1996年度日本オペレーションズ・リサーチ学会
秋季研究発表会

主成分分析の経済多重時系列への応用

01300450　日本大学　高橋 磐郎　TAKAHASHI Iwaro
02401540　*日本大学　中本 和彦　NAKAMOTO Kazuhiko

1　はじめに

統計的データの中に、時系列データとみなされるものは非常に多い。特に経済データ、中でも国民経済全体に関するマクロデータと呼ばれるものは皆そうである。そこで本研究において、このような経済時系列データに、主成分分析を導入し、その有効性について検討する。

2　主成分分析

2．1　主成分分析の意義

主成分分析の主な目的は、多数の（s 種の）観測可能な項目のもつ情報を、できるだけ少数の変量、いわゆる主成分、に集約することにある。そのためたとえば第１主成分は、その分散が最大になるように、各項目のウエイトを決めるという原則で求められている。しかし、この第１主成分は、同時に、それと各項目の観測値の相関係数の２乗和が最大となるという性質をもっている。また、この第１主成分を説明変数とし、各項目の観測値を被説明変数とした場合の誤差の２乗和が最小になるという性質を持っている。以上から少なくとも第１主成分は与えられたデータの情報を最も合理的に集約したものと考えることができる。

2．2　主成分分析によるデータの予測

第１主成分のウエイトは相関行列 R の最大固有値に対する（長さ１の）固有ベクトルの成分であり、じつは第２主成分のウエイトは R の第２番目に大きな固有値に対する（長さ１の）固有ベクトルの成分なのである。

定理１　第１主成分を z 、そのウエイトを $u_1, \cdots, u_s$ とし、第２主成分を y 、そのウェイトを $v_1, \cdots, v_s$ とすると、

$$z = u_1 x_1 + u_2 x_2 + \cdots + u_s x_s$$
$$y = v_1 x_1 + v_2 x_2 + \cdots + v_s x_s$$

となる。

定理２　次のような回帰式

$$x_{ik} = \alpha z_k + \beta y_k + e_k, (k = 1, \cdots, n)$$

において、α, β の最小２乗推定はそれぞれ u_i, v_i である。つまり x_{ik} は、

$$\hat{x}_{ik} = u_i z_k + v_i y_k, (k = 1, \cdots n, i = 1, \cdots s)$$

によって予測することができる。

このことは、x_{ik} の情報が第１、第２主成分に集約されていることを示している。もちろん第３、第４主成分以降と高次の主成分までとれば、予測の誤差は少なくなる。

3　時系列モデル

不規則に変動する時系列データを解析するために、確率の概念が導入されることが多い。とくに時点ごとに得られる値を確率変数の実現値とみなして、その時系列を生み出す確率的な構造をモデル化したものは確率過程と呼ばれる。確率過程を時系列の解析を進めるための基礎としたとき、時系列を表現するモデルは、単に時系列データを観測しそれを記述するためのモデルといった漠然とした意味でのモデルではなくて、確率的な構造を表現することを

その本質的な特徴とするモデルとなる。このように定義されるモデルはしばしば**時系列モデル**と呼ばれる。そこで以下に、基本的な１変量の時系列モデルを示す。

3．1　自己回帰（AR）モデル

確率過程$x(t)$が、

$$x(t)=\sum_{m=1}^{p}a_m x(t-m)+u(t)$$

で表されるとき、これを**p次の自己回帰(AR)モデル**といい、以下３つの条件を満足する。

$$E[u(t)]=0$$
$$E[u(t)u(s)]=0(t\neq s),\sigma_u{}^2(t=s)$$
$$E[u(t)x(t-m)]=0$$

4　対象データと分析方法

本研究では、対象データ（時系列データ）として、景気動向指数のデータを用いる。この景気動向指数を表している説明変数$x_i(t)$を以下に示す。

$x_1(t)$生産指数	$x_6(t)$労働投入量指数
$x_2(t)$資財出荷指数	$x_7(t)$百貨店販売額
$x_3(t)$原材料消費指数	$x_8(t)$商業販売額指数
$x_4(t)$電力使用量	$x_9(t)$中小企業売上高
$x_5(t)$稼働率指数	$x_{10}(t)$有効求人倍率

分析方法として、以下（Ⅰ）**method 1**，（Ⅱ）**method 2**に示す。

（Ⅰ）method 1

時系列データを、

$$x_i(t)=a_{i1}x_i(t-1)+a_{i2}x_i(t-2)+\cdots+u(t)$$
$$(i=1,\cdots,s)$$

で表現し、そして最小２乗法によって、各$x_i(t)$に対するARモデルの係数$a_{i1},a_{i2},\cdots$の推定を直接行うことによって、ARモデルの係数の推定値$\hat{a}_{i1},\hat{a}_{i2},\cdots$を求め、原データ$x_i(t)$の予測を行う。

（Ⅱ）method 2

時系列データに対して、まず最初に主成分分析を行う。第１，２主成分を各$x_i(t)$のウエイトu_i,v_iを用いて表現すると、

$$z(t)=u_1x_1(t)+\cdots+u_sx_s(t)$$
$$y(t)=v_1x_1(t)+\cdots+v_sx_s(t)$$

となる。ここで各主成分に対するARモデル、

$$z(t)=c_1z(t-1)+c_2z(t-2)+\cdots+e(t)$$
$$y(t)=d_1y(t-1)+d_2y(t-2)+\cdots+d(t)$$

に対して、最小２乗法によって各係数c_i,d_i（$i=1,2,\cdots$）の推定を行うことによって、各係数の推定値$\hat{c}_i,\hat{d}_i$を求める。そして定理２より、

$$\hat{\hat{a}}_{i1}=u_i\hat{c}_1+v_i\hat{d}_1$$
$$\hat{\hat{a}}_{i2}=u_i\hat{c}_2+v_i\hat{d}_2$$

によって、（Ⅰ）**method 1**の$a_{i1},a_{i2},\cdots$の推定を行い、$\hat{a}_{i1},\hat{a}_{i2}$との比較を行う。$\hat{\hat{a}}_{i1},\hat{\hat{a}}_{i2}$の方が推定の誤差が少ないことが期待される。

5　おわりに

本研究では、主成分分析の経済多重時系列データへの適用の有効性について、提案した。

今後、各方面の分野にも適用されることが考えられる。

参考文献

［1］廣松 毅，浪花 貞夫：「経済時系列分析」朝倉書店（１９９０）

［2］高橋 磐郎，小林 竜一，小柳 芳雄：「統計解析」培風館（１９６９）

［3］奥野 忠一 他：「多変量解析法」日科技連出版社（１９７１）