p)$。t は時点で $t=1, \cdots, n$ とする。$x_{0t}\equiv y_t$, $\boldsymbol{x}_t=(x_{0t}, \cdots, x_{pt})$ とおく。以下では理論的基礎をはっきりさせるため，形式的に y_t, $\boldsymbol{x}_t$ を確率変数とし，$\boldsymbol{x}_t$ の母数との関係で選択の系列を議論する。

(仮定) (1)各系列 x_{it} は，平均 $E(x_{it})=\mu_{it}$, 分散 $\mathrm{Var}(x_{it})=E(x_{it}-\mu_{it})^2$, 共分散 $\mathrm{Cov}(x_{it}, x_{jt})=E(x_{it}-\mu_{it})(x_{jt}-\mu_{jt})$ をもつ確率変数で，分散，共分散は時間に関係なく一定とする。すなわち

$$\mathrm{Var}(x_{it})=\sigma_i^2\equiv\sigma_{ii}, \mathrm{Cov}(x_{it}, x_{jt})=\sigma_{ij} \tag{2.1}$$

(2) x_{it} は t に関して互いに無相関

$$\mathrm{Cov}(x_{it}, x_{is})=0 \quad (t\neq s) \tag{2.2}$$

とする。

仮定(1)は多くのデータ系列に対して満たされていると考えてよい。他方仮定(2)は，必ずしも満たされず，それが満たされない場合，その情報を利用した指標化があり，それを後に示す。(2.2)をチェックする方法としてはアンダソンの検定がある。以下上の仮定のもとに系列の選択と指標化

の方法を述べる。

① まず $x_{0t} \equiv y_t$ の変動に興味をもつため，各変数を

(2.3) $z_{it} = (x_{it} - \mu_{it})/\sigma_i \qquad (i=0, 1, \cdots, p)$

と基準化する。従って z_{it} は平均 $E(z_{it})=0$，分散 $\mathrm{Var}(z_{it})=1$ となる。z_{0t} の変動を指標化することを考え，$z_{1t}, \cdots, z_{pt}$ のうち比較的少数の q 個の変数で

(2.4) $z_{0t} \simeq a_1 \tilde{z}_{it} + \cdots + a_q \tilde{z}_{qt}$

と z_{0t} の変動を指標化することを狙う。ここで $\tilde{z}_{1t}, \cdots, \tilde{z}_{qt}$ は $z_{1t}, \cdots, z_{pt}$ のうち z_{0t} の指標のために選択される系列である。このため

(2.5) $\boldsymbol{z}_t = (z_{0t}, z_{1t}, \cdots, z_{pt})'$

を定義する。そして $\boldsymbol{z}_t$ に因子分析的発想に基づく主成分分析をかけ，

(2.6) $\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{f}_t, \qquad \boldsymbol{f}_t = (f_{0t}, \cdots f_{pt})'$

と表現されたとする。ここで $\boldsymbol{f}_t$ は

(2.7) $E(\boldsymbol{f}_t)=0, \qquad \mathrm{Cov}(\boldsymbol{f}_t) = E(\boldsymbol{f}_t \boldsymbol{f}_t') = \boldsymbol{I}$

をもつ確率変数で，$\boldsymbol{A}$ は

(2.8) $\boldsymbol{\Omega} \equiv \mathrm{Cov}(\boldsymbol{z}_t) = \boldsymbol{A}\,\mathrm{Cov}(\boldsymbol{f}_t)\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{A}'$

となる $(p+1)\times(p+1)$ の行列である。z_{jt} は x_{jt} を基準化した変数であるから，(2.8)の $\boldsymbol{\Omega}$ は x_{jt} の相関行列である。(2.6)(2.7)(2.8)から明らかなように，$\boldsymbol{A}$ と $\boldsymbol{f}_t$ は一意的には決まらない。実際，$(p+1)\times(p+1)$ の任意の直交行列 $\boldsymbol{Q}$ に対して

(2.9) $\hat{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{Q}, \qquad \tilde{\boldsymbol{f}}_t = \boldsymbol{Q}'\boldsymbol{f}_t$

と変換した $\hat{\boldsymbol{A}}, \tilde{\boldsymbol{f}}_t$ も (2.6)～(2.8)を満たす。以下の議論は1つの $\boldsymbol{A}$ と $\boldsymbol{f}_t$ の選択に基づいているが，後にみるようにその選択の仕方は結果に影響しない。また以下の議論では $\boldsymbol{\Omega}$ したがって $\boldsymbol{A}$ は既知とする。もちろん実際には $\boldsymbol{\Omega}$ および $\boldsymbol{A}$ はデータから推定する。さて $\boldsymbol{z}_t$ の第 i 要素は

(2.10) $z_{it} = \alpha_{i0} f_{0t} + \alpha_{i1} f_{1t} + \cdots + \alpha_{ip} f_{pt} \qquad (i=0, 1, \cdots, p)$

と書け，z_{it} の変動は，互いに無相関で分散1の変量 f_{jt} の変動を α_{ij} 倍(符号も含めて)受けているとみることができる。実際(2.7)と(2.10)より

$$1 = \mathrm{Var}(z_{it}) = \alpha_{i0}^2\,\mathrm{Var}(f_{0t}) + \cdots + \alpha_{ip}^2\,\mathrm{Var}(f_{pt}) = \alpha_{i0}^2 + \alpha_{i1}^2 + \cdots + \alpha_{ip}^2$$

がでる。したがって $|\alpha_{ij}| \leq 1$ であり，分散でみた z_{it} の変動の $\alpha_{ij}^2 \times 100\,\%$ が f_{jt} の変動の影響によるものである。それゆえ(2.10)のウェイト α_{ij} の絶対値は，f_{jt} の変動の z_{it} の変動に与える影響の強さを示している。

② 参照変数 y_t を基準化した変量 z_{0t} のウェイト $\alpha_{0j}(j=0, 1, \cdots, p)$ に注目する。上の議論から

(2.11) $\mathrm{Var}(\alpha_{0j} f_{jt})/\mathrm{Var}(z_{0t}) = \alpha_{0j}^2$

であることに注意して，α_{0j}^2 を大きい順に並べる。j は f_{jt} の番号にすぎないので適当に番号をつけかえることで，あるいは同じことだが(2.6)で $\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{P}\boldsymbol{P}'\boldsymbol{f}_t$，ただし $\boldsymbol{P}$ は適当な順列行列，と表わし，$\boldsymbol{A}\boldsymbol{P}$ および $\boldsymbol{P}'\boldsymbol{f}_t$ をそれぞれ新しく $\boldsymbol{A}$ および $\boldsymbol{f}_t$ と定義することで，一般性を失うことなく

(2.12) $\alpha_{00}^2 \geq \alpha_{01}^2 \geq \cdots \geq \alpha_{0p}^2$

と仮定できる。したがって z_{0t} の変動は，第0変量 f_{0t} の影響を最も強く受け，第1変量 f_{1t} の影響を次に強く受ける，というように以下順に影響が小さくなっていく。次に z_{0t} の変動を構成する f_{jt} のうち，小さな不規則的変動としての影響しか与えないもの(すなわち $|\alpha_{0j}|$ が小さい f_{jt})を無視することにして，分散でみた z_{0t} の変動を70～80 % 説明する f_{jt} を大きい順にとりだす。すなわち

(2.13) $\eta_0 \equiv \alpha_{00}^2 + \alpha_{01}^2 + \cdots + \alpha_{0q}^2$

が0.7～0.8となる q をとり

(2.14) $z_{0t}^* = \alpha_{00} f_{0t} + \alpha_{01} f_{1t} + \cdots + \alpha_{0q} f_{qt}$

とおく。このとき

(2.15) $\mathrm{Var}(z_{0t}^*)/\mathrm{Var}(z_{0t}) = \eta_0$

となり，z_{0t}^* は z_{0t} の変動を $\eta_0 \times 100\,\%$ 近似しているとみることができる。z_{0t} に対して新しく z_{0t}^* を定義する理由は，z_{0t} の変動を f_{jt} の変動からみたとき，z_{0t} にとって不規則的な小さな変動(それは収集系列が多ければ多いほどこのような変動は多くなる)を取り除くことにある。主成分分析では，$p+1$ 個の変数 $\boldsymbol{z}_t$ の変動は，$p+1$ 個の互いに無相関な変数 $\boldsymbol{f}_t$ の変動で完全に記述されることになるが，そのうち z_{0t} に体系的に大きな影響を与える変数に注目することで，z_{0t} の変動をよりはっきり把握するのである。なお①で(2.6)～(2.8)を満たす $\boldsymbol{A}$ と $\boldsymbol{f}_t$ の選択が一意的でないことを述べたが，ここで(2.12)の条件を課し

たので，その選択の幅が非常に小さくなる。実際(2.12)で厳密な不等号が成立すると，(2.9)において(2.6)～(2.8)および(2.12)を満たす直交行列 $\boldsymbol{Q}$ は，対角要素に $+1$ または -1 をもつ対角要素だけである。実際の多くのデータでは，(2.12)の不等号が確率1で成立するため，$\boldsymbol{A}$ と $\boldsymbol{f}_t$ の一意性は符号の変換を除けば確保されることになる。

③　番号を付けかえられた f_{jt} に対して，参照変数以外の他の変量 z_{it} も，再び(2.10)と同じ形に表現される。ただしいまの場合，基準化した参照変数 z_{0t} に 70～80％の影響を与えるものが，最初の$(q+1)$個の変量$(f_{0t}, f_{1t}, \cdots, f_{qt})$である。$z_{ot}$ の変動と大まかな点で似ている系列を $z_{jt}(j=1, \cdots, p)$の中から選択するため，(2.14)と同様に

(2.16)　$z_{jt}{}^*=\alpha_{i0}f_{0t}+\alpha_{i1}f_{1t}+\cdots+\alpha_{iq}f_{qt} \quad (i=1, \cdots, p)$

とおき，z_{0t} と z_{it} の距離を

(2.17)　$d(z_{0t}, z_{it})\equiv E(z_{0t}{}^*-z_{it}{}^*)^2$

で定義する。(2.7), (2.14)および(2.16)から

(2.18)　$d(z_{0t}, z_{it})=\Sigma_{j=0}{}^q(\alpha_{0j}-\alpha_{ij})^2$

と計算される。この距離が小さい順に系列 $z_{it}(i=1, \cdots, p)$(したがって x_{it})を並べる。ここで添数 i は与えられた系列 x_{it} の番号であるから，一般性を失うことなく

(2.19)　$d(z_{0t}, z_{1t})\leq d(z_{0t}, z_{2t})\leq\cdots \leq d(z_{0t}, z_{pt})$

とすることができる。そこでこの距離の小さい順に系列を選択する。選択する系列数をいくつにするかについての客観的基準を，この論文では与えることができなかった。ただ系列数を少なくすると(たとえば1系列 z_{1t} のみ)，参照変数 z_{0t} により近くなるから参照変数と対応した変動をよりうまく作りだせると期待されよう。しかしその場合，選択された系列に特有な変動の影響が指標に現れやすく指標が不安定になりやすい。他方，選択系列数をふやしていくと，作られる指標は安定していくかもしれないが，指標の変動が参照変数の変動と次第に異なっていくだろう。その意味で，指標の目的と安定性の間にミニマックスポイントを与えるような基準が考えられよう。しかし安定性の尺度を考えつくことができなかったため，ここでは実際に計算される距離に基づいて，たとえば距離が0.1未満，0.2未満等，いくつかのケースを試み，安定性との勘案で適当な系列数を選択するといった試行錯誤的なものを提案しておく。また問題によっては，(2.19)で距離と距離の間隔が一様でなく，ある番号 i で大きくなっている，すなわち $d(z_{0t}, z_{it})-d(z_{0t}, z_{i-1,t})$ が大きくなっていることもありえよう。その場合 $i-1$ までの系列を選択してもよい。最後に技術的な点として，(2.18)からわかるように α_{0j} を $-\alpha_{0j}$ に，α_{ij} を $-\alpha_{ij}$ におきかえても(2.18)の距離は変らない。このことは，①, ② で述べた $\boldsymbol{A}$ と $\boldsymbol{f}_t$ の一意性の問題は結果に影響を与えないことを意味する。実際②の段階では，$\boldsymbol{A}$ と $\boldsymbol{f}_t$ を符号変化を与える対角行列 $\boldsymbol{Q}$ で $\boldsymbol{AQ}, \boldsymbol{Q}'\boldsymbol{f}_t$ に変換しうることを述べた。このことは $\boldsymbol{A}$ の各列の符号を任意に変化させてもよいことを意味するが，この変化に対して(2.18)の距離は不変である。

④　③で述べた視点から，仮に k 個の系列 $z_{1t}, \cdots, z_{kt}$ が選択されたとしよう。この k 個の選択系列を指標にまとめる方法として，$(z_{1t}, \cdots, z_{kt})$を主成分分析にかけ，その第1主成分でもって目的とする指標とする。すなわち

(2.20)　$w_t(k)=\beta_1z_{1t}+\cdots+\beta_kz_{kt}$

ただし$(\beta_1, \cdots, \beta_k)'$ は$(z_{1t}, \cdots, z_{kt})$の分散行列 $\boldsymbol{\Omega}_k$ の最大固有値に対する固有ベクトルで，$\beta_1{}^2+\cdots+\beta_k{}^2=1$ を満たすものである。なお，この指標化の段陽では z_{0t} を用いていないことに注意せよ。

⑤　系列数 k の妥当性，あるいは指標の目的性と安定性をチェックするため，(2.19)の距離の大きさをみながらいくつかの k を選び，④に述べた方法で指標化し，グラフに書いて参照変数 z_{0t} の動きと比較する。

以上がいわゆる一致指標作成法であり，k 期先を予測するための先行指標を作るには，上の $x_{0t}\equiv y_t$ を $x_{0t}{}^k\equiv y_{t+k}$ でおきかえればよい。明らかに為替レートなどをこのような方式で予測することは，回帰式の予測と異なる。上の方式による予測は共変関係による予測であって，因果関係による予測でない。

実際のデータに対しての具体的な方法は，上の議論から明らかであるが後に述べる。その前に若干のコメントをしておく。

1.　上のプロセスで恣意性が残っている部分は，(2.13)のη_0が0.7～0.8となるqを選択する部分と，(2.19)に基づいて系列を選択するときの系列数kの部分である。

2.　指標は参照変数の値を直接用いない。したがって(2.16)のz_{1t}^*はα_{ij}したがって$\boldsymbol{\Omega}$に依存するので，z_{it}^*をもとに(2.20)のように指標化できない。

3.　系列の選択が参照変数の動きに近いものから順に行われるので，④の指標化では常に第1主成分が目的とするものになる。またその場合の第1主成分の説明率(寄与率)の大きさは，指標の良し悪しには直接的には関係しない。

4.　(2.17)のz_{0t}とz_{it}の距離の定義は，直接的な距離

$$\bar{d}(z_{0t}, z_{it})=E(z_{0t}-z_{it})^2=\Sigma_{j=0}{}^p(\alpha_{0j}-\alpha_{ij})^2$$

でなく，z_{0t}^*とz_{it}^*の距離としたのが上の系列選択法の特徴である。

5.　(2.20)の$w_t(k)$はz_{0t}の変動を指標化したものであるが，$z_{0t}=(y_t-\mu_{0t})/\sigma_0$であるから，$z_{0t}$を$w_t(k)$でおきかえると

(2.21)　$$y_t \simeq \mu_{0t}+\sigma_0 w_t(k)$$

となり，実際の予測(その場合はすでに述べたようにy_tはy_{t+k}となる)ではy_tのトレンドμ_{0t}とy_tの分散の推定値が必要となる。ここではμ_{0t}は簡単な代数式$a+bt$やae^{bt}等を仮定してy_tの系列から推定することを提案しておく。

6.　最初のデータ収集系列は，季節変動は必ずしも除去する必要はないが，不規則変動は移動平均等で除いておくことが望ましい。また，対数等適当な変換をほどこしておいてもよい。さらに，たとえばx_{1t}と$\log x_{1t}$を異なるデータ系列として一緒に入れてかまわない。

次に仮定(2)が満たされていない場合を簡単に扱っておく。この場合，z_{1t}とz_{is}とは系列相関をもつことになるが，規範的な考え方を求めるために$\boldsymbol{z}_t$に正規分布を仮定する。すなわち$\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{z}_1{}' \\ \vdots \\ \boldsymbol{z}_n{}'\end{pmatrix}$とおき

(2.22)　$$\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{\Phi}\otimes\boldsymbol{\Omega})$$

を仮定して論をすすめる。ここで$\boldsymbol{\Omega}$は前と同じように各$\boldsymbol{z}_t$の分散行列であるが，$\boldsymbol{\Phi}$に系列相関を表わす($\boldsymbol{Z}$の各列の)分散共分散行列である。$\boldsymbol{\Phi}$が全く未知であるとパラメータ数がデータ数より多くなり推定不能となる。ここでは，$\boldsymbol{\Phi}$は既知の直交行列で対角化可能であるとする。この仮定は，一階の系列相関等で近似的に満たされるが，$\boldsymbol{Z}$の共分散行列が(2.22)のように書けるということは各$\{z_{it}\}$の系列相関のパターンが共通であるという強い仮定を意味する。さらにパラメータを減らすため，$\boldsymbol{\Phi}$の固有値λ_iに対し，$\lambda_i=\lambda_i(\boldsymbol{\theta})(\boldsymbol{\theta}: q\times 1, q$一定)とする。この場合，

(2.23)　$$\boldsymbol{W}\equiv\boldsymbol{\Omega}^{-1/2}\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{w}_1{}' \\ \vdots \\ \boldsymbol{w}_n{}'\end{pmatrix}$$

とし，$\boldsymbol{w}_t$に対し上記①～④のプロセスをふめばよい。もちろん実際のデータでは，$\boldsymbol{\Omega}$はその推定値でおきかえる。

最後に，上で述べた方法を実際のデータに適用することを考える。$\boldsymbol{x}_t, x_{0t}\equiv y_t$をこれまで通りとし，不規則変動調整済とする。もちろん必要ならば季節調整もしておく。以下の①～④のプロセスは，上記①～④に対応する。

①　各系列$x_{it}(i=0, 1, \cdots, p)$からトレンドを除去する。すでに述べたようにトレンドは簡単な代数曲線を最小2乗法で求めるものとする。x_{it}からトレンド除去した変量をX_{it}とし，X_{it}の算術平均と分散

(2.24)　$$\bar{X}_i=\Sigma_{t=1}{}^n X_{it}/n,$$
$$s_i{}^2=\Sigma_{t=1}{}^n(X_{it}-\bar{X}_i)^2/n$$

を求め，基準化変量

(2.25)　$$z_{it}=(X_{it}-\bar{X}_i)/s_i \qquad (i=0, 1, \cdots, p)$$

を作る。そして$\boldsymbol{z}_t=(z_{0t}, z_{it}, \cdots, z_{pt})'$の分散行列($(y_t, x_{1t}, \cdots, x_{pt})$の相関行列)

(2.26)　$$\boldsymbol{R}=\Sigma_{t=1}{}^n\boldsymbol{z}_t\boldsymbol{z}_t{}'/n$$

を計算する。$\boldsymbol{R}$の対角要素は1である。$\boldsymbol{R}$は(2.

8)の $\boldsymbol{\Omega}$ の推定値とみなす。$\boldsymbol{R}$ に基づく主成分分析をかけ

(2.27)　$\boldsymbol{z}_t=\boldsymbol{A}\boldsymbol{f}_t$

となる $\boldsymbol{A}$ と $\boldsymbol{f}_t$ を求める。ただし $\boldsymbol{A}=(a_{ij})$ は $\boldsymbol{R}=\boldsymbol{A}\boldsymbol{A}'$ を満たす。

② (2.27)の第1行に注目し，

(2.28)　$a_{00}{}^2\geq a_{01}{}^2\geq\cdots\geq a_{0p}{}^2$

となるように $\boldsymbol{f}_t$ の要素と $\boldsymbol{A}$ の要素を並べかえておく。そして

(2.29)　$h=a_{00}{}^2+a_{01}{}^2+\cdots+a_{0q}{}^2$

が0.7～0.8となる q を選択する。

③ 距離を計算するため，

(2.30)　$z_{it}{}^*=a_{i0}f_{0t}+\cdots+a_{iq}f_{qt}$　$(i=0,\cdots,p)$

とおく。このとき(2.18)の距離

(2.31)　$d(z_{0t},z_{it})=\Sigma_{t=1}{}^n(z_{0t}{}^*-z_{it}{}^*)^2/n$
$=\Sigma_{j=0}{}^q(a_{0j}-a_{ij})^2$

と計算される。ここで $\Sigma_{t=1}{}^n\boldsymbol{f}_t\boldsymbol{f}_t'/n=\boldsymbol{I}$ を用いた。したがって(2.30)の $z_{it}{}^*$ を求めることなく(2.31)の右辺を(2.27)の $\boldsymbol{A}$ から計算すればよい。ただしその場合の $\boldsymbol{A}$ は，(2.28)を満たすように全体を並べかえたものである。z_{it} を距離の小さい順に並べる。一般性を失うことなく

(2.32)　$d(z_{0t},z_{1t})\leq d(z_{0t},z_{2t})\leq\cdots$
$\leq d(z_{0t},z_{pt})$

とし，この距離がたとえば0.1未満の系列を$(z_{1t},\cdots,z_{kt})$とする。

④ ③で選択された系列$(z_{1t},\cdots,z_{kt})$に主成分をかけ，その第1主成分でもって求める指標とする。

⑤ 安定性をチェックするため，距離が0.2未満あるいは0.3未満の場合の系列の選択等，いくつかの場合の指標をグラフに書き，参照変数 z_{0t} の動きと比較する。

§3　サーベイデータの利用

経済的指標の中には，経済活動(アクティヴィティ)それ自体を客観的に表現しようとするもののほかに，生計費指数，福祉指標等のように，社会構成員の満足の度合など，福祉厚生に密接に関係した指標がある。このような指標においては，嗜好選択等各個人の心理的側面とその個人をとりまく経済的社会的環境の関係を無視できないばかりか，社会全体の満足度を平均的にみる社会的厚生それ自体を，経済的社会的環境に関するデータの系列で把握しようとする指標も少なくない。たとえば物価指数の理論では，異時点間の個人の効用水準を比較し，それを価格データの系列で表現しようとするのであるから，この類の指標である。しかしそこでは，与えられた系列に対して無数にある系列からどの系列を選択するかについての議論は極めて少ない。さらに嗜好したがって効用関数の形状が時間的に不変であるという仮定をはずすと，ラスパイレス指数あるいはパーシェ指数の理論的基礎が弱くなるとともに，個人間で嗜好の変化に差があるのが一般であるからアグリゲーションの問題を生じさせる。系列選択の問題は，以下に述べる嗜好の変化の問題とともに，サーベイデータの利用法の中で議論しうるであろうと考える。同様に福祉指標，景況感を表わす景気指標等，社会構成員の心理的満足度，好況不況感，社会的厚生といったようなものを，社会経済の環境の側から把握しようとする場合，以下のアプローチは有効である。

まず作成しようとする指標に対して，十分計画されたサーベイが四半期毎(もちろん月毎でも半年毎でもよい)になされているとする。このサーベイは N 人(企業等)の標本からなり，次の3つの選択枝

A 良くなった，B 変らない，C 悪くなった

に対してなされているとする。この場合，たとえば「良くなった」という意味は

(1) 今期に期待しているものと比較して「良くなった」

(2) 前期と比較して「良くなった」

の2通り考えることができよう。もちろんこれは，指標で表現しようとする対象に関するサーベイの標本計画の問題である。のちの議論では(1)と(2)が区別される。次に，社会的経済的環境とサーベイの応答との関係を設定する。まず§2で述べた指標作成のプロセスⅠで収集される p 種のデータ系列を各期 t に対して

(3.1)　$\boldsymbol{x}_t=(x_{1t},\cdots,x_{pt})$　$t=1,\cdots,n$

として $p\times 1$ のベクトルで表わす。ただしデータは，一般性を失うことなく月別に与えられているとし，$t=1$ はサーベイデータが得られている最初の時点，$t=n$ はその最後の時点とする。したがって四半期毎にサーベイが行われていると想定しているから，$t=1, 4, 7, \cdots, n$ の月では，$\boldsymbol{x}_t$ と N 個のサーベイデータ $(D_{1t}, \cdots, D_{Nt})$ が同時に観測されていることになる。ここで，D_{jt} は第 j 解答者（人または企業）の時点 t での値で，A と応答している場合 $D_{jt}=1$, B と応答している場合 $D_{jt}=0$, また C と応答している場合 $D_{jt}=-1$，の値をとるものとする。さらに第 j 解答者が D_{jt} と応答した裏には，指標の対象と考えられる第 t 月の社会的経済的環境を表わす(1.1)のデータ $\boldsymbol{x}_t$ があって，効用等嗜好や心理的感覚的なものを通してそれに反応したものと仮定する。すなわち

(3.2) $\quad D_{jt}=D_{jt}(\boldsymbol{x}_t)$

とする。この場合，第 j 解答者は必ずしも $\boldsymbol{x}_t$ を観察する必要はなく，$\boldsymbol{x}_t$ は期間(月)t の指標の対象となっている社会的経済的環境を示すのみで，第 j 解答者はその環境の中で活動をしたとすれば十分である。

さて(1)の場合，各個人(企業)は平均的な社会的経済的環境の動向を過去の経験等を通じて感知しており，そこから期待されるものと比較して D_{jt} と反応(応答)したと仮定する。そして，その平均的な当該環境の動向を，$\boldsymbol{x}_t$ の平均的な傾向を示すトレンドで代理(近似)し，(3.2)の関数を

(3.3) $\quad D_{jt}=f_j(\boldsymbol{x}_t-\boldsymbol{\mu}_t)$
$\quad\quad =f_j(x_{1t}-\mu_{1t}, \cdots, x_{pt}-\mu_{pt})$

と表わす。ここで f_j は時間に対して不変であると仮定しており，$\boldsymbol{\mu}_{it}$ は第 i 系列のトレンド，を示す。(3.3)は，第 j 解答者の反応が，社会経済環境の平均的傾向からの乖離に依存している，ことを示している。さらに解答者は，社会的経済的環境が季節的に変化することを知っているだろうし，また細かな不規則変動を許容してこのような変動には反応しないだろうから，(3.3)の $\boldsymbol{x}_t$ の各要素 x_{it} は，季節変動と不規則変動を調整した変数とも考えることができる。また $\boldsymbol{x}_t$ の要素がすべて正であれば，(3.3)の代りに

(3.4) $\quad D_{jt}=g_j(x_{1t}/\mu_{1t}, \cdots, x_{pt}/\mu_{pt})$
$\quad\quad =\tilde{g}_j(\log x_{1t}-\log \mu_{1t}, \cdots, \log x_{pt}-\log \mu_{pt})$

とスペシファイすることもできる。以下簡単化のために，議論を(3.3)の場合に限るが，(3.4)の場合も同様である。さらに解答者は経験的に各 x_{it} の変動の大きさを知っていると想定して，各要素 x_{it} の標準偏差 σ_i で $x_{it}-\mu_{it}$ を割った値

$\quad z_{it}=(x_{it}-\mu_{it})/\sigma_i$

(基準化した変数)の変化に応答していると考えることもできよう。その場合(3.3)は

(3.5) $\quad D_{jt}=f_j(\boldsymbol{z}_t)=f_j(z_{1t}, \cdots, z_{pt})$

と表わすことができる。以上がサーベイデータ D_{jt} と環境 $\boldsymbol{x}_t$ との間の想定である。この想定は後にふれるロジット・プロビット分析に似ている。

次に時点 t の N 個のサーベイデータ $(D_{1t}, \cdots, D_{Nt})$ を，DI（ディフュージョン・インデクス）としてアグリゲートすることを考える。すなわち

(3.6) $\quad d_t=\Sigma_{j=1}{}^N D_{jt}/N$

の形にアグリゲートする。単純に加算するのは，各反応が独立であるとみるばかりか，各個人を同等(平等)に扱うことである。D_{jt} の値は ± 1 または 0 であるから，d_t は期待していたものと比べて「良くなった」と答えた人から「悪くなった」と答えた人を差し引いた数の割合である。この d_t がサーベイデータに基づく参照変数であり，これが指標の表現対象の代理変数である。サーベイデータが四半期毎に行われる場合，月別データに直すため d_t の欠けている月に対しては直線補間するものとする。その結果，(3.5)と(3.6)より $t=1, \cdots, n$ に対して

(3.7) $\quad d_t=\Sigma_{j=1}{}^N f_j(\boldsymbol{z}_t)/N=F(\boldsymbol{z}_t)$

が近似的に成立することになる。このとき問題は，p 種のデータ系列から比較的少数の q 系列を選択し，それに基づいて参照変数 d_t に対応する動きをする指標を作成することである。その場合，指標化の段階では参照変数の値を用いない。それは社会構成員がその経済活動の中で意識的にも無意識的にも感じとっていることを，対象となっている社会的経済的環境を表わすデータ系列の方から把握し指標化することが目的であるからである。

もちろん d_t の値を使う立場もありえよう。その場合，社会平均的な構成員の気持を環境 $\boldsymbol{x}_t$ の中に投影することになろう。ここでは指標化の立場から(3.7)を線型近似して

(3.8)　$d_t \simeq \alpha_1 z_{1t} + \cdots + \alpha_p z_{pt} + \mu_0$

と表現し，d_t を参照変数として第2節で述べた方法によって，系列の選択(II)，その指標化(III)，を行うことができる。(3.6)の d_t 自体1つの指標であるのに，d_t を参照変数として別な変数を作るのは

（i）d_t と当該環境を表わす系列 $\boldsymbol{x}_t$ と(3.7)の関係が想定でき，d_t の動きを $\boldsymbol{x}_t$ の動きと対応をつけること自体興味あること，

（ii）社会経済環境の変化を先どりして d_t の変化の予測可能性を確保しうること，

（iii）サーベイデータは一般に速報性に欠けること

等が挙げられよう。(ii)および(iii)の理由は，景気指標にとって重要である。

次に，サーベイデータが1期前と比較して良し悪しを問う(2)のケースを考察する。この場合，第 j 解答者の応答 D_{jt} は，一期前の当該環境 $\boldsymbol{x}_{t-3}$ と今期の環境 $\boldsymbol{x}_t$ の差に依存すると考えられよう。すなわち，

(3.9)　$D_{jt} = D_{jt}(\boldsymbol{x}_t - \boldsymbol{x}_{t-3})$

である。ここで(1)の場合と同様に，$\boldsymbol{x}_t$ の要素は季節変動と不規則変動を調整したものと考えられると同時に，今期の環境 $\boldsymbol{x}_t$ と1期前の環境 $\boldsymbol{x}_{t-3}$ の差に対して期待される差があって，その差に比べて大きく変化したとき反応 D_{jt} の値が1または -1 をとると想定しよう。そしてこの期待される差を，$\boldsymbol{x}_t$ のトレンドの今期と1期前の差 $\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-3}$ で代理(近似)できるものとする。このとき(3.9)は

(3.10)　$D_{jt} = h_j(\boldsymbol{x}_t - \boldsymbol{x}_{t-3} - (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-3}))$　$(t \geq 4)$

と表現される。このあとの議論は(1)の場合と同様であるが，基準化変量 $\boldsymbol{z}_t$ を用いると

(3.11)　$D_{jt} = \bar{h}_j(\Delta \boldsymbol{z}_t)$　　$(\Delta \boldsymbol{z}_t = \boldsymbol{z}_t - \boldsymbol{z}_{t-3})$

のように $\boldsymbol{z}_t$ の変化(変動幅)に依存することになる。したがって系列の選択，指標化は $\Delta \boldsymbol{z}_t$ によってなされることになる。たとえば，景気感を表わす指標を作成する場合，1期前と比較して景気が良くなったと思うかどうか，についてサーベイがなされているとすれば，そのときには $\Delta \boldsymbol{z}_t$ に基づいて系列が選択され，指標化される。その場合，変化 $\Delta \boldsymbol{z}_t$ に基づいて系列を選択されるため，直観的には意味のつきにくい系列が選択される可能性がある。しかしそれは上の理由によって正当化されよう。また実際にも景気感というのは，水準 $\boldsymbol{z}_t$ でなく，その変化 $\Delta \boldsymbol{z}_t$ と密接に関連している部分が大きいだろう。変化 $\Delta \boldsymbol{z}_t$ に基づく指標化のメリットは，水準 $\boldsymbol{z}_t$ に基づく場合に比べて，第2節で述べた系列相関の問題をさける度合が大きい。

最後にロジット・プロビット分析にふれる。通常ロジット・プロビット分析では，応答 D_{jt} は2項確率変数で $+1$ または0をとるものとする。ここでは D_{jt} は $+1$(良くなった)，-1(悪くなった)の2つの値をとるものとし，通常のように

(3.12)　$P(D_{jt}=1)$
$$= G_j(\beta_0 + \beta_1 x_{1t} + \cdots + \beta_p x_{pt})$$

$P(D_{jt}=-1) = 1 - P(D_{jt}=1)$ とする。ここで一般性を失うことなく $x_{1t}, \cdots, x_{pt}$ はそれを基準化した変量 $z_{1t}, \cdots, z_{pt}$ でおきかえることができる。そして $D_{jt}=1$ となる j の個数を Q_t とすると，Q_t/N を $P_t \equiv \Sigma_{j=1}{}^N P(D_{jt}=1)/N$ の推定値と考える。すなわち

(3.13)　$Q_t/N \simeq \Sigma_{j=1}{}^N G_j(\beta_0 + \beta_1 z_{1t} + \cdots + \beta_p z_{pt})/N$
$$\equiv G(\beta_0 + \beta_1 z_{1t} + \cdots + \beta_p z_{pt})$$

とする。さらに G に対してたとえばロジスティク分布関数(ロジット分析)や正規分布関数(プロビット分析)を仮定する。そして

(3.14)　$G^{-1}(Q_t/N) \simeq \beta_0 + \beta_1 z_{1t} + \cdots + \beta_p z_{pt}$

により，$\beta_0, \cdots, \beta_p$ を推定する(回帰式として)。ここではあくまでも $\boldsymbol{x}_t$ のもとで(3.13)の良い$(D_{jt}=1)$と反応する社会平均的割合を求めているにすぎない。その意味で求める指標化とは別な方向にある。加えて，上記のアプローチでは，最初に適当な系列が選択されている必要があろう。

§4　回帰分析での変数選択への応用

回帰分析では式(1.1)の説明変数 $x_{1t}, \cdots, x_{pt}$ の選択にあたって，t 検定，F 検定，あるいは情報量基準等が用いられる。しかし，説明変数の数 p が多い場合，これらの方式は，多重共線性のために不安定になるし，また変数選択の順序のとり方に関係して，後退消去法等，採用する方法に依存して選択される説明変数が異なる(たとえば Draper and Smith [7] をみよ)。加えてこれらに共通する欠点としては，与えられた説明変数を全体として眺め，被説明変数への貢献度に関して順序を付ける方式がない。あえていえば，各変数の t 値が，その役割を果すともいえるが，すでに述べたように，P が大きいとき多重共線性により結果は不安定になりがちである。

この節では，第2節で述べた指標作成時の系列選択の方法を，上記の欠点をカバーする変数選択の方法として提案する。第2節で述べた方法では，参照変数(目的変数)あるいは被説明変数 $y_t \equiv x_{0t}$ を説明変数$x_{1t}, \cdots, x_{pt}$ と一緒に主成分分析にかけ，被説明変数と強い「共変関係」にある説明変数を距離を通じてとりだした。もちろん説明変数は，そこでは指標の選択対象となっている系列であった。しかし第2節の議論は，トレンド除去後の変数を基準化したもの $z_{it}=(x_{it}-\mu_{it})/\sigma_i$ に基づいており，トレンド μ_{it} の推定に関して問題が残るし，また回帰分析では $z_{0t}=(y_t-\mu_{0t})/\sigma_0$ の変動より y_t 自体の変動に興味があるため，トレンド除去後の変数に主成分分析をかけるよりも，生のデータ $(y_t, x_{1t}, \cdots, x_{pt})$ に主成分分析をかけた方がよいともいえよう。しかし多くの経験は，とくに y_t のトレンドが強いとき，その変動が全体をひっぱって，必ずしもうまくいかないことを教えている。実際 $(y_t, x_{1t}, \cdots, x_{pt})$ がトレンドからの乖離の変動でみる限り強い共変関係にあっても，生のデータの変動そのものでみる限り弱い共変関係となる場合がしばしばある。このトレンドの問題は，経済分析における回帰分析全体として重要な問題であるにもかかわらず，あまり認識されておらず，議論も少ない。それは，トレンド自体がある意味で無定義用語であって，この問題を扱おうとするとどうしても議論がアドホックにならざるを得ない。第2節では，指標による y_t の予測方式を

$$(4.1)\qquad \hat{\mu}_{0t}+\hat{\sigma}_0 w_t(k)$$

と提案した。(2.20)および(2.21)をみよ。ここで $\hat{\mu}_{0t}$ は y_t のトレンドであり，$\hat{\sigma}_0$ は y_t の標準偏差，$w_t(k)$ は選択された系列$(z_{1t}, \cdots, z_{kt})$に基づく指標である。このような見方が回帰分析でも役に立つことはいうまでもない。すなわち，トレンドが強い被説明変数をもつ回帰式では，トレンドを別に説明し，残りを$(x_{1t}-\mu_{1t}, \cdots, x_{pt}-\mu_{pt})$で説明する。これは $\mu_{it}=a_i+c_it$ のときには，(1.1)を

$$(4.2)\qquad y_t=\mu_{0t}+b_1(x_{1t}-a_1-c_1t)+\cdots +b_p(x_{pt}-a_p-c_pt)+u_t$$

$$(4.3)\qquad \mu_{0t}=b_0+[b_1a_1+\cdots+b_pa_p] +[b_1c_1+\cdots+b_pc_p]t$$

と書きかえ，先に $\hat{\mu}_{it}=\hat{a}_i+\hat{c}_it\,(i=1, \cdots, p)$ と $\hat{\mu}_{0t}$ を求め，それを差し引いて主成分分析をかける第2節の方式を提案する。これは，アドホックの指摘はまぬがれないが，トレンドが安定しているときは，とくに有効であると思われる。

若干脱線もしたが，本節で提案する方式をまとめておく。

(1)　トレンドを考慮して，トレンド変数 $z_{p+1,t} \equiv t$(あるいは $z_{p+1,t} \equiv \log t$)を加え，生の変数$(y_t, z_{1t}, \cdots, z_{p+1,t})$に主成分分析をかけ，第2節で述べた②のステップの系列選択を行い，距離(2.17)(実際のデータに対しては(2.31))に基づいて距離の小さい順に選択順位をつける。

(2)　この結果もしトレンド $z_{p+1,t}$ が第1順位にきた場合，(4.2)(4.3)のようにトレンドを取り除くことにする。そして第2節の $\boldsymbol{z}_t=(z_{0t}, \cdots, z_{pt})$ に基づいて再び主成分分析をかけ，(2.21)(あるいは(2.31))の距離に基づいて距離の小さい順に選択順位をつけ，小さい方から決定係数とのかねあいで k 個の説明変数 $z_{1t}, \cdots, z_{kt}$ を選択し，z_{0t} を$(z_{1t}, \cdots, z_{kt})$に回帰する。その回帰式を $w_t(k)$ とし，(4.1)式でもって y_t の回帰式とする。ここで第2節では，$w_t(k)$ は指標であるので，$w_t(k)$ の決定では z_{0t} の値を用いていないことに注意する。

(3)　(1)でもしトレンド $z_{p+1,t}$ が第1順位でない

場合，説明力(決定係数)とのかねあいで距離の小さい順に k 個の説明変数を選択し，y_t をそれに回帰する。その場合第2順位以降にトレンドが選ばれるかもしれない。その場合，そのままそれを用いる。

最後に2点コメントする。第1は，トレンドについてである。経済理論における時間の概念は，経済構造のダイナミズム等に関して本質的に経済現象の理解の仕方に関係する。ここでは深く立ち入ることができないが，何らかの「変化」を理解しようとするとき，その変化の概念は時間と空間の概念に従属すると考えられる。その意味でたとえば，消費 y_t が所得 $\boldsymbol{x}_t$ の関数であり，それが時系列的変化の中で理解されるとき，時間はその関数に密接に関係する。経済理論では $y_t=f(x_t)$ という形で，f として時間から自由な関数を求め，それによって時間に関して不変な経済行動を把握しようとする。このような理解の方向は基本的に正しいと思われるが，問題は如何なる形でこのような不変性が確保されるかということである。われわれの経済行動はいろいろな経済的非経済的要素に反応している。それは経済政策に影響されるだろうし，あるいは将来の経済環境の変化に対する期待にも大きく依存する。この種の要素に対する不変性は，合理的期待形成の議論の中で大きな問題となり，経済学の理解の仕方に積極的な方向を与えた。もう1つの要素として，非経済的要素の変化に対するわれわれの経済行動の反応も見逃すことはできない。そして，この要素の1つの代理変数として常に時間というものが考えられよう。すなわち，消費と所得の間の不変的な経済構造を把握しようとするとき，$y_t=f(x_t)$ とするのではなく，$y_t=f(x_t,t)$ とすることによって求める経済行動の不変性をより正確に記述できよう。その意味で，常に時間は説明変数の中に加えてもさしつかえないと考える。そして上で提案したトレンドの導入と処理は，このような形で理解される。

第2点としては，回帰分析への主成分分析の応用は，多く説明変数の中に潜む多重共線性をさけるものが多い。刈屋[1]では沢山の回帰があるとき，説明し残した残差の中から共通の説明変数をさがすという応用を指摘している。しかし上記のような説明変数選択への応用は寡聞にして知らない。

(一橋大学経済研究所)

参考文献

[1] 刈屋武昭(1978)「要因の発見と指標化」溝口・刈屋編『統計学』第4章，青林書院。

[2] ――(1979)『回帰分析の理論』岩波書店。

[3] 経済企画庁調査局『景気動向調査四半期概況』。

[4] 通産省(1968)『鉄鋼業景気動向指標』。

[5] 山一証券経済研究所(1979)『YRIインデックス』。

[6] ――(1982)『新YRIインデックス』。

[7] Draper, N.R. and Smith, H. (1966) *Applied Regression Analysis.* Wiley, New York (中村訳『応用回帰分析』森北出版)。