

赤池情報量基準(AIC)を用いた時系列データの重回帰分析

Multiple Regression Analysis of Time Series Data with Akaike Information Criterion (AIC)

森 下 伊三男
Isao Morishita

要 旨

何らかの因果関係のある二つの時系列データについて、両者の関係が重回帰モデルで表される場合の解析方法について考察した。このようなモデルでは、一般に、一方の時系列データを目的変数（従属変数）、その時系列データの変動要因となる他方の時系列データを説明変数（独立変数）、その説明変数に掛かる係数及び定数項が未知パラメータ（回帰係数）となっている。そして、ある時刻の目的変数の値は、その時刻以前のあらゆる時刻での説明変数と回帰係数を用いた1次式で表現される。重回帰分析では、説明変数の実際の値と重回帰式により計算される計算値との残差二乗和が最小となるように、最小二乗法を用いて回帰係数が決定される。しかし、一般に、説明変数を多く用いれば最小二乗法における残差二乗和をより小さくすることができる。従って、そのような重回帰分析のみの結果からは、導入される説明変数の数の決定は不可能である。そこで、採用される説明変数を選択し、その最適な組み合わせを得る為に、本稿では、赤池の情報量基準（AIC: Akaike Information Criterion）を導入した方法について論じ、また、実際の時系列データを用いた解析事例を紹介する。

1 はじめに

時系列データを解析する時には、一般に、以下の様な手法が多く用いられている。

単一の時系列データの解析には、自己回帰モデルを考えて、自己相関関数やパワースペクトルを求めたり、あるいは周波数解析を行う。その結果、対象となる時系列データに内在するさまざまな特性を見い出すことができる〔例えば、1, 2〕。前者の自己相関関数やパワースペクトルからは、その時系列データの持っている周波数特性をくみ取ることができる。しかし、特定の周波数成分について得られる情報はその大きさ（power）のみであり、位相についての情報は得られない。一方、後者の周波数解析からは、注目している周期成分の大きさ（amplitude）とその位相を得ることができる。ただし、データの中に何らかの周期成分が存在していても、途中でその位相が変化する場合、周波数解析では良い

結果が保証されないという欠点もある。

二つの時系列データの関係を解析するには、相互相関関数とクロススペクトルを用いた分析が行われる〔例えば、3, 4, 5〕。二つの時系列データ間に、何らかの“遅れ”を伴った相関関係が存在すれば、相互相関関数を調べることで、その“遅れ”の大きさと相関の度合いを簡単に知ることができる。また、クロススペクトルを求めることで、各周波数成分での相互の関連の大きさ（power）及び位相の情報も得ることができる。

一方、二つの時系列データの関係に明らかな因果関係が考えられる場合、そのようなスペクトル分析よりも、次のようなモデルを考えた方がより適した解析方法となることが多い。即ち、原因となる何らかの入力時系列データ $x(t)$ に対し、定係数線形系を考慮して重み関数 $h(\tau)$ を導入し、時系列データ $y(t)$ がその結果出力となるような次式及び図1で示されるようなモデルである。

$$y(t) = y_0 + \int_0^{\infty} h(\tau) x(t-\tau) d\tau \quad (1)$$

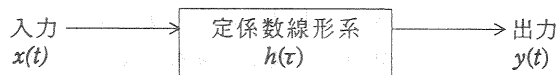


図1. 定係数線形系のモデル

これはいわゆる畳み込み積分 (Convolution Integral) であり、この積分方程式を解いて $h(\tau)$ を求めることで両者の基本的な動特性を調べることができる。ただし、ここでは、時系列データ $y(t)$ を定数成分（直流成分）と変成分（交流成分）に分けることができるとし、単純な畳み込み積分に定数項 y_0 を導入してある。従って、交流成分（ $y(t)-y_0$ ）だけを取り出して、それを新たな $y(t)$ と置き換えれば、畳み込み積分の項だけになる。

このモデルでは、データは時間に関し連続関数として与えられている。しかし、一般に解析されるデータは、ある期間についての平均値等の離散化されたデータであることが多い。そこで、本稿では、初めに上の式で表現されるようなモデルを、時間的平均値に対応した形に変形し離散化する。そして、結果出力となる方を目的変数とし、原因となる方を説明変数とした重回帰モデルの式となるように変形する。その変形後は、重み関数 $h(\tau)$ を求めることは、重回帰モデルの回帰係数を求めることと同等になる。

その様な重回帰モデルにおける回帰係数の決定は、一般に最小二乗法によって行われる。即ち、目的変数となる時系列データの実際の値と、モデルによって計算された値との差（残差）の二乗をすべてのデータについて足し合わせた残差二乗和が最小となるように回帰係数が決定される。時系列データの場合、一般の重回帰分析と違い、説明変数そのもの

は互いに独立ではなく、どれだけの数の回帰係数を導入すべきなのか又は導入できるのか、モデルを考える上で重要なポイントとなる。なぜなら、一般に、回帰係数を多く導入すればするほど、残差二乗和を小さくすることができるからである。従って、残差二乗和が小さいことは回帰係数の値を決定する時には必要となるが、回帰係数導入の必然性には全く関与していないことになる。最小二乗法を用いる限り、その様な回帰係数の必要性についての答えは得られない。以下、本稿では、定式化の後、実際の解析で問題となる回帰係数の取り扱いについて、赤池の情報量基準（AIC: Akaike Information Criterion）の考え方を導入する。そして、具体的な時系列データの例を取り上げて、AICを用いた回帰係数の決定について説明する。

2 モデルの定式化と赤池の情報量基準の導入

2-1 モデルの定式化

初めに、1で紹介した定係数線形系の時間的平均値に対する定式化から始める〔6〕。

一般に、時々刻々と変化していくデータに対し、我々はある時間間隔をとってその間の時間平均値を用いることが多い。そこで、先に導入した二つの時系列データ $x(t)$, $y(t)$ について、時刻 $t_n - \Delta t/2$ から $t_n + \Delta t/2$ ($n = \dots, -1, 0, 1, \dots$) までの時間間隔 $\Delta t (= t_n - t_{n-1})$ の平均値をとることとし、それらを x_n , y_n とする。すると、式(1)を時間 t で積分することによって、

$$y_n \Delta t = \int_{T_n}^{T_n'} y(t) dt = y_0 \Delta t + \int_{T_n}^{T_n'} dt \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau, \quad (2)$$

となる。ここで、 $T_n = t_n - \Delta t/2$, $T_n' = t_n + \Delta t/2$ とした。また、以下の変形の都合で τ についての積分の下限を $-\infty$ としたが、因果関係の要請より、 $\tau < 0$ の時には $h(\tau) = 0$ であるとする事で本質的な問題とはならない。この積分の変数変換 ($t - t' = \tau$, $d\tau = -dt'$) をしてやると、次のような式を得る。

$$\begin{aligned} y_n \Delta t &= y_0 \Delta t + \int_{T_n}^{T_n'} dt \int_{-\infty}^{\infty} h(t-t') x(t') dt' \\ &= y_0 \Delta t + \int_{T_n}^{T_n'} dt \left[\sum_{j=0}^{\infty} \int_{T_n - \tau_j}^{T_n' - \tau_j} h(t-t') x(t') dt' \right], \end{aligned} \quad (3)$$

ここで、 $\tau_j = j\Delta\tau$ とし、 $\tau < 0$ では $h(\tau) = 0$ なので $j < 0$ の領域は Σ から除くことにした。つぎに、 $T_n - \tau_j$ から $T_n' - \tau_j$ の期間の $x(t')$ の平均値を x_{n-j} と表すとする、式(3)は

近似的に次のように展開できる。

$$y_n \Delta t = y_0 \Delta t + \sum_{j=0}^{\infty} x_{n-j} \int_{T_n}^{T_n'} dt \int_{T_n - \tau_j}^{T_n' - \tau_j} h(t-t') dt' \quad (4)$$

$$= y_0 \Delta t + \sum_{j=0}^{\infty} x_{n-j} \left[\int_{\tau_{j-1}}^{\tau_j} (\tau - \tau_{j-1}) h(\tau) d\tau + \int_{\tau_j}^{\tau_{j+1}} (\tau_{j+1} - \tau) h(\tau) d\tau \right] \quad (5)$$

$$= y_0 \Delta t + (1/2) h_0 x_{n-0} \Delta t \Delta \tau + \sum_{j=1}^{\infty} h_j x_{n-j} \Delta t \Delta \tau. \quad (6)$$

ここで、 $\Delta \tau$ は先の Δt に等しく、 $h_j (j \neq 0)$ は式(4)の二重積分の平均値を表すものとする。また、 $j=0$ の場合には、 $\tau < 0$ で $h(\tau)=0$ となるので、二重積分の積分区間の内で半分の区間では $h(t-t')$ はゼロとなる。その結果、 $j=0$ の場合、 $h(t-t')$ の二重積分による平均値はゼロではない区間だけを考えた場合の平均値に対して半分の値になっている。そこで、平均操作である二重積分の結果を $(1/2)h_0$ とおくことにする。このようにすれば、 $h(t-t')$ のゼロでない区間についての平均値が h_0 と一致することになり、式の解釈がきわめて自然となる。また、式(4)の二重積分は式(5)の様に単純な積分の和として表現することができる。この式を見ると、各積分の区間がそれぞれ重なっていることが分かる。例えば、 h_0 の積分区間は $0 \sim \Delta \tau$ 、 h_1 では $0 \sim 2\Delta \tau$ 、 h_2 では $\Delta \tau \sim 3\Delta \tau$ 、...という具合である。しかし、それぞれの重なっている区間が平均値に及ぼす影響はそれほど大きくないと考えられるので、単に平均値 h_j はそれぞれ次のような区間の平均値であるとみなしても差し支えはない。

$$\begin{aligned} j=0 & : 0 < \tau < \Delta \tau / 2 \\ j>0 & : (j-1/2)\Delta \tau < \tau < (j+1/2)\Delta \tau \end{aligned}$$

最後に、式(6)を Δt で割り、 $(1/2)h_0 \Delta \tau$ 、 $h_j \Delta \tau$ をそれぞれ f_0 、 $f_j (j=1, 2, \dots)$ と置き換えてやると、次式のような表現になる。

$$y_n = y_0 + \sum_{j=0}^{\infty} f_j x_{n-j} \quad (7)$$

この式は重回帰モデルを表す式であり、回帰係数(定数項 y_0 と各係数 $f_j (j=0, 1, 2, \dots)$)は最小二乗法を用いて決定される。すなわち、重回帰モデルで計算される値 $y_{n, \text{CAL}}$ と実際に得られている値 $y_{n, \text{OBS}}$ との残差二乗和 Δ^2 が最小となるような y_0 と $f_j (j=0, 1, 2, \dots)$ が求められることになる。時系列データ y_n のデータ数を N とすると残差二乗和 Δ^2 は次のように表される。

$$\Delta^2 = \sum_{n=1}^N \{y_{n,CAL} - y_{n,OBS}\}^2 \quad (8)$$

ここで、もし式(7)の f_0 のみを取り上げ、 $f_j (j \geq 1)$ を考慮しない場合は x_n と y_n の単相関を考えていることになる。また、 $f_0 \sim f_\infty$ の中からどれか一つだけを取り上げて残りの係数をすべてゼロとした場合には、“遅れ”を考えた単相関を調べることになる。

時系列データに対する一般の重回帰モデルでは、ある時点での y_n に影響を及ぼすのは、その時刻より有限の時間 $j, \Delta\tau$ 以前までと仮定する事が多い。その場合には式(7)の Σ の上限は ∞ ではなく j に置き換えられる。そして y_n は $x_k (k=n, n-1, \dots, n-j)$ の値にのみ依存することになり、各係数 $f_j (j=0, 1, 2, \dots, j)$ が決定されることになる。

2-2 情報量基準の導入

以上で、二つの時系列データに対する式(1)の積分方程式から重回帰分析の式(7)を導入することができた。しかし、一般的に j を大きくとればとるほど式(8)の Δ^2 の値は小さくなっていく。従って、十分に意味のある j の値を、他の情報から得ることができる場合は良いのだが、通常は、 j の上限すら分からないことが多い。

そこで、次に、赤池の情報量基準(AIC)の導入を行う[7]。重回帰モデルで、データの数 N 、未知パラメータ(回帰係数)の数を $m (=j+2)$ とすると、AICは次のように表される。

$$AIC = N(\log 2\pi + 1) + N \log N \Delta^2 + 2m \quad (9)$$

一般に、AICを用いる場合、AICの値そのものに意味があるのではなく、未知パラメータの数 m を変えたときにAICがどのように変化するかを調べる事が重要である。もし、ある回帰係数を導入することによってAICが減少するのなら、少なくとも導入された回帰係数はモデルにとって必要であると判断される。逆に、導入によってAICが増加する場合には、その回帰係数の導入は不適当であるということになる。そこで、式(9)から未知パラメータの数 m に依存する項のみを取り出し、新たなAICの値として以下の式を本稿では用いることにする。

$$AIC = N \log \Delta^2 + 2m \quad (10)$$

この式からはっきり分かるように、AICの増加要因としては残差二乗和 Δ^2 の増加と未知パラメータの数 m の増加があげられる。残差二乗和 Δ^2 の増加はモデルによる計算値が実際の観測値を再現できないことによるものである。従って、もし、データの数 N とパラメータの数 m が等しく、異なる説明変数を用いたモデル間での比較(例えば、 x_n の候補として異なる二つの時系列データを用い、どちらの時系列データが y_n を再現するのにより有効であるかを調べる場合など)を考えた場合、 N や m の値は同じであるから、残差

二乗和 Δ^2 の小さい方が AIC も小さくより有効であることが分かる。この場合、モデルの良否について、AIC を用いた結果と最小二乗法を用いた結果とは一致する。しかし、未知パラメータの数 m が異なる場合はそう単純ではない。未知パラメータの数 m の増加は一般的に残差二乗和 Δ^2 を減少させるものであり、単に残差二乗和 Δ^2 の大きさだけからではどちらが良いか判断できない。その点で、AIC では未知パラメータの数 m が増加要因となっている。従って、不要な係数の導入は残差二乗和 Δ^2 のみを調べたのでは分からず、AIC を導入することによって初めてその不必要性が分かることになる。更に、 x_n として異なる二つの時系列データを比較する場合でも、AIC をより小さくする時系列データの方が y_n を説明するのにより適当であると考えることができる。当然のことながら、同じ時系列データ x_n を用いた重回帰分析の場合、係数のあらゆる組み合わせを考えて、その中で AIC が最小となるような組み合わせが最適解ということになる。

本稿では、以下、実際の時系列データを用い、その様な最適な係数の組み合わせの選択を試みる。

3 解析事例

ここでは、具体的なデータを用いて、AIC を利用した重回帰分析の最適解を求める例を紹介する。目的変数として用いる時系列データは、1950 年 1 月から 1992 年 10 月までの約 43 年間 (514 カ月) にわたって記録された宇宙線強度 $I(t)$ (データの総数 $N=514$) である。宇宙線は宇宙から地上に降り注いでいる放射線で、我々地上の生物に対して重要な環境の要素となっているものである。また、説明変数となる時系列データは太陽のウォルフ黒点数 $R(t)$ である。ウォルフ黒点数は、地球のあらゆるエネルギーの源となっている太陽活動の大きさを表すもので、総体に大きな値を持つほど太陽は活発に活動をしている傾向にあると考えられている。良く知られているように、太陽の活動の大きさは約 11 年の周期で変動しており、それに伴って宇宙線強度も逆相関を示すような変動をしている。ただし、正確に連動した変動をしているのではなく、宇宙線強度の変動に若干の“遅れ”がみられ、更にその“遅れ”は単純に何カ月という具合に表現できないことも知られている。従って、両者の関係は式 (1) の積分方程式で表されるモデルが成立すると考えられている [6]。即ち、

$$I(t) = I_0 + \int_0^\infty h(\tau) R(t-\tau) d\tau \quad (11)$$

と表現することができる。ここで実際に扱うそれぞれのデータとしては、1 カ月間の観測の平均値 (月平均値) であり、それらをそれぞれ I_n, R_n と表すものとする。時系列データが離散化されていることに伴って、上の式 (11) の積分方程式を 2-1 で述べた様に離散化して、次のような関係式を得る。

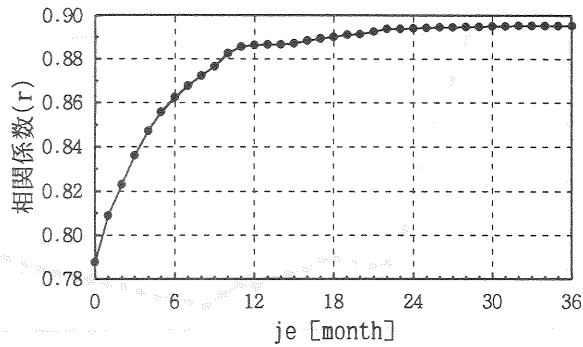


図2 . 上限 j_e を変化させたときの相関係数(r)

$$I_n = I_0 + \sum_{j=0}^{j_e} f_j R_{n-j} \quad (12)$$

ここで問題になるのが、 Σ の上限 j_e の値の決定及び上限 j_e より小さな j に対するあらゆる係数 f_j の必要性の吟味である。まず、上限 j_e をいろいろと変えて重回帰分析をしたときの様子を図2に示す。図2の縦軸には、重回帰分析で得られた係数を用いた計算値と実際に観測された観測値との合いさ加減を表す値として、両者の相関係数(r)を示してある。また、横軸には上限 j_e をとってある。上限 j_e を大きくすればするほど計算値は観測値を十分に再現する事ができる様になる為、相関係数は単調増加関数となっている。図3には上限 j_e を36にしたときに得られる係数 f_j を示す。全体に係数 f_j が負の側に多いのは、太陽活動度と宇宙線強度が逆相関関係を示していることを意味している。この図を見ると、 j の値が23以上の係数 f_j の値は全体的に小さく、上下のばらつきも見られ、それらが十分に意味のある値を示しているか否かの判断ができなくなっている。一方、 j の値が22以

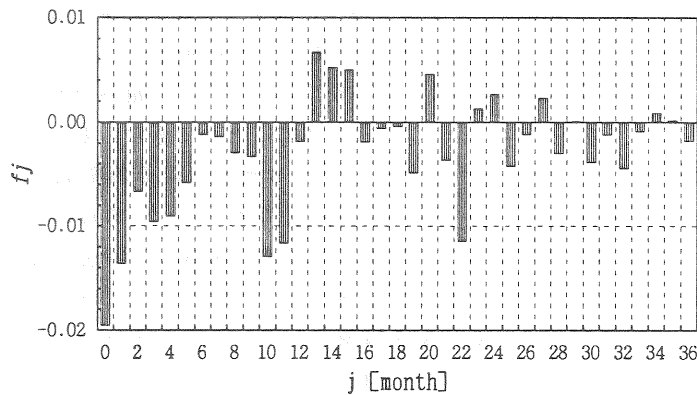


図3 . 回帰係数 f_j (上限 j_e を36にした場合)

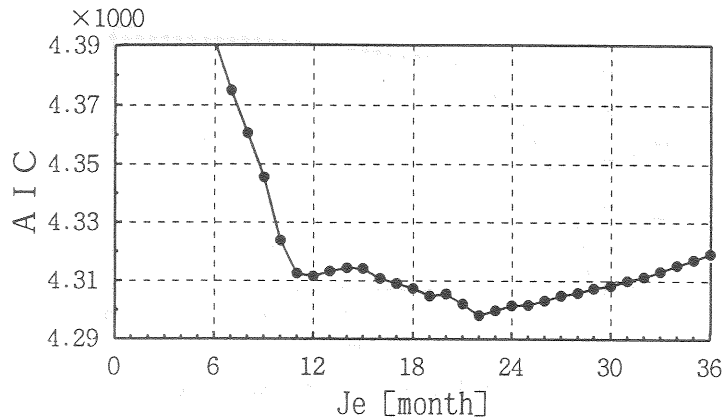


図4. 上限 j_e を変化させたときのAICの値

下の領域でも、係数 f_j は大小さまざまであり、すべてに意味があるとは判断しかねる状態である。

この様に、単に重回帰分析を行っただけでは、どの係数が本当に意味のあるものなのかを判断することができない。そこで、上で述べた上限 j_e を変化させた時の重回帰分析のそれぞれの結果に対し、求めたAICを図4に示す。縦軸はAIC、横軸は上限 j_e である。この図から分かるように、AICは $j_e=12$ の近辺まで急激に減少し、その後若干の増加の後、再び減少し、 $j_e=22$ で最小になる。その後は上限 j_e が大きくなると共にAICも増加していく。この結果、少なくとも Σ の上限 j_e の値は22とすれば十分であることがわかる。更に、 $j_e=12$ 以降でのAICのゆっくりした増加は、そこでの係数 f_j は不要である可能性の高いことをうかがわせる。そこで、次に、不要と思われる係数 f_j を除いたあらゆる組み合わせについて重回帰分析を行い、それぞれのAICの値を求めた。

この分析は、要するに“虱潰し”の手法であり、大変な手数が必要となる。しかし、最近のパーソナルコンピュータの発達により、応用ソフトを利用して（多少CPU時間はかかるものの）簡単に求めることができる。この例では、すべての組み合わせをチェックした結果、最終的にいくつかの f_j のグループを見い出すことができた。それらのグループをIグループ、Mグループ、Oグループと名付けるとそれらはそれぞれ、Iグループが $j=0,1,3,4$ 、Mグループが $j=10,11$ 、Oグループが $j=22$ であった。これらのグループ以外のどの f_j を入れてもACIは大きくなった。参考までに、“虱潰し”手法の最終段階におけるAICの減少の様子を図5に示す。図5には各グループ間のすべてのコンビネーションを縦軸にとり、それらに対するAICの値を横軸として、図の右部分に示してある。この図で、各列の■印の付いたグループが組み合わせに採用されたグループを表している。す

赤池情報量基準(AIC)を用いた時系列データの重回帰分析

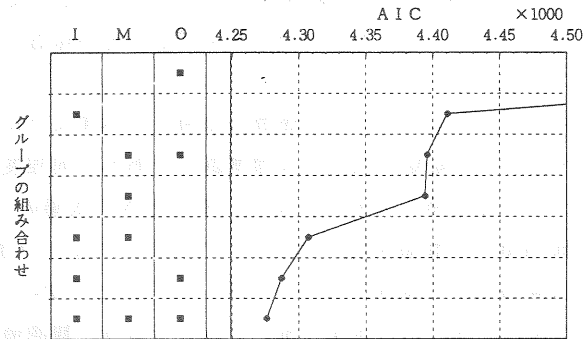


図 5 . 三つのグループのあらゆる組み合わせと AIC

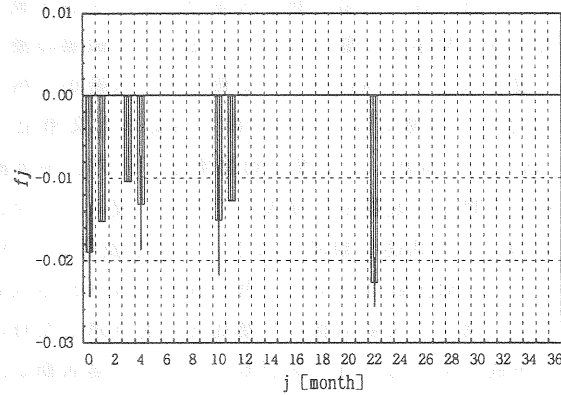


図 6 . 最適化された回帰係数 f_j

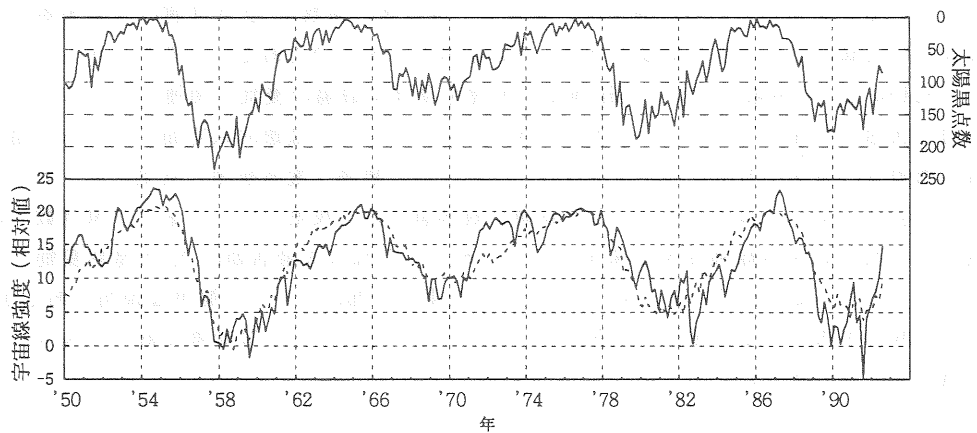


図 7 . 太陽のウォルフ黒点数（上）と宇宙線強度（下、実線：観測値、点線：計算値）

すべてのグループを採用したときにAICが最小となることが分かる。従って、それ以外のどの係数 f_i を導入しても、あるいは上の三つのグループからどの係数 f_i を削除しても必ずAICは増加することになる。

結局、この例の時系列データに対しては、係数 f_i は図6に示すように決定された。この図の一部の係数 f_i に付いている縦線は代表的な標準誤差であり、同程度の誤差が他の係数 f_i にも付いているがここでは省略してある。また、そのときの太陽のウォルフ黒点数 R_n 及び宇宙線強度 I_n の観測値と計算値を図7に示す。ここで、ウォルフ黒点数 R_n は図の上部に示してあり、宇宙線強度との関係が分かりやすいように上下逆にプロットしてある。また、宇宙線強度 I_n は図の下部に相対強度(%)で示してあり、観測値は実線、最適化モデルによる計算値は点線で示してある。

最後に、ここでの解析事例に関する物理的な意味について簡単に述べておく。まず、データの一つであるウォルフ黒点数は太陽活動の大体の大きさを表す値で、その実体は主に太陽の磁場の擾乱の大きさに関連した量になっている。この磁場の擾乱は太陽風と呼ばれる高速プラズマ(速さはおよそ400 km/s)に乗って、太陽から四方八方に広がっている。従って、ある時に発生した擾乱は1カ月で約7 a.u. (天文単位: 平均的な太陽地球間距離で約 1.5×10^8 km)だけ移動した位置に到達する。一方、宇宙線の方は、その磁場の擾乱に逆らって銀河系空間より太陽系に侵入してくる。侵入した宇宙線は磁場の擾乱の影響を受けつつ地球近傍にまで到達し観測される。これを表現したのが式(11)である。従って、重回帰モデルでの回帰係数はウォルフ黒点数で代表される太陽活動が宇宙線強度にどのように影響するか、過去に太陽を離れた擾乱までを考慮した時の影響の大きさを示していることになる。係数が全体的に負の側が多いのは、太陽活動が激しい時には宇宙線は侵入しづらくその強度は一般に小さくなるからである。

また、最適化された係数は何カ月前に太陽を離れた擾乱が宇宙線強度に大きな影響を及ぼしているかを示している。例えば、図6で一番右側の係数は22カ月前に太陽を離れた擾乱が宇宙線強度に大きな影響を及ぼしている事を示すものである。このことから、太陽から約130 a.u.の距離の所に宇宙線強度に影響を及ぼす磁場の擾乱に関連した何らかの物理的な状態が存在していることが予想できる。即ち、そこには銀河系空間に対する太陽系磁気圏領域の境界があることが予想され、これは他の観測(電波観測)結果とも一致するものである。また、図7に見られる計算値と観測値の差(残差)は、ウォルフ黒点数では表現しきれない宇宙線強度を意味するものであり、これは太陽活動の22年周期変動と関係のあることが知られている[9]。以上は、結果の一部についての簡単な説明であるが、本稿は解析手法についての報告であるので、更に詳しい点については参考文献[6, 8, 9]を参照していただきたい。

4 まとめ

本稿では、二つの時系列データを用いた重回帰分析の手法について、赤池の情報量基準(AIC)に基づいた未知パラメータ数の最適化について論じた。

一般に、時系列データ以外のデータについての重回帰モデルでは、ある目的変数となるデータに対し、様々な種類の説明変数となるデータが導入され、その線形結合でモデルが表されている。それらのデータ間の関係を調べる為に最小二乗法によって未知パラメータ(回帰係数)の値を決定し、しかるべき誤差も付けることができる。それによって、各説明変数と目的変数の個々の関係もそれぞれ解釈できることにもなる。

しかし、重回帰分析の対象となるデータが、ここで取り上げたような時系列データの場合、説明変数は互いに独立ではなく、未知パラメータ相互の関係も独立という保証はない。従って、未知パラメータを決定する時には、できるだけ長期のデータを用いることと未知パラメータの数を最適化することの二つの要件が必要となる。長期のデータが必要となることは、回帰係数の決定が最小二乗法によって行われるからである。即ち、未知パラメータを十分な精度で決定するには、説明変数の独立性のなさを補う必要がある。その為には、データの数を十分に多くしなければならないということである。一方、未知パラメータの最適化については、最小二乗法だけでは解決することができない為、AICを導入する必要がある。これらの二つの要件を満たすことにより、二つの時系列データの間をより細かく調べることができ、両者に内在する因果関係をより明確に知ることができる。

本稿では、その例として、十分長期にわたって観測されている二つの時系列データについてその解析を試みた。その結果は太陽圏物理学の面から見ても十分に満足のいくものであり、採用した解析手法の有効性を如実に示すものである。

最後に、解析に対する計算機環境について述べる。本稿のような重回帰分析はパーソナルコンピュータの発達により、応用ソフトを利用して簡単に実行できる環境となってきた。本稿の解析でもすべて応用ソフトで実行できた。そのようなソフトは設定できる説明変数の数も50以上が可能となっている。ただし、“氾濫し”の方法はかなり手数がかかり、また回帰分析の計算時間そのものも説明変数が多くなるとかなり長めになる。

“氾濫し”の手法に関しては、本稿の解析の場合、毎回キーボードより設定し直して、回帰分析を実行した。しかし、未知パラメータの最適化を含めた重回帰分析を応用ソフトを用いて一度に実行するマクロの作成を現在検討しているところである。

また、計算時間に関しては、例えば、cpuが80386(16MHz)の場合、データ数が514、未知パラメータ数が75の解析に約30分の計算時間が必要であった。必要となる計算時間は未知パラメータ数の二乗で増えていく為、未知パラメータの数が多い場合、cpuが80486あるいはpentium以上のパーソナルコンピュータ、さらにはワークステーションで計算を実行する必要性が生じる可能性もある。この点について、応用ソフトの場合はアルゴリズム

にユーザーが手を加えることができない為、計算時間をより短くする必要があるときには、より演算速度の早い計算機を利用するしか解決方法はない。例えば、汎用大型計算機のパッケージソフトを利用するのも良いかもしれない。ただし、更に計算時間を短くする必要がある場合は、自らアルゴリズムを作成し、工夫することが必要となるであろう。

参考文献

- [1] Harvey, A.C.; Time Series Models, Phillip Allan (1981).
- [2] Blackman, R.B. and Turkey, J.W.; The Measurement of Power Spectra from the Point of View of Communications Engineering, Dover (1958).
- [3] Chatterjee, S. and Price, B.; Regression Analysis by Example, John Wiley & Sons (1977).
- [4] 日野幹雄; 「スペクトル解析」朝倉書店 (1977).
- [5] Bendat, J.S. and Piersol, A.G.; Random Data : Analysis and Measurement Procedures, John Wiley & Sons (1971).
- [6] Nagashima, K. and Morishita, I.; Long Term Modulation of Cosmic Rays and Inferable Electromagnetic State in Solar Modulating Region, Planetary and Space Science, Vol.28, pp. 177-194 (1980).
- [7] 坂元慶行、石黒真木夫、北川原四郎; 「情報量統計学」共立出版 (1989).
- [8] Morishita, I. and Nagashima, K.; The Extent of the Heliosphere Responsible for the Solar Cycle Modulation of the Galactic Cosmic-Rays, Inferred from their Ground Level Observations during 1950-1992, Proc. of 23rd Int. Conf. on Cosmic Rays (Calgary), Vol.3, pp.571-574 (1993).
- [9] Nagashima, K. and Morishita, I.; Twenty-two Year Modulation of Cosmic Rays Associated with Polarity Reversal of Polar Magnetic Field of the Sun, Planetary and Space Science, Vol.28, pp.195-205 (1980).

(経営学部情報管理学科助教授)