

社会システム分析のための統合化プログラム 1 1

－ 時系列分析 －

福井正康, 王迎春*, 王晶*, 石丸敬二**

福山平成大学経営学部経営学科

*福山平成大学大学院経営学研究科経営情報学専攻

**福山大学経済学部経済学科

概要

我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は新たに時系列分析に関するプログラムを作成した。時系列分析は時間の経過と共に変化する変数の変動の予測を目的とする分析である。我々は変動の分解モデルと呼ばれる手法を用いて、データを傾向変動 1（近似モデル）＋傾向変動 2（予測モデル）＋季節変動（振幅変動×周期変動）＋残差に分解するモデルを考えた。

キーワード

College Analysis, 社会システム分析, 統計, 時系列分析

URL: <http://www.heisei-u.ac.jp/ba/fukui/>

1. はじめに

我々はこれまで教育用社会システム分析ソフトウェアの一環として様々なプログラムを作成してきたが、この論文では時系列分析について紹介する。時系列分析は時間の流れとともに変化する変数の過去のデータから、未来の値を予測する手法である。例えば企業の売上予測、在庫の受注予測、株価の変動など時系列的に変化するデータがこの分析の対象である。

分析方法には大きく分けて、古くから考えられてきた予測モデルという方法とデータの変動をいくつかの典型的な変動に分解する変動の分解モデルという方法がある。予測モデルには、予測値にこれまでの変動の差分を使う差の平均法、過去のデータにウェイトを付けて使う指数平滑法やブラウン法、過去の最も似た状況を探す最近隣法、重回帰分析を活用する ARIMA などがあるが、これらはデータ数が少なく周期性を見抜くことが困難なデータに適用されることが多い。

一方変動が周期性を持っているようなデータに対しては変動の分解モデルが適用される。これは変動を「傾向変動」、「季節変動」、「循環変動」、「残差」などに分け、それぞれの特徴をとらえて予測値を求めるもので、長期的な予測もある程度可能な手法である。傾向変動はデータの平均的な変動を表し、予測には移動平均や回帰を基礎とした近似モデルが利用される。一般に季節変動は周期が一定の変動で、循環変動は周期が変化する変動を表す。

本来予測モデルと変動の分解モデルは別々に開発されたものであるが、後者の傾向変動に例えば ARIMA の結果を利用するなどということも可能であるため、我々のプログラムでは2つの手法を組み合わせて使うことができるようにしている。本来変動の分解モデルの傾向変動については、移動平均や線形近似、対数近似などの近似手法が利用されることが多いので、傾向変動を2つに分けて、「傾向変動1」としてこれらの近似手法を、「傾向変動2」として先に述べた予測モデルを用いることにする。もちろんどちらか1つを選んでよい。これらの分解の後、必要があればデータの周期的な変動の分解を行う。

周期的な変動には季節変動と循環変動があるが、循環変動についてはまだプログラムに組み込んでいない。また、季節変動を「振幅変動」と振幅が一定の「周期変動」の積に分解し、これらをまとめて以下のモデルとする。

$$\text{データ変動} = \text{傾向変動1} + \text{傾向変動2} + \text{振幅振動} \times \text{周期変動} + \text{残差}$$

プログラムでは振幅変動の平均が1に近くなるように設定し、周期変動の意味を理解し易くしている。

2. 時系列分析の方法

時間を過去から未来へ等間隔で区切ったとき、ある時点 t ($t = 1, \dots, N$) でのある変数 X の値を x_t とする。時系列分析はこの変数の変化を分析し、モデルを作成して今後の予測を行うことを目的とする。以後このデータ書式を用いて予測モデルと変動の分解モデルの理論について説明

する。

2.1 データの前処理

時系列分析では、データをそのままの形で使うより、何らかの変換を加えてから分析を進める方がよりはっきりとした結果を得られることがある。ここではよく利用されるデータの変換について述べる。

変数が値の増大とともに変動の大きさも大きくなっていくような場合は、元の変数の対数をとって新しい変数とすると分析が容易になる場合がある。また、比率や確率のように $[0,1]$ 区間の値の場合は、以下のロジット変換によって値域が $(-\infty, \infty)$ の時系列に変換できる。

$$\text{対数変換} \quad z_t = \log_e x_t$$

$$\text{ロジット変換} \quad z_t = \log_e \left(\frac{x_t}{1-x_t} \right)$$

また、時系列データの差分を使って新しい変数を作り出すことも行われる。

$$\text{差分 (} i \text{ 期)} \quad z_t = x_t - x_{t-i}$$

$$\text{差分比 (} i \text{ 期)} \quad z_t = x_t / x_{t-i}$$

2.2 予測モデル

時系列データの周期性が明らかでない場合やデータの数が周期性を見るのに十分でない場合、予測モデルと呼ばれる方法を用いて時系列データの予測が行われる。これからは図 2.2.1 のデータを用いて各種の予測モデルを紹介する¹⁾。これらのモデルは基本的に t 時点までのデータを元に $t+1$ 時点での予測値を求めるもので、長期の予測には向かない。

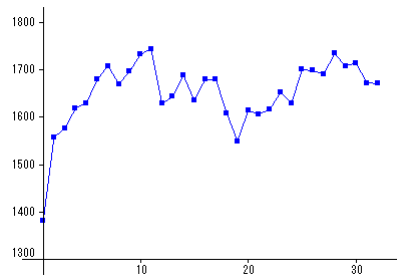


図 2.2.1 時系列データ

2.2.1 差の平均法 (差分法)

これは $t+1$ 時点の予測値 y_{t+1} を t 時点のデータ x_t とこれまでの 2 時点間の差分の平均で与えるものである。

$$y_{t+1} = x_t + A_t$$

ここに

$$A_t = \frac{(x_2 - x_1) + (x_3 - x_2) + \cdots + (x_t - x_{t-1})}{t-1} = \frac{x_t - x_1}{t-1}$$

差の平均法を用いた予測を図 2.2.2 に示す。これを見るとデータが上下している場合、残差の平均は相殺され、予測値は 1 期前の値と余り変わらない様子が見える。この手法はデータに上昇傾向や下降傾向が見られる場合に適用できる。

2 期以上の予測には実測値の代わりに予測値を使うことにすれば、予測は直線的に伸びて行く。

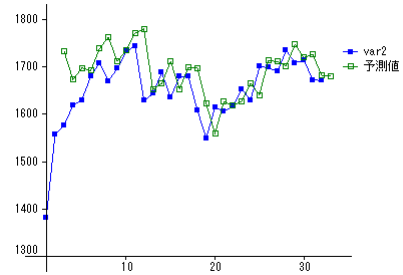


図 2.2.2 差の平均法を用いた予測

2.2.2 指数平滑法

この方法は $t+1$ 期の予測値 y_{t+1} を t 期の実測値 x_t と予測値 y_t を使って以下のように与えるものである。

$$y_{t+1} = \alpha x_t + (1 - \alpha)y_t \quad \text{但し、} y_1 = x_1 \text{ (または } y_2 = x_1 \text{) とする。}$$

ここに α は $0 < \alpha < 1$ のパラメータである。またこの式は以下のように書き換えると、指数平滑の意味が分かり易い。

$$\begin{aligned} y_{t+1} &= \alpha x_t + (1 - \alpha)[\alpha x_{t-1} + (1 - \alpha)y_{t-1}] \\ &\dots \\ &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots + \alpha(1 - \alpha)^{t-2} x_2 + (1 - \alpha)^{t-1} x_1 \end{aligned}$$

これを見ると α の値が小さいほど過去からの影響を受けやすくなっていることが分かる。これは今期以前の指数平滑値を次期の予測値とするものである。この方法を用いて時系列データの変動を $\alpha=0.74$ として予測した結果を図 2.2.3 に示す。パラメータの値は図 2.2.4 のようにパラメータの値を変えて残差の平均を調べ、最小値をとることによって求めた。

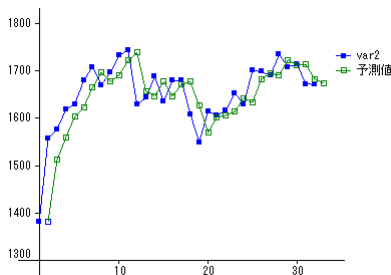


図 2.2.3 指数平滑法による予測 ($\alpha=0.74$)

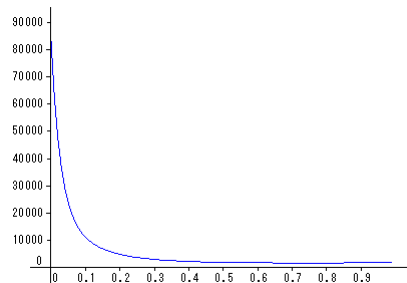


図 2.2.4 パラメータの推定

差の平均法と同様この場合も変動は平滑され、予測値は 1 期前の実測値に近い値になっている。また 2 期以上先の予測値は、実測データを予測データで置き換えると同じ値が続く。この予測値を見ると 1 期前の実測値にかなり引きずられていることが分かる。指数平滑法も上がり下がりのあるデータには向かない。

2.2.3 ブラウン法（ブラウンの2重指数平滑法）

指数平滑法は単純に今期までの指数平滑値を予測値としたものであって、予測値の精度については考慮されていない。この精度を考慮した方法がブラウン法（2重指数平滑法）である。

ここで比較のために指数平滑法の公式を少し書き換えておく。

$$y_{t+1} = u_t$$

$$u_t = \alpha x_t + (1 - \alpha)u_{t-1} \quad t \text{ 時点の } x \text{ の指数平滑値（} t+1 \text{ 時点の } x \text{ の予想値）}$$

ブラウン法は、指数平滑法で予測される $t+1$ 期の予測値 u_t に、この予測値と指数平滑法による u_t の予測値 v_{t-1} との差（の m' 倍）を足して来期を予測するものである。指数平滑を2度行うので2重指数平滑法と呼ばれる。

$$y_{t+1} = u_t + m'(u_t - v_{t-1})$$

$$u_t = \alpha x_t + (1 - \alpha)u_{t-1} \quad t \text{ 時点の } u \text{ の値（} t+1 \text{ 時点の } x \text{ の予想値）}$$

$$v_{t-1} = \beta u_{t-1} + (1 - \beta)v_{t-2} \quad t-1 \text{ 時点の } v \text{ の値（} t \text{ 時点の } u \text{ の予想値）}$$

ここに m, α, β はパラメータである。

この式を分かり易く表現すると以下となる。

$$x \text{ の補正予測値} = t+1 \text{ 時点の } x \text{ の予測値} + m'(t \text{ 時点の } u \text{ の値} - t \text{ 時点の } u \text{ の予測値})$$

$$= t+1 \text{ 時点の } x \text{ の予測値} + t+1 \text{ 時点の予測補正項}$$

実際の計算では、参考文献1に従い、 $m' = 1$,

$\alpha = \beta$ としており、

$$y_{t+1} = a_t + b_t$$

$$a_t = 2u_t - v_t$$

$$b_t = \frac{\alpha}{1 - \alpha}(u_t - v_t)$$

以下の初期値をおいている。

$$u_1 = v_1 = x_1$$

$$a_1 = x_1, \quad b_1 = [(x_2 - x_1) + (x_4 - x_3)]/2$$

このため予測値は、 $t = 5$ から求める。

ブラウン法による最適なパラメータでの予測を図2.2.5に示す。ここでも明らかなように増加・減少のあるデータに対してブラウン法はあまり有効とは言えない。

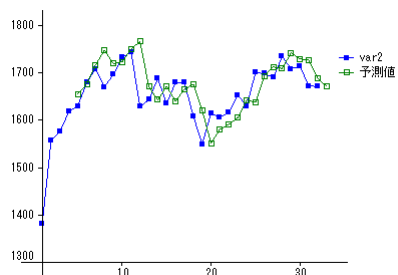


図 2.2.5 ブラウン法による予測 ($\alpha = 0.42$)

2.2.4 最近隣法

最近隣法は現在とその1期前のデータに似た過去のデータを探して、次期のデータの予測値を決めるものである。

最近隣法は以下の形で予測を行う。現在とその1期前のデータを x_t, x_{t-1} とし、過去のデータ

x_{t-m}, x_{t-m-1} との距離 d_m を以下のように考える。

$$d_m = \sqrt{(x_t - x_{t-m})^2 + (x_{t-1} - x_{t-m-1})^2}$$

距離の最小値 d_{\min} を求め、距離がその 1.62 倍未満のデータを集める。

$$S = \{d_m \mid d_m < 1.62 \times d_{\min}\}$$

この 1.62 は黄金分割比と呼ばれ、実用上多く使われる¹⁾。その集めた距離の逆数を利用して重み w_m ($d_m \in S$) 計算する。但し、距離が 0 の場合はある小さな値 (このソフトの場合は 0.0001) としている。

$$w_m = \frac{1/d_m}{\sum_{d_k \in S} 1/d_k}$$

この重みを使って予測値 y_{t+1} を以下のように求める。

$$y_{t+1} = \sum_{d_m \in S} w_m x_{m+1}$$

実際に最近隣法を用いた予測は図 2.2.6 のようになる。

この方法はデータの上がり下がりの変動が大きいほど有効で、上昇下降傾向があるデータには向かない。また過去の似た状況を探すことから、一般に過去のデータが多いほど予測の精度は上がる。

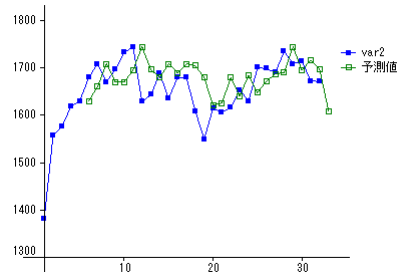


図 2.2.6 最近隣法による予測

2.2.5 自己相関モデル (ARIMA)

このモデルには 3 つのパラメータ p, d, q があり、これらのパラメータを用いて、 $ARIMA(p, d, q)$ と表される。以後各パラメータについて説明し、最後に全体を見渡す。

最初にパラメータ d について述べる。これはデータの差分の回数である。差分は傾向変動などを取り除く 1 つの手段である。 $x_t^{(1)}$ を 1 回の差分、 $x_t^{(2)}$ を 2 回の差分とするとそれぞれ元のデータを用いて以下のように表される。

$$x_t^{(1)} = x_t - x_{t-1}$$

$$x_t^{(2)} = x_t^{(1)} - x_{t-1}^{(1)} = x_t - 2x_{t-1} + x_{t-2}$$

d 回の差分データに対して $ARMA(p, q)$ モデルを適用する手法が $ARIMA(p, d, q)$ モデルである。但し、 d 回の差分データでは利用できるデータが、 $d+1$ 期から t 期までとなる。

MA モデル

次にパラメータ q について考える。このパラメータは $MA(q)$ と呼ばれるモデルのパラメータである。このモデルは $t \geq t_0$ に対して以下の仮定が基礎になっている。

$$x_t = b_1 u_{t-1} + b_2 u_{t-2} + \dots + b_q u_{t-q} + b_0 + u_t$$

ここに $u_t, u_{t-1}, \dots, u_{t-q}$ は各時点のホワイトノイズである。特に $b_0 = 0$ の場合が教科書などに載っ

ている。

1 期先の予測値 y_{t+1} を実測値 x_{t+1} からホワイトノイズ u_{t+1} を引いたものと定義すると以下のような関係が得られる。

$$\begin{aligned} y_{t+1} &= x_{t+1} - u_{t+1} \\ &= b_1 u_t + b_2 u_{t-1} + \dots + b_q u_{t-q+1} + b_0 - u_{t+1} + u_{t+1} \\ &= b_1(x_t - y_t) + b_2(x_{t-1} - y_{t-1}) + \dots + b_q(x_{t-q+1} - y_{t-q+1}) + b_0 \end{aligned}$$

計算手順はまず $t < t_0$ の間のノイズ $x_t - y_t$ の初期値を決める。我々はこれを $N(0,1)$ の正規乱数としている。次にこれらの初期値を用いて $t = t_0$ の場合に上式から重回帰分析を用いて予測値 y_{t_0+1} を求める。但し、計算が可能ためには初項の時期をずらしたデータの組が q 個必要であり、少なくとも $t_0 > 2q$ でなければならない。我々はこれを $t_0 = 2q + 2$ にしている。ここで得た予測値 y_{t_0+1} を使って、上式を用いて再度重回帰分析を行うことによって新しい予測値 y_{t_0+2} を得る。これを繰り返して行くことで、最終的な予測値 y_{t+1} を得る。

この処理では長期予測は不可能である。長期予測のためには実測値の代わりに予測値を用いるしかないが、そうすると説明変数が 0 になって行き、前の予測値が続くようになる。

MA(1) と MA(2) による予測グラフを図 2.2.7a と図 2.2.7b に示す。

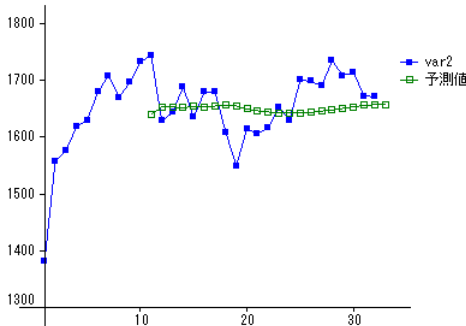


図 2.2.7a MA(1) モデルによる予測

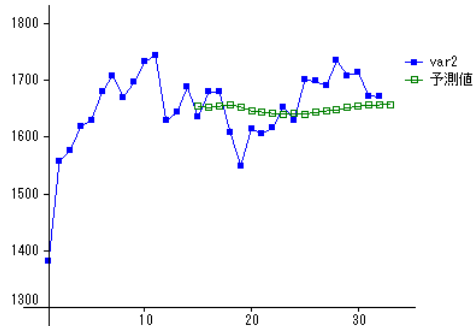


図 2.2.7b MA(2) モデルによる予測

AR モデル

パラメータ p は AR(p) と呼ばれるモデルのパラメータである。このモデルは以下の仮定が基礎になっている。

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} + a_0 + u_t$$

ここに u_t は t 時点のホワイトノイズである。特に $a_0 = 0$ の場合が教科書などによく載っている。

1 期先の予測値 y_{t+1} を実測値 x_{t+1} からホワイトノイズ u_{t+1} を引いたものと定義すると $t \geq t_0$ に対して以下のような関係が得られる。

$$y_{t+1} = a_1 x_t + a_2 x_{t-1} + \dots + a_p x_{t-p+1} + a_0$$

計算は重回帰分析を用いるが、手順は過去の予測値を使う必要がないので MA モデルと比べると

単純である。但し、計算が可能ためには初項の時期をずらしたデータの組が p 個必要であり、少なくとも $t_0 > 2p$ でなければならない。我々はこれを $t_0 = 2p + 2$ にしている。

この処理でも長期予測は不可能である。長期予測のためには実測値の代わりに予測値を用いるしかないが、 a_i が殆ど変わらない状況では例えば $p = 1$, $|a_1| < 1$ の場合、

$$y_n = a_1 y_{n-1} + a_0, \quad \lim_{n \rightarrow \infty} y_n = a_0 / (1 - a_1)$$

となり、前の予測値に近い値が続くようになる。AR(1)と AR(2) による予測グラフをそれぞれ図 2.2.8a と図 2.2.8b に示す。

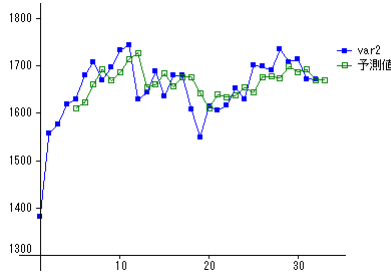


図 2.2.8a AR(1) モデルによる予測

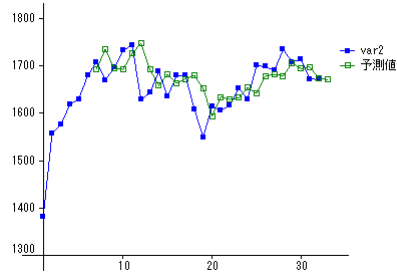


図 2.2.8b AR(2) モデルによる予測

ARIMA モデル

ここではこれまで学んできたモデルを複合した場合を考える。今 d 回の差分データを $x_t^{(d)}$ とすると、ARIMA(p, d, q) モデルは $t \geq t_0$ で以下のように表される。

$$x_t^{(d)} = \sum_{i=1}^p a_i x_{t-i}^{(d)} + \sum_{i=1}^q b_i u_{t-i}^{(d)} + c + u_t$$

これを用いて予測値 $y_{t+1}^{(d)}$ は以下ようになる。

$$y_{t+1}^{(d)} = \sum_{i=1}^p a_i x_{t-i+1}^{(d)} + \sum_{i=1}^q b_i (x_{t-i+1}^{(d)} - y_{t-i+1}^{(d)}) + c$$

計算手順は、まず $t < t_0$ 以前のノイズ $x_t^{(d)} - y_t^{(d)}$ を標準正規乱数で初期化する。後は MA モデルの場合と同様に、 $t = t_0$ の場合の予測値 $y_{t_0+1}^{(d)}$ を重回帰分析で求めて、これを利用してさらに次の予測値を求める方法をとる。但し計算が可能ためには、上式に必要なデータが $r = \max(p, q)$ 個、それを時期をずらして $p + q$ 期分必要とすることから、少なくとも $t_0 > r + p + q + d$ でなければならない。我々は少し大きくとって、以下としている。

$$t_0 = r + p + q + d + 2$$

計算が可能であることで上のような条件を付けたが、計算の正確さを考えると十分でない。MA モデルでは計算の初期値を乱数で与えているので、 t_0 の近くの推定値は良い近似ではない。我々は値が安定するまで待つ必要がある。そのため、誤差の計算や表示に利用するのは実際には経験

的に以下にしている。

$$t_0 = 2p + d + 2 \quad q = 0 \text{ の場合}$$

$$t_0 = (r + p + q + d + 2) + (2q + 5) \quad q > 0 \text{ の場合}$$

これで $t_0 + 1$ 期からの予測値 $y_{t+1}^{(d)}$ が求められたが、これは差分を d 回取ったデータの予測値である。我々はこれを元のデータに戻す必要がある。データ間に

$$x_{t+1}^{(d-1)} = x_t^{(d-1)} + x_{t+1}^{(d)}$$

の関係があることから、これを以下のように拡張する。

$$y_{t+1}^{(d-1)} = x_t^{(d-1)} + y_{t+1}^{(d)}$$

即ち、以下のように求められる。

$$y_{t+1} = y_{t+1}^{(0)} = x_t^{(0)} + y_{t+1}^{(1)} = x_t^{(0)} + x_t^{(1)} + y_{t+1}^{(1)} = \dots = \sum_{i=0}^d x_t^{(i)} + y_{t+1}^{(d)}$$

ARIMA(1,0,1), ARIMA(1,1,1) による予測グラフを図 2.2.9a と図 2.2.9b に示す。

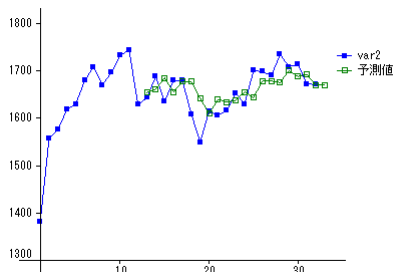


図 2.2.9a ARIMA(1,0,1) モデルによる予測

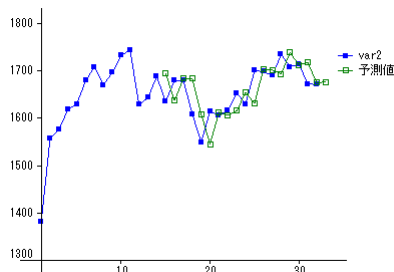


図 2.2.9b ARIMA(1,1,1) モデルによる予測

差分を入れると 1 期前の実測値に差分の予測値を足すことになり、やはり 1 期前の状態に引きずられるようである。

2.3 変動の分解モデル

具体的なイメージを持ってもらうために、今後しばらく図 2.3.1 のデータを元にして話を進める^{2),4)}。

データは様々な要因で変動するが、我々は大きくこれを、傾向変動 T 、季節変動 S 、循環変動 C 、残差変動 R に分ける。ここに傾向変動は長期にわたる継続的な変化で、季節変動は周期が

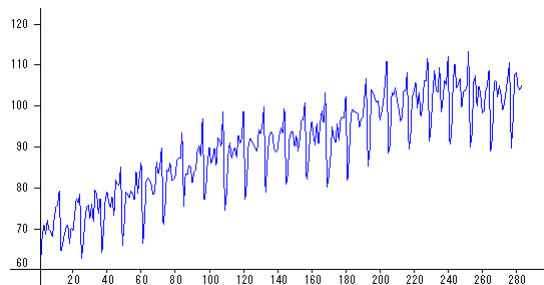


図 2.3.1 時系列データ decomp_food

一定の変化、循環変動は周期が一定でないものの周期性が認められる変化、残差変動は観測誤差などのゆらぎである。一般に変数 X はこれらの変動の関数として以下のように表される。

$$X = f(T, S, C, R)$$

この一般の関係の中で、実際の分析のためには様々な仮定を置くことが多い。我々のプログラムでは周期が変化する循環変動については考えず、それぞれの変動の合計で表される以下の加法モデルを採用している

$$X = T + S + R$$

但し、傾向変動には通常、移動平均や回帰近似が利用されるが（これを近似モデルと呼ぶ）、我々は傾向変動を2つに分け、近似モデル T_1 と 2.1 節で述べた予測モデル T_2 の和と考える。これによって予測モデルだけの処理も変動の分解モデルと合わせた処理も可能になる。また季節変動について、振幅の変化も考え、季節変動を振幅変動 A と振幅一定の季節変動 S' （以後これを周期変動と呼ぶ）の積に分解する。ここで振幅変動には回帰近似を用い、周期変動の意味を分りやすくするため、大きさの平均を 1 に近くなるようにとる。これらを合わせて、我々のプログラムでは以下のようなモデルを扱う。

$$X = T_1 + T_2 + A \times S' + R$$

以後 2.1 節で述べた予測モデル T_2 を除いて、それぞれの変動の分解について詳細に説明する。

2.3.1 傾向変動の分解

傾向変動の抽出は主に移動平均法による方法と最小 2 乗法の手法を応用した方法（回帰分析はこれに含まれる）がある。 n 期の移動平均法では時点 t のデータの値を以下のようにして、データの平滑化を図る。

$$d_t = \frac{1}{2m+1} \sum_{i=-m}^m x_{t+i} \quad n = 2m+1 \text{ の場合}$$

$$d_t = \frac{1}{2m+2} \left\{ \sum_{i=-m}^m x_{t+i} + \frac{1}{2} (x_{t-m-1} + x_{t+m+1}) \right\} \quad n = 2m+2 \text{ の場合}$$

これは中心法と呼ばれる方法であるが、移動平均を予測に用いる場合には、以下のような方法が使われる。我々はこの方法を用いる。

$$d_t = \frac{1}{n} \sum_{i=-n}^{-1} x_{t+i}$$

また、時間のずれに対して重み係数を掛ける場合もある。データに周期性がある場合、この方法では傾向変動に周期成分が残るが、移動平均を行ったデータに再度移動平均を行うとさらになめらかな傾向が得られる。但し、移動平均では時系列データの前後、または前が使えなくなるので、ある程度データ数も必要である。我々のプログラムでは複数回の移動平均は考えていない。

予め大雑把なデータの変化を近似的につかんでおくことは重要である。最小2乗法の手法を応用した近似手法の中で線形回帰分析を利用するものは計算が容易である。よく使われる線形回帰の方法には以下のようなものがある。

1次近似 $d_t = at + b$

対数近似 $d_t = a \log t + b$

べき乗近似 $d_t = bt^a$

指数近似 $d_t = be^{at}$

多項式近似 $d_t = a_p t^p + a_{p-1} t^{p-1} + \dots + a_1 t + a_0$

ここにべき乗近似と指数近似については両辺の対数をとって線形回帰分析を行う。また、多項式近似は重回帰分析を用いてパラメータの推定を行う。例として2次式による近似結果を図2.3.2に示す。このデータについては以下の2次曲線が最良である。

$$y = -0.00029t^2 + 0.211t + 68.002$$

これら以外の近似には非線形最小2乗法など他の方法を利用する。

この傾向変動の結果を元データから分離するには、我々のモデルでは引き算を用いる。

$$y_t = x_t - d_t$$

この2次曲線を傾向変動として取り除くと図2.3.3の結果となる。この段階での実測値と予測値の相関係数の2乗(決定係数) R^2 は0.7904である。

予測モデルの分解については、2.2節で述べたので省略する。

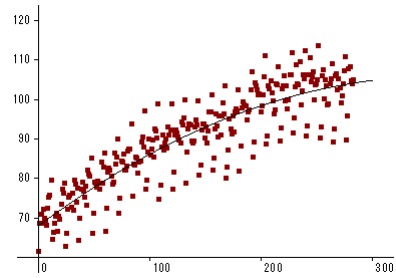


図 2.3.2 2次曲線の当てはめ

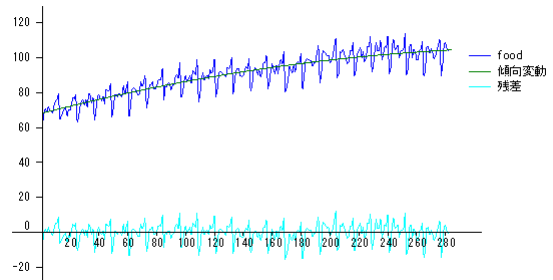


図 2.3.3 傾向変動の分離

2.3.2 振幅変動の分解

振幅変動の推定は以下の振幅変動データに対して近似曲線を考えることによって与えることにする。

振幅変動データ = 傾向変動の残差の絶対値

÷ 傾向変動の残差の絶対値の平均値

これによって振幅変動の値はほぼ1に近い値となり、周期変動を平均的な振幅を持つ季節変動と意味付けることができるようになる。図2.3.4に近似直線を求

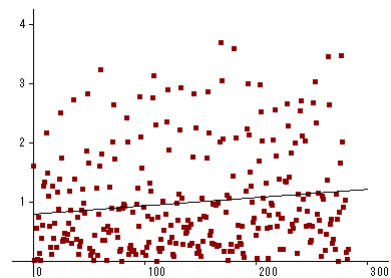


図 2.3.4 振幅変動の推定

める図を示す。振幅変動を分離した残差は傾向変動残差÷振幅変動推定値で与えられる。

2.3.3 周期変動の分解

周期変動のスペクトル抽出は傾向変動と振幅変動を除去したデータ y_t にどのような周波数成分が含まれるかを知る重要な処理である。最初に時間的なラグの影響を見るために自己相関係数を求め、ラグの値によってそれをプロットするコレログラムを作成する。

自己相関係数 r_k ($k=1,2,\dots,L < N-1$) は以下の式により求められる。

$$r_k = \frac{s_k^2}{s_0^2}, \quad \text{ここに } s_k^2 = \frac{1}{N-k} \sum_{t=k+1}^N (x_t - \bar{x}_{k+1}^N)(x_{t-k} - \bar{x}_1^{N-k}), \quad \bar{x}_a^b = \sum_{t=a}^b x_t$$

図 2.3.5 に最大周期を 70 にしたコレログラムを示す。これによると変動の周期は 12 であることが分かる。

次にこのコレログラムに対してその周波数成分を見ると周期性がより明確になる。このような問題には関数のフーリエ (Fourier) 展開という手法が用いられるが、ここでは参考のために期間 $2L$ を周期に持つ関数 $f(x)$ のフーリエ展開の公式を与えておく。

$$f(x) = \frac{a_0}{2L} + \frac{1}{L} \sum_{k=1}^{\infty} (a_k \cos k\pi x/L + b_k \sin k\pi x/L)$$

$$a_k = \int_{-L}^L f(x) \cos k\pi x/L dx, \quad b_k = \int_{-L}^L f(x) \sin k\pi x/L dx$$

この式は関数を周波数 $f_k = k/2L$ ($k=1,2,3,\dots$) の正弦波成分の合計で表したもので、各成分の強さは係数 a_k と b_k で与えられる。

我々の時系列データでは関数が離散的であるため、離散フーリエ変換という手法を利用する。 n を時系列データ x_t の周期として、離散フーリエ展開の公式を以下に与える。

$$x_t = \frac{1}{n} \sum_{k=0}^{n-1} (a_k \cos 2\pi kt/n + b_k \sin 2\pi kt/n) \tag{1}$$

$$a_k = \sum_{t=1}^n x_t \cos 2\pi kt/n, \quad b_k = \sum_{t=1}^n x_t \sin 2\pi kt/n$$

この公式を自己相関係数 r_i に対して適用する。自己相関係数は $r_i = r_{-i}$ であるため、 $-m \leq i < m$ の範囲で偶関数である。その際には周期を $2m$ として、以下の形で与えられる。

$$r_i = \frac{1}{2m} \sum_{k=-m}^{m-1} (a_k \cos 2\pi k/2m + b_k \sin 2\pi kt/2m) = \frac{1}{m} \sum_{k=0}^{m-1} a_k \cos \pi kt/m$$

$$a_k = \sum_{i=-m}^{m-1} r_i \cos 2\pi ki/2m = 2 \sum_{i=0}^{m-1} r_i \cos \pi ki/m$$

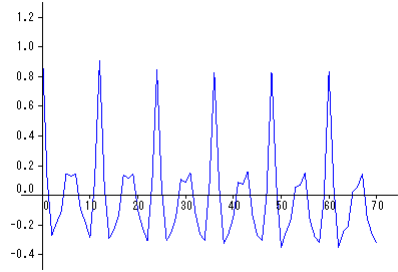


図 2.3.5 コレログラム

この量 a_k を周波数 $f_k = k/2m$ の生スペクトルと呼び、これをラグごとに表したグラフをピリオドグラムという。実用上は生スペクトルより、平滑化という処理を行ったピリオドグラムがよく用いられる²⁾。

実際のデータに対する平滑化したピリオドグラムを図 2.3.6 に示す。これを詳細に見るとまず、周波数 0.167 (周期 6 : これらは別に表示されるデータから読み取れる) に大きなピークがあり、同様に周波数 0.25 (周期 4)、周波数 0.33 (周期 3)、周波数 0.08 (周期 12) などにもピークがある。これらの全体的な周期は、ここに現れた周期の重ね合わせ (最小公倍数、但し時系列の長さの半分より小さいこと) と考えると周期 12 である。

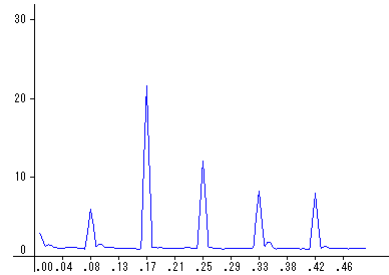


図 2.3.6 ピリオドグラム

この変動の分離には一般の離散フーリエ変換の式 (1) を利用するが、上で考えた周期を n として残差 y_t に適用し、周期変動 u_t を得る。

$$u_t = \frac{1}{n} \sum_{k=0}^{n-1} (a_k \cos 2\pi kt/n + b_k \sin 2\pi kt/n)$$

$$a_k = \sum_{t=1}^n y_t \cos 2\pi kt/n, \quad b_k = \sum_{t=1}^n y_t \sin 2\pi kt/n$$

時系列のデータには周期性があると言っても、各周期間には揺らぎが見られる。しかし上の計算では時系列中どの 1 周期を考えればよいのか分からない。そこで実際の計算には特定の 1 周期を選ぶのではなく、各周期中の同一時点の残差の平均 \bar{y}_t を用いて計算を行った。

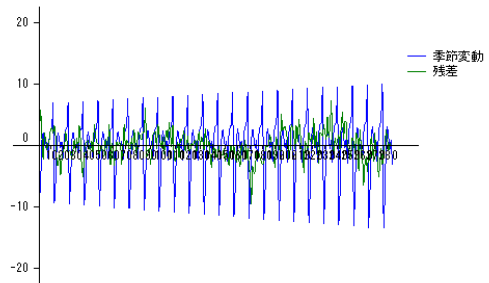


図 2.3.7 季節変動の分解

このようにして季節変動を除去した結果が図 2.3.7 である。ここでは除去した季節変動と残差のみ示してある。この段階での実測値と予測値の R^2 は 0.9647 である。

もう少し詳細に残差の周波数をながめて (タイムラグ 200 まで) 図 2.3.8 でピリオドグラムを描いてみる。

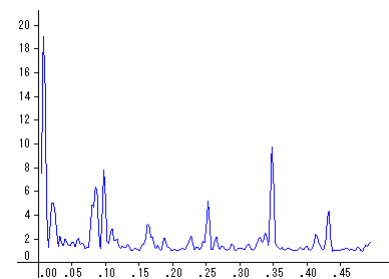


図 2.3.8 残差のピリオドグラム

これを見ると、0 の近くにピークがあり、これは周期 130 近傍のピークであることが分かる。残差の

標準偏差を最小にするように選んでやると、周期は 129 となる。そこでこの周期変動を差し引いて、最終的に図 2.3.9 の分解になる。最終的な実測値と予測値の R^2 は 0.9838 となる。振幅変動を分離しない場合の R^2 は 0.9830 であり、この場合振幅変動の分解の効果はわずかである。

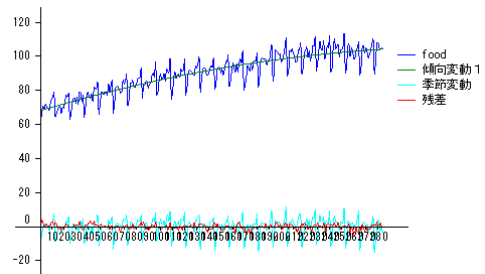


図 2.3.9 時系列データの分解

実はこの残差にはまだ周期性が残っており、これに対して周期性の分離を行い、さらに残差を小さくできる。実際、例えば 91,90,41 と周期性を取り除いていくと実測値と予測値の R^2 は 0.9941 と大きくできる。これを見ると予測精度が上がっているように思われるが、すでに周期成分 129 を入れているのでこのデータの数 283 個から見れば、わずかに 2 周期分を用いて予測を行っていることになる。3 周期目はそれ以前と少しずれることを考えると、いくら残差が小さくできたからといって予測が正しくなる保証はない。ある程度のところで止めておくべきであろう。

さて分解がうまくいき、これ以上分解が難しくなる場合もある。そのとき残差の自己相関係数は 0 に近い値となり、ピリオドグラムは平坦に近くなる。このような波をホワイトノイズと呼ぶ。ホワイトノイズの検定には、Ljung-Box 検定が用いられる。それには、利用するデータ数を t 、ラグ i の母相関係数と標本相関係数をそれぞれ ρ_i 、 r_i として、以下の関係が利用される。

$$\text{帰無仮説: } \rho_1 = \rho_2 = \dots = \rho_m = 0$$

$$Q = t(t+2) \left\{ \frac{r_1^2}{t-1} + \frac{r_2^2}{t-2} + \dots + \frac{r_m^2}{t-m} \right\} \sim \chi_m^2$$

2.3.4 変動の分解モデルによる予測

時系列データの変動の分解は、データにある程度の周期性があること、その数が最低でも 2 周期分以上あることが条件で可能となる。また傾向変動 2 (予測手法) を使うと長期予測は難しい。これまで見てきたデータについて 100 期先までの長期予測を試みよう。見易くするために $t=200$ からのデータを図 2.3.10 に表示する。

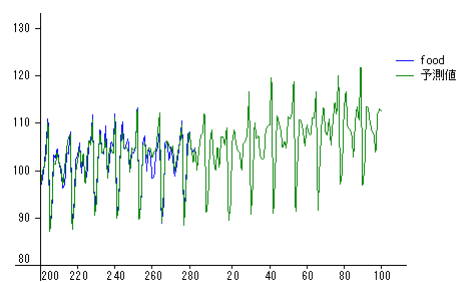


図 2.3.10 時系列データと長期予測

3. プログラムの動作

ここでは具体的に実行画面を見ながらプログラムの動作について説明する。時系列分析のメニュー画面を図 3.1 に示す。それぞれのボタンの出力結果については2章の図で示しているのので、ここではメニューの使い方に焦点を絞って説明する。

最初に変数選択ボタンで分析対象の変数を選択するが、単独で選択しても、時間を指定する変数と同時に選択してもよい。変数を2つ選択する場合、分析対象の変数を目的変数として先に選択する。入力されたデータを見るためには「データ表示」コン

ボボックスの形式を選んだ後、「描画」ボタンをクリックする。データの表示形式には元データ、対数変換、差分、差分期間比がある。このプログラムでは自動的にこれらでデータを変換して分析を実行することはできないが、結果をデータに貼り付けて実行することは可能である。グラフの横軸目盛間隔は右上の「グラフ横軸間隔」テキストボックスで、時間の出力範囲は「出力範囲」テキストボックスで指定できる。グラフのデータポイントの有無はデータ点チェックボックスで選択できる。

変動の分解モデルでの実際の分解は、変動の分解グループボックス内で必要な項目をチェックし、「実行」ボタンをクリックすることで実行できる。特に周期変動の分解では、周期テキストボックスに分解する周期を入力する。周期はカンマ区切りで複数入力できる。「残差の検定」ボタンをクリックすると、変動の分解残差について Ljung-Box 検定が実行される。「係数」や「評価」のコマンドボタンはそれぞれの分解で最適なパラメータを確認するために用いられる。

メニュー左側に並んだコンボボックスでは、傾向変動や振幅変動の分解のメニューが示される。傾向変動1（近似モデル）のコンボボックスには、移動平均、1次近似、対数近似、べき乗近似、指数近似、多項式近似、非線形近似へ、の項目が含まれている。移動平均の期間や多項式近似の次数は、横のテキストボックスで指定する。非線形最小2乗法へを選択すると、すでに設定済みかどうかのメッセージの後、未設定の場合は非線形最小2乗法の分析メニューが表示される。ここで得た結果は傾向変動1の値となる。傾向変動2（予測モデル）のコンボボックスには、差の平均法、指数平滑法、ブラウン法、最近隣法、ARIMAの項目が含まれている。これらの分析のパラメータは横や下にあるテキストボックスで指定する。振幅変動のコンボボックスには1次近似、対数近似、べき乗近似、指数近似、多項式近似の項目が含まれている。多項式近似の次数は横のテキストボックスで指定する。

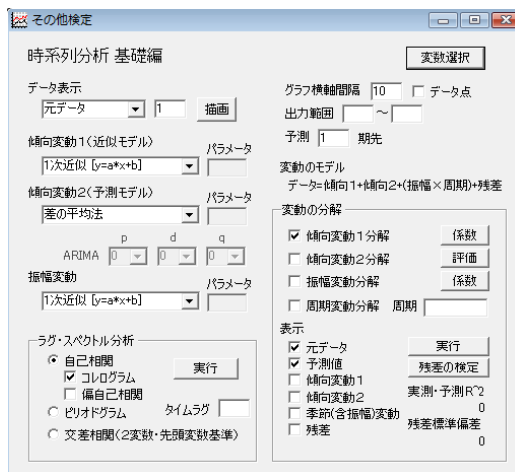


図 3.1 時系列分析メニュー

周期変動の周期は左下のラグ・スペクトルグループボックスで調べる。必要なラジオボタンやチェックボックスを選び、実行ボタンでそれぞれのグラフが表示される。タイムラグテキストボックスでは詳細な検討のためのコレログラムのタイムラグやピリオドグラムの変動の最大を与える。特に指定がなければ、データで利用できる最大値が使われる。

4. おわりに

我々は時系列分析について基本的な事柄のみを用いてプログラムを作成した。しかし1つ1つの手法は単純でも、それらを組み合わせることで実用的な分析も可能になっているように思う。このプログラムは1年前にプロトタイプが完成していたが、分析方法がまとまりに欠けていたことやメニューが大きくなり過ぎて見づかったことで公表できずにいた。今回手法を組み合わせ、画面も整理できたので論文にまとめることにした。

分析手法を組み合わせ分ったことであるが、予測モデルは個々に利用すると予測能力はそれほど高いとはいえないが、変動の分解モデルと組み合わせると、振幅変動の除去の役割を担うものもあり、思った以上に役に立つ。但し、長期の予測には向かないので状況に応じて利用する必要がある。どのような時系列データが我々のプログラムで予測可能か検討しておくことは今後の教育にも役に立つ。

新しい分析の開発以外にも我々はプログラムに細かな修正を加えている。特に多変量解析では結果の連携の重要性を考えるようになった。例えば、因子分析の因子得点を他の分析で使用し、クラスター分析の分類を使って分散分析を行うなどすることで応用範囲は広がってくる。そのため機能として、因子得点や重回帰分析の予測値、クラスター分析の分類などを欠損値も含めて表示し、エディタに追加できるようにしている（現在、重回帰分析、数量化Ⅰ類、判別分析、主成分分析、因子分析に適用している）。また因子分析には解を求める手法として主成分分析による方法を加え、因子負荷量の表示順を変数の変数の分類に役立つように並べ替える機能を追加した。これらはゼミでの SPSS との機能比較や利用者からの指摘によるところが大きい。

これまで文系学部における数理系教育の方法について考えてきて、徹底した実践教育の必要性を主張してきた⁵⁾。しかし大学院における授業については、理論的な基礎を固めればよい、程度のことを考えるだけであった。ただ学部の時代（もっと言えば高校時代）から数学に苦手意識を持っている大学院生に数学の知識を急に教えたとしても限界がある（もちろん例外はあるし、ある程度の数学的理解は不可欠でもある）。しかし次第に分析の連携を意識し始めるにつれ、大学院での教育はこのような連携方法の教育こそ重要であると考えようになってきた。実際研究では分析の連携は日常的に行われているが、大学院教育としてこの姿勢を通した授業はあまり多くないように思われる。幸い我々のプログラムは統計だけでなく、経営科学の分野も含んだ統合型

である。こういった考え方には向いているのかも知れない。今後は、新しい分析手法を加える一方で、どのような連携があり得るのかも考えて教材作成を行いたい。

参考文献

- 1) 高橋玲子他著, 上田太一郎監修, Excel で学ぶ時系列分析と予測, オーム社, 2006.
- 2) 北川源四郎, 時系列解析入門, 岩波書店, 2005.
- 3) 石村貞夫, SPSS による時系列分析の手順 [第2版], 東京図書, 2006.
- 4) この論文では、統計数理研究所のホームページの中の Web Decomp のサンプルデータを使用させていただきました。 (<http://ssnt.ism.ac.jp/inets/inets.html>)
- 5) 日本教育情報学会第 25 回年会論文集, (2009) (於立命館大学, 2009/8/23) .

Multi-purpose Program for Social System Analysis 11

- Time Series Analysis -

Masayasu FUKUI, Yingchun WANG*, Jing WANG* and Keiji ISHIMARU**

Department of Business Administration, Faculty of Business Administration,
Fukuyama Heisei University

* Graduate School of Business Administration,
Fukuyama Heisei University

** Department of Economics, Faculty of Economics,
Fukuyama University

Abstract

We have been constructing a unified program on the social system analysis for the purpose of education. This time we created a new program of time series analysis. We thought a time series decomposition model as Trends1 (approximation model) + Trends2 (prediction model) + Seasonal variation (amplitude variation \times periodic variation) + Residual.

Keywords

College Analysis, social system analysis, OR, statistics, time series analysis

URL: <http://www.heisei-u.ac.jp/ba/fukui/>