社会システム分析のための統合化プログラム２７
― 局所重回帰分析・パネル重回帰分析 ―

福井正康・*岩村忠昭・尾崎誠
福山平成大学経営学部経営学科
*元 川崎製鉄（JFE スチール）株式会社

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。今回は予測曲面の形状を仮定しない予測手法である局所重回帰分析と複数変数の時系列データによるパネル重回帰分析についてのプログラムを紹介する。特にパネル重回帰分析では予測の精度を上げるために、傾向変動及び周期変動分解による時系列分析の予測値を変数として加えることを考えた。これにより２つの手法の良い部分を活かすことができると考える。その際、傾向変動としては局所回帰分析を使って予測値を求める方法を取り入れている。
キーワード：College Analysis:　多変量解析，局所重回帰分析，パネル重回帰分析，時系列分析
URL：　http://www.heisei-u.ac.jp/ba/fukui/

１．はじめに

我々がこれまで College Analysis の中で作成してきた、重回帰分析や非線形最小２乗法は、パラメータを含んだ関数形を仮定して、最小 2 乗法によりそのパラメータを推定する手法である。このうち重回帰分析はモデルとして簡単であるが、線形の影響だけに限られ、非線形最小２乗法は、適合する関数形が見つけられる場合は有力であるが、関数形を見つけること自体が簡単ではない。今回作成した局所重回帰分析は、関数形の仮定をせずに、非線形の説明変数の影響を含んだデータの予測を行うことができる興味深い手法である。但し、関数形が指定されないので、一度モデルを決めてしまうと後はその中に数値を当てはめるだけというわけにはいかず、説明変数のデータが与えられる度に計算をやり直さなくてはならない。また、そもそもモデルという考え自体もあまり適当ではない。

局所重回帰分析は文字通り、予測したい要求点の近傍のデータに大きなウェイトをかけて局所的に実行する重回帰分析である。どの程度の距離まで影響を考えるかはバンド幅という数値でユーザーが指定する。そのため、要求点によって回帰式が大きく異なるし、その回帰式で与えられる予測値は、要求点が与えられた後で一から計算される。

パネル重回帰分析は、ある変数の変動の前触れは何かの変数に現れているという考えに基づく予測手法である。ある変数とその他の変数の現在及び過去のデータで未来のデータを予測するが、過去に１期ずつさかのぼりながらデータを作り、重回帰分析を適用する。しかし、データによっては傾向変動や周期変動の分解モデルの方が良い結果を与える場合もある。そのため、我々のパネル重回帰分析には、変数の１つとして変動の分解モデルの予測値をパネルデータに追加する機能を加えてある[2],[3]。

２．局所重回帰分析

これまでの重回帰分析や非線形最小 2 乗法の予測手法は、パラメータを含んだ関数形を仮定し、最小 2 乗法によってパラメータの値を定め、予測関数を確定するものであった。しかし、局所重回帰分析は要求点を与えることによって、その近傍の点による重回帰分析の結果から直接予測値を求める方法で、関数形を必要としない予測手法である。

2.1　局所重回帰分析の理論

標準的な重回帰分析は、目的変数 $y_\lambda\ (\lambda=1,2,\cdots,N)$ と、説明変数 $x_{i\lambda}\ (i=1,2,\cdots,p)$ の線形結合 $Y_\lambda=\sum_{i=1}^{p}b_i x_{i\lambda}+b_0$ の差の 2 乗和 L を最小にするようにパラメータ $b_i\ (i=0,1,2,\cdots,p)$ を決定する。ここに L は以下で与えられる。

$$L=\sum_{\lambda=1}^{N}\left(y_\lambda-Y_\lambda\right)^2=\sum_{\lambda=1}^{N}\left(y_\lambda-\sum_{i=1}^{p}b_i x_{i\lambda}-b_0\right)^2$$

これに対して局所重回帰分析は、各データに対してウェイト w_λ をかけて以下の L' を最小化する。

$$L'=\sum_{\lambda=1}^{N}w_\lambda\left(y_\lambda-Y_\lambda\right)^2=\sum_{\lambda=1}^{N}w_\lambda\left(y_\lambda-\sum_{i=1}^{p}b_i x_{i\lambda}-b_0\right)^2$$

この解は、$\mathbf{b}={}^t\begin{pmatrix}b_0 & b_1 & b_2 & \cdots & b_p\end{pmatrix}$ として、以下のように求めることができる。

$$\mathbf{b}=\left({}^t\mathbf{\Omega\Pi\Omega}\right)^{-1}{}^t\mathbf{\Omega\Pi y} \tag{2.1}$$

ここに、

$$\mathbf{y}={}^t\begin{pmatrix}y_1 & y_2 & \cdots & y_N\end{pmatrix},$$

$$\mathbf{\Omega}=\begin{pmatrix}1 & x_{11} & x_{21} & \cdots & x_{p1}\\ 1 & x_{12} & x_{22} & \cdots & x_{p2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_{1N} & x_{2N} & \cdots & x_{pN}\end{pmatrix},\quad \mathbf{\Pi}=\begin{pmatrix}w_1 & 0 & \cdots & 0\\ 0 & w_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & w_N\end{pmatrix}$$

要求点 x_i^r の予測値 Y^r は、以下のように与えられる。

$$Y^r=\sum_{i=1}^{p}b_i x_i^r+b_0 \tag{2.2}$$

ウェイト w_λ は以下のように求める。まず、説明変数についての要求点 x_i^r とバンド幅（調整パラメータ） $p\ (>0)$ を定める。要求点は局所重回帰分析のウェイトの中心を表す点である。次に標準化された観測点 $\tilde{x}_{i\lambda}=\dfrac{x_{i\lambda}-\overline{x}_i}{\sigma_i}$ と標準化された要求点 $\tilde{x}_i^r=\dfrac{x_i^r-\overline{x}_i}{\sigma_i}$ との間のユークリッド距離

$$\Gamma_\lambda = \sqrt{\sum_{i=1}^{p}\left(\tilde{x}_{i\lambda} - \tilde{x}_i^r\right)^2}$$

を求める。但し、標準化の際の標準偏差は不偏分散からのものとする。

この距離 Γ_λ について、その平均を $\overline{\Gamma}$、不偏分散からの標準偏差を σ_Γ とし、これらを用いて、ウェイト w_λ を以下のように定義する。

$$w_\lambda = \exp\left[-\left(\Gamma_\lambda / p\sigma_\Gamma\right)^2\right] \tag{2.3}$$

これによって要求点の近傍の点にウェイトをかけて最小 2 乗法の解を求めることになる。

標準化偏回帰係数については、標準化されたデータ $\tilde{y}_\lambda, \tilde{x}_{i\lambda}$ を用いて、以下のように求めることもできる。

$$\tilde{\mathbf{b}} = \left({}^t\tilde{\mathbf{\Omega}}\mathbf{\Pi}\tilde{\mathbf{\Omega}}\right)^{-1}{}^t\tilde{\mathbf{\Omega}}\mathbf{\Pi}\tilde{\mathbf{y}} \tag{2.4}$$

ここに、

$$\tilde{\mathbf{y}} = {}^t\left(\tilde{y}_1 \quad \tilde{y}_2 \quad \cdots \quad \tilde{y}_N\right), \qquad \tilde{y}_\lambda = \frac{y_\lambda - \overline{y}}{\sigma_y} \quad \text{（不偏分散を用いた標準化）}$$

$$\tilde{\mathbf{\Omega}} = \begin{pmatrix} 1 & \tilde{x}_{11} & \tilde{x}_{21} & \cdots & \tilde{x}_{p1} \\ 1 & \tilde{x}_{12} & \tilde{x}_{22} & \cdots & \tilde{x}_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \tilde{x}_{1N} & \tilde{x}_{2N} & \cdots & \tilde{x}_{pN} \end{pmatrix}, \qquad \mathbf{\Pi} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N \end{pmatrix}$$

別の書式で書くと以下となる。

$$\tilde{b}_i = \frac{\sigma_i}{\sigma_y} b_i, \qquad \tilde{b}_0 = \frac{1}{\sigma_y}\left(b_0 + \sum_{i=1}^{p} b_i \overline{x}_i - \overline{y}\right) \tag{2.5}$$

この関係は、以下のように求めることができる。

$$\begin{aligned} \frac{Y_\lambda - \overline{y}}{\sigma_y} &= \frac{1}{\sigma_y}\left(\sum_{i=1}^{p} b_i x_{i\lambda} + b_0 - \overline{y}\right) \\ &= \sum_{i=1}^{p} \frac{\sigma_i}{\sigma_y} b_i \frac{x_{i\lambda} - \overline{x}_i}{\sigma_i} + \frac{1}{\sigma_y}\left(b_0 + \sum_{i=1}^{p} b_i \overline{x}_i - \overline{y}\right) \end{aligned}$$

通常の重回帰分析では $\overline{y} = \overline{Y} = \sum_{i=1}^{p} b_i \overline{x}_i + b_0$ であるから、標準化された定数項は 0 になるが、局所重回帰分析では一般に $\overline{y} \neq \overline{Y}$ であるので、標準化された定数項は 0 にならない。

偏回帰係数と標準化偏回帰係数の関係は、(2.5)式とは逆に以下のように書くこともできる。我々のプログラムではこの関係を利用している。

$$b_i = \frac{\sigma_y}{\sigma_i}\tilde{b}_i\,, \qquad b_0 = \sigma_y \tilde{b}_0 - \sum_{i=1}^{p} \frac{\sigma_y}{\sigma_i}\tilde{b}_i \bar{x}_i + \bar{y} \tag{2.6}$$

局所重回帰分析はバンド幅（調整パラメータ） p が無限大になるとウェイトがすべて 1 になり、通常の重回帰分析に近づく。

局所重回帰分析は要求点の近傍で成り立つ近似手法であるので、通常の RMSE や重相関係数の指標は使えず、その信頼性を求める指標は 1 個抜き交差検証法（HOOCV：Leave-One-Out Cross-Validation）を用いて与える。即ち、データ中の 1 点を抜き、その説明変数の座標 $x_{i\lambda}$ を要求点とし、残りの点で局所重回帰分析を行い、要求点の予測値 Y_λ を求める。元々この点には実測値 y_λ があるので予測の誤差が求められる。

局所重回帰分析の精度の指標はこの実測値と予測値を利用し、通常の重回帰分析の RMSE や重相関係数の定義を用いて以下のように与える。もちろんこの指標はバンド幅に影響される。

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{\lambda=1}^{N}\left(y_\lambda - Y_\lambda\right)^2}\,,$$

$$\text{重相関係数} = \frac{\sum_{\lambda=1}^{N}\left(y_\lambda - \bar{y}\right)\left(Y_\lambda - \bar{Y}\right)}{\sqrt{\sum_{\mu=1}^{N}\left(y_\mu - \bar{y}\right)^2 \sum_{\nu=1}^{N}\left(Y_\nu - \bar{Y}\right)^2}} \tag{2.7}$$

局所重回帰分析は、バンド幅や 1 個抜く点によって必ずしも予測値が求められるとは限らない。そのため、RMSE や重相関係数の値は求められた点だけを用いて計算することもある。

2.2　プログラムの利用法

メニュー［分析－多変量解析等－予測手法－局所重回帰分析］をクリックすると図 2.1 に示すような局所重回帰分析の実行メニューが表示される。

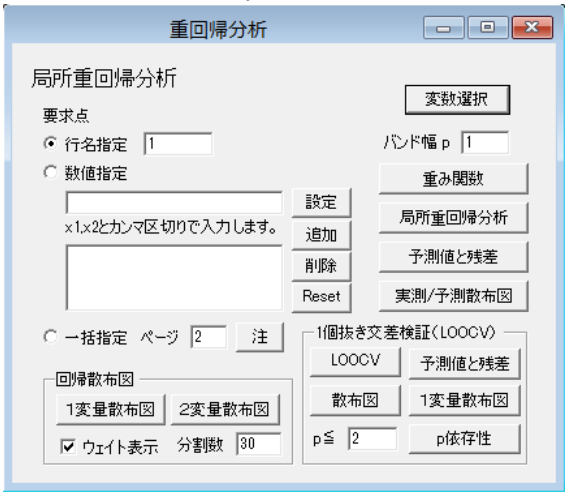


図 2.1　実行メニュー

通常の重回帰分析と同様に「変数選択」で、目的変数、説明変数の順番に変数を選ぶ。要求点は、「行名指定」でデータから選択するか、「数値指定」で外部から入力する。行名指定は、データのレコード名の部分の表示で指定する。レコード名が見当たらない場合は、実行の際にメッセージが表示される。数値指定の場合は、テキストボックスに説明変数の値をカンマ区切りで入力する。複数の要求点を調べることが必要であるので、プログラムには入力した値を保存しておく機能が付いている。テキストボックスに書いた要求点のデータは、「追加」ボタンで下のリストボックスに追加保存される。リストボックスのデータは選択して、「設定」ボタンでテキストボックスに呼び戻すことができる。また、選択して「削除」ボタンで 1 つだけリストから削除でき、「Reset」ボタンですべて削除することができる。変数選択の場合と同じ要領で活用できる。

バンド幅を適当な値（ここでは 1）に設定し、適当な行名を指定して「局所重回帰分析」ボタンをクリックすると、図 2.2 のような分析結果が得られる。

偏回帰係数と要求点

	偏回帰係数	標準化係数	要求点	標準化点
▶ 説明1	0.3812	0.1982	31	−0.9408
説明2	0.4113	0.2146	19	−1.2370
切片/要求予測値	51.6478	−0.5032	71.2791	−0.9552

図 2.2　偏回帰係数の出力結果

重回帰式による推測結果と各観測点のウェイト値は「予測値と残差」ボタンで図 2.3 のように表示される。

予測値と残差

	実測値	予測値	残差	ウェイト
▶ 1	66	71.2791	−5.2791	1.0000
2	89	82.2936	6.7064	0.4037
3	73	75.1613	−2.1613	0.6670
4	80	76.4653	3.5347	0.5477
5	75	74.8603	0.1397	0.9383
6	79	67.4070	11.5930	0.8994
7	81	80.4278	0.5722	0.0679
8	66	72.2221	−6.2221	0.9224
9	70	75.4421	−5.4421	0.7975
10	78	77.7895	0.2105	0.6696

図 2.3　実測値と予測値

実測値と予測値の関係は「実測/予測散布図」をクリックすると、図 2.4 のように表示される。

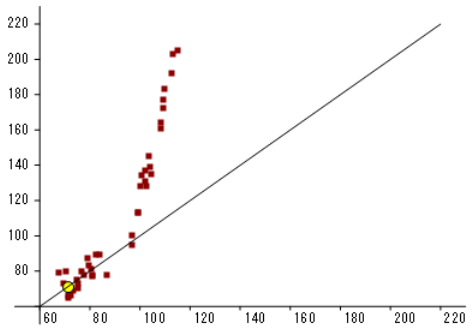


図 2.4　実測/予測値散布図 1

図中の黄色い点は要求点で、直線は実測と予測が同じであるとする直線である。要求点近傍の点の予

測がうまく行っている状況が見える。

偏回帰係数は、要求点とバンド幅に大きく影響を受ける。要求点を変更したときの結果を図 2.5 に示す。今度は別の点の予測がうまく行っている。

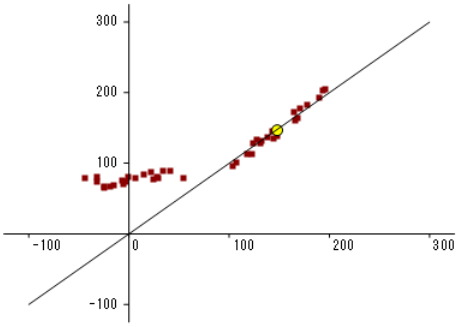


図 2.5　実測/予測値散布図２

実際の x,y 軸の上で回帰直線を引いてみる。変数を目的変数と説明変数を１つにして、「1 変量回帰散布図」を描くと図 2.6 のようになる。2 つの図は要求点を変えて描いている。

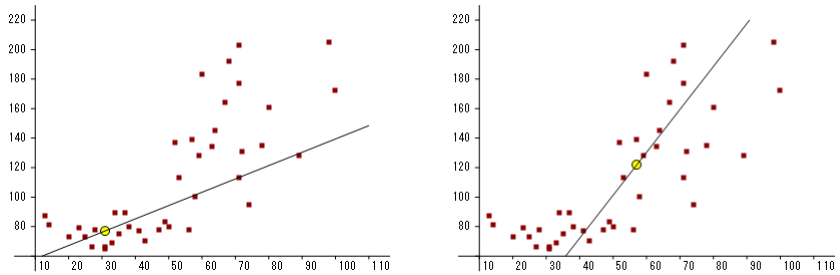


図 2.6　１変量回帰散布図（p=1）

これは、データの散布図であり、図中の直線は局所回帰直線である。要求点によって局所回帰直線が変化しているのが分かる。

また、実際の x,y,z 軸上で局所回帰平面を描いてみる。変数を目的変数と説明変数を２つにして、「２変量回帰散布図」を描くと図 2.7 のようになる。2 つの図は要求点を変えて描いている。

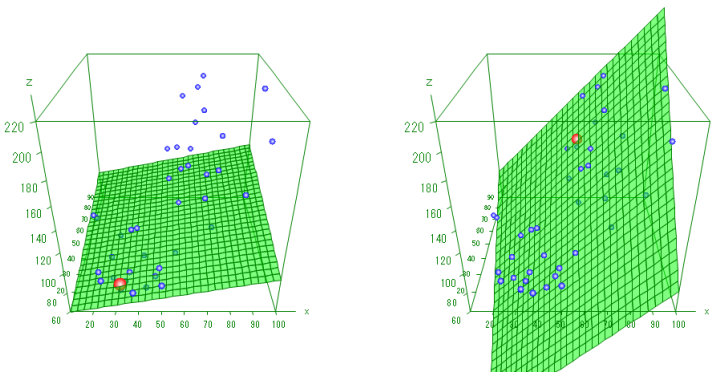


図 2.7　２変量回帰散布図（p=1）

次にバンド幅を p=0.5 と p=5 にし、説明変数の数を 1 つにして、1 変数回帰散布図を描く。結果を図 2.8 に示す。

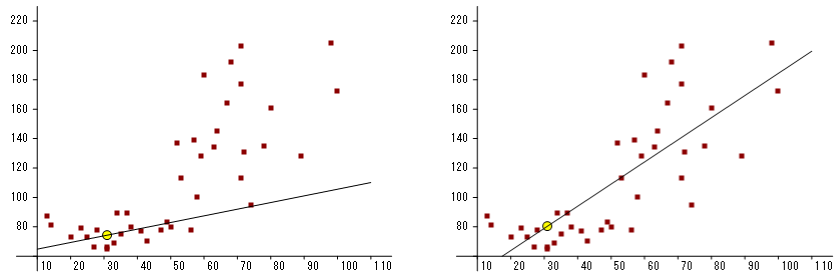


図 2.8　図 6 左の要求点で p=0.5（左）と p=5（右）の 1 変量回帰散布図

バンド幅の値により、局所性が大きく変更を受けていることが分かる。右側の図は通常の回帰直線に近い。

分析メニューで「重み関数」ボタンをクリックすると 2 変数グラフ描画メニューが表示される。その中の「グラフ描画」ボタンをそのままクリックすると、図 2.9 左のような実際の重み関数のグラフ（この場合は 1 変量）が表示される。2 変量の場合は図 2.9 右のようなグラフになる。

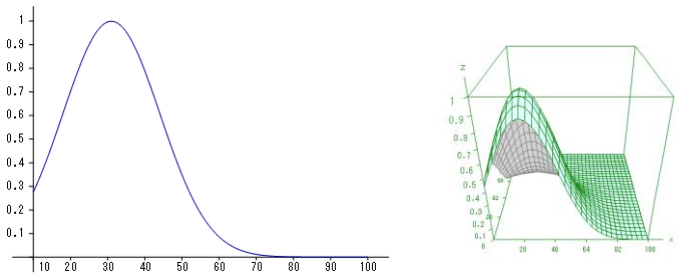


図 2.9　重み関数グラフ（左は 1 変量、右は 2 変量）

これまでは要求点を 1 点だけ指定したが、現実の分析では多くの要求点を一度に与えて予測値を求めることも考えられる。実行メニューで、要求点の「一括指定」ラジオボタンを選択すると、別のページに与えられた複数の要求点のデータから一括で予測値を求めることもできる。要求点のページは「一括指定」ラジオボタン右側の「ページ」テキストボックスに与える。デフォルトは 2 ページ目になっているので必要なら変更する。要求点のページの例を図 2.10 に示す。

図 2.10　要求点の一括指定

ここで注意することは、変数名を必ず正確に（全角半角や大文字小文字の区別を付けて）指定することである。分析では変数選択の数や順番が要求点の指定通りとは限らないので、プログラムでは変数名を探して順番等を合わせるようにしている。

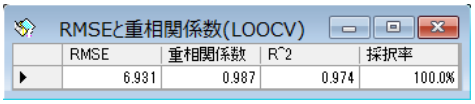
一括指定した要求点を用いた場合は、重回帰式の偏相関係数などは重要でないので、結果は要求点と予測値を表形式で与える。要求点指定に空欄がある場合は、予測値の欄が空欄になる。予測値の出力例を図 2.11 に与える。


要求点と予測値

	予測値	説明1	説明2
1	71.279	31	19
2	84.700	34	43
3	75.704	25	34
4	78.734	50	14
5	74.550	35	24
11	82.866	13	55
12	71.279	31	19
13	80.777	41	33

図 2.11　要求点一括指定の出力

局所重回帰分析の予測精度を与えるために、1 個抜き交差検証（LOOCV）を用いた RMSE と重相関係数を与える。「LOOCV」ボタンをクリックすると図 2.12 のような結果が表示される。

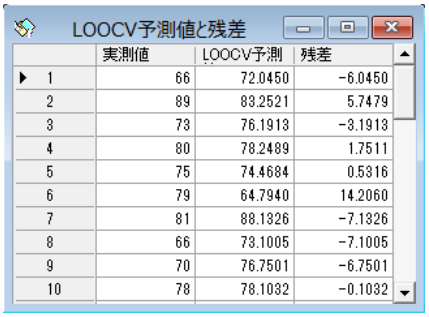

RMSEと重相関係数(LOOCV)

RMSE	重相関係数	R^2	採択率
6.931	0.987	0.974	100.0%

図 2.12　1 個抜き交差検証による RMSE と重相関係数

ここで採択率は、1 個抜いたデータで計算ができない場合があるので、計算できるデータ点の割合を示したものである。

この求めた予測値と実測値の具体的な値は 1 個抜き交差検証中の「予測値と残差」ボタンをクリックすることで図 2.13 のように与えられる。予測値が求められなかった部分は空白になっている。


LOOCV予測値と残差

	実測値	LOOCV予測	残差
1	66	72.0450	-6.0450
2	89	83.2521	5.7479
3	73	76.1913	-3.1913
4	80	78.2489	1.7511
5	75	74.4684	0.5316
6	79	64.7940	14.2060
7	81	88.1326	-7.1326
8	66	73.1005	-7.1005
9	70	76.7501	-6.7501
10	78	78.1032	-0.1032

図 2.13　1 個抜き交差検証による実測値と予測値

この関係は 1 個抜き交差検証中の「散布図」ボタンで、実測/予測散布図として図 2.14 のように与えられる。

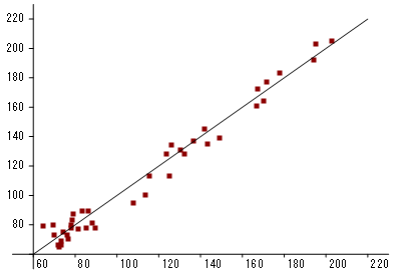


図 2.14　1 個抜き交差検証による実測/予測散布図

説明変数による予測値と実測値の関係は、1 変量の場合「1 変量散布図」をクリックして図 2.15 のように与えられる。この図の場合、特別に説明変数を 1 個だけにした。

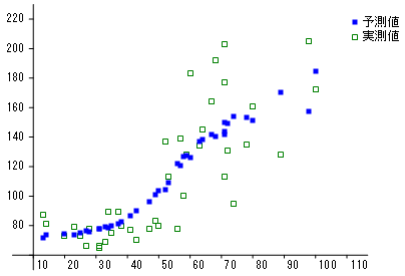


図 2.15　1 個抜き交差検証による 1 変量散布図

バンド幅によって、RMSE や重相関係数の値は変化する。「p 依存性」ボタンをクリックすると、RMSE のバンド幅 p の値による変化が図 16 のように示される。

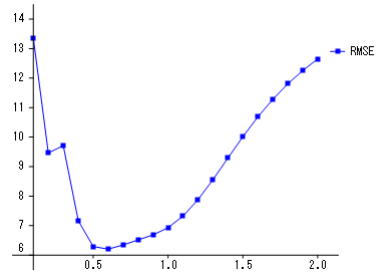


図 2.16　バンド幅の値による RMSE の変化

ここで、p=0.3 のところで値が急に大きくなっているが、この部分は 1 個抜き交差検証ですべての点が利用できなかった部分である。この場合は、p=0.6 当たりで適合が良さそうである。

３．パネル重回帰分析

3.1　パネル重回帰分析の理論

変数i（$i=1,\cdots,p$）、時刻t（$t=0,\cdots,T$）の時系列データ$x_{i,t}$があるとき、その中から時刻tを含めてr期分のそれ以前のデータを取り出す。それらのデータを説明変数とし、時刻$t+a$（$a\geq 1$）のある変数dのデータ$x_{d,t+a}$を目的変数として予測する重回帰分析をパネル重回帰分析という。これはa期先の予測である。

予測値を $X_{d,t+a}$ とすると予測式は以下のように与えられる。

$$X_{d,t+a} = \sum_{i=1}^{p}\sum_{j=0}^{r-1} b_{i,j} x_{i,t-j} + b_0 \tag{3.1}$$

係数 $b_{i,j}, b_0$ は以下の量 L を最小化することによって求める。

$$L = \sum_{t=a+r-1}^{T} \left(x_{d,t} - \sum_{i=1}^{p}\sum_{j=0}^{r-1} b_{i,j+1} x_{i,t-a-j} - b_0 \right)^2 \tag{3.2}$$

今、目的変数と説明変数をそれぞれ以下のように定義する。

$y_\lambda = x_{d,\lambda+a+r-2}$ （$\lambda = 1, \cdots, T-a-r+2$）

$z_{\alpha,\lambda} = z_{i+pj,\lambda} = x_{i+pj,\lambda+r-2-j}$ （$i = 1, \cdots, p,\ j = 0, \cdots, r-1,\ \alpha = 1, \cdots, pr$）

これは、例えば $\lambda = T-a-r+2$ として、ある変数の時刻 T の予測を行うのに、すべての変数の時刻 $T-a \sim T-a-(r-1)$ のデータを使うことを表している。

回帰式の係数を b_α にして(3.2)式を書き変えると、以下のような式になる。

$$L = \sum_{\lambda=1}^{T-a-r+2} \left(y_\lambda - \sum_{\alpha=1}^{pr} b_\alpha z_{\alpha,\lambda} - b_0 \right)^2 \tag{3.3}$$

これから、偏回帰係数 $\mathbf{b} = {}^t\begin{pmatrix} b_0 & b_1 & b_2 & \cdots & b_s \end{pmatrix}$, $s = pr$ は以下のように求めることができる。

$$\mathbf{b} = \left({}^t\mathbf{\Omega}\mathbf{\Omega} \right)^{-1} {}^t\mathbf{\Omega}\mathbf{y} \tag{3.4}$$

ここに、

$$\mathbf{y} = {}^t\begin{pmatrix} y_1 & y_2 & \cdots & y_N \end{pmatrix},\quad N = T-a-r+2$$

$$\mathbf{\Omega} = \begin{pmatrix} 1 & z_{11} & z_{21} & \cdots & z_{s1} \\ 1 & z_{12} & z_{22} & \cdots & z_{s2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{1N} & z_{2N} & \cdots & z_{sN} \end{pmatrix}$$

時系列分析ではデータが時間の経過とともに明らかになっていくので、現在のすべてのデータから求めたパラメータを使って、過去の各時間の予測を行うことはその時点のデータの影響を強く受け過ぎるという難点がある。そこで、過去の予測を行う際には、その時点までのデータから計算されたパラメータを用いることとし、これによって実測値と予測値の相関を求めることにする。これは一種の交差検証になっている。プログラムにはこの交差検証を付け加えている。

パネル重回帰分析には、他の分析で予測した結果を組み込むことができる。そこで時系列分析の結果をデータとして組み込むことを考えてみた。時系列分析は、傾向変動と周期変動を分解するモデルを考える[2],[3]。データの不規則な大きな変動も考える必要があるので、傾向変動には自然に傾向を求めることができる局所回帰分析を採用した。そのためバンド幅によって局所的な回帰式に影響を与える範囲を限定することができる。また周期変動については、分解する周期（周波数）を複数指定できるようにしている。

3.2　プログラムの利用法

パネル重回帰分析のデータは複数変数の時系列データである。その例を図 3.1 に示す。

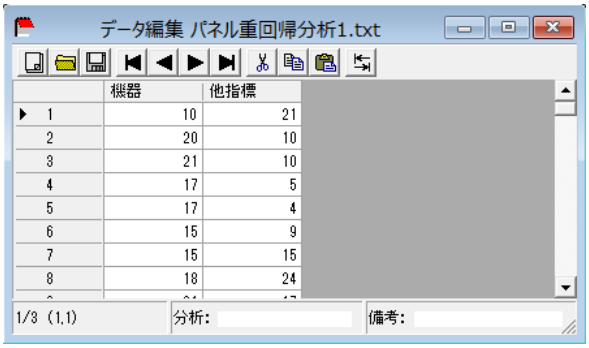


図 3.1　パネル重回帰分析のデータ

メニュー［分析－多変量解析他－予測手法－パネル重回帰分析］を選択すると図 3.2 のようなパネル重回帰分析の実行メニューが表示される。

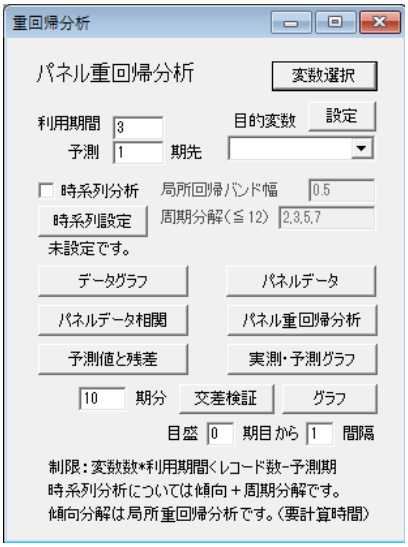


図 3.2　実行メニュー

使用するデータをすべて「変数選択」ボタンで選ぶが、変数間の時間的な影響を調べるツールとして使うことも考えているため、通常の重回帰分析のように目的変数を最初に選択することはしない。目的変数は、変数選択した候補をコンボボックスに読み込んだ後で、その中から「設定ボタン」で選択する。コンボボックスの選択肢の中には単独の変数の他に「すべて」というものがあり、選択したすべての変数を目的変数にして、素早く結果を求めるときに利用する。ボタンによってはこれが使えないものもある。

この分析では、何期分のデータを利用するか、何期先の予測をするかを設定する。それに応じて、「パネルデータ」ボタンでは時系列データを通常の重回帰分析の形式に変形して出力する。出力結果をそのまま重回帰分析のデータとして利用することもできる。変数「機器」を目的変数とし、3 期分のデータを利用し、1 期先の予測をする場合の計算用データを図 3.3 に示す。

パネルデータ

	機器	機器_1	他指標_1	機器_2	他指標_2	機器_3	他指標_3
▶ 4	17	21	10	20	10	10	21
5	17	17	5	21	10	20	10
6	15	17	4	17	5	21	10
7	15	15	9	17	4	17	5
8	18	15	15	15	9	17	4
9	21	18	24	15	15	15	9
10	26	21	17	18	24	15	15
11	23	26	4	21	17	18	24

図 3.3　計算用データ

この中で「機器」は目的変数で、左に月単位で与えられている。また、例えば「他指標_2」は変数「他指標」の 2 期前のデータを表している。

図 3.3 の計算用データの各変数間の相関係数は、「パネルデータ相関」ボタンをクリックすることで図 3.4 のように与えられる。

パネルデータ相関係数

	機器	機器_1	他指標_1	機器_2	他指標_2	機器_3	他指標_3
▶ 機器	1.000	0.824	−0.052	0.833	0.105	0.834	−0.146
機器_1	0.824	1.000	−0.095	0.817	−0.054	0.825	0.102
他指標_1	−0.052	−0.095	1.000	−0.093	−0.056	−0.019	−0.137
機器_2	0.833	0.817	−0.093	1.000	−0.099	0.807	−0.060
他指標_2	0.105	−0.054	−0.056	−0.099	1.000	−0.094	−0.060
機器_3	0.834	0.825	−0.019	0.807	−0.094	1.000	−0.113
他指標_3	−0.146	0.102	−0.137	−0.060	−0.060	−0.113	1.000

図 3.4　パネルデータ相関出力結果

このデータを使った重回帰分析の詳細は、「パネル重回帰分析」ボタンで図 3.5 のように与えられる。

パネル重回帰分析

	偏回帰係数	標準化係数	t検定値	自由度	確率値
▶ 機器_1	0.3258	0.3200	3.4654	90	0.0008
他指標_1	0.0243	0.0112	0.2531	90	0.8008
機器_2	0.3579	0.3433	4.1283	90	0.0001
他指標_2	0.3891	0.1784	4.0746	90	0.0001
機器_3	0.3166	0.2978	3.3981	90	0.0010
他指標_3	−0.2432	−0.1121	−2.3768	90	0.0196
切片	−1.2488	0.0000	−0.3919	90	0.6960
重相関・寄与率	0.912	0.833			

図 3.5　目的変数を「機器」とした場合のパネル重回帰分析結果

目的変数を「すべて」に設定すると、「パネル重回帰分析」ボタンで図 3.6 のような結果になる。

パネル重回帰分析（すべて）

	機器:偏回帰:	標準化	確率値	他指標:偏回	標準化	確率値
▶ 機器_1	0.3258	0.3200	0.0008	−0.1127	−0.2414	0.2784
他指標_1	0.0243	0.0112	0.8008	−0.0719	−0.0718	0.4983
機器_2	0.3579	0.3433	0.0001	0.0622	0.1300	0.5159
他指標_2	0.3891	0.1784	0.0001	−0.1324	−0.1324	0.2105
機器_3	0.3166	0.2978	0.0010	0.0234	0.0480	0.8198
他指標_3	−0.2432	−0.1121	0.0196	0.0167	0.0168	0.8826
切片	−1.2488	0.0000	0.6960	16.8657	0.0000	0.0000
重相関・寄与率	0.912	0.833		0.195	0.038	

図 3.6　目的変数をすべてとした場合のパネル重回帰分析結果

これは各変数を目的変数にして、偏回帰係数、標準化偏回帰係数、確率値、重相関係数、寄与率を出力している。どの変数の何期前のデータが重要であるか、標準化係数や確率値を見ることで知ることができる。

目的変数を「機器」とした場合の実測値、予測値、残差は、「予測値と残差」ボタンをクリックすることで図 3.7 のように求められる。ここで一番下の予測値は、1 期先（設定で変更可能）の予測値で、実測値はまだない。

予測値と残差

	機器	予測値	残差
94	57.000	56.017	0.983
95	62.000	64.020	-2.020
96	78.000	59.132	18.868
97	61.000	66.715	-5.715
98	70.000	73.932	-3.932
99	69.000	70.957	-1.957
100	66.000	65.528	0.472
1期先		70.471	

図 3.7　目的変数を「機器」とした場合の予測値と残差結果

また、目的変数を「すべて」とした場合の実測値、予測値、残差は、同様にして図 3.8 のように求められる。

予測値と残差（すべて）

	機器	予測値	残差	他指標	予測値	残差
94	57.000	56.017	0.983	4.000	13.589	-9.589
95	62.000	64.020	-2.020	7.000	13.965	-6.965
96	78.000	59.132	18.868	23.000	14.173	8.827
97	61.000	66.715	-5.715	21.000	10.749	10.251
98	70.000	73.932	-3.932	12.000	11.852	0.148
99	69.000	70.957	-1.957	15.000	11.333	3.667
100	66.000	65.528	0.472	18.000	12.551	5.449
		70.471			12.274	

図 3.8　目的変数をすべてとした場合の予測値と残差結果

実測値と予測値について、結果をグラフで表示するためには、「実測・予測グラフ」ボタンをクリックする。実行結果は図 3.9 に示す。

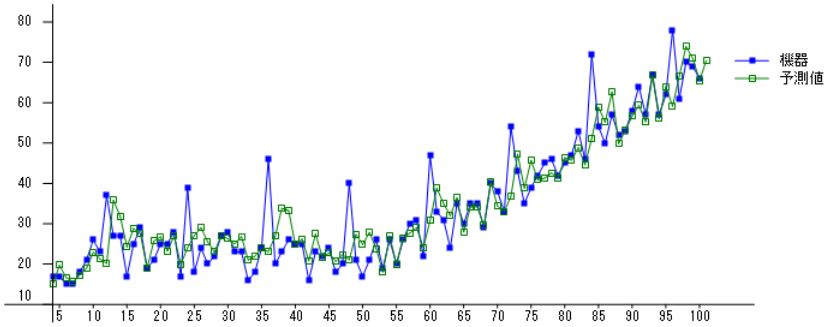


図 3.9　実測値と予測値グラフ

ここにデータの名前は 5 期ごとに設定している。

我々がこれまで求めてきた各時点の予測値は、全体の結果を使って求めた係数から計算して得られた値である。それゆえ、この係数には各時点の実測値の結果が含まれている。そのためこれらのデータは厳密には予測値ではない。これを補正するためには、予測値は各時点のそれより過去のデータから求めるべきであろう。この考え方は交差検証の考え方に通じる。「期分」のテキストボックスに予測したい期間の数値を入れ、「交差検証」ボタンをクリックすると、過去のデータからだけで作られた予測値と残差が図 3.10 のように表示される。但し、表示期間を 50 期分にしている。

	機器	予測値	残差
94	57.000	54.447	2.553
95	62.000	63.934	-1.934
96	78.000	57.643	20.357
97	61.000	68.532	-7.532
98	70.000	74.934	-4.934
99	69.000	71.341	-2.341
100	66.000	65.481	0.519
R･R^2	0.907	0.824	

図 3.10　目的変数を「機器」とした場合の 50 期分の交差検証結果

目的変数をすべてにして同様の結果を得ることもできる。「グラフ」ボタンをクリックすると、図 3.10 の結果をグラフ化することができる。結果を図 3.11 に示す。

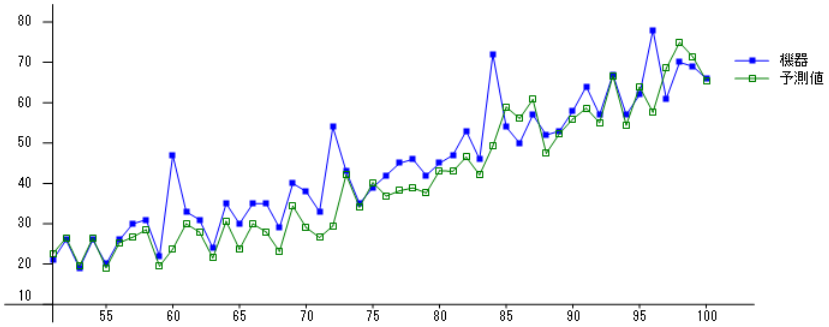


図 3.11　交差検証での実測値と予測値

純粋なパネル重回帰分析の結果は以上であるが、我々はさらに予測精度を上げるために、傾向変動や周期変動の分解を考える時系列分析の予測値をパネルデータに加え、２つの分析の良い部分を組み合わせることにした。ここで、傾向変動の分解には局所回帰分析（説明変数が１つであるから、このような呼び方にした）を用いている。図 3.2 の実行メニューの「時系列分析」チェックボックスにチェックを入れると、「局所回帰バンド幅」と「周期分解（≦12)」のテキストボックスが利用できるようになる。バンド幅の値はデフォルトでほぼ良い結果が得られるが、例えば 12 ヶ月周期が明らかな場合には、周期分解に 12 を含める。周期分解のためのデータ数は最低でも最大周期の２倍必要なので、周期は適当に小さくという意味で「(≦12)」の指摘を加えてある。しかし、この範囲に縛られる必要はない。ここでは 12 を指定している。

時系列分析を加えた場合、データの数によっては計算時間がかかる場合があるので、最初に「時系列設定」のボタンをクリックする。「計算が終わりました。」の表示が出たら、以後はその結果を元に計算される。「パネルデータ」ボタンをクリックすると、図 3.12 のように最後の列に時系列分析の予測値が追加される。但し、計算が可能な途中からの挿入となる。プログラムはこの部分を利用して計算をする。

パネルデータ

	機器	機器_1	他指標_1	機器_2	他指標_2	機器_3	他指標_3	機器_ts
22	28	25	1	25	22	21	17	
23	17	28	5	25	1	25	22	
24	39	17	4	28	5	25	1	
25	18	39	15	17	4	28	5	32.405752
26	24	18	9	39	15	17	4	26.560121
27	20	24	13	18	9	39	15	21.476063
28	22	20	23	24	13	18	9	20.140135
29	27	22	25	20	23	24	13	22.462548
30	28	27	16	22	25	20	23	20.830945
31	23	28	15	27	16	22	25	24.610999

図 3.12　時系列分析を加えた計算用データ

重回帰分析では、変数の数が増えると寄与率の値は増加するので、前以上の結果は期待できるが、増加の程度は、元のデータの性質による。例えば周期性が強いデータならば、時系列分析の変数の効果が強く効いてくる。

これ以降の分析は時系列分析を含めない場合と同様であるので、図 3.13 と図 3.14 に交差検証の結果のみを示しておく。データがそろってきた最後の方の数値はよく合っている。

交差検証

	機器	予測値	残差
94	57.000	56.524	0.476
95	62.000	59.821	2.179
96	78.000	78.351	-0.351
97	61.000	63.500	-2.500
98	70.000	72.022	-2.022
99	69.000	71.466	-2.466
100	66.000	67.657	-1.657
R･R^2	0.976	0.952	

図 3.13　時系列分析を加えた交差検証結果

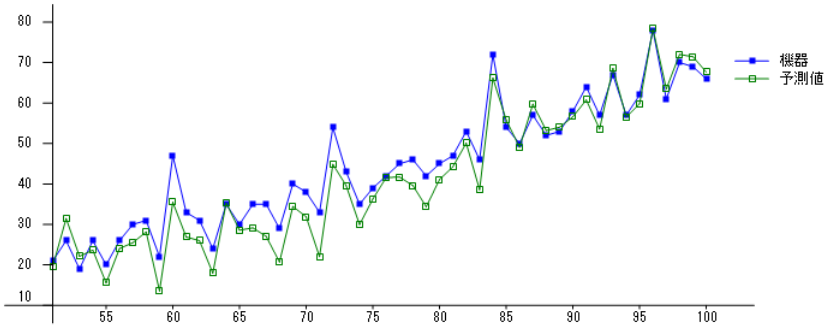


図 3.14　時系列分析を加えた交差検証での実測値と予測値

４．おわりに

局所重回帰分析は予測する関数の形を仮定しない分析であり、利用者の要求点に対する予測値をその都度計算をさせて求める分析である。そのため結果にモデルとしての価値は殆どなく、予測のみが与えられる現実的な手法である。計算自体も非線形最小２乗法のようにニュートン・ラフソン法等の数値解法を用いる必要はなく、ほぼ重回帰分析と同等な計算を実行するだけである。ただ、要求点が多くあるとその数だけ重回帰分析を繰り返すことになり、計算量は増えて行く。

時系列分析の傾向変動の分解では、これまで関数形を仮定したり、移動平均を用いて平滑化を行ってきたりしたが、関数形の仮定はうまくその形になる保証はなく、移動平均は過去の一定期間のデータの平均を予測値とするので、データの変動の傾向を取り入れるには十分でない。その点局所回帰分析は要求点を予測時点にすれば、その近傍のデータの傾向を簡単に予測に取り入れることができる。また、バンド幅の設定によって影響の範囲も調節が可能である。傾向変動の推定には適した手法と言えよう。但し、時系列の各時点で予測値を局所回帰分析によって計算するので、傾向変動の計算にはデータ数分の局所回帰分析の計算が必要である。

我々のパネル重回帰分析では、傾向変動の分解に周期変動の分解を加えて予測値を出し、パネルデータにその値を加えている。一時点分の予測値の計算には、その時点までのデータ個数回の局所回帰分析を実行している（平均はデータ数÷２）。さらに全期間の予測値の推定には、過去から現在までのデータ数分の繰り返しが必要であり、局所回帰分析の計算回数はほぼ、時系列データ個数の２乗÷２となる。これにはある程度計算時間を必要とし、瞬時に結果は求められない。そのため、我々は時系列の計算だけ分離して予め計算しておく方法を採用し、結果の表示時間の短縮化を図っている。

我々のパネル重回帰分析では、データのレコード数は、「変数選択の個数×利用する期間数＋予測する期」に比べて十分大きくないと信頼できる結果は得られない。また、時系列分析を加える場合、レコード数は周期変動の最大周期の２倍より十分大きく取る必要もある。どちらの制約が強くなるかは、状況によるが、変数数 5 個、利用する期間 5 期、予測する期 5 期先と考えると、データの再編成でレコード数が 30 は少なくなる。また、月別データとして 12 ヵ月周期を考えると、最低でも 24 ヵ月以上のレコード数は必要となる。このようなことを考えると、データとしては 50 期以上のデータを利用するのが望ましいだろう。

データ数がこのように制約を受けることから、再編成したデータの中から、さらに必要なデータを選別して利用することを考えてもよいが、これは今後の課題とする。

パネル重回帰分析では、どのようなデータを選ぶかが非常に重要である。プログラムはこれらを選びやすいように作ったつもりであるが、実際に運用してみないと成果のほどは明らかでない。今後、様々な例で利用して、どの程度の結果を出すのか十分検討する必要がある。

謝辞

局所重回帰分析のプログラムの開発に際して、JFE スチール株式会社の茂森弘靖氏にご協力をいただきました。心より感謝致します。

参考文献

[1] W.S.Cleveland and S.J.Delvin, Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, Journal of the American Statistical Association, Vol.83, No.403, (1988) 596-610.

[2] 福井正康, 王迎春, 王晶, 石丸敬二, 社会システム分析のための統合化プログラム 11 －時系列分析－, 福山平成大学経営研究, 6 号, (2010) 81-98.

[3] 北川源四郎, 時系列解析入門, 岩波書店, 2005.

Multi-purpose Program for Social System Analysis 27
- Local and Panel Multiple Regression Analysis -

Masayasu FUKUI, *Tadaaki IWAMURA and Makoto Ozaki
Department of Business Administration, Faculty of Business Administration,
Fukuyama Heisei University
* Former Kawasaki Steel (UFJ Steel) Co., Ltd.

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we create a program of the local multiple regression analysis which is a prediction method that does not assume the shape of the prediction curved surface. Also, we create a program of the panel multiple regression analysis with time-series data of multiple variables. Especially, in the panel multiple regression analysis, we adopt the method adding the prediction value which is predicted by the time series analysis of variation decomposition model. This is considered to be able to take advantage of good part of the two approaches.

Keywords: College Analysis, multivariate analysis, local regression analysis, panel regression analysis, time series analysis
URL: http://www.heisei-u.ac.jp/ba/fukui/