

# 多変量時系列に対する主成分・因子分析

統計数理研究所 川崎能典

(受付 2000 年 10 月 16 日; 改訂 2001 年 4 月 2 日)

## 要 旨

主成分分析や因子分析において標準的に想定されるデータセットの代表的な形式は、異なる被験者で観察される値が測定項目ごとに並べられているものと言えるだろう。多変量時系列においては、個体を表す添え字を時間の添え字に読み替えることで、形式的には主成分分析・因子分析を行うことは可能であるが、このような形式的な適用については古くから問題点が指摘されてきた。本稿の目的は、多変量時系列解析における主成分分析・因子分析に理論的に妥当性を与える 2 つの切り口から、これまでの主要な結果を整理・紹介することにある。ひとつは、時系列の離散 Fourier 変換によって漸近的に独立なデータに変換し、古典的な主成分・因子分析の枠組みに帰着させる方法であり、第 2 は観測されない因子過程に直接モデルで表現を与える方法である。最後に、時間領域での主成分分析が同時点での相関構造のみに着目した方法であることを指摘し、ラグ付き潜在因子を考慮した動的因子モデルの解析例を示す。

キーワード：多変量時系列，主成分分析，因子分析，周波数領域，時間領域。

## 1. はじめに

主成分分析や因子分析はさまざまな個別科学の分野においてデータの縮約等の目的で非常に広範に用いられている手法である。特に社会科学の分野で、個体ごとに測定項目が得られているような多変量データの解析においては一般的な手法と言えるが、時系列解析の場面ではそれほど頻りに持ち出されることがない。その最も大きな理由は、独立同一の多変量正規分布からの標本という前提で構築される主成分分析・因子分析の推測理論は、その前提が時系列データの特性とは相容れないからであろう。

例えばここで非常に簡単な一因子の因子分析モデルを取り上げてみよう。いま  $p$  個の変量  $x_i$  ( $i = 1, \dots, p$ ) が唯一の共通因子  $z$  と独自因子  $\epsilon_i$  によって

$$x_i(t) = \lambda z(t) + \epsilon_i(t), \quad t = 1, \dots, n$$

と表されるものとしよう。観測不能な因子  $z(t)$  と  $\epsilon_i(t)$  はすべての  $i$  と  $t$  で互いに無相関、 $z(t)$  の分散は 1 に基準化されているものとする。まず、標準的な多変量解析の設定で、 $t$  を個体を表す添え字だと見なそう。このとき  $t$  に関する無相関性は妥当性を持つように思われる。しかし、もし  $x_i(t)$  が時系列であれば、一般に  $x_i(t)$  と  $x_i(t+s)$  は相関を持つので、 $t$  に関する無相関性の仮定はほとんど常に妥当性を欠く。このような状況にも形式的に独立同一分布を前提とした推測理論を適用し検定を行ったりすることは、誤った結論を導きかねない。

本稿の目的は、多変量時系列解析における主成分分析・因子分析に理論的に妥当性を与える 2 つの切り口から、これまでの主要な結果を紹介することである。ひとつは、時系列の離散 Fourier 変換によって漸近的に独立なデータに変換し、古典的な主成分・因子分析の枠組みに帰着させる方法である。これは第 2 節で報告する。第 2 の切り口は、観測されない因子過程に直接モデルで表現を与える方法である。このようなモデルを本稿では動的因子モデルと呼ぶ。第 3 節に於いては、時間領域における動的因子モデルに対する 3 種類の接近法について紹介する。第 4 節では、時間領域における主成分分析——より正確には、ラグ構造を無視した上での時間領域における分析——に関して幾つかの注意点を喚起し、ラグ構造を考慮することの重要性を示す解析例を報告する。第 5 節にはまとめを記す。

なお、本稿で触れることができなかった論点をここで挙げておく。ひとつは時変分散モデルである。ポラティリティのモデリングが盛んになるにつれて、多変量の枠組みも登場し、実際に市販のソフトウェアにもプロシジャが組み込まれているケースが多い。このような状況下では、多変量の時変分散プロセスで次元落ちを考慮する仕事が、本来なら本稿の視野に入ってくるはずである。もうひとつは、独立成分分析である。独立成分分析は音声信号の分離問題のような非ガウスの i.i.d. 系列に対して識別能力の高い方法であるが、近年では時系列的な従属性を持つ信号に対しても枠組みが拡大されつつあるようである。しかし、これら 2 点に関しては、紙幅の都合から取り上げることができなかったことを予めお断りしておきたい。

## 2. 周波数領域での解析法

時系列に対して主成分・因子分析を行うための最初の切り口は、古典的な主成分・因子分析が適用可能な形に時系列データを変換することである。ここで用いられる変換が、離散 Fourier 変換 (discrete Fourier transform) である。離散 Fourier 変換に基づく時系列理論は、Box and Jenkins (1970) 等で一般的になった時間領域での解析より歴史が古い。その意味では古典的であるが、主成分分析の適用を想定する場合に限らず定常時系列全般に関して解析的な結論を導く際の強力な道具となる。

時系列の近接する観測値の間には一般に相関がある。しかし、離散 Fourier 変換は周波数でインデックスづけられた量となり、そのインデックスで近接する変数どうしであっても、漸近的には独立である。具体的には、いま定常時系列  $x_t (t = 1, \dots, n)$  が観測されたとしよう。このとき、 $\nu_k = k/n (k = 0, 1, \dots, n)$  に対して  $x_t$  の離散 Fourier 変換

$$X(\nu_k) = n^{-1/2} \sum_{t=1}^n x_t \exp(-2\pi i t \nu_k)$$

は漸近的に独立に正規分布に従う(詳細は Brillinger (1981) の定理 4.4.1 を参照)。このとき、スペクトル密度行列  $f_x(\nu)$  が、古典的な主成分・因子分析における共分散行列の役割を果たす。主成分分析の観点からは標本共分散に相当する  $\hat{f}_x(\nu)$  があればよいが、因子分析で最尤法を用いる際には漸近正規性が理論の根拠となる。

以降 2.1 節では、定常時系列に対する主成分分析法として Brillinger (1981) にまとめられている結果を紹介する。2.2 節で述べる周波数領域での因子分析は、Priestly et al. (1974), Priestly and Subba Rao (1975), Geweke (1977) 等の結果である。2.3 節では、周波数領域での主成分・因子分析の応用例について簡単に紹介する。

## 2.1 周波数領域での主成分分析

### 2.1.1 古典的主成分分析

最初は時系列から離れて、独立同一分布の仮定の下で  $p$  変量のデータ  $x = (x_1, \dots, x_p)'$  が得られたとし、いま  $x$  を線形結合によって最も情報量のあるスカラーに変換したいとする。そこで、線形結合

$$(2.1) \quad y = c'x = c_1x_1 + \dots + c_px_p$$

のうち、 $y$  の分散を最大にするようなベクトル  $c$  を求めよう。ただし  $c'c = 1$  という制約をつけておく。 $\Sigma_x$  を  $x$  の共分散行列としてここでの問題を形式的に記せば、

$$(2.2) \quad \max c' \Sigma_x c$$

となるようなベクトル  $c (\neq 0)$  に関心がある。いま、行列  $\Sigma_x$  の固有値を降順に並べて  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  とし、それらに対応する長さ 1 に基準化された固有ベクトルを  $e_1, e_2, \dots, e_p$  としよう。このとき (2.2) に対する解は、 $c = e_1$  によって与えられ、このとき線形結合  $y_1 = e_1'x$  が最大分散  $\text{var}(y_1) = \lambda_1$  を達成する。すなわち、

$$\max c' \Sigma_x c = e_1' \Sigma_x e_1 = \lambda_1$$

である。このとき  $y_1$  のことを第 1 主成分と呼ぶ。同様の手続きで今度は、 $\text{cov}(y_1, y_2) = 0$  かつ  $c'c = 1$  のもとで  $\text{var}(y_2)$  を最大にする線形結合  $y_2 = c'x$  を第 2 主成分と呼ぶ。このようにして第  $p$  主成分まで定義される。主成分が分散のうちのどの程度の割合を説明しているかが、各主成分の重要度の尺度となる。 $\text{tr}(\Sigma_x) = \lambda_1 + \dots + \lambda_p$  より、 $\text{var}(y_k)/\text{tr}(\Sigma_x) = \lambda_k / \sum_{j=1}^p \lambda_j$  を第  $k$  主成分の寄与率と言う。実際の統計解析においては、繰り返し測定で得られた  $x_1, \dots, x_n$  に基づいて、母分散  $\Sigma_x$  を標本分散  $S_x = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$  で置き換えて主成分分析を行うことになる。

### 2.1.2 時系列での主成分分析

さて、いま  $p$  変量定常時系列  $x_1, \dots, x_t$  が観測されたとする。一般性を失うことなく平均はゼロと仮定し、 $p \times p$  のスペクトル密度行列  $f_x(\nu)$  が存在するとしよう。 $f_x(\nu)$  は複素数値、非負定値エルミート行列である。通常的主成分分析における議論((2.1)式、(2.2)式)からの類推で、ある固定した  $\nu$  において、複素数ベクトル  $c(\nu)$  によって与えられる複素数値一変量過程  $y_t(\nu) = c(\nu)^* x_t$  を考える ( $c^*$  は、複素数行列あるいはベクトル  $c$  の複素共役転置を表す)。 $c(\nu)^* c(\nu) = 1$  のもとで、周波数  $\nu$  におけるスペクトル密度が最大になるような重み係数  $c(\nu)$  を求めたい。ここで、周波数  $\nu$  における  $y_t(\nu)$  のスペクトル行列は  $f_y(\nu) = c(\nu)^* f_x(\nu) c(\nu)$  と表されるから、問題は

$$(2.3) \quad \max c(\nu)^* f_x(\nu) c(\nu)$$

となる複素数ベクトル  $c(\nu) (\neq 0)$  を求めることに帰着する。

いま、 $f_x(\nu)$  の固有値を  $\lambda_1(\nu) \geq \lambda_2(\nu) \geq \dots \geq \lambda_p(\nu)$  とし、それらに対応する長さ 1 に基準化された固有ベクトルを  $e_1(\nu), e_2(\nu), \dots, e_p(\nu)$  としよう。このとき (2.3) に対する解は、 $c(\nu) = e_1(\nu)$  によって与えられ、 $y_t(\nu) = e_1(\nu)^* x_t$  が最大分散を達成する。すなわち、

$$\max c(\nu)^* f_x(\nu) c(\nu) = e_1(\nu)^* f_x(\nu) e_1(\nu) = \lambda_1(\nu)$$

である。このとき  $y_t(\nu)$  を改めて  $y_{t1}(\nu)$  と書き、これを周波数  $\nu$  における第 1 主成分と呼ぶ。同様にして  $y_{tk}(\nu)$  ( $k = 2, \dots, p$ ) が定義できるのは古典的な主成分分析と同様である。また、ここまでの議論は任意の周波数  $\nu$  で成り立つ。

### 2.1.3 データの縮約と主成分系列

このような道具を考えることの動機付けとして、以下のような問題を考えることができる (Brillinger (1981))。いま、我々がいる地点からは遠く離れた所で多変量時系列データが収集されているとしよう。観測地点から何らかの手段でデータがオンラインで転送されてくるのだが、そのための回線の本数が限られているとする。このとき、ある時点で観測されるデータ  $x_t$  の情報量をなるべく損なわないよう、より低次元のデータ——例えば一変量過程  $y_t$ ——に集約しなければならない。

$$y_t = \sum_{j=-\infty}^{\infty} c_{t-j}^* x_j$$

ここでフィルタ  $\{c_j\}$  は絶対総和可能 ( $\sum_{j=-\infty}^{\infty} |c_j| < \infty$ ) とする。受け手の側では逆にこの  $y_t$  を元に、絶対総和可能なフィルタ  $\{b_j\}$  によって

$$\hat{x}_t = \sum_{j=-\infty}^{\infty} b_{t-j} y_j$$

として信号を復元する。できるだけ精度の良い近似にするためには、このフィルタを

$$E \{ (x_t - \hat{x}_t)^* (x_t - \hat{x}_t) \}$$

が最小となるように決めればよい。

いま、 $b(\nu)$  と  $c(\nu)$  を、それぞれ  $\{b_j\}$ 、 $\{c_j\}$  の Fourier 変換としよう。すなわち、

$$c(\nu) = \sum_{j=-\infty}^{\infty} c_j \exp(-2\pi i j \nu)$$

あるいは逆に

$$(2.4) \quad c_j = \int_{-1/2}^{1/2} c(\nu) \exp(2\pi i j \nu) d\nu$$

とする ( $b(\nu)$ 、 $\{b_j\}$  についても同様に定義する)。Brillinger (1981) 定理 9.3.1 によれば、この問題に対する解は、 $c(\nu)$  として (2.3) を満たすものを選び、 $b(\nu) = \overline{c(\nu)}$  と取ることである。すなわち、 $c(\nu) = e_1(\nu)$  として  $b(\nu) = \overline{e_1(\nu)}$  とすればよい。フィルタ (重み係数) は Fourier 逆変換 (2.4) 式によって与えられる。このようにして得られる系列を例えば  $y_{t1}$  と書いて、これを第 1 主成分系列 (first principal component series) と呼ぶ。  $y_{t2}$  から  $y_{tp}$  までも同様に定義できる。

### 2.1.4 統計的推測

実際には  $f_x(\nu)$  は見本過程  $x_1, \dots, x_n$  から推定しなければならない。まずデータからピリオドグラム行列  $I_n(\nu_j)$  を求め、それを適当なラグ・ウィンドウで平滑化したものを推定値  $\hat{f}_x$  とする。

$$\hat{f}_x(\nu_j) = \sum_{\ell=-(L_n-1)/2}^{(L_n-1)/2} h_{\ell,n} I_n(\nu_j + \ell/n)$$

ここで重み係数の項数  $L_n$  は、標本数  $n$  に依存する適当な奇数で、 $n \rightarrow \infty$  のとき  $L_n/n \rightarrow 0$  と仮定する。ウィンドウは  $\ell$  に関して対称で、各重み係数  $h_{\ell,n}$  は正値、総和は 1 である。重み係数の二乗和を  $\eta_n^{-2} = \sum_{\ell=-(L_n-1)/2}^{(L_n-1)/2} h_{\ell,n}^2$  と定義しておこう。適当な条件の下で、固有値と固有ベクトルの同時分布

$$\left( \eta_n \left[ \hat{\lambda}_1(\nu_j) - \lambda_1(\nu_j) \right] / \lambda_1(\nu_j), \quad \eta_n \left[ \hat{e}_1(\nu_j) - e_1(\nu_j) \right]' \right)'$$

は、規格化定数として  $\sqrt{n/L_n}$  を掛けた上で  $n \rightarrow \infty$  のとき平均ゼロの多変量正規分布に収束する。固有値は標準正規分布に収束し、固有値と固有ベクトルは漸近的に独立、固有ベクトルの漸近共分散行列は

$$\Sigma_{e_1}(\nu_j) = \eta_n^{-2} \lambda_1(\nu_j) \sum_{\ell=2}^p \lambda_\ell(\nu_j) \{\lambda_1(\nu_j) - \lambda_\ell(\nu_j)\}^{-2} \mathbf{e}_\ell(\nu_j) \mathbf{e}_\ell(\nu_j)^*$$

で与えられる。ここでは第1主成分に対応する固有ベクトルについてのみ結果を例示しているわけだが、その分布が他の固有値・固有ベクトルにも依存していることがわかる。更に上の結果から  $\hat{e}_1$  の各成分の信頼区間を与えることができる。 $\hat{e}_1(\nu) = (\hat{e}_{11}(\nu), \dots, \hat{e}_{1p}(\nu))'$  とし、 $s_j^2(\nu)$  を  $\hat{\Sigma}_{e_1}(\nu)$  の第  $j$  対角成分とすると、

$$\frac{2|\hat{e}_{1j}(\nu) - e_{1j}(\nu)|^2}{s_j^2(\nu)} \quad (j = 1, \dots, p)$$

は漸近的に自由度 2 の  $\chi^2$  分布に従う。

## 2.2 周波数領域での因子分析

### 2.2.1 古典的因子分析

前節と同様、最初は時系列を離れて独立同一分布からの繰り返し観測が可能な状況から議論を始める。 $x$  を平均ゼロ、分散  $\Sigma_x$  の  $p \times 1$  確率変数ベクトルとする。因子分析モデルにおいては、観測値  $x$  がごく少数の観測されない共通因子  $z = (z_1, \dots, z_q)'$  によって説明されると考える。すなわち

$$x = \mathcal{B}z + \epsilon$$

というモデルを考える。ここで  $\mathcal{B}$  は  $p \times q$  の因子負荷行列、 $z$  は  $q \times 1$  の因子ベクトルで  $E(z) = 0$ 、 $E(zz') = I_q$  とする。誤差項  $\epsilon$  は因子と独立でその共分散行列は対角  $D = \text{diag}(d_1^2, \dots, d_p^2)$  と仮定する。重回帰モデルとの違いは  $z$  が観測されないことにある。この因子モデルは  $x$  の共分散構造の観点から次のように記述することもできる。

$$(2.5) \quad \Sigma_x = \mathcal{B}\mathcal{B}' + D$$

つまり、 $x$  の共分散行列は、ランクが  $q (\leq p)$  の対称非負定値行列と非負定値対角行列との和になっている。

さて、標本  $x_1, \dots, x_n$  が与えられたもとで因子分析モデルのパラメータを推定する方法はいくつかある。最も実行が容易なのは主成分法 (principal component method) であろう。 $S_x$  を標本共分散行列とし、 $(\hat{\lambda}_i, \hat{e}_i)$  をそこから計算される固有値・固有ベクトルの組 (ただしここで、 $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$ ) とする。主成分法とは、 $\hat{\lambda}_{q+1}, \dots, \hat{\lambda}_p$  を捨てて

$$\hat{\mathcal{B}} = \left( \sqrt{\hat{\lambda}_1} \hat{e}_1; \sqrt{\hat{\lambda}_2} \hat{e}_2; \dots; \sqrt{\hat{\lambda}_q} \hat{e}_q \right)$$

と取る方法である。 $\hat{D}$  は  $S_x - \hat{\mathcal{B}}\hat{\mathcal{B}}'$  の第  $j$  対角成分を  $\hat{\delta}_j^2$  とし  $\hat{D} = \text{diag}(\hat{d}_1^2, \dots, \hat{d}_p^2)$  で推定する。もし  $D$  が単位行列の定数倍であれば、因子分析は主成分分析に帰着するので (丘本 (1986) p. 24), この事実が主成分法に根拠を与える。

第2の方法は、 $z$  と  $\epsilon$  が多変量正規分布に従うという仮定を追加して最尤法を用いることである。定数項を無視した対数尤度の主要部は

$$-2 \ln L(\mathcal{B}, D) = n \ln |\Sigma_x| + \sum_{j=1}^n \mathbf{x}_j \Sigma_x^{-1} \mathbf{x}_j$$

となる． $\mathcal{B}$  と  $D$  は  $\Sigma_x = \mathcal{B}\mathcal{B}' + D$  として対数尤度の中に入っている．しかし，因子の回転 (factor rotation) の自由度があるため，このままでは最尤法の解は一意的ではない．そこで通常は  $\mathcal{B}D\mathcal{B}'$  が対角行列になるなどの制約を置いて一意性を確保する．尤度の最大化は数値的最適化によって行われる．

### 2.2.2 時系列の因子分析

主成分分析の時と同様， $x_t$  をスペクトル密度行列  $f_x(\nu)$  を持つ  $p$  変量定常時系列とする．周波数領域での時系列のモデリングでは，独立同一分布の場合の共分散行列  $\Sigma_x$  に代わるものがスペクトル密度  $f_x(\nu)$  であることに再び注意すれば，古典的な因子分析におけるモデル (2.5) からの類推で，ここでのモデルを

$$f_x(\nu) = \mathcal{B}(\nu)\mathcal{B}(\nu)^* + D(\nu)$$

と定式化できる．ここで  $\mathcal{B}(\nu)$  は  $p \times q$  の複素数値行列で  $\text{rank}(\mathcal{B}(\nu)) = q \leq p$ ， $D(\nu)$  は実数値非負定値対角行列である．

最も簡単な場合は，多変量時系列が一個の要因で決まっていると考えるモデルである．すなわち， $x_t = (x_{t1}, \dots, x_{tp})$  のとき

$$(2.6) \quad x_{tj} = c_j s_{t-\tau_j} + \epsilon_{tj}, \quad j = 1, \dots, p$$

というモデルを考えることである． $c_j \geq 0$  は要素ごとの因子負荷を， $\tau_j$  は要素ごとの位相シフトを表す．共通因子  $s_t$  と独自因子  $\epsilon = (\epsilon_{t1}, \dots, \epsilon_{tp})'$  が独立で， $\epsilon_t$  のスペクトル密度行列  $D_\epsilon(\nu)$  が対角としよう．このとき， $x_{tj}$  の離散 Fourier 変換

$$X_j(\nu) = n^{-1/2} \sum_{t=1}^n x_{tj} \exp(-2\pi i t \nu)$$

によってモデル (2.6) の周波数領域版

$$(2.7) \quad X_j(\nu) = a_j(\nu)X_s(\nu) + X_{\epsilon_j}(\nu)$$

が得られる．ここで  $a_j(\nu) = c_j \exp(-2\pi i \tau_j \nu)$ ， $X_s(\nu)$  と  $X_{\epsilon_j}(\nu)$  はそれぞれ共通因子  $s_t$  と独自因子  $\epsilon_{tj}$  の離散 Fourier 変換である．(2.7) 式の個々の要素を並べることで，古典的な一因子の因子分析モデルの複素数版が以下のように得られる．

$$\begin{pmatrix} X_1(\nu) \\ \vdots \\ X_p(\nu) \end{pmatrix} = \begin{pmatrix} a_1(\nu) \\ \vdots \\ a_p(\nu) \end{pmatrix} X_s(\nu) + \begin{pmatrix} X_{\epsilon_1}(\nu) \\ \vdots \\ X_{\epsilon_p}(\nu) \end{pmatrix}$$

あるいはもう少し簡潔に記せば

$$(2.8) \quad \mathbf{X}(\nu) = \mathbf{a}(\nu)X_s(\nu) + \mathbf{X}_\epsilon(\nu)$$

となる．(2.8) から  $x_t$  のスペクトル密度について以下の関係式が導かれる．

$$f_x(\nu) = \mathbf{b}(\nu)\mathbf{b}(\nu)^* + D_\epsilon(\nu)$$

ここで  $\mathbf{b}(\nu)$  は  $p \times 1$  の複素数値ベクトルで， $\mathbf{b}(\nu)\mathbf{b}(\nu)^* = \mathbf{a}(\nu)f_s(\nu)\mathbf{a}(\nu)^*$  を満たす．

### 2.2.3 統計的推測

前節に例示した一因子モデルを以下のように拡張しておこう．

$$(2.9) \quad x_t = \sum_{j=-\infty}^{\infty} A_j s_{t-j} + \epsilon_t$$

ここで  $\{A_j\}$  は  $p \times q$  の実数値因子負荷行列,  $s_t$  は  $q \times 1$  の共通因子を表す定常過程で各因子は独立とする. すなわち,  $s_t$  のスペクトル行列 ( $q \times q$ ) は対角行列  $f_s(\nu) = \text{diag}(f_{s1}(\nu), \dots, f_{sq}(\nu))$  となる. 一方,  $\epsilon_t$  はホワイトノイズで  $s_t$  とは独立と仮定する. このとき  $\epsilon_t$  のスペクトル行列 ( $p \times p$ ) は対角行列  $D_\epsilon(\nu) = \text{diag}(f_{\epsilon 1}(\nu), \dots, f_{\epsilon p}(\nu))$  となる. 更に  $A_j$  の行列ノルムは有限と仮定する. このとき

$$(2.10) \quad \Lambda(\nu) = \sum_{t=-\infty}^{\infty} A_t \exp(-2\pi i t \nu)$$

とし, またこれによって  $\mathcal{B}(\nu) = \Lambda(\nu) f_s^{1/2}(\nu)$  と定義しておけば,  $x_t$  のスペクトル行列は

$$(2.11) \quad f_x(\nu) = \Lambda(\nu) f_s(\nu) \Lambda(\nu)^* + D_\epsilon(\nu) = \mathcal{B}(\nu) \mathcal{B}(\nu)^* + D_\epsilon(\nu)$$

という関係式を満たす. モデルの識別性のためにすべての  $\nu$  に対して  $f_s(\nu) = I_q$  とおけば  $\mathcal{B}(\nu) = \Lambda(\nu)$  となる. それでも因子の回転に関する恣意性は残る.

モデル (2.11) のパラメータ推定法として, まず主成分法が考えられるのは前節と同じである.  $\hat{f}_x(\nu)$  を  $f_x(\nu)$  の推定値として,  $(\hat{\lambda}_j(\nu), \hat{e}_j(\nu))$ ,  $j = 1, \dots, p$  を,  $\hat{\lambda}_j$  の降順に並べた  $\hat{f}_x(\nu)$  の固有値・固有ベクトルの組としよう. このとき, 古典的主成分分析と同様に

$$\hat{\mathcal{B}}(\nu) = \left( \sqrt{\hat{\lambda}_1(\nu)} \hat{e}_1(\nu); \sqrt{\hat{\lambda}_2(\nu)} \hat{e}_2(\nu); \dots; \sqrt{\hat{\lambda}_q(\nu)} \hat{e}_q(\nu) \right)$$

によって行列  $\mathcal{B}$  を推定できる.  $\hat{f}_x(\nu) - \hat{\mathcal{B}}(\nu) \hat{\mathcal{B}}(\nu)^*$  の第  $j$  対角成分を  $\hat{f}_{\epsilon_j}(\nu)$  として, 独自因子のスペクトル密度行列は  $\hat{D}_\epsilon(\nu) = \text{diag}(\hat{f}_{\epsilon 1}(\nu), \dots, \hat{f}_{\epsilon p}(\nu))$  と推定できる.

主成分法の代わりに最尤法を使うこともできる.  $X(\nu_j)$  を周波数  $\nu_j = j/n$  における  $x_1, \dots, x_p$  の離散 Fourier 変換とし, 同様に  $X_s(\nu_j)$ ,  $X_\epsilon(\nu_j)$  を共通因子と独自因子の離散 Fourier 変換としよう. このとき適当な条件の下で,  $\ell = 0, \pm 1, \pm 2, \dots, \pm(L_n - 1)/2$  に対して

$$(2.12) \quad X(\nu_j + \ell/n) \approx \Lambda(\nu_j) X_s(\nu_j + \ell/n) + X_\epsilon(\nu_j + \ell/n)$$

となる ( $\Lambda(\nu_j)$  は (2.10) で与えられる). (2.12) は, 周波数  $\nu_j$  の局所的な近傍で  $n \rightarrow \infty$  のとき上の近似が妥当性を持ち, 周波数領域で (2.9) に対応するモデルの標本版が局所的に構成できることを示している.

適当な条件の下では  $\{X(\nu_j + \ell/n); \ell = 0, \pm 1, \pm 2, \dots, \pm(L_n - 1)/2\}$  は漸近的に独立に, 平均ゼロ, 共分散  $f_x(\nu_j)$  の多変量複素正規分布に従うことから, 近似的な尤度 (の  $-2$  倍) を

$$-2 \ln L(\mathcal{B}(\nu_j), D_\epsilon(\nu_j)) = n \ln |f_x(\nu_j)| + \sum_{\ell=-(L_n-1)/2}^{(L_n-1)/2} X^*(\nu_j + \ell/n) f_x^{-1}(\nu_j) X(\nu_j + \ell/n)$$

で与えることができる. 未知母数は  $f_x(\nu_j) = \mathcal{B}(\nu_j) \mathcal{B}(\nu_j)^* + D_\epsilon(\nu_j)$  として尤度関数の中に入っている. さて, ここでの最尤法に関して 2 点注意しておこう. 第一は,  $X(\nu_j + \ell/n)$  の分布は漸近的には独立に共分散  $f_x(\nu_j)$  の正規分布に従うものの, 有限の  $n$  に対しては  $f_x(\nu_j)$  そのものからの独立なサンプルとは言えない. 従って, 分布族に基づく厳密な意味での最尤法 (古典的な場合) と同一ではないことに注意が必要である. 第二に, ここで用いた離散 Fourier 変換の漸近正規性は, 主成分分析を正当化する際に用いられた Fourier 周波数における漸近正規性ではなく, ある Fourier 周波数近傍での漸近正規性になっていることに注意が必要である.

### 2.3 応用

本節では応用例を簡単に紹介する. まず, この分野での枠組みを確立していった先駆的な業績における数値例として Geweke (1977) と Brillinger (1981) を取り上げる. Brillinger の例は

個別科学的な含意は若干乏しいが、Geweke が取り扱っている例は景気動向指数であり、これは議論の枠組みこそ様々に異なれ、現在でもなお時系列の因子分析で頻繁に取り上げられている。一方、周波数領域における主成分・因子分析は、関心のある共通成分の周波数がはっきりしていれば使いやすい枠組みと言えるので、そうした例として Shumway and Stoffer (2000) と Young and Pedregal (1999) をその後で取り上げる。

Geweke (1977) は、時系列に対する周波数領域での因子分析においては、理論的な貢献の大きい論文として引用されることが多いが、この論文では一時解雇率、労働時間、離職率、鉱工業生産指数に対し、一因子モデルをあてはめ、その適否を論じている。ここでの因子は「需要」であると考えられる。一般に経済時系列ではトレンドや季節変動といった非定常成分が支配的であるので、時系列の定常性を前提にした理論を適用するには注意が必要である。Geweke (1977) ではこうした明らかな非定常成分に対応する周波数を外して因子分析を行っている。Geweke (1977) の枠組みの応用としては他に Geweke and Singleton (1981) がある。

Brillinger (1981) の第 9 章には、ヨーロッパの 13 の都市とアメリカの 1 都市 (New Haven) の合計 14 都市で観測された月中平均気温からなる多変量時系列データ (21 年分 252 点) に対して主成分分析を行った解析例がある。主成分系列のスペクトル密度をプロットすると、どの主成分系列についても原点をピークになだらかに減衰する形になっており、第 1 主成分から徐々にスペクトルのレベルが低下してゆく。分析結果からは、第 1 主成分が欧州 13 都市の平均気温に対応していて、第 2 主成分が New Haven に対応していることが伺える。ここでは事前に着目する特定の周波数が明確にあったわけではないが、分散レベルの違いから異質なデータを検出できる例になっている。もっとも、例の構成を知っていれば予想できる結果とも言えるが、多変量時系列のクラスタリングという、より具体的な応用目的に照らせば示唆に富んだ例である。

ところで、周波数領域における主成分・因子分析は、第 3 節に述べる時間領域での方法に比べると、実例が乏しい。例が少ない理由のひとつは、関心ある周期や周波数が決まっていれば初めて積極的に援用できる枠組みとなっている点に求められるであろう。例えば多変量時系列に共通のトレンドが含まれているかどうかを調べたいのであれば、関心ある周波数はゼロであるし、外界からの刺激に対する脳の反応を見たいのであれば、その実験の持つ繰り返しの周期に関心が集まるだろう。あるいは、事前に明らかでなくても、共通の因子や成分が存在するかどうかの検討から出発して、もしあるとすればその基調的周波数成分はどこなのか、という解析の中から徐々に周波数が絞り込まれて来る場合もあるであろう。以下の二つは、このような意味で関心のある周期がはっきりしている例である。

Shumway and Stoffer (2000) には fMRI (functional magnetic resonance imaging, 脳機能磁気共鳴画像法) によって得られたデータの解析に、周波数領域での主成分分析、因子分析を行った例が取り上げられている。複数の被験者の手をブラシでこするという刺激を 32 秒間与えては 32 秒間休む、というサイクルを繰り返し、大脳皮質、視床、小脳の複数箇所を fMRI によりデータを採取する。被験者で平均をとり、観測箇所に関する多変量時系列を分析している。データは 256 秒間とられており、この中に「刺激-休止」のサイクルは 4 回入っている。サンプリングレートが 2 秒に 1 回になっているので、 $\nu = 4/128$  が関心のある周波数ということになる。期待されるように、第 1 主成分系列のスペクトル密度は周波数  $4/128$  にピークがあり、この周波数での主成分ベクトルに対する信頼区間から、8 つの測定箇所の中でブラシによる刺激に反応していないと見なされる箇所が 1 箇所あったことが報告されている。

Young and Pedregal (1999) では、アメリカ合衆国のマクロ経済時系列データ (失業率、国民総生産、消費、政府支出、民間資本投資、いずれも四半期で季節調整済を伸び率に変換) で興味深い分析を行っている。関心は、政府支出や民間資本投資が失業率に及ぼす影響にある。個別のスペクトル解析を行うと、4 年ないし 8 年のサイクルが顕著に見られる。 $\hat{f}_x(\nu)$  の第 1 主

成分と第 2 主成分を  $\nu$  を横軸にプロットすると、前者が 4 年サイクルでピークを迎え、後者は 8 年サイクル近辺でピークを迎える。4 年サイクルの周波数における第 1 主成分ベクトルは政府支出を除くすべてにウェイトがあり、逆に 8 年サイクルにおける第 2 主成分ベクトルは政府支出が大きなウェイトを占めている。コヒーレンシー (ある固定した周波数におけるスペクトル密度の相関) のなさを見ても、政府支出は他の経済変数に対して外生性の強い変数と言えるであろう。

ところで、経済学における応用では、多変量時系列が共通のトレンド成分を持つか否かという問題に特別な関心が払われている。換言すればこれはゼロ周波数に関心が集中しているということであり、その意味では、周波数領域での分析に向けた問題といえる。しかし実際には、共通トレンドを取り扱う枠組みは時間領域における手法が多い。従ってそれらは次節以降の時間領域でのモデリングの中で取り上げることとし、ここでは文献として Phillips and Ouliaris (1988) を挙げておくのにとどめる。

### 3. 時間領域での因子分析

前節に紹介された手法の骨子は、時系列を離散 Fourier 変換することで漸近的な独立性を確保し、古典的な枠組みに近づけるといえるものであった。本節で扱う第 2 の切り口はより直接的で、観測されない因子に何らかの形で時系列的な表現を与えるものである。このように、観測されない因子自体に時系列的なダイナミクス表現を与えるモデルを、動的因子モデル (dynamic factor model) と呼び、本節では幾つかの観点からその推定法や応用、相互の関連などを整理する。動的因子モデルは、心理学での応用を背景に発展が促されてきたという側面があるので、3.1 節ではその文脈に沿って、因子分析の直接的な拡張と言える共分散構造分析の観点から記述する。

一方、時系列モデリングの立場からすれば、動的因子分析モデルは 70 年代終わりから 80 年代前半で研究が進められた、観測不能要素モデル (unobserved component model) や構造時系列モデル (structural time series model) と密接な関連があり、実際その特殊形と理解することができる (3.2 節)。ここでのアプローチが有利な点は、非定常因子に対して多項式モデルを含むより広いクラスのモデルを想定できることである。一方、多変量 ARMA モデルに基づく動的因子分析も可能である。これについては 3.3 節で触れる。3.4 節では、因子過程を明示的にモデル化しない方法として、縮小ランク回帰モデルと誤差修正モデルについて言及する。

#### 3.1 共分散構造拡張型の動的因子分析

##### 3.1.1 計量心理学における歴史的経緯

心理学実験等で観察される多変量経時データに対して通常の意味での因子分析をあてはめることについては、Holtzman (1962), Anderson (1963) で批判が浴びせられていた。論点の一つは、「因子時系列が現在の観測値に影響を及ぼすとしても、それが同時点での影響のみにとどまる可能性は少なく、むしろ変量のうちの幾つかは、ラグを伴って因子の影響を受けるはずだ」というものである。一方、本稿 2.1 節に展開された議論を心理学に適用するには大きな現実的問題がある。それは、離散 Fourier 変換に基づいて推論を行うには、心理学における経時データは時間方向にあまりにデータが少なすぎるからである。事情は心理学に比べれば若干ゆるやかなるものの、データ点数の不足は経済学においても懸念されることである。Brillinger (1981) での解析例は気象データであり、社会科学におけるデータに比べれば圧倒的にデータ数が恵まれている。ある周波数  $\nu$  におけるスペクトル密度の推定値  $\hat{f}_x(\nu)$  は  $\nu$  の近傍で定義される (2.1.4 節参照) ため、ある程度データ数が確保できないと、考察される周波数も限られるばか

りかラグ・ウィンドウによる平滑化自体が困難になり、 $\hat{f}_x(\nu)$  の良い推定が得られないという問題がある。適当なバンド幅の選択に関しては、観測されるすべての周波数に興味があるのがそれとも一部でしかないのかに依存するので、簡明な指針を示すのは難しいが、例えば Tukey ((1978), 26-27) が参考になる。

### 3.1.2 共分散構造モデリング

こうした論点を踏まえて時間領域で動的因子モデルの推定を提案したのが Molenaar (1985) であるが、拡張のアイディアは直接的なものである。出発点として  $n$  変量  $p$  因子の、通常の意味での因子分析モデルを

$$(3.1) \quad x(t) = \mathbf{B}z(t) + \epsilon(t), \quad t = 0, 1, \dots, n$$

と書く。因子分析の文脈では、 $z(t)$  は共通因子、 $\mathbf{B}$  は因子負荷、 $\epsilon(t)$  は独自因子である。異時点間での共通因子の共分散は

$$\text{cov}(z(t), z(t-u)) = \Xi(u)$$

と仮定する。一方独自因子に变量間での相関はなく、同一変量に対する独自因子の異時点間での相関は時差 ( $u$ ) のみの関数、すなわち

$$\text{cov}(\epsilon(t), \epsilon(t-u)) = \mathbf{D}(u) = \text{diag}(d_1(u), \dots, d_p(u))$$

と仮定する。ここで (3.1) 式を  $s$  次のラグまで含めて

$$x(t) = \sum_{u=0}^s \mathbf{B}(u)z(t-u) + \epsilon(t), \quad t = 0, 1, \dots, n$$

と拡張すると、このとき  $x(t)$  の共分散関数  $\Sigma_x(u)$  は

$$(3.2) \quad \Sigma_x(u) = \sum_{v=0}^s \sum_{w=0}^s \mathbf{B}(v)\Xi(u+w-v)\mathbf{B}(w)^T + \mathbf{D}(u), \quad u = 0, \pm 1, \dots$$

となる。通常の因子分析では  $u = 0$  の場合のみが問題となる。

モデル (3.2) にベクトル・行列表現を与えて同時構造方程式に帰着させることで、(3.2) が形式的には古典的因子分析の自然な拡張になっていることを示そう。いま

$$\begin{aligned} \mathbf{x}' &= (x(t)', \dots, x(t-a)') \\ \mathbf{z}' &= (z(t)', \dots, z(t-a-s)') \\ \boldsymbol{\epsilon}' &= (\epsilon(t)', \dots, \epsilon(t-a)') \end{aligned}$$

としよう。ただしここで  $a \geq s$  である。更に

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}(0) & \mathbf{B}(1) & \dots & \mathbf{B}(s) & 0 & \dots & 0 & 0 \\ 0 & \mathbf{B}(0) & \dots & \mathbf{B}(s-1) & \mathbf{B}(s) & \dots & 0 & 0 \\ \vdots & & \ddots & & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{B}(0) & \dots & \dots & \mathbf{B}(s-1) & \mathbf{B}(s) \end{pmatrix}$$

とすると、動的因子モデルは

$$\mathbf{x} = \mathbf{B}\mathbf{z} + \boldsymbol{\epsilon}$$

と簡潔に表記できる。この式から共分散構造として

$$\Sigma_x = \mathbf{B}\Xi\mathbf{B}' + \mathbf{D}$$

が導かれる．ただしここで

$$\Sigma_x = \begin{pmatrix} \Sigma_x(0) & & & & \\ \Sigma_x(1) & \Sigma_x(0) & & & \\ \vdots & & \ddots & & \\ \Sigma_x(a) & \Sigma_x(a-1) & \cdots & \Sigma_x(0) & \end{pmatrix},$$

$$\Xi = \{\Xi(i-j); i, j = 1, \dots, a+s+1\}$$

$$D = \{D(i-j); i, j = 1, \dots, a+1\}$$

である．多変量時系列データから計算されるのが  $\Sigma_x$  で，これは通常の変量解析の文脈で言えば標本共分散行列に相当する部分である ( $s=0$  なら通常の因子分析に帰着する)．Molenaar (1985) が提案した方法とは，いま上に述べたような意味で動的因子モデルを構造方程式 (Jöreskog (1978), Bollen (1989), Jöreskog and Sörbom (1998)) の形に書き換え， $\Sigma_x$  がウィシャート分布に従うと仮定して共分散構造分析の枠組みでパラメータを最尤法によって推定するものである．

従って，Molenaar (1985) の最尤法は本来の意味での最尤法ではなく，疑似最尤法であることに注意しておく．これは，動的因子モデルにおいては共分散行列が精密 (exact) にウィシャート分布に従うとは言えないからである．それでもウィシャート分布が真の分布であるかのように仮定して最尤法を実行しているという意味で，ここでの最尤法は疑似最尤法になっている．これは時系列で言えば，正規分布の仮定が厳密に当てはまるという保証がない状況で正規分布に基づく尤度を用いる方法 (Gaussian MLE) と同じである．疑似最尤法については，例えば Gourieroux et al. (1984) を参照されたい．

一方，潜在変数である時系列因子 (因子の軌跡) については，推定したパラメータを所与としてカルマンフィルタで推定を行っている．今日的視点からすれば，尤度計算自体もカルマンフィルタを利用することが可能である．(ここで考察されているモデルに対するマルコフ表現の与え方はほぼ明らかであり，カルマンフィルタの公式自体もすでに時系列解析の文脈では確立された内容であるので，本稿ではこれらについて改めて紙幅を費やさない．具体的な状態空間表現については Molenaar の原論文を，カルマンフィルタについては Jazwinski (1970), Anderson and Moore (1979), Harvey (1989) 等のテキストを，比較的最新の成果までを含んだものとしては Kitagawa and Gersch (1996) を参照されたい．)

ここまでは定常時系列の範囲でのみ議論が展開されているが，Molenaar et al. (1992) では時間に依存する平均値関数を導入して非定常の場合にも対処する方法を提案している．具体的には， $t \in \{1, 2, \dots, n\}$  の平均と分散をそれぞれ  $\bar{t}$ ,  $S(t)$  として  $\tau(t) = (t - \bar{t})/S(t)$  という基準化されたトレンド関数を用い， $f(t) = \gamma\tau(t) + z(t)$  というように，線形トレンドと定常過程  $z(t)$  の和で潜在因子を表す．いま  $q$  個の潜在因子が仮定されているとすると， $\gamma$  は  $q \times 1$  ベクトルで，個々の (非定常) 因子系列におけるトレンドの傾きは異なる値をとることができる．これによって必要となる変更は形式においてはわずかである．すなわち，(3.2) 式の代わりに，

$$\Sigma_x(u) = \sum_{v=0}^s \sum_{w=0}^s \mathcal{B}(v)(\gamma\gamma' + \delta(u+w-v)I_q)\mathcal{B}(w)^T + D(u), \quad u = 0, \pm 1, \dots$$

を考察すればよい ( $\delta$  はクロネッカーのデルタ関数である)．

ここでは時間の決定論的関数をモデルに含めることで，変数の定常化を行わずに成長曲線や学習曲線をモデリングする方法を提案しているが，パラメータ推定にもカルマンフィルタのような逐次フィルタを利用するのであれば，3.2 節に述べるように確率的トレンドを要素にモデリングすることも可能である．

### 3.2 構造時系列モデルによる動的因子分析

トレンドの扱いは個別科学における物事のとらえ方に大きく依存している．例えば，前節で述べた心理学におけるデータでは，時間方向のデータ長が短いこともあるのだろうが，トレンド構造は比較的単純である．基本的に成長曲線のような非定常効果には多項式近似以上に複雑なモデリングが必要とされることは希であろう．これに対して，四半期ないし月次で 100 点から 300 点のデータの中にしばしば構造変化を伴う経済時系列ではトレンドの動きは単調とは言えず，決定論的な関数のあてはめで事足りるケースは少ない．このような時には，局所的に直線あるいは曲線をあてはめるモデル(確率的トレンド)が適切である．本小節では，構造時系列モデル (structural time series model) の枠組みで共通の確率的トレンドを扱う方法を解説する．

#### 3.2.1 構造時系列モデル

観測される(おそらくは個々に非定常な) $p$ 変量時系列を  $x_t$  ( $t = 1, \dots, n$ ) と記す． $q \times 1$  ベクトル  $z_t$  を観測されない共通のトレンド成分とし，データの生成過程を次のように仮定する．

$$(3.3) \quad \begin{aligned} x_t &= Bz_t + z_0 + \epsilon_t, & \text{var}(\epsilon_t) &= \Sigma_\epsilon \\ z_t &= z_{t-1} + \beta + \eta_t, & \text{var}(\eta_t) &= \Sigma_\eta \end{aligned}$$

ここで  $B$  は因子負荷行列 ( $p \times q$ ) で，ドリフト項  $\beta$  は時間不変である． $z_0$  の先頭  $q$  個にはゼロ制約を置き，残りの要素は無制約でこれを  $\bar{z}$  と書くことにする．通常の因子分析と同様，因子の回転に関してこのモデルは識別性を欠いている．従って， $\Sigma_\eta = I$  と仮定した上で  $B$  の上三角部分をゼロと置くなどの制約下でモデルを推定し，その後解釈が容易な方向に因子の回転が行われる．識別性の問題さえクリアできれば，この種の構造時系列モデルに標準的な議論を経由して，尤度計算，パラメータ推定，状態推定が行われる．モデルに対する状態空間表現の与え方やカルマンフィルタの公式などについては，3.1 節と同じ理由でここでは繰り返さない．3.1 節にあげた文献を参照されたい．

#### 3.2.2 共和分との関係

観測ベクトル  $x_t$  の各要素が一階の差分の後で弱定常過程に従う，すなわち， $x_t$  の各要素  $x_{i,t}$  ( $i = 1, \dots, p$ ) について  $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$  が弱定常過程に従うとしよう．別の見方をすれば， $x_t$  は弱定常過程の累積過程として定義されているということであり，平均が確率的に変化する非定常過程となっている．このときもし  $x_t$  の線形結合  $\gamma'x_t$  が定常過程と見なすことができれば，多変量非定常過程  $x_t$  には共和分 (cointegration) が存在すると言い， $\gamma$  を共和分ベクトルと言う．直感的には，多変量過程  $x_t$  の各要素が共通のトレンド成分を持っているために，その差や線形結合が定常残差をもたらす，ということである．

3.2.1 節に定義した構造時系列モデルにおいては，レベル変数  $x_t$  の線形結合で定常過程に帰着するものが  $p - q$  個存在，すなわち  $p - q$  個の共和分が存在する．ここで共和分ベクトルは因子負荷行列の零化空間に含まれる  $p - q$  本のベクトル(すなわち  $AB = 0$  なる  $A$  の各行)である．従って観測方程式 (3.3) は

$$(3.4) \quad Ax_t = Az_0 + A\epsilon_t$$

となり， $Ax_t$  は  $(p - q) \times 1$  変量の定常過程となる．ここではより単純な構造である，平均  $Az_0$ ，共分散行列が  $A\Sigma_\epsilon A'$  の多変量ホワイトノイズ過程に帰着している．

簡単な例 ( $p = 2, q = 1$ ) で考えよう．このときモデルは

$$(3.5) \quad \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ \theta \end{pmatrix} z_t + \begin{pmatrix} 0 \\ \bar{z} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

$$z_t = z_{t-1} + \beta + \eta_t$$

となる(ここで  $\text{var}(\eta_t) = \sigma_\eta^2$  とする). 共和分ベクトルは方向しか決まらないので,  $A = (1, \alpha)$  と基準化しておく.  $1 + \alpha\theta = 0$  でなければならないことから,  $\alpha = -1/\theta$  となる. (3.5) に共和分ベクトルをかけて,

$$x_{1t} = (1/\theta)x_{2t} + (-1/\theta)\bar{z} + \epsilon_t$$

を得る. ただしここで,  $\epsilon_t = \epsilon_{1t} - \epsilon_{2t}/\theta$  である. このように, 非定常成分を持つ動的因子モデルではレベル変数に線形制約があり, 動的因子モデルから共和分モデルが導けることがわかる.

モデルの推定においてはカルマンフィルタを利用した予測誤差分解形で対数尤度を評価し, 未知母数を数値的最適化の対象とすればよい. 細かな話としては,  $\bar{z}$  や  $\beta$  を状態変数の中に組み込むなどの工夫も考えられる. それ以外にも, 尤度を時間領域・周波数領域のどちらで与えるか, 状態の初期分布をどう与えるのが適当か, などの論点もあるが, これらは動的因子分析モデルに特有の問題というより時系列全般にわたる問題であるので, 本稿では紙幅の都合上議論を振り返らない.

動的因子分析のために構造時系列モデルを用いる長所としては, 因子が非定常過程であってもモデリングが容易なこと, 観測不能な因子の軌跡やその分布もモデルあてはめの副産物として同時に得られること, などがあげられる. 一方, 短所と思われるのが, 因子数の決定に関して厳密な理論が展開しにくいことである. もちろん, 情報量規準を用いたプロシジャを定義することは可能であろうが, 3.4 節に記す枠組みの方が因子数の決定に関しては議論を構築しやすいであろう.

### 3.3 多変量 ARMA モデルによる動的因子分析

3.2 節では構造時系列モデルを利用したが, 背後にある因子プロセスを ARMA 過程で記述する枠組みももちろん存在する. 因子過程が 2 次元以上であれば, 当然多変量 ARMA 過程となる. ここでは Peña and Box (1987) を中心に, 関連する研究結果を紹介する.

#### 3.3.1 正準解析と因子モデル

$X_t$  を  $p \times 1$  の定常ベクトル値時系列とし, 以後は  $x_t = X_t - \mu_X$  (ただし  $\mu_X = E(X_t)$ ) を観測値とみなす. ここで次のようなモデルを想定する.

$$x_t = \mathcal{B}z_t + \epsilon_t$$

ただし,  $z_t$  は  $q \times 1$  次元 ( $q \leq p$ ) の観測不能な因子,  $\mathcal{B}$  は  $p \times q$  次元の因子負荷行列で  $\text{rank} \mathcal{B} = r$ ,  $\epsilon_t$  は  $p$  次元のホワイトノイズ過程で, フルランクの共分散行列  $\Sigma_\epsilon$  を持つとする.  $\mathcal{B}$  の取り方に自由度がありすぎると問題なので, ここでは  $\mathcal{B}'\mathcal{B} = I_q$  と仮定して, 直交変換による任意性にとどめておく.

更に, 観測不能な因子  $z_t$  は以下のような多変量 ARMA 過程に従うと仮定する.

$$\phi_z(L)z_t = \theta_z(L)a_t$$

ここで,  $\phi_z(L)$ ,  $\theta_z(L)$  はラグ作用素  $L$  に関するそれぞれ  $p_z$  次,  $q_z$  次の多項式, すなわち

$$\begin{aligned}\phi_z(L) &= I_q - \phi_z(1)L - \cdots - \phi_z(p_z)L^{p_z} \\ \theta_z(L) &= I_{q_z} - \theta_z(1)L - \cdots - \theta_z(q_z)L^{q_z}\end{aligned}$$

とする. このとき,  $z_t \sim \text{VARMA}(p_z, q_z)$  と書くことにしよう. ここでは  $|\phi_z(L)| = 0$  及び  $|\theta_z(L)| = 0$  の根がすべて単位円の外にあると仮定する. このとき  $z_t$  は定常かつ反転可能である. また,  $a_t$  は正規分布に従うホワイトノイズ列で, 平均はゼロ, 共分散行列  $\Sigma_a$  は正値定符号とする. 最初は  $r \times r$  行列  $\phi(\cdot)$ ,  $\theta(\cdot)$  がすべて対角である場合から検討を始めよう. これを独立因子モデル (uncoupled-factors model) と呼ぶ.

Peña and Box (1987) の結果を簡単に説明すれば,  $x_t$  に対するモデルは, 一見自由度がありすぎるように見えるが, 推定すべきパラメータは実は多くの制約を満たさなければならず, 結果としてモデルの識別は容易である, というものである. 特に, (a) AR 係数 ( $\phi_x(h)$ ) はすべて,  $\mathcal{B}$  の生成する部分空間  $S(\mathcal{B})$  を張る共通の  $q$  個の固有ベクトルを持ち, (b) 純粋 MA 表現したときの係数行列  $\psi_x(h)$  はすべてランクが  $q$  で, その列ベクトルは  $S(\mathcal{B})$  に属する. 換言すれば, (a) は  $x_t$  の AR 係数が強い制約下にあることを示し, (b) は MA 係数がそれらの AR 係数にアンカーされていることを示している.

この結果は, 独立因子モデルの特徴付けとしてはよいが, 実際のモデル識別の手続きには, 以下の議論によって共分散行列を利用した方がよい.  $\Gamma_x(h) = E(x_{t-h}x_t')$  を観測値  $x_t$  の共分散行列,  $\Gamma_z(h) = E(z_{t-h}z_t')$  を観測不能因子  $z_t$  の共分散行列としよう. このとき

$$\begin{aligned}\Gamma_x(0) &= \mathcal{B}\Gamma_z(0)\mathcal{B}' + \Sigma_\epsilon \\ \Gamma_x(h) &= \mathcal{B}\Gamma_z(h)\mathcal{B}', \quad h \geq 1\end{aligned}$$

であり,  $h \geq 1$  のときの  $\Gamma_x(h)$  のランクは, 共通因子数  $q$  に等しい. 因子どうしがすべてのラグで独立で  $\Sigma_a$  が対角であれば  $\Gamma_z(h)$  も対角である. このとき,  $\Gamma_x(h)$  は  $h \geq 1$  に対して対称,  $\mathcal{B}$  の列ベクトルは固有値  $\gamma_i(h)$  に対応する  $\Gamma_x(h)$  の固有ベクトルである(一方  $\{\gamma_i(h)\}$  は  $\Gamma_z(h)$  の対角成分でもある). 従って分析の手始めは, 観測値から  $\Gamma_x(1)$ ,  $\Gamma_x(2)$ , ... 等を計算し, それらの固有値と固有ベクトルを調べてみることになる.

潜在因子の数に見当がついたら, 今度は因子の解釈が容易な構造に変換するのが望ましい. そこで, 因子の影響だけを取り出すように以下のような変換を行う.  $\mathcal{B}^-$  を負荷行列  $\mathcal{B}$  の一般化逆行列とすると  $z_t = \mathcal{B}^-x_t - \mathcal{B}^-\epsilon_t$  となる. 次に,  $B$  を  $(p-q) \times p$  行列としてデータに対する変換行列  $M$  を

$$M = \begin{pmatrix} \mathcal{B}^- \\ B \end{pmatrix}$$

で定義すると, 変換後の観測ベクトル  $\tilde{x}_t$  は

$$\tilde{x}_t = Mx_t = \begin{pmatrix} \mathcal{B}^-x_t \\ Bx_t \end{pmatrix} = \begin{pmatrix} z_t + \mathcal{B}^-\epsilon_t \\ B\mathcal{B}z_t + B\epsilon_t \end{pmatrix} = \begin{pmatrix} \tilde{x}_{1t} \\ \tilde{x}_{2t} \end{pmatrix}$$

となる. 因子の影響だけを分離して取り出すには,  $B\mathcal{B} = O$  となるように  $B$  を取ればよい ( $B\mathcal{B}'$  のゼロ固有値に対応する  $p-q$  個の固有ベクトルを  $B$  の行とすればよい).  $\mathcal{B}^-$  の選び方には任意性があるが,  $\Sigma_\epsilon = I$  と仮定できる時には  $\tilde{x}_{1t}$  と  $\tilde{x}_{2t}$  を無相関にすることができる, という理由からムーア・ペンローズ一般逆行列 (すなわち  $\mathcal{B}^- = (\mathcal{B}'\mathcal{B})^{-1}\mathcal{B}'$ ) に利点がある. このような変換を, 正準変換 (canonical transformation, Box and Tiao (1977)) と呼ぶ.

独立因子モデルにおける, 因子のプロセスが無相関という仮定は, 現実的には受け入れがたい. Peña and Box (1987) は, 因子どうしの相関を考慮したモデルを, 非独立因子モデル

(coupled-factors model) と呼ぶ。このとき、 $\Gamma_x(h)$  および  $\phi_x(h)$  の固有ベクトルはもはや  $B$  の列とはならず、独立因子モデルの議論に帰着させるには、 $\Gamma_z(h)$  および  $\phi_z(h)$  を対角化する行列が必要である。しかし、実際にはそれらの行列を知らなくても正準変換が構成でき、因子数の決定は独立因子モデルのときと同様に、 $\Gamma_x(h)$  ( $h \geq 1$ ) や  $\phi_x(1) - \theta_x(1)$  のランク等を手がかりに行えることが示される。因子間の相関の程度問題と言えるかもしれないが、Peña and Box (1987) に示された実際例(スペインにおける複数の州の月次小麦価格の分析)を見る限りでは、実際のモデル識別や因子の抽出に大きな困難はないように見受けられる。

### 3.4 因子過程を明示的に与えない方法

前節までの共分散構造分析型的手法、構造時系列モデル、正準変換を利用した手法は、観測不能な因子過程に明示的に確率過程を与えていることでは共通している。本節では、因子過程を明示的に与えない方法で、動的因子モデルと関連のある方法を取り上げる。ひとつは縮小ランク自己回帰モデルであり、もうひとつは誤差修正モデルである。説明の便宜上、多変量自己回帰モデルを出発点として議論を進めるが、前者は通常の回帰モデルの文脈で論ずることも可能であるし、後者は単一方程式の中で議論される場合もある。

#### 3.4.1 縮小ランク自己回帰モデル

形式的には観測値のみによってダイナミクスが記述される時系列モデルの枠組みで、フィードバックの作用素がランク落ちしていると捉えるものとして、縮小ランク回帰モデル (reduced rank regression) がある。 $x_t$  を  $p \times 1$  次元の観測ベクトルとしよう。 $x_t$  が従う  $p$  変量自己回帰過程

$$(3.6) \quad x_t = \sum_{j=1}^s \Phi_j x_{t-j} + \epsilon_t$$

が縮小ランク回帰になっているとは、ラグ作用素  $I - \Phi(L) = \sum_j \Phi_j L^j$  が  $p \times q$  ( $q \leq p$ ) 次元の作用素  $A(L) = A_1 L + \dots + A_{s_1} L^{s_1}$  と  $q \times p$  次元の作用素  $B(L) = B_1 L + \dots + B_{s_2} L^{s_2}$  によって

$$x_t = A(L)B(L)x_t + \epsilon_t = \sum_{u=1}^{s_1} \sum_{v=1}^{s_2} A_u B_v x_{t-u-v} + \epsilon_t$$

と表されるときを言う(ただしここで、 $s_1 + s_2 = s$  である)。Reinsel (1983) は  $s_2 = 0$  の場合、すなわち

$$x_t = \sum_{j=1}^s A_j B_0 x_{t-j} + \epsilon_t$$

におけるパラメータ推定とモデル選択を議論している。この場合のモデルをインデックスモデルとしばしば呼ぶ。 $B_0 x_{t-j}$  がさまざまな変量を集約した「指数」に対応するという意味合いからであろう。Velu et al. (1986) は逆に  $s_1 = 0$  の場合、すなわち

$$x_t = A_0 \sum_{j=1}^s B_j x_{t-j} + \epsilon_t$$

におけるパラメータ推定とその漸近理論を議論し、更に Box and Tiao (1977) 流の正準変換と縮小ランク回帰モデルとの関連性に言及している。Velu et al. (1986) に記されているように、これらのアプローチは殆ど同じで、「予測可能な構造を持つ  $q$  個の変量を求める」という観点から、共分散行列の固有値の大きい順に注目するか、「不要な成分の個数  $p - q$  を決める」という観点から固有値の小さい順に着目するかの違いでしかない。ただ、Box and Tiao (1977) の枠組みはプロシジャとしては理にかなったものではあったが、Velu et al. (1986) のように推測理論

の枠組みで本格的に議論されたものではなかったと言えよう．更に Ahn and Reinsel (1988) では入れ子型 (nested) 縮小ランク回帰モデルの推定論が展開されている．入れ子型というのは，(3.6) 式の  $\Phi_j$  が

$$\text{rank}(\Phi_j) = q_j \geq \text{rank}(\Phi_{j+1}) = q_{j+1}, \quad j = 1, \dots, s-1$$

という制約を満たすことを言う．現在から遠い時点に関しては必ずしもすべてのチャンネルの信号が意味を持たないとか，あるいは時間のインデックスを空間的なインデックスとして同様の物理的解釈が可能であろう．

### 3.4.2 誤差修正モデル

多変量 AR 過程 (3.6) 式において，観測ベクトル  $x_t$  の各要素が一階の差分の後で弱定常過程に従うとしよう．すなわち， $x_t$  の各要素  $x_{i,t}$  ( $i = 1, \dots, p$ ) について  $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$  が弱定常過程に従うとしよう．一方， $x_t$  の定常性の如何によらず，(3.6) を変形して次に示すような誤差修正表現 (error correction representation, Engle and Granger (1987)) を与えることができる．

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{s-1} \Phi_j^* \Delta x_{t-j} + \epsilon_t$$

$$\Phi_j^* = - \sum_{i=j+1}^s \Phi_i, \quad \Pi = - \left( I - \sum_{j=1}^p \Phi_j \right)$$

このモデルは通例以下のように解釈される．観測ベクトルの変化 (伸び率) は，過去の伸び率によってだけでなく，レベル変数の線形結合  $\Pi x_{t-1}$  によっても説明される．もし  $\Pi = O$  ならこの多変量プロセスは  $p$  本の個別な非定常成分を持ち，またもし  $\Pi$  がフルランクであればもともと  $x_t$  は定常でなければならないので，いずれも考慮対象から外れる．関心があるのは， $\Pi$  のランクが非零で観測次元より落ちている場合である．このとき  $\Pi x_{t-1}$  はレベル変数間に成り立つ線形制約であり，経済学的な文脈では経済変数間の長期的な均衡関係であると解釈される．実際には  $\Pi x_{t-1}$  は恒等的にゼロということはありません確率的にバラつく．このバラつきは「長期的均衡からの乖離」とも言うべき「誤差」であり，その「誤差」が現在の伸び率にフィードバックしてくる形式になっているところから，誤差修正モデルと呼ばれている．

このように，誤差修正モデルでは明示的に因子過程に興味があると言うよりは，背後にある因子過程の次元と，観測変数に関してそこから導かれる線形制約式に関心がある．データから因子過程の次元を決定する方法として最もよく知られた方法が Johansen の尤度比検定 (Johansen (1988)) であり，経済学においてはこの接近法が主流である．Johansen 法については既に多くのレビューやテキストが存在する．例えば Johansen (1995), Hamilton (1994) の 20 章などがよいと思われる．共通トレンドの数の決定に関するものとしては，他に Stock and Watson (1988) がある．

観測値  $x_t$  が非定常の場合は，頻度論者の立場からは誤差修正モデルを経由するのが常套手段である．しかし，ベイズ統計の立場に立てば，レベル変数のまま縮小ランク回帰を考察することが可能である．ベイズ的な観点から縮小ランク回帰に取り組んだものとして Geweke (1996) を挙げておく．Gibbs sampler の適用を容易にするように各パラメータの周辺分布の形を決めるのは自明ではなく，事前分布の入れ方に工夫が必要である．

#### 4. 時間領域分析とラグ構造

ここまでの議論は、時系列が持つ系列相関をどう取り込みながら多変量データの次元を縮約してゆくかという観点から様々な手法を眺めてきた。ここでは、主成分・因子分析、周波数領域・時間領域という区分けで生じる4つの象限で言うと最後に残った、時間領域での主成分分析について議論する。4.1節に述べるように、多変量時系列に対する時間領域での主成分分析の有効性は、ノンパラメトリックなトレンド推定や、インデックス作成に代表されるデータ圧縮などに限られる。古典的な主成分分析の方法論を時系列データにそのまま適用することは、時間方向の相関を無視して同一時点での確率変数の相関のみに着目することに等しいからである。4.2節では、ラグ構造なしの一因子モデルが時間領域での主成分分析にほぼ対応することを利用し、ラグを伴う動的因子モデリングの重要性を強調するための事例として、2変量の経済モデルを考察する。

##### 4.1 時間領域での主成分分析と二段階法

標準的な多変量解析の枠組みでは、データはある多変量分布からランダムかつ独立に抽出された標本であると仮定されている。このとき、第1の観測値から第 $n$ の観測値までが、どのような順序で得られようと議論に変更はない。なぜなら、観測行列内で個体に関する順序を入れ替えても共分散行列は本質的に不変だからである。しかしながらこのような状況は、すべての多変量データに当てはまる特徴ではない。時系列や空間データなどでは順序とそれに伴う相関構造が重要であり、順序を変更してしまえば統計解析の結果も変わってしまう。

逆に考えれば、時点を止めて得られる観測値の組内での、いわば同時点における相関が唯一の関心事であれば、系列相関を情報として活用する必要はない。これは、多変量時系列データの「時系列」という側面を捨象して、「相関を有する確率変数の組が たまたま 時間の流れに沿って観測されたにすぎない」とする考え方と言える。

具体的な応用事例を3つ取り上げよう。第一に、景気指数などのインデックスを作成する問題が挙げられる。Theil (1960) や Kloek and de Wit (1961), 刈屋 (1986) などがこの問題を論じている。第二に、主成分分析を通じてトレンドを取り出すなど、データの平滑化を行いたい場合などがあげられるだろう。このような事例に関する文献としては例えば Ahamad (1967) が挙げられるが、分析の前段階で見逃していた自明な要因の検出に終わることもしばしばである。Ahamad (1967) の解析は、1950年から63年までのイングランドとウェールズにおける18種の犯罪に関する年間発生件数のデータに主成分分析を行ったものであるが、主成分分析に先立って人口で犯罪件数を補正するなどの処理を行わなければ、13歳から19歳までの青少年人口の増加率とほぼ同じ曲線が第一主成分として出てきてしまう。第三に、株価収益率のように、同時点での複数銘柄間のショックの相関が非常に強く、一方でその影響がすぐに減衰する(と考えられている)ものについては、同時点における相関のみに注目する理由があると言える (Kariya (1993))。

このように、主成分分析を平滑化やデータ圧縮の一手段として用いている範囲ではまだ問題が少ないが、主成分分析で得られた結果をあたかも新たなデータのように取り扱う二段階法には、大いに問題があると思われる。すなわち、最初はあたかも独立なサンプルが得られたかのような立場で時系列から主成分を抽出しておいて、その後で主成分の系列に時系列モデルをあてはめる、という手法である。この方法論は、それが実際の応用の場でどの程度役に立つかどうかは別として、論理的に破綻している。最初に形式的に主成分分析を行う段階では、分析者はあくまで同時点での構造にしか興味がないという立場にある。たまたま時間の流れに沿って多変量データがやってきたのだ、と。にもかかわらず分析者は、主成分分析が終わった途端に今度はそれらにダイナミクスを認める立場へと豹変するのである。

同じような事例は、主成分分析を離れたところでも見つかる。例えば、各時点で重回帰分析を行って、得られた係数を事後的に新たな「観測値」だと思ってそれに時系列モデルをあてはめる、などという手続きも同様である。これも、回帰分析を行う段階ではパラメータ自体は固定された真の値があると思っているのに、時点ごとに回帰分析が終わった後ではそれに確率的変動を認めるという一貫性のなさが問題である(例えば Haugen and Baker (1996))。それでも、こうした「事後処理」によって得られた結果を平滑化してから解釈したい、という動機であればまだ分析の意味は認められるかもしれないが、これを予測に用いるとなるとはなはだ疑問である。係数列にあてはめた時系列モデルのよさは、あくまで仮想的な「観測値」の空間で測られた良さであり、元のデータの空間での予測性とは何の関係もないからである。

計量心理学では、多水準分析 (multilevel analysis) において同様の問題が指摘されている。ここで言う問題とは、例えば、地区、学校、その中の生徒、というように様々なサンプリングレベルが存在するときに、ある学校のある生徒たちから得られたデータに対して何らかの統計解析を行い、そこで推定されたパラメータを一段上のレベルである学校の特性値として(つまりデータとして)扱うことである。この種の問題に対しては階層的なモデルを与えることが直接的な解法になるが、詳しくは Hox and Kreft (1994) および同論文所収の特集号を参照されたい。

#### 4.2 数値例

前節に述べたように、多変量時系列に対する時間領域での主成分分析の有効性は、ノンパラメトリックなトレンド推定や、インデックス作成に代表されるデータ圧縮などに限られるものである。主成分分析は本質的に座標の変換に過ぎないので、3.1.1 節冒頭に述べたような、ラグ構造を無視した古典的因子分析に加えらるる批判もまた同様に避けられない。そこで本小節では、ラグ構造なしの一因子モデルが時間領域での主成分分析に対応することに注目し、ラグを伴う動的因子モデリングの重要性を強調するための実例として、以下のような2変量の経済モデルを考えよう。

$p_t$  ( $t = 1, \dots, T$ ) を消費者物価指数とし、 $y_t$  は他の何らかの経済変数を表すものとする。ここで、 $p_t$  と  $y_t$  はおのおの個別に確率的トレンド ( $\mu_{1,t}$ ,  $\mu_{2,t}$ ) を持つが、トレンド周りの定常成分 ( $s_t$ ) は共通で、各変数への因子負荷は異なると仮定する。更に確率的トレンドは1階のランダムウォークに従い、共通の定常成分は2次のARモデルで表現されるとする。すなわちここで想定しているモデルは、

$$\begin{aligned}
 \mu_{1,t} &= \mu_{1,t-1} + v_{1,t}, & v_{1,t} &\sim N(0, \tau_1^2) \\
 \mu_{2,t} &= \mu_{2,t-1} + v_{2,t}, & v_{2,t} &\sim N(0, \tau_2^2) \\
 s_t &= \phi_1 s_{t-1} + \phi_2 s_{t-2} + v_{3,t}, & v_{3,t} &\sim N(0, \tau_3^2) \\
 p_t &= \mu_{1,t} + c_1 s_t + w_{1,t}, & w_{1,t} &\sim N(0, \sigma_1^2) \\
 y_t &= \mu_{2,t} + c_2 s_t + w_{2,t}, & w_{2,t} &\sim N(0, \sigma_2^2)
 \end{aligned}
 \tag{4.1}$$

と明示的に表される。モデルの推定に先立って  $\{p_t\}$ ,  $\{y_t\}$  の平均をゼロに調整する。更に差分系列を適当に規格化することにより、 $\tau_1^2 = \tau_2^2$  としてよい。また、一般性を失うことなく  $\tau_3^2 = 1$  と仮定する。

$y_t$  として失業率を考えれば、そのトレンドは自然失業率と見なすことができ、そこからの乖離が超過需要の指標となる。それが物価の変動をある程度説明するという図式は、フィリップス曲線の定式化である。本数値例の目的は、そのような経済理論の実証ではなく、共通因子にラグ構造を持たせる必要性を強調することである。従って、(4.1) 式を若干拡張して、

$$p_t = \mu_{1,t} + c_1 s_{t-\ell} + w_{1,t}, \quad (\ell = 0, 1, 2, \dots)$$

表 1. パラメータ推定値.

パラメータ	ラグ 0	ラグ 5
$\tau^2$	$0.83 \times 10^{-2}$	$0.90 \times 10^{-2}$
$\sigma_1^2$	$0.14 \times 10^{-1}$	$0.15 \times 10^{-2}$
$\sigma_2^2$	$0.25 \times 10^{-2}$	$0.18 \times 10^{-2}$
$\phi_1$	1.47	0.84
$\phi_2$	-0.55	-0.52
$c_1$	-0.08	0.10
$c_2$	-0.04	0.56
対数尤度	-135.47	-133.29

と置き換えよう．すなわち， $y_t$  の変動を説明するのに  $s_t$  は即座に反応するが， $p_t$  (物価) に対してはラグを伴って作用する可能性を考慮する．更に  $y_t$  には，経済活動の先行指標となり得るものとして，ここでは建築受注 (民間計) を取り上げる．原データは図 1 の上側 2 枚のパネルにプロットしてある．ともにデータ期間は 1982 年第 1 四半期から 1999 年第 4 四半期までである．元々は月次データであったものを四半期に変換している．ただし，消費者物価指数については，月次の段階で消費税導入時と引き上げ時のレベルシフトをあらかじめ除去した後に四半期に変換してある．

実際の推定結果が表 1 に報告されている．ここでの分析は，2 つの時系列に 3 つの観測不能要素が入っており，単純な主成分分析が適用できる状況ではないが， $\ell = 0$  の場合がその状況に近いと言える．一方， $\ell = 5$  のモデルのほうが対数尤度で 2 弱改善している．改善度は小さいように見えるが，元々のデータはトレンド項が支配的で，推定されたパラメータ ( $\tau^2$ ) に大きな差がないことを考えれば，ある程度納得がいく．

ラグ 0 のモデルでは，トレンドモデルはラグ 5 の場合とあまり変わらないものの，推定された共通 AR 成分の分散が一桁大きい．また，AR 係数 ( $\phi_1, \phi_2$ ) も非定常に近い値となっている ( $\phi_1 + \phi_2 \approx 1$ )．実際，図 1 の下側 2 枚のパネルを見ると，ラグ 0 で推定された共通 AR 成分は局所的にトレンドを持っているように見える．このことは，共通因子の働きにラグ構造を考慮しなかったために，適切な成分が抽出できなかったことを示している．

## 5. まとめ

多変量時系列データに対する主成分分析・因子分析についてこれまでの主な結果を整理した．分析を時間領域で行うか周波数領域で行うかが手法を分類する重要な軸となる．周波数領域での主成分・因子分析は，時系列の離散 Fourier 変換を利用する．時間をインデックスとしたとき近接する観測値にある相関も，周波数をインデックスとすると漸近的に独立性が確保され，独立同一分布ケースでの主成分・因子分析の複素変数版を考えることで多変量時系列に対する分析 (スペクトル密度行列の固有値分解) が可能になる．

時間領域での因子分析は，観測されない因子に何らかの形で時系列的な表現を与える手法として概括できる．接近法としては，従来の因子分析にラグ構造を入れて共分散構造分析の観点から解析するもの，構造時系列モデルの枠内で共通因子を与えるもの，共通因子を多変量 ARMA 過程で記述するものなどがある．これらはすべて明示的に因子に対してモデルを与えるものである．明示的に共通因子をモデル化しない場合は，作用素のランク条件に議論が集約される．そのような手法としては縮小ランク回帰モデルや誤差修正モデルがある．一方，古典

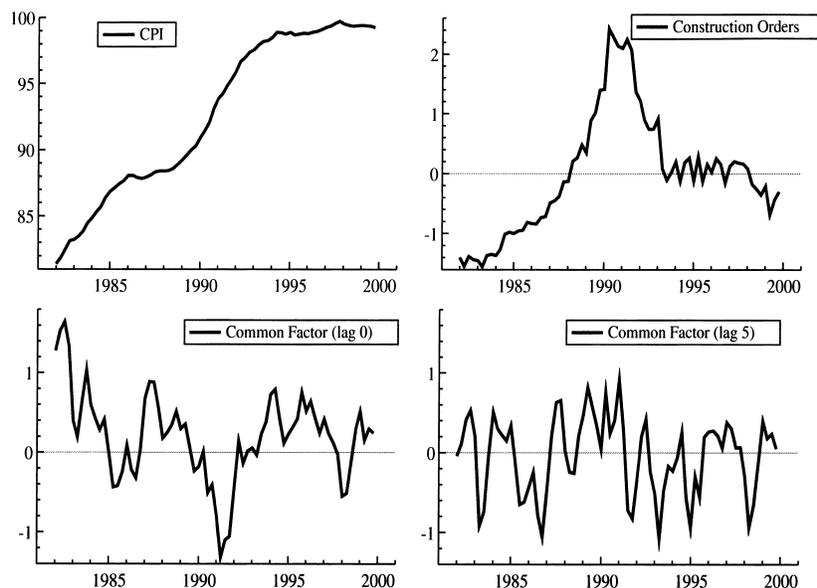


図 1. (上左) 消費者物価指数, (上右) 建築受注, (下左) ラグなしで推定された動的共通因子, (下右) ラグ 5 期として推定された動的共通因子.

的な主成分分析の方法論を時系列データにそのまま適用することは, 時間方向の相関を無視して同一時点での確率変数の相関のみに着目することに等しく, その有効性はノンパラメトリックなトレンド推定やインデックス作成に代表されるデータ圧縮等に限られる. このような立場で一度主成分分析を行った結果に新たにモデルを与える手法には注意が必要である.

## 謝 辞

執筆を助めてくれた編集委員の伊庭幸人氏に感謝します. また, 二人の査読者の丁寧なコメントのおかげで内容は大幅に改善されました. ここに記して感謝いたします. 最後に, 誠実に討論してくれた前田忠彦氏(統計数理研究所)に感謝いたします. 同氏との討論がなければ, 本稿は完成し得ませんでした.

## 参 考 文 献

- Ahamad, B. (1967). An analysis of crimes by the method of principal components, *Appl. Statist.*, **16**, 17–35.
- Ahn, S. K. and Reinsel, G. C. (1988). Nested reduced rank autoregressive models for multiple time series, *J. Amer. Statist. Assoc.*, **83**, 849–856.
- Anderson, T. W. (1963). The use of factor analysis in the statistical analysis of multiple time series, *Psychometrika*, **28**, 1–25.
- Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*, Wiley, New York.

- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, California.
- Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series, *Biometrika*, **64**, 355–365.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory* (expanded edition), Holden-Day, San Francisco, California.
- Engle, R. F. and Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing, *Econometrica*, **55**, 251–276.
- Geweke, J. F. (1977). The dynamic factor analysis of economic time series model, *Latent Variables in Socio-Economic Models* (eds. D. J. Aigner and A. S. Goldberger), 365–383, North-Holland, Amsterdam.
- Geweke, J. F. (1996). Bayesian reduced rank regression in econometrics, *J. Econometrics*, **75**, 121–146.
- Geweke, J. F. and Singleton, K. (1981). Latent variable models for time series: A frequency domain approach with an application to the permanent income hypothesis, *J. Econometrics*, **17**, 287–304.
- Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory, *Econometrica*, **17**, 287–304.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, Massachusetts.
- Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns, *Journal of Financial Economics*, **41**, 401–439.
- Holtzman, W. H. (1962). Methodological issues in P-technique, *Psychological Bulletin*, **59**, 243–256.
- Hox, J. J. and Kreft, I. G. G. (1994). Multilevel analysis methods, *Sociological Methods and Research*, **22**, 283–299.
- Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.
- Johansen, S. (1988). Statistical analysis of cointegration vectors, *J. Econom. Dynam. Control*, **12**, 231–254.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices, *Psychometrika*, **43**, 443–477.
- Jöreskog, K. G. and Sörbom, D. (1998). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*, Scientific Software International, Chicago, Illinois.
- 刈屋武昭 (1986). 『計量経済分析の考え方と実際』, 東洋経済新報社, 東京.
- Kariya, T. (1993). *Quantitative Methods for Portfolio Analysis; MTV Approach*, Kluwer, Dordrecht.
- Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*, Springer, New York.
- Kloek, T. and de Wit, G. M. (1961). Best linear and best linear unbiased index numbers, *Econometrica*, **29**, 602–616.
- Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series, *Psychometrika*, **50**, 181–202.
- Molenaar, P. C. M., de Gooijer, J. G. and Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series, *Psychometrika*, **57**, 333–349.
- 丘本 正 (1986). 『因子分析の基礎』, 日科技連, 東京.
- Peña, D. and Box, G. E. P. (1987). Identifying a simplifying structure in time series, *J. Amer. Statist. Assoc.*, **82**, 836–843.

- Phillips, P. C. B. and Ouliaris, S. (1988). Testing for cointegration using principal component analysis, *J. Econom. Dynam. Control*, **12**, 205–230.
- Priestly, M. B. and Subba Rao, T. (1975). The estimation of factor scores and Kalman filtering for discrete parameter stationary process, *Internat. J. Control*, **21**, 971–975.
- Priestly, M. B., Subba Rao, T. and Tong, H. (1974). Application of principal components analysis and factor analysis in the identification of multivariable systems, *IEEE Trans. Automat. Control*, **19**, 730–734.
- Reinsel, G. C. (1983). Some results on multivariate autoregressive index models, *Biometrika*, **70**, 145–156.
- Shumway, R. H. and Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*, Springer, New York.
- Stock, J. H. and Watson, M. W. (1988). Testing for common trends, *J. Amer. Statist. Assoc.*, **83**, 1097–1107.
- Theil, H. (1960). Best linear index numbers of prices and quantities, *Econometrica*, **28**, 464–480.
- Tukey, J. W. (1978). Can we predict where “time series” should go next?, *Directions in Time Series* (eds. D. R. Brillinger and G. C. Tiao), 1–31, Institute of Mathematical Statistics, Ames.
- Velu, R. P., Reinsel, G. C. and Wichern, D. W. (1986). Reduced rank models for multiple time series, *Biometrika*, **73**, 105–118.
- Young, P. C. and Pedregal, D. J. (1999). Macro-economic Relativity: Government spending, private investment and unemployment in the USA 1948–1998, *Structural Change and Economic Dynamics*, **10**, 359–380.

## Principal Component and Factor Analysis for Multiple Time Series

Yoshinori Kawasaki

(The Institute of Statistical Mathematics)

The use of the principal component analysis and the factor analysis (PCA/FA) on multivariate time series is discussed. The standard results on PCA/FA rely on the assumption of independent sampling from an identical multivariate distribution, which is almost always inappropriate in time series settings. In this paper, we review two approaches in PCA/FA that is accommodated to serially correlated observations. The first attempt is to use discrete Fourier transformation of a time series. Asymptotic independence of DFT enables the ordinary PCA/FA technique valid in the frequency domain. The second methods are more straightforward in a sense that they give an explicit time series model to the latent factor series, which can be called dynamic factor analysis. We detail several styles of modeling dynamic factor, namely the simultaneous structural equation, the structural time series models and the canonical transformation of time series. Reduced rank regression models and error correction models are also related to dynamic factor model but they do not assume any explicit model for latent factor series. It is also mentioned that the principal component analyses in the time domain think little of the serial correlation of a time series. A numerical example in the final section highlights the importance of considering a lagged common factor, which will not be detected by a simple-minded application of PCA.

---

Key words: Multivariate time series, principal component analysis, factor analysis, frequency domain, time domain.