A numeric recipe for Econometrics
part-2

Masato Shimura

2006 年 3 月 16 日

目次

第 1 章

重回帰モデルとその応用

多変数の重回帰モデルの正規方程式は行列で表すと次のようになる。（独立変数が 2 変数の例、単回帰の拡張にすぎない）

$$\begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum Y \\ \sum x_1 Y \\ \sum x_2 Y \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{pmatrix}$$

$$\beta = \begin{pmatrix} a \\ b \end{pmatrix} \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

1.1　対数回帰/The Log Linear Model

線形とは、原因と結果が何らかの意味で比例的であると言うことである（山口　昌哉）

経済データでは対数線型を取り扱うことが多い。これは、

1. 対数にすると変化の量でなく率で表され、非線形データでも多くが対数線型関係になる。。
2. 係数が、elastically を表す。elastically（弾性値）とは、x が 1% 増加したときに、y が何 % 増加するかを示し、単位を取捨した係数で表されるので、都合がよい。(単位の差は、定数項にあらわれる。)

半対数線型は、Y, X のどちらかが、比率や % になっている場合に、その対数を取らずに回帰する場合などに用いられる。

Model	Form	$Slope = \frac{dY}{dX}$	$Elasticaly = \frac{dY}{dX} \cdot \frac{X}{Y}$ 弾性
Linear	$Y = B_1 + B_2 X$	B_2	$B_2 \left(\frac{X}{Y}\right)$
Log-linear 対数線形	$lnY = B_1 + B_2 lnX$ $(Y = B_1 X^{B_2})$	$B_2 \left(\frac{Y}{X}\right)$	B_2
Log-linear 半対数線形	$lnY = B_1 + B_2 X$ $(Y = e^{B_1 + B_2 X})$	$B_2(Y)$	$B_2(X)$
Linear-log 対数線形	$Y = B_1 + B_2 lnX$	$B_2 \left(\frac{1}{X}\right)$	$B_2 \left(\frac{1}{Y}\right)$
Reciprocal 逆数	$Y = B_1 + B_2 \left(\frac{1}{X}\right)$	$-B_2 \left(\frac{1}{X^2}\right)$	$-B_2 \left(\frac{1}{XY}\right)$

Working Example 製造業の生産関数 (コブ · ダグラス生産関数)

1980-2002 の製造業のデータ。総労働時間 (LAB) は就業者数 (単位 1000 万人) に月平均労働時間を、資本稼働率 (CAP) は民間資本ストックに稼働率を掛ける。

```
a1=. |: (1{|: a),(;"1) ,.*/ (L:0) 1 0 1 0 <;.(1)   2 3 4 5 { |: a

'key GDP CAP LAB'  plot (1980+i.23);|: stand  a1
 pd 'eps temp\mori_01.eps'
```

```
   reg2    ^.   2 1 0{"1 a1

1.27786 0.25785 0.7074
   reg2  reg_q_ad  ^.   2 1 0{"1 a1
```

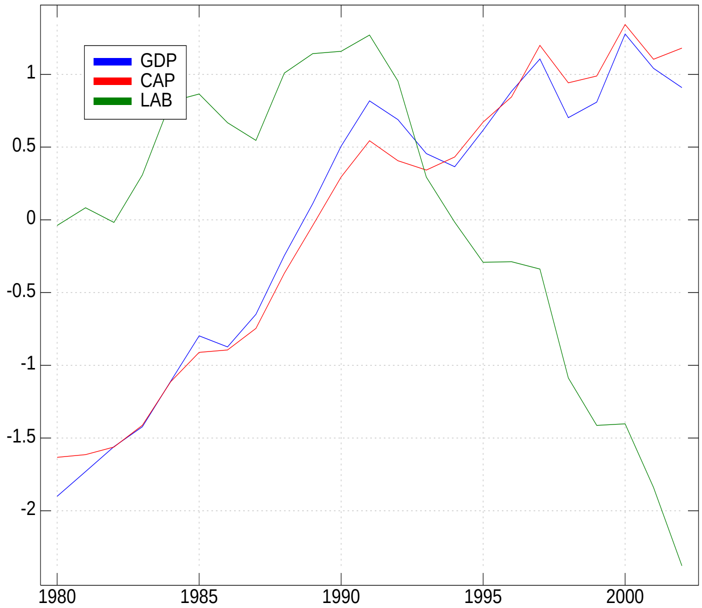

図 1.1　製造業の時系列データ

```
+-----------+----------------------+
|f=         |1.27786 0.25785 0.7074 |
+-----------+----------------------+
|co. of det:|99.3993               |
+-----------+----------------------+
|AIC:       |_188.238              |
+-----------+----------------------+
|t=:        |3.46684 5.43342 53.2258|
+-----------+----------------------+
```

$lnY = 1.27786 + 0.25785 lnLAB + 0.7074 lnCAP$

製造業の GDP は資本財を 1% 増加させれば 0.707% 増え、労働力を 1% 増加させれば 0.25% 増加する。相関係数は非常に高く t 値も十分である。図が示すように、バブル期の過剰設備投資が経済の変調で過剰設備となっており、その調整を労働面に求めていった軌跡がよく示されている。

多変数回帰はこのグラフのように足並みが揃っていないデータのモデル化が可能である。 グラフは多様な数値を一枚に表せるように規準化している。

```
'key Y Y^ ' plot (1980+i.23); |: reg2  reg_Q_ad  ^.  2 1 0{"1 a0
```

```
pd 'eps temp\mori_02.eps'
```

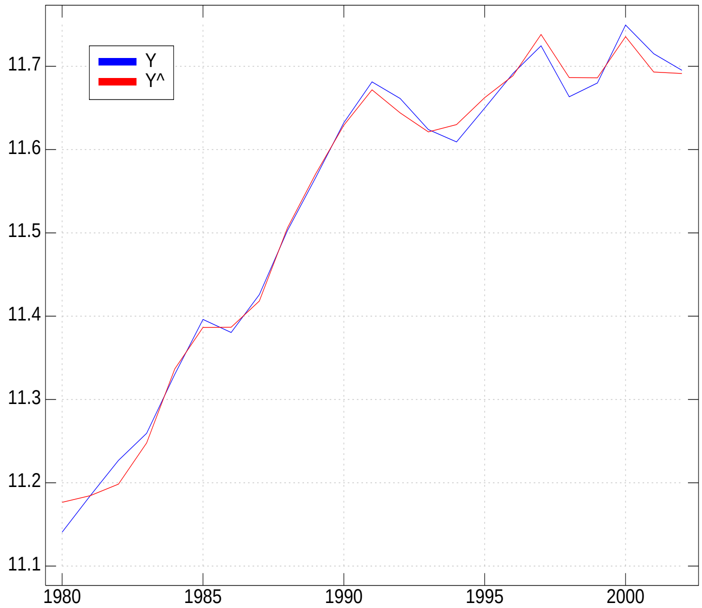

図 1.2　製造業の GDP の実測値と推計値

Year	製造業ＧＤＰ	民間資本ストック	稼働率	就業者	労働時間
80	68921	145910	1.104	1.367	178.3
81	72010	154206	1.053	1.385	177.6
82	75123	162466	1.022	1.38	176.9
83	77606	170363	1.036	1.406	177.9
84	83306	180196	1.096	1.438	180.4
85	88973	192710	1.098	1.453	179.2
86	87589	203003	1.048	1.444	177.8
87	91663	212704	1.049	1.425	178.6
88	99001	225110	1.109	1.454	180.9
89	105492	241073	1.131	1.484	178.9
90	112676	258955	1.143	1.505	176.6
91	118334	280096	1.119	1.55	172.8
92	115995	295824	1.027	1.569	167

93	111750	305737	0.979	1.531	163.2
94	110107	313455	0.975	1.496	163.2
95	114669	322394	1	1.456	164.2
96	119524	331200	1.01	1.445	165.5
97	123571	344180	1.044	1.442	165.2
98	116237	357011	0.956	1.382	162.4
99	118177	362712	0.95	1.345	162.4
00	126691	371963	0.993	1.321	165.5
01	122401	381234	0.925	1.284	164
02	119993	382896	0.935	1.222	164.2

(出典 森棟 P99)

両項	*reg* *reg*1	`reg=: %.1&,.@]` `reg1=: 4 : 'x. %. 1,. y.'`
単項	*reg*2	`reg2=:3 : '({:"1 y.)%. 1,.( }:"1 y.)'`

*1

1.2　モデル選択

1.2.1　AIC によるモデル選択

AIC　(Akaike　Information Criteria) 赤池情報量基準はモデル選択の重要な指標である。

２００３年の暮れに横浜で、AIC 誕生３０周年のシンポジウムが開催された。AIC（A Information Criteria) は統計数理研究所の赤池弘次により創られた理論である。セメントキルンのオートメーション化を図っていた秩父セメントの中川東一郎に協力する過程で、数理理論として最適な制御方程式を曖昧さを残さないで決定するという現実の課題の下で開発され、多くの現場で使い込まれた頑強な手法である。

*1 reg reg2 は関数型定義、 reg1 は明示的定義である。

モデルを概観するには, グラフを描いてみるのが一番良い。

次に, 残差平方和とこれを基準化した決定係数でモデルを計ることが出来る。

重相関モデルには, 変数を増やしていくと多重共線性が生じ, 残差平方和を小さくするモデルが必ずしも真のモデルとはいえないと言うやっかいな性質がある。多重共線性は, いわばデータの派閥であり, その弊害は何事にも共通である。

最尤推定量を求めるには、尤度関数

$LL = logL = \sum_{1=1}^{n} log f(x_i, \theta)$

を最大にすることである。L の対数を取ったものを対数尤度 LL と、最大対数尤度を MLL とあらわす。

対数尤度を θ に関して微分したものを最尤方程式とする。

$\frac{\partial}{\partial\theta} = \sum_{i=1}^{n} log f(x_i, \theta) = 0$

を解いて得られる解の中から LL を最大にする θ を選べばよい。

正規分布の確率密度関数の対数尤度は

$LL = \sum_{i=1}^{n} \frac{(x_i - u)^2}{2\sigma^2} + (u$ を含まない項)

最尤方程式は

$\frac{\partial}{\partial u} LL = \sum_{i=1}^{n} \frac{x_i - u}{\sigma^2} = -\frac{1}{\sigma^2}[\sum x_i - nu] = 0$

解き進めて u に対する最尤推定を求めると

$u = \frac{1}{n} \sum_{i=1}^{n} x_i = M(x)$

σ^2 に対する最尤推定量は

$\sigma^2 = \frac{1}{n}[\sum_{i=1}^{n}(x_i - u)^2] = V(x)$

1.2.2 情報量規準 AIC のアルゴリズム

残差平方和　$Q = (y - Xb)'(y - Xb) = y'y - (y'X)b$

AIC 検定は,AIC の値が小さいほうのモデルの方がより良く当てはまるので, 重相関のモデル選択に有用である。AIC はクライテリアであり, 数値の比較で, 小さい方のモデルの方が優れているという決定を行う。

数値自体には意味はないので, モデル選択時には,AIC の値を添えればよい。*reg_q_ad* は回帰係数, 決定係数,t-value,AIC を同時に計算する。

MLL=-(2/n)log(Q/n)

AIC=: nlog(Q/n)+2(k+1)

1.2.3　AIC によるモデル選択 · 竹内 · 鈴木関数の利用

2005 年 8 月の蓼科でのサマーセミナーで竹内が変数を選択する場合に、多くのクライテリアで判別する方法を提示し、翌月に、鈴木が *AIC* のみを用いる簡単な 2 の関数を示した。

part_1 現実の問題として ESRI(内閣府経済社会総合研究所) が毎月発表する景気動向指数 (Leading Coinsist Lagged) の月次指標の項目が 32 ある。これから重相関に用いる指標を 10 項目程度選びだそうとすると、

$_{32}C_{10}$

10!32 =64512240 の組み合わせが出来る。竹内鈴木関数で選択しなければ収拾がつかないし、計算時間から AIC のみとする。

竹内・鈴木関数のテスト

サンプルデータ (R) から 9 変数を選択する場合の 9 個の組み合わせと AIC 値 (ソート後・最初の 10 個)

竹内・鈴木関数はデータ *X* は横長の形で受け付けるようになっている。縦長のデータを用いるときは (| :) で rotate しておく。

*find_best*0 は AIC の小さい順に並べる。

```
   10{.  find_best0  Y1 compare 9;R
+-------------------------+-------+
|1 1 0 1 0 1 1 1 0 0 1 1 1|209.464|
+-------------------------+-------+
|1 1 0 1 0 1 1 1 1 0 1 0 1|209.542|
+-------------------------+-------+
|1 1 0 1 1 1 1 1 0 0 1 0 1|209.602|
+-------------------------+-------+
|1 1 0 1 0 1 1 1 0 1 1 0 1|209.603|
+-------------------------+-------+
|1 1 1 1 0 1 1 1 0 0 1 0 1|209.614|
+-------------------------+-------+
|1 1 0 1 0 1 0 1 1 0 1 1 1|209.667|
+-------------------------+-------+
|1 1 1 1 0 1 0 1 1 0 1 0 1|209.691|
```

```
+-------------------------+-------+
|1 1 0 1 1 1 0 1 1 0 1 0 1|209.705|
+-------------------------+-------+
|1 1 0 1 0 1 0 1 0 1 1 1 1|209.714|
+-------------------------+-------+
|1 1 0 1 0 1 0 1 1 1 1 0 1|209.718|
+-------------------------+-------+
```

find_best は組み合わせの範囲を指定して、各組毎の、AIC の最小値を表示する。*find_best*0 を見ても、AIC の値は多少うつろうが、並べてみて、0 の谷の深いところを外していけば、大綱では概ね妥当な選択が出来ている。

組み合わせは、${}_mC_n$ の n が大きいと out_of_memory を頻発する。予選リーグを行い、L C LG 毎に相関の少ないものを先に外していくと計算量がずっと少なくなって、妥当な結果が得られる。

7 変数から 12 変数の場合の AIC 最小の組み合わせを選ぶ。

```
   (7;12)  find_best  Y1; R
+--+-------------------------+-------+
|7 |1 1 0 1 0 1 0 1 0 0 1 0 1|209.881|
+--+-------------------------+-------+
|8 |1 1 0 1 0 1 1 1 0 0 1 0 1|209.615|
+--+-------------------------+-------+
|9 |1 1 0 1 0 1 1 1 0 0 1 1 1|209.464|
+--+-------------------------+-------+
|10|1 1 0 1 0 1 1 1 0 1 1 1 1|209.395|
+--+-------------------------+-------+
|11|1 1 0 1 1 1 1 1 0 1 1 1 1|209.36 |
+--+-------------------------+-------+
|12|1 1 0 1 1 1 1 1 1 1 1 1 1|209.357|
+--+-------------------------+-------+
```

1.3 Refelence

Script

```
NB. J.Takeuchi discuss best fit using various criteria
NB. At Tateshina seminor Aug 2005,
NB. G.Suzuki simplified using AIC
NB. next 2 script was written by G.Suzuki Sept/2005
   stat_reg=:3 :0
regb=:[%.1:,.] NB.regression coefficient
regp=:(1:,.])+/ .*regb NB.predicted value
regq=:[:+/[:*:[-regp NB.sum of residuale
regcd=:100"_*1:-regq%[:+/[:*:(]-+/%#)@[
mat=:[:%.(|:+/ .*])@(1:,.]) NB.inverse of data matrix
resvar=:regq%[:-/[:$1:,.] NB.residual varience
regt=:regb%[:%:resvar*[:(<1 0)&|:mat@]
mll=:>:@^.@((o.2)"_*regq%#@[)*#@[%_2: NB.Mll
regaic=:+:@(1:+#@(1:,:]))-2:*mll
'program set of regression model'
)
NB.    stat_reg ''
NB. program set of regression model
NB.    $ R=:A,B,C,D,E,F,G,H,I,J,K,L,:M
NB. 13 87

compare=:4 :0
s=.((>{.y.)=+/"1 t)#t=.#:i.2^#y=.>{:y.
r=.,:(k{s);x.regaic|:(k=.0){u=.s#y
while. k<<:#s
  do. r=.r,(k{s);x.regaic|:(k=.k+1){u
end.
)

NB. hereinafter 2 script is written by M.Shimura
NB. combinient for using in actual
```

```
find_best0=:3 : 0
NB. Usage u. Y1 compare 9;R
TMP0=: y.
(/:  {:"1 TMP0){ TMP0
)

find_best =: 4 : 0
NB. Usage: (7;12)  find_best  Y1; R
Y=: ;{. y.
ZONE=:  ({. ; x.) + i. -/ |. ; x.
ANS=: 0;0
COUNTER=: 0
while. COUNTER < # ZONE do.
X=: (COUNTER { ZONE);;{: y.
TMP1=: {. find_best0 Y compare X
ANS=: ANS,TMP1
COUNTER=. >: COUNTER
end.
({@> ZONE),.((,(# ZONE),2)$  2 }. ANS)
)
```

1.3.1　ダービン・ワトソン検定

自己相関のテストとして、ダービンとワトソンにより創られたダービン・ワトソンテストがある。

系列相関のテスト。系列相関がない場合は, 値が２に近くなる。

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{{}_{t=1}^{n}e_t^2}$$

```
  dw  a0
1.64706
```

Durbin は２００３年に来日した。London School of Economics で教鞭を執っており、

時系列分野でも業績を上げている

1.3.2　t 値

Guiness brewery の化学, 数学の技師であった William Sealey Gosett(1876-1937) が少量のサンプルで出来るビールの品質管理の研究を Student のペンネームで発表したものが $t-distribution$　$t-$分布 である。

$$\frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n}}}$$

標本分散　s^2

Working Example

母集団の平均 u が既知である場合の標本の検定

ある大学の水泳部員の肺活量
3740 3680 3800 4100 3720 3900 3700 4500 3780 3880
一般成人男子の肺活量３７００ｃｃと比べて多いといえるか.
「出典」和達他　キーポイント確率・統計　ｐ１１３

J による計算

```
  3700 test_t DT
   3880
63022.2
2.26739
```

有意水準５%の t 検定の自由度 9 の値＝ 1.833（片側検定なので 0.10（１０%）の値を見る）

帰無仮説は棄却され水泳部員の肺活量は大きいといえる.

経過と説明

dat3 の $\bar{x} = 3880$, s^2=63022.2

$$\frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n}}} = \frac{3880-3700}{\sqrt{63022.2/10}} = 2.267$$

1.3.3　p-value

最近の経済学などでは、t 値と併せて、p 値が記載されることが多い。

p value(i.e. probability value) は t-値から求められる。

p-value は帰無仮説ののとで、検定統計量の値を超える確率を示し、帰無仮説を棄却する最少の有意水準を示す。

D.Gjarati [**?**]p value is defined as the lowest significance
level of at which a null hypothesis can rejected

例えば、自由度 (df) ８の t-value が 5.86 であった場合、t-table の α の値は 0.0010 では 5.041 である。　$tcdf$ で求めた値は 5.041 よりも大きく α は 0.001 より小さくなる。ｔ値から計算すると、ｐ値は 0.000189233 でエラーの確率は概ね 0.02 %、$\frac{2}{10000}$ であるといえる。

```
      8 tcdf 5.86
0.999811

  1- 8 tcdf 5.86
0.000189233   NB. p-value
```

[**?**]Issue:Asano and Nakamura table1-3

1.4　トレンド変数をを入れる

浅野 · 中村の分析 (為替レートの変化と CPI) をフォローする。

EXR Exchange rate(1970-1992)

CPI Consumer Price Index(Same)

nr 1970 を１としたトレンド変数

トレンド変数やダミー変数には ln は付けない。従って,$reg0$ を用いる。

$reg0$ のデータ形式は,X,.Y であり, データマトリクスから選択の後, 所要の変換を行う。この部分は,handmade である。

```
'key EXC CPI WAGE' plot ({."1 asa); |:   stand  }."1   asa
   pd 'eps temp_asamo_01.eps'
```

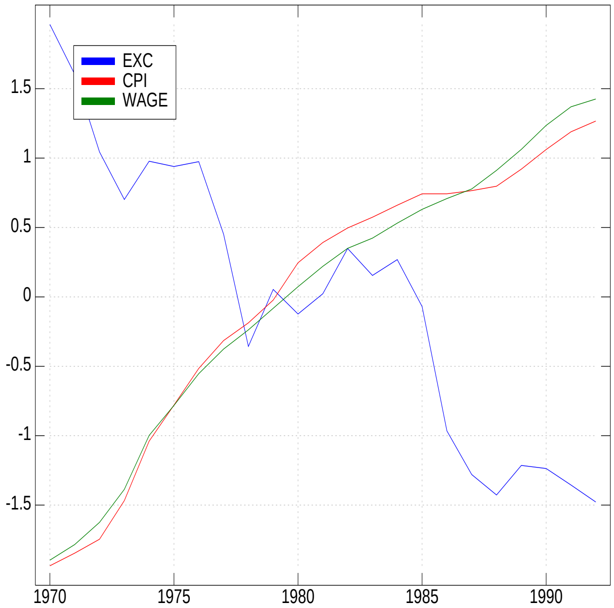

図 1.3　EXC CPI WAGE1

```
 reg2  (1{"1 asa) ,.(>:i.# asa),.2{"1 asa
26.4252 0.0381768 3.5465
```

asa	Exchange	CPI	Wage
1970	360	35.1	19.2
1971	335.47	37.1	21.8
1972	297.23	39.3	25.5
1973	273.8	45.4	31
1974	292.64	54.8	40
1975	290.04	60.5	45
1976	292.43	66.3	50.3
1977	256.74	70.7	54.4
1978	201.44	73.5	57.6
1979	229.5	77.1	61.2
1980	217.4	83	64.8
1981	227.38	86.2	68.2
1982	249.66	88.5	71.2
1983	236.39	90.2	72.9
1984	244.19	92.1	75.4
1985	221.09	93.9	77.7
1986	159.83	93.9	79.5
1987	138.33	94.4	81.1
1988	128.27	95.1	84.2
1989	142.82	97.8	87.7
1990	141.3	100.9	91.7
1991	133.18	103.7	94.8
1992	124.8	105.4	96.1

次のようにコントロールとしてトレンド変数を挿入する。

```
      dat
 EXR    nr  CPI
   360  1  35.1
335.47  2  37.1
297.23  3  39.3
 273.8  4  45.4

 141.3 21 100.9
133.18 22 103.7
 124.8 23 105.4
```

重相関の場合, データの組み合わせと, 対数や逆数の選び方は, 幾通りもある。その都度回帰プログラムを用意してもよいが, 腕の良い料理人は, 一本の包丁で多くの料理をこなすように, *reg*2 一本を用いて、データの方をを変型させる。

為替とトレンド変数の回帰

プレーンな重回帰

```
   reg2  (1{"1 asa) ,.(>:i.# asa),.2{"1 asa
26.4252 0.0381768 3.5465
```

$CPI = 26.4252 + 0.0381768EXR + 3.5465YEAR$

Reciprocal

```
   reg2  (%1{"1 asa) ,.(>:i.# asa),.2{"1 asa
51.4814 _5487.67 4.42389
```

$CPI = 51.4814 - 5487.67\frac{1}{EXR} + 4.42389YEAR$

対数回帰

トレンド変数は対数を用いない。

```
   reg2  (^.1{"1 asa) ,.(>:i.# asa),.^.2{"1 asa
1.50601 0.377514 0.0639988
```

$lnCPI = 1.50601 + 0.377514lnEXR + 0.0639988YEAR$

次に,*reg_q_ad* を適用して, 決定係数,AIC,t 値を見てみよう。
[*2]

```
   reg2 reg_q_ad (1{"1 asa) ,.(>:i.# asa),.2{"1 asa
+-----------+-----------------------+
|f=         |26.4252 0.0381768 3.5465|
+-----------+-----------------------+
|co. of det:|92.5341                |
+-----------+-----------------------+
|AIC:       |88.3799                |
+-----------+-----------------------+
|t=:        |1.3826 0.689366 6.21416 |
+-----------+-----------------------+
   reg2 reg_q_ad (%1{"1 asa) ,.(>:i.# asa),.2{"1 asa
+-----------+-----------------------+
```

[*2] *reg_q_ad* と副詞にした. 副詞は関数 (動詞) を左引数にとる。即ち回帰関数を色々と変えて, 引数とできるので回帰係数毎に *reg_q_ad* を作成する必要はない。

```
|f=         |51.4814 _5487.67 4.42389|
+-----------+------------------------+
|co. of det:|95.478                  |
+-----------+------------------------+
|AIC:       |76.8481                 |
+-----------+------------------------+
|t=:        |13.2416 _3.71547 11.9488|
+-----------+------------------------+
```

```
   reg2 reg_q_ad (^.1{"1 asa) ,.(>:i.# asa),.^.2{"1 asa
+-----------+--------------------------+
|f=         |1.50601 0.377514 0.0639988|
+-----------+--------------------------+
|co. of det:|86.6931                   |
+-----------+--------------------------+
|AIC:       |_90.3275                  |
+-----------+--------------------------+
|t=:        |1.12053 1.66249 5.7685    |
+-----------+--------------------------+
```

2 番目の回帰式 (逆数) のトレンドの係数から, 期間中 EXR が一定ならば,4.42point/Year 国内物価が上昇したことを示している。また, 円高は, 国内物価を押し下げる効果があること。為替の限界効果は, $-5487.67 \div EXR^2$ で次のようになり, 円高の方が限界効果が大きい。

```
    e,. _5487.67 % e^2
 90  _0.67749
100 _0.548767
110 _0.453526
120 _0.381088
130 _0.324714
140 _0.279983
```

推計

reg_Q_ad も汎用になるように副詞（関数を引数に取る関数）タイプとした。

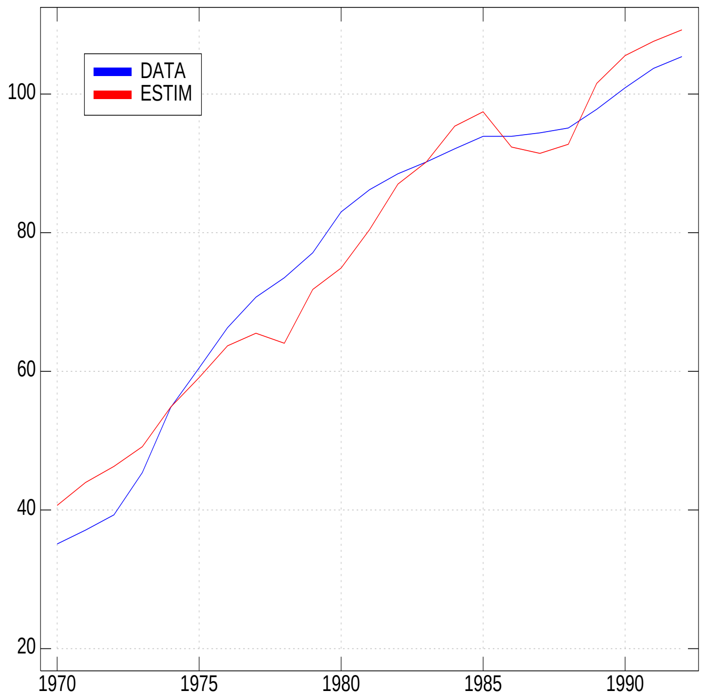

図 1.4　Estimate Reciplocal

1.5　ダミー変数を用いたモデル

パラメーターの構造変化（バブル, リセッション）や数量化できない質的属性（災害, 戦争）などを含んだデータの構造変化や特性を考慮し, ダミーを付けることにより, 回帰方程式を何本かに分けることができる。

月データで, 四季や繁忙期を分けるのにも用いられる。

ダミーデータを用いた場合も, 重相関と同じように回帰できる。

四季のダミーを入れる	*mk_dummys*	*x. mk_dummy4 y.* 0 Spring 1 Summer 2 Autumn 3 Winter
月のダミーを入れる	mk_dummy12	8 1 2 *mk_dummy*12 y.

Working Example Canada の宝石売り上げのシーズンデータ。クリスマスシーズンが突出している。

```
36 44 45 106 38 46 47 112 42 49 48 118 42 50 51 118
```

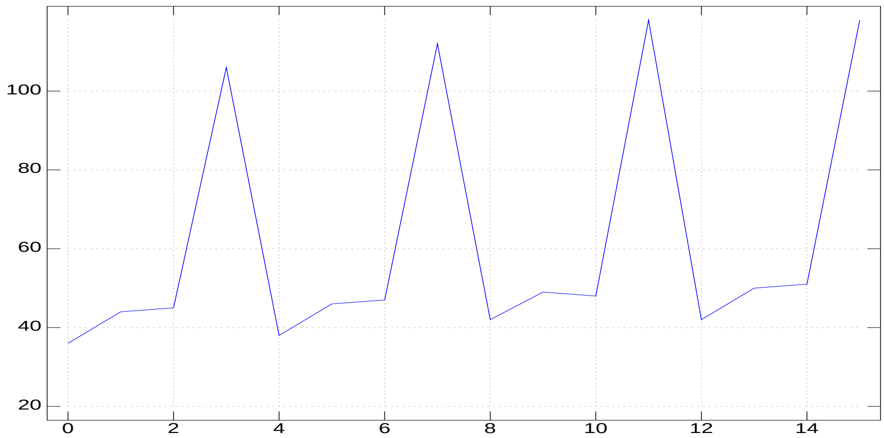

図 1.5　宝石の売り上げ

1.5.1　ダミー変数を作る

```
1 2 3 mk_dummy4 W2
 0 0 0 0   36
 1 0 0 1   44
 2 0 1 0   45
 3 1 0 0  106
 4 0 0 0   38
```

```
 5 0 0 1  46
 6 0 1 0  47
 7 1 0 0 112
 8 0 0 0  42
 9 0 0 1  49
10 0 1 0  48
11 1 0 0 118
12 0 0 0  42
13 0 0 1  50
14 0 1 0  51
15 1 0 0 118
```

指定した 4 半期を分けて, フラグを立てる.
四半期は０　１　２　３　で会計区分や四季（春夏秋冬）に対応する。

```
 |: 1 2 3 =/ 0 1 2 3
0 0 0
1 0 0
0 1 0
0 0 1
```

X データにサイズを合わせ,X と結合する.

```
   regx 1 2 3 mk_dummy4 DAT
35.6 0.65 72.05 6.95 7.1

 linefit_reg   1 2 3 mk_dummy4 DAT
   pd 'eps temp\wonacotto_1.eps'
```

*3

*3 オリジンを０としているので, 教科書と回帰係数が異なることがある。

$$f = 35.6 + 0.65Q_1 + 72.05Q_2 + 6.95Q_3 + 7.1Q_4$$

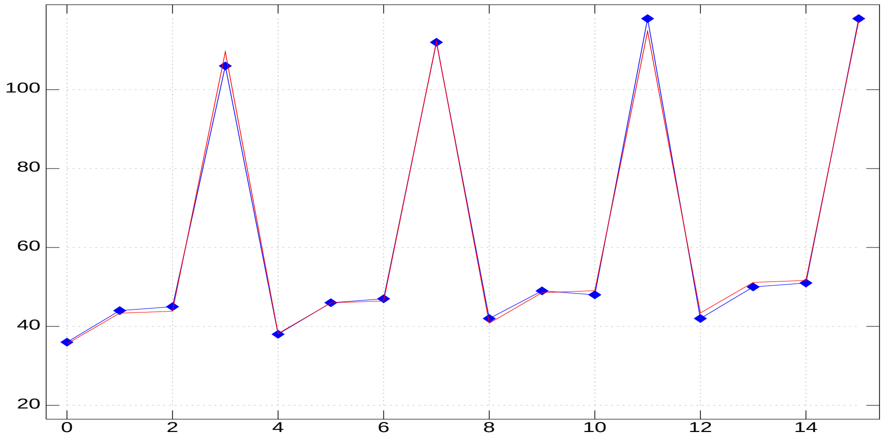

図 1.6　宝石の売り上げと推計値

クリスマスシーズンのみが突出しているので、3 のみにタグを付けても良い。

```
 |: 3 mk_dummy0 W2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
0 0 0 1 0 0 0 1 0 0  0  1  0  0  0  1

        regx   3 mk_dummy4 DAT
39.69 0.734756 67.1972

linefit_reg    3 mk_dummy4 DAT
   pd 'eps temp\wonacotto_2.eps'
```

b

1.5.2　月次 data の tag

Working Example　東北自動車道栃木、福島県境付近の下り線の交通量 (1996/4-2006/1) 最新のデータから 12 の倍数を取り、月平均した上で、月別の構成をあらわしている。次のダミーを付ける参考となる。何パーセントかの指標を決めれば自動でタグを付けることもできる。

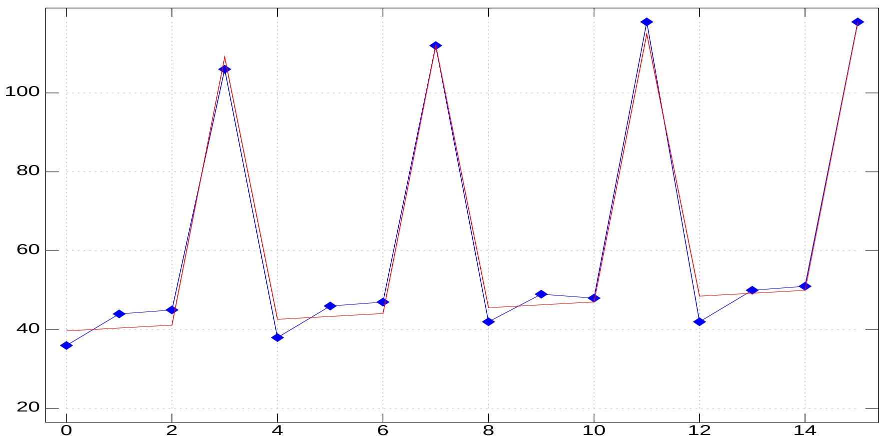

図 1.7　宝石の売り上げと推計値

ave_year a1	8 1 2 mk_dummy12 a1
NB. Nasu traffic	
nr month average	

```
 ave_year a1
 NB. Nasu traffic
nr month  average
-----------------
   ave_year a1
 0   4   99.6
 1   5  105.2
 2   6   91.3
 3   7  102.4
 4   8  164.3
 5   9  100.8
 6  10  105.5
 7  11   97.2
 8  12   90.9
 9   1   72.3
10   2     77
11   3   93.5
```

```
 8  1  2  mk_dummy12  a1
--------------------
  0  0  0  0  17256
  1  0  0  0  17953
  2  0  0  0  15911
  3  0  0  0  17306
  4  1  0  0  29298
  5  0  0  0  17955
  6  0  0  0  17682
  7  0  0  0  16949
  8  0  0  0  16749
  9  0  1  0  13284
 10  0  0  1  13891
 11  0  0  0  17358
```

```
 reg0   8  1  2  mk_dummy12  a1
17330.4 _19.3525 10805.5 _4215.43 _3552.77
```

```
  linefit_reg 8 1 2 mk_dummy12 ,a1
 pd ' eps temp\reg_month.eps'
```

$17330.4 -_1 9.3525x + 10805.5AUg - 4215.43Jan - 3552.77Feb$

月次のデータの対数を取ると、回帰係数に、月当たりの増減率が現れる。１２０乗すれば１０年間の増減となる。

```
   reg0  8 1 2 mk_dummy12 ^. a1  NB. ^. is log
9.76025 _0.00119763 0.512065 _0.302989 _0.247201

   (>:_0.00119763)^120
0.866059                        NB. decrease 14%
```

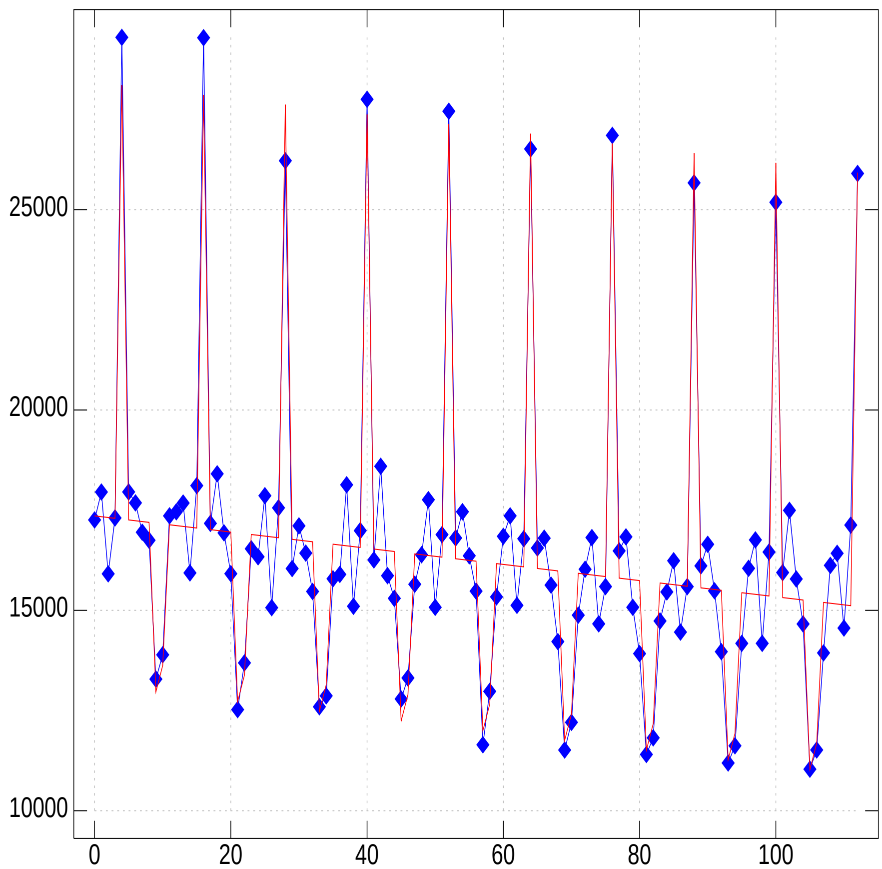

図 1.8　8,1,2 月 dummy

1.6　ラグ変数を用いる

第 2 章

季節変動とスムージング

2.1　移動平均 ·Moving Average

2.1.1　単純移動平均

mav	Moving Average	m. mav n.
mav_c	Centred mav auto classified	m. mav_c n.
*mav*12	12 month mav centred	mav12 n.
vmav	moving average for vertical data	m. vmav n.
*vmav*12	12 month vmav	vmav12 n.
		m. 右引数＝次数 n. 左引数＝ Data

J は、移動平均を次の簡潔なイディオムで表現できる。

```
mav=: +/\ % [
```

これは

```
mav=:(# %~ +/)\
```

でも表現できる。

経過と解説

J の infix（\) の機能は強力であり左引数（x.）で指定した数での移動平均のためのずらした組み合わせを簡単に作る。

```
  ] s=. 10?.30
   6 27 9 16 28 19 26 10 20 14

3<\s
+------+-------+------+-------+-------+-------+-------+--------+
|6 27 9|27 9 16|9 16 28|16 28 19|28 19 26|19 26 10|26 10 20|10 20 14|
+------+-------+------+-------+-------+-------+-------+--------+

   3+/\s
42 52 53 63 73 55 56 44
```

これを $n = 3$ で割れば単純移動平均が求められる。

```
   3 mav s
14 17.3333 17.6667 21 24.3333 18.3333 18.6667 14.6667
```

中心化

移動平均は,x.（左パラメーター) が偶数の場合には, 原データとの対比のために中心化がよく行われる。通常は奇数、偶数を自動判別するこの *mav_c* 使えばよい。

```
mav=:+/\ % [
mav_c=: 2&mav@mav ‘ mav @. (2&|@[)   NB.centred MA

   s=.10?.20

  a , (2 mav a),: 2 mav_c a

 6   3     19     15   10    14    0    7    12  17
 4.5  11     17   12.5   12     7  3.5  9.5  14.5   0
7.75  14  14.75  12.25  9.5  5.25  6.5   12     0   0
```

[J **の文法**]

Infix(\) の左には Box(<) を持ってくることが多いが、ダイレクトに計算式を入れてもよく、簡略化できる。

~は *reflex* で fork に合わせている。

`@.` 以下の条件式で偶数奇数の判定を行い、偶数の場合は, 中心化を行っている。

?.　乱数の生成 (?)　?. はシードを固定した乱数でいつも同じ乱数が生成される。

2.1.2　12 ヶ月移動平均

*mav*12 は *mav* の引数を１２に固定して組こみ中心化を行っている。

```
mav=: +/\ % [
NB. Season adjustment using 12 month

mav12=: [: 2&mav 12&mav

      2<\ 12 mav i. 24
+-------+-------+-------+-------+--------+---------+---------+---------+
|5.5 6.5|6.5 7.5|7.5 8.5|8.5 9.5|9.5 10.5|10.5 11.5|11.5 12.5|12.5 13.5|
+-------+-------+-------+-------+--------+---------+---------+---------+...
   mav12 i. 24
6 7 8 9 10 11 12 13 14 15 16 17
```

2.1.3　加重移動平均

wmav	3 month weighted moving average	wmav n.
*mav_w*13	13 項 weighted moving average	mav_w13 n.
spencer	spencer's 15 point moving average	spencer n.
henderson	henderson n point moving average	m.(次数) henderson n. 次数は奇数を指定

一次

３ヶ月の移動平均値を更に移動平均する場合。一層のスムージングがはかれる。

$$= \frac{\frac{Q_{t-2}+Q_{t-1}+Q_t}{3}\ \frac{Q_{t-1}+Q_t+Q_{t+1}}{3}\ \frac{Q_t+Q_{t+1}+q_{t+2}}{3}}{3}$$
$$= \frac{1}{9}Q_{t-2} + \frac{2}{9}Q_{t-1} + \frac{3}{9}Q_t + \frac{2}{9}Q_{t+1} + \frac{1}{9}Q_{t+2}$$
$$= \frac{1}{9}[1 \quad 2 \quad 3 \quad 2 \quad 1]$$

これは、上記のウエイトによる５ヶ月の加重移動平均と同じである。四季の変動の特徴をよく捕らえることができる。

補完

加重 3 ケ月移動平均の両端値の補完ウェイトとして, 同じ方法によるとウェイトは次のとおりとなる.

$$\mathrm{Q}_{t+1} = \frac{1}{9}Q_{t+3} \to \frac{1}{9}\left(\frac{Q_t+Q_{t+1}+Q_{t+2}}{3}\right)$$
$$\mathrm{Q}_{t+1} = \frac{3 \quad 7 \quad 10 \quad 7}{27}$$
$$\mathrm{Q}_{t+2} = \frac{2}{9}Q_{t+3} \to \frac{2}{9}\left(\frac{Q_t+Q_{t+1}+Q_{t+2}}{3}\right) + \frac{1}{9}\left(\frac{Q_t+Q_{t+1}+Q_{t+2}}{3}\right)$$
$$\mathrm{Q}_{t+2} = \frac{2 \quad 3 \quad 4}{9}$$

cpl_mavw3 は両端の補完を行う。plot_w3 は cpl_mavw3 を組み込んだ.

に近似させて、原系列の傾向、変動要素を抽出している。

2.1.4　Spencer 移動平均

スペンサー１５ポイント移動平均はセンサス局法 X-10 で用いられた。
ウエイトを次のように置いており、隣接に重心をおくので、三角形や低位の数値が破壊されないでよく残っている。

これは、次のウエイトに依っている。

$$\frac{1}{4}\left[\ -3 \quad 3 \quad 4 \quad 3 \quad -3\ \right]$$

$$\frac{1}{5}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
$$\frac{1}{4}\begin{bmatrix} -3 & 3 & 4 & 3 & -3 \end{bmatrix}$$
$$\frac{1}{4}\begin{bmatrix} -3 & 3 & 4 & 3 & -3 \end{bmatrix}$$

これは、実際には、次のウエイトになる。 **[?]** **[?]**

Spencer のウエイトは、15 項の移動平均に限定されており、加重１５項移動平均のパラメーターと一致する。

$$a_0 = \frac{74}{320}, a_{\pm1} = \frac{67}{320}, a_{\pm2} = \frac{46}{320}, a_{\pm3} = \frac{21}{320},$$
$$a_{\pm4} = \frac{3}{320}, a_{\pm5} = \frac{-5}{320}, a_{\pm6} = \frac{-6}{320}, a_{\pm7} = \frac{-3}{320}$$

```
wts=: (|.,}.) 74 67 46 21 3 _5 _6 _3%320
```

Script

```
NB. ===============================
NB. C.Reiter Fractals Visuarization and J (2nd Edition 2000)
NB. ==============================
cos=: 2&o.
randunif=: (?%<:)@:($&2147483647) : ({.@[+({:-{.)@[*$:@])
randsn=: cos@+:@o.@randunif * %:@-@+:@^.@randunif

wts=: (|.,}.) 74 67 46 21 3 _5 _6 _3%320
locspen=: (+/ . *)&wts
spencer=: 15"_ locspen\ (7: # {.),] ,7:#{:
walk=: +/\ randsn 100
```

2.1.5 Henderson 移動平均

センサス局法 X-11,X-12 の採用する季節調整法に採り入れられている手法で、基本的には、全変動を、Trendcycle(TC)、季節変動 (S)、不規則変動 (I) の３つの成分に分けると

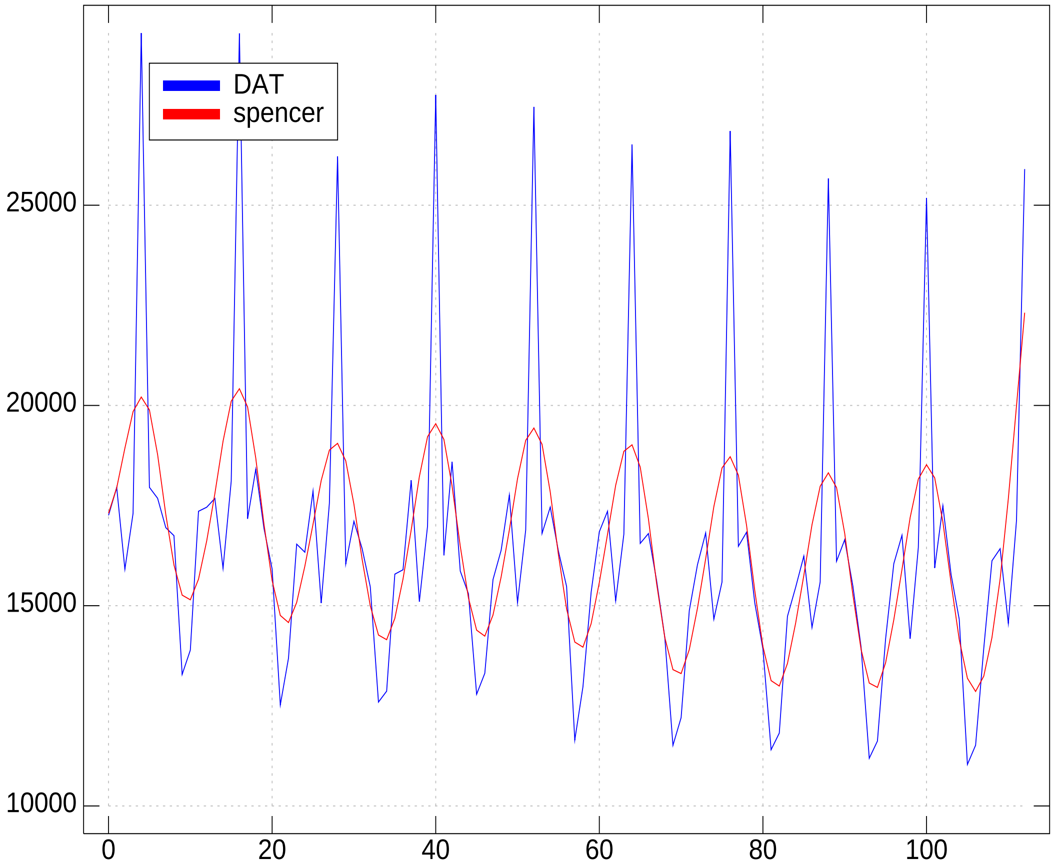

図 2.1　東北道那須付近の月平均交通量

きの、TC を求めるのに使用される重み付き移動平均法である。

$$H^n(B) = \sum_{i=-(n-1)/2}^{(n-2)/2} h_i^{(n)} B^i, \quad i = 0, \pm 1, \pm 2, \cdots, \pm(n-1)/2$$

B , 時間の遅れ、　F、時間の前のパラメータ

B is B(back)

F is F(Foward)

$\mathrm{h}_i^{(n)}$重み係数　$m = (n+3)/2$

$$\mathrm{h}_i^{(n)} = \frac{315((m-1)^2 - i^2)(m^2 - i^2)((m+1)^2 - i^2)((3m^2 - 16) - 11i^2)}{8m(m^2-1)(4m^2-1)(4m^2-9)(4m^2-25)}$$

5 項のウエイト

$$\mathrm{H}^{(5)}(B) = -\frac{21}{286}B^2 + \frac{84}{286}B + \frac{160}{286} + \frac{84}{286}F - \frac{21}{286}F^2$$

7 項のウエイト

$$\mathrm{H}^{(7)}(B) = -\frac{42}{715}B^3 + \frac{42}{715}B^2 + \frac{210}{715}B + \frac{295}{715} + \frac{210}{715}F + \frac{42}{715}F^2 - \frac{42}{715}F^3$$

Spencer や Henderson の移動平均は、原系列を追いながら滑らかさを実現していること

ら項では滑らかになる。Spencer

ような滑らかさを持っている。

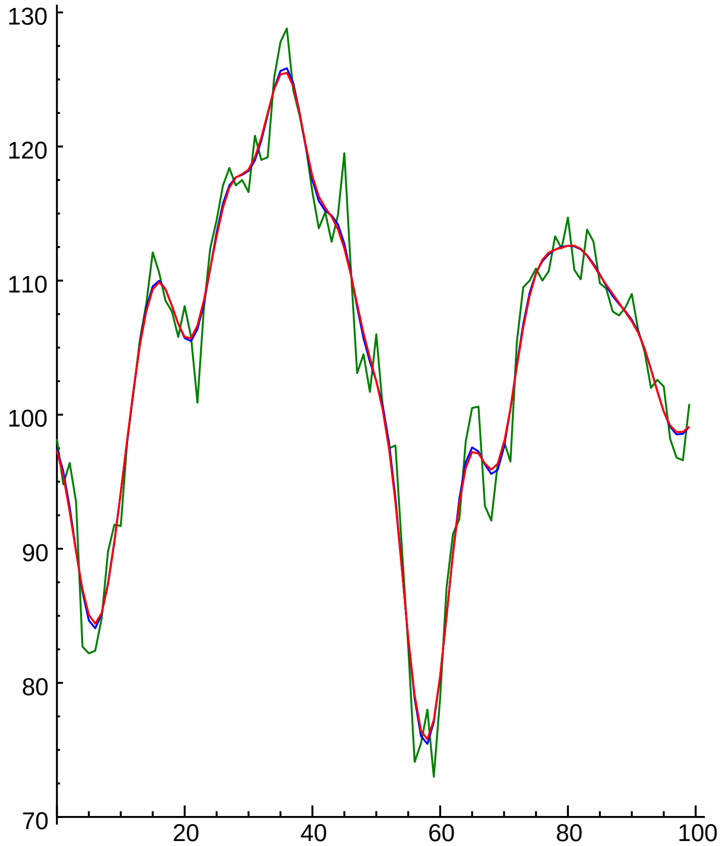

図 2.2　Stock Price of IBM NY Market
青=Spencer, 赤=Henderson

IBM Stock Price of NY Market 最初の１００ポイントを Spencer と Henderson の１５ポイントで plot してみると、 ほとんど線が重なる。

henderson は項数が自由に取れる (ただし奇数のみ)。ここでは 23 項を取ったので spencer よりは滑らかである。　交通量やデパートの売り上げのデータでは、８月や 12 月のピークが大きく、高次になるほど、追従より滑らかさを　求める。保守的 (Concerbative) であるといわれる所以である。

```
'key DATA henderson' plot |:  a,. 23 henderson ,a
   pd 'eps temp\henderson.eps'
```

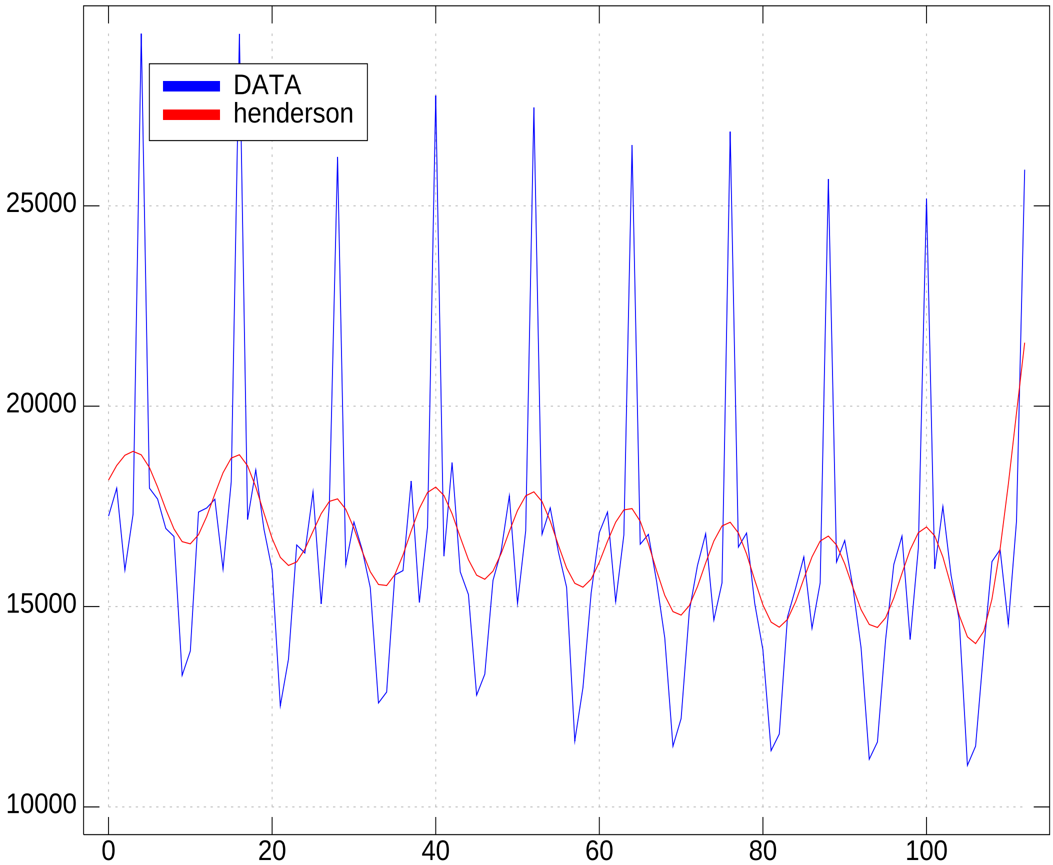

図 2.3　東北道那須付近の月平均交通量

Script

スクリプトは竹内による。[?]　henderson の移動平均の項数は任意であり、１３－１５項では Spencer と極めて類似した結果となる。

このスクリプトは、x. の項数は奇数を指定する。偶数の場合は、切り上げか切り捨てるようにスクリプトに手を加えないとエラーになる。

```
    henderson=:4 : 0
NB. Usage: i.e. 15&henderson y. //or 7 henderson y.
wt=: x. hnwgt i:-:<:x.
x. (+/ . *)&wt\((<.@-:x.)#{.y.),y.,(<.@-:x.)#{:y.
```

```
)
   hnwgt=:4 : 0
m=: -: x. +3 NB. (n+3)/2
nm=: 315*(((m-1)^2)-*:y.)*((m^2)-*:y.)*(((m+1)^2)-*:y.)*(((3**:m)-16)-11**:y.)
hi=:nm%8*m*((m^2)-1)*((4**:m)-1)*((4**:m)-9)*((4**:m)-25)
)
```

補項

移動平均は前後に欠項が出るが、この欠項を補完する方法として、次のような方法が用いられている。

簡易法として、$2P+1$ 項の移動平均を行う場合に、前後に各 p 項の平均値を連結して移動平均を行うと欠項が埋められる。先に移動平均を行った後に、p 項の平均値を連結する方法もある。spencer,henderson はこの方法に依っている。

反復移動平均法と言われる加重移動平均法の場合は、欠項の部分に、短縮加重として、残った部分の合計に、残ったパラメーターを適用する方法である。

３項加重移動平均の cpl はこの方法に依っている。

ここで、１２月中心化移動平均の欠項補完を考えてみる。中心化１２ヶ月移動平均は、`0.5 1 1 1 1 1 1 1 1 1 1 1 0.5` のウエイトの１３項加重移動平均と同じである。

前後に、６項づつの欠項が出るので、前後のデータを１２から順次減じ、７項まで取る。

```
      0 1 2 3 4 5 6 7 8 9 0 10 11 12 13 14
                    *
    7 8 9 10 11 12 13
```

7 8 9 10 11 12 と６項の欠損を取り、ｎで割る。残ったウエイトをそのままにしておくと結果が過大になるので、補項にはウエイトを乗せないこととすると厳密性に欠ける。

比較のために、加重１３項移動平均に短縮加重を行う *mav_w*13 を作成した。グラフでは近似している。US.Census 局の *reg_arima* などの様に、時系列で補完するのでなく、短縮補完する方法は、補完部分の値は参考程度にとどめた方がよい。

移動平均の PLOT

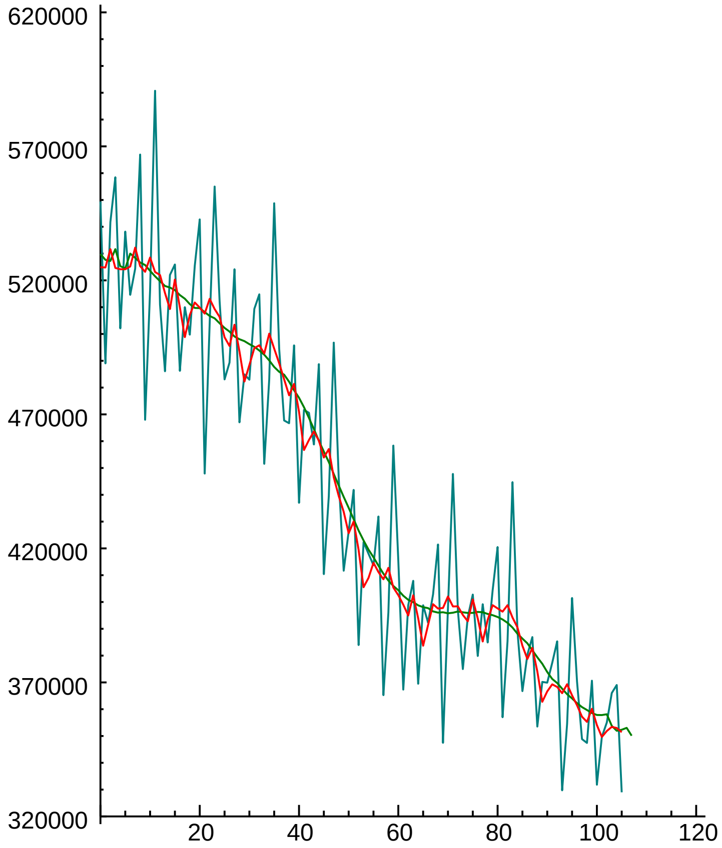

図 2.4　plot_w13

plot_wmav	weighted mav	plot_wmav n.
*plot_mav*12	12 months mav	plot_mav12 n.
*plot_mav*12*cpl*	12 months mav	plot_mav12cpl n.
*plot_w*13	13 weighted mav and mav12cpl	plot_w13 n.
plot_spencer	spencer	plot_spencer n.
plot_henderson	henderson 15 項	plot_henderson n.
*plot_henderson*2	henderson 任意の項 and spencer	m. plot_henderson2 n.
plot_comb_all	henderson 任意の項と and spencer ,mav12cpl	m. plot_comb_all n. henderson は次数は奇数のみ 31 33 付近は中心化１２月移動平均と類似　両端は計算方法の差によるばらつきがある。

移動平均を求め原データと同時にグラフにあらわす関数を用意した.
・任意の次数（例では３＝四半期）と原系列を plot する
偶数の場合はグラフにあらわすため中心化の処理を行っている。

2.2　12 カ月のトレンドと deseasonalize

saj	季節変動指数を求める.whole months	saj n.(List)
$saj2$	季節変動指数を求める。last 3 years	saj2 n.(List),usualy using saj2
adj_season	季節変動指数を求め、回帰と Best Fitting を行う	adj_season n.(List)
$adj\_season2$	季節変動指数を求め、回帰と Best Fitting を行う.saj2	adj_season2 n.(List),usualy using adj_season2
$plot\_adj\_season$		
$plot\_adj\_poly$		

2.2.1　12 月の season

移動平均法により季節変動指数を求め、季節変動を除去することで、トレンドをもとめる。

移動平均後 5 年分取るには (12 月では移動平均で最初に 1 年分失われる)6 年必要であり、5 年分を 12 月ずつ並べ、最大値と最小値をカットして 3 年分の平均を求めて季節変動指数とする。$saj2$ は最新のデータから 12 の倍数を取るようにしてある。

```
(4 5 6 7 8 9 10 11 12 1 2 3),.  12 saj2 a1
NB. 東北自動車道栃木ー福島県境付近の交通量の季節変動指数
 4  1.00729
 5  1.05477
 6  0.916349
 7  1.00835
```

```
 8   1.65713
 9   1.03282
10   1.05852
11   0.973519
12   0.888617
 1   0.72057
 2   0.754038
 3   0.928036
```

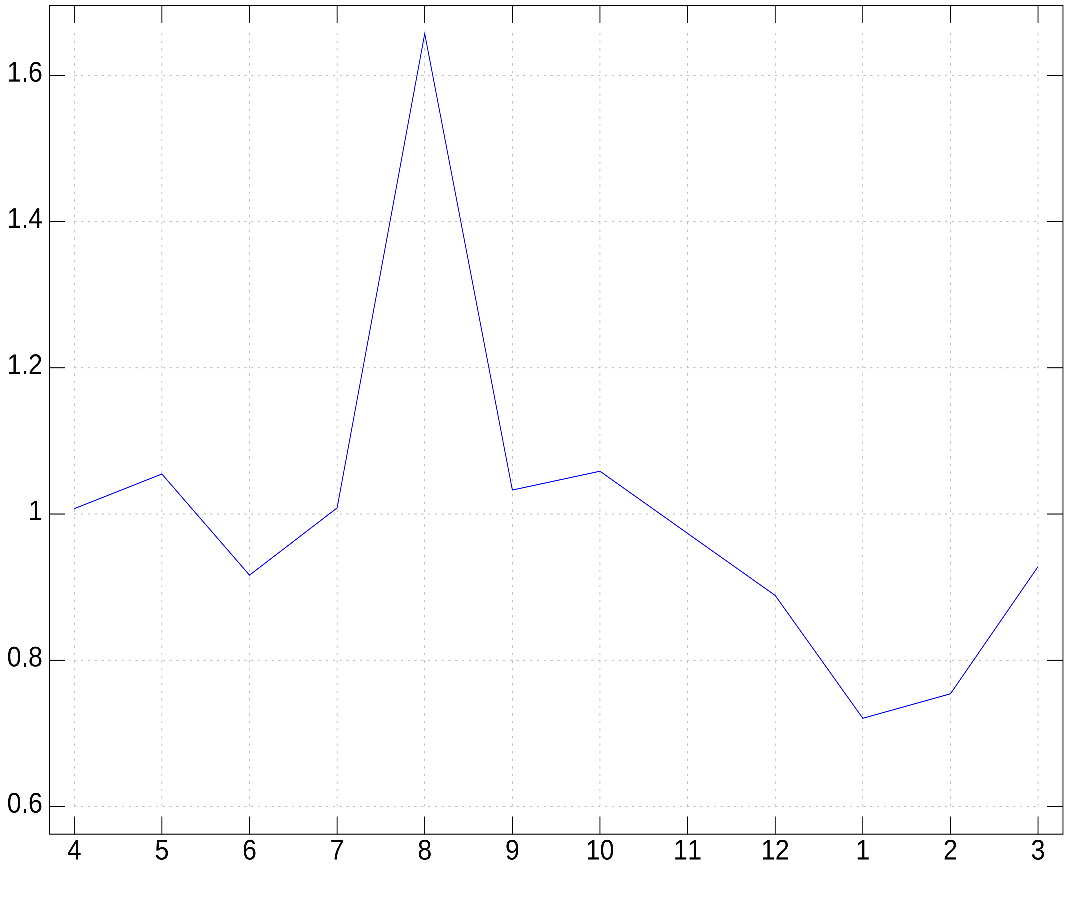

図 2.5　東北道那須付近の交通量の季節変動

deseason したデータを OLS や低次の多項式で回帰するとトレンドが求まる。$adj\_season2$ は OLS で deseason とトレンドを求める。

```
12{. 12 adj_season2 a1
```

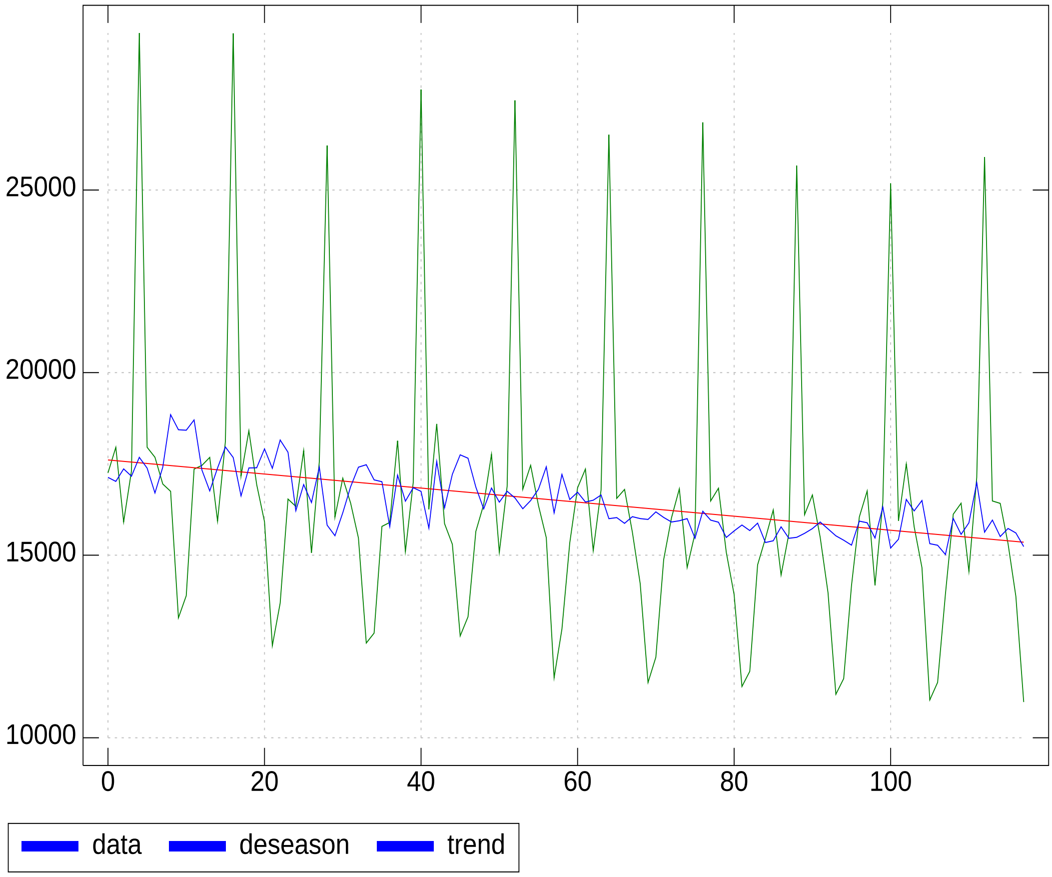

図 2.6　東北道那須付近の交通量の deseason と trend

```
data   deseason trend
-----------------------
17256 17131.1 17607.9
17953 17020.8 17588.6
15911 17363.5 17569.4
17306 17162.7 17550.1
29298   17680 17530.9
17955 17384.5 17511.6
17682 16704.4 17492.4
16949   17410 17473.2
16749 18848.4 17453.9
13284 18435.4 17434.7
13891 18422.1 17415.4
17358   18704 17396.2
```

トレンドは低次の多項式の方が滑らかになる。多項式は両端が次数により上下するので注意を要する。

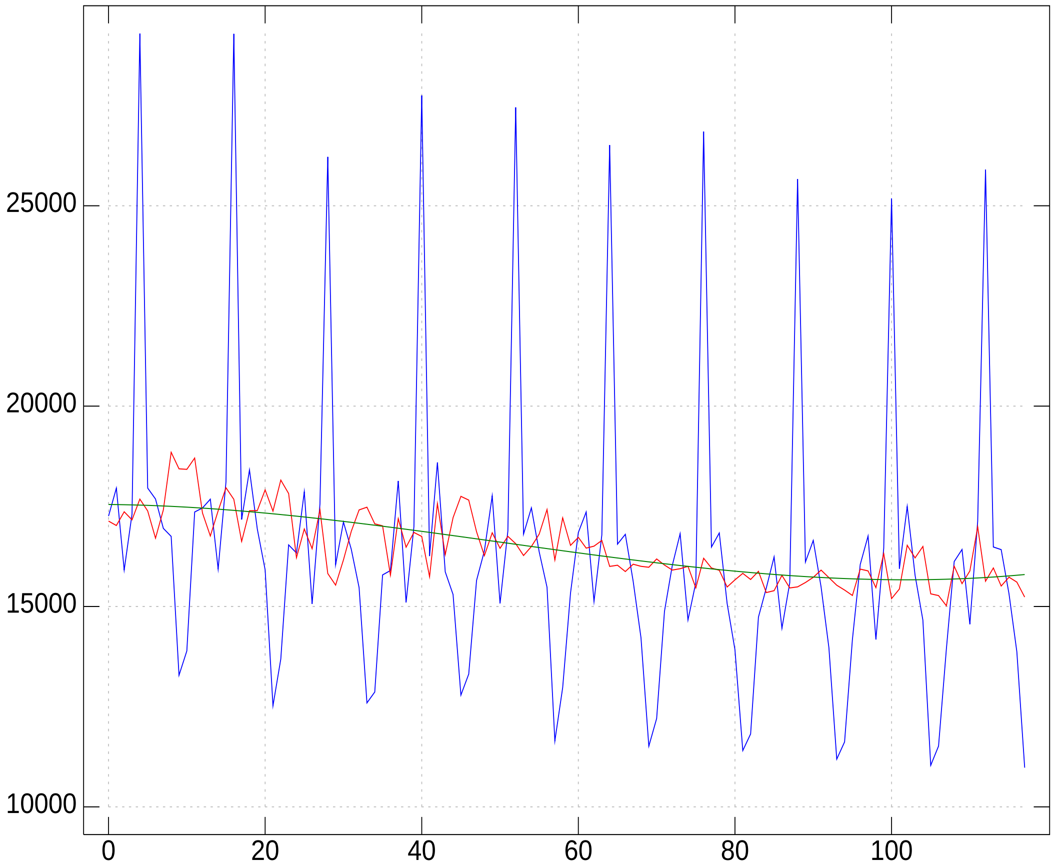

図 2.7　東北道那須付近の交通量の deseason と trend

2.2.2　**弾性値** Elasticity

データを自然対数に変換して、最少自乗法で回帰を行うと、トレンド項に変化率が求まる。長期の増減率と見なせる。

通季のデータを自然対数に変換し、回帰係数（定数項と elastically）を求めたもの。

```
 reg0 ^. 1{("1)  12 adj_season2 a1
9.77655 _0.00115986


(1+ _0.00115986)^120    NB. 120 is 10 years
```

```
0.869998
```

2.3 指数平滑化

2.3.1 指数平滑化

計算が簡単でよく使い込まれた便利な方法に指数平滑化がある。任意の $0 < w < 1$ のウェイトを与えるとウェイトにより平滑化を行う。ウエイトを決める方法はないので、グラフを見ながら適当な値を捜す。

$\bar{x}_1 = x_1$

$\bar{x}_2 = w x_1{}^- + (1 - w) x_2$

$\bar{x}_3 = w x_2{}^- + (1 - w) x_3 = w^2 x_1 + w(1 - w) x_2 + (1 - w) x_3$

```
plot |:  a1,. 0.95 smexp a1
  pd 'eps temp\smexp0.eps'
```

2.3.2 Holt's Method

ホルト・ウインターの 2 段階指数平滑化

Holt's two-parameter double exponential smoothing model

$S_1 = \alpha Y_1 + (1 - \alpha)(S_{t-1} + bt - 1)$

$b_t = \beta(S_1 - S_{t-1}) + (1 - \beta) b_{t-1}$

$F_{t+m} = S_1 + b_t m$

	Starting base
α = level smoothing constant S_1 = smoothed at end of period t	 Y_1

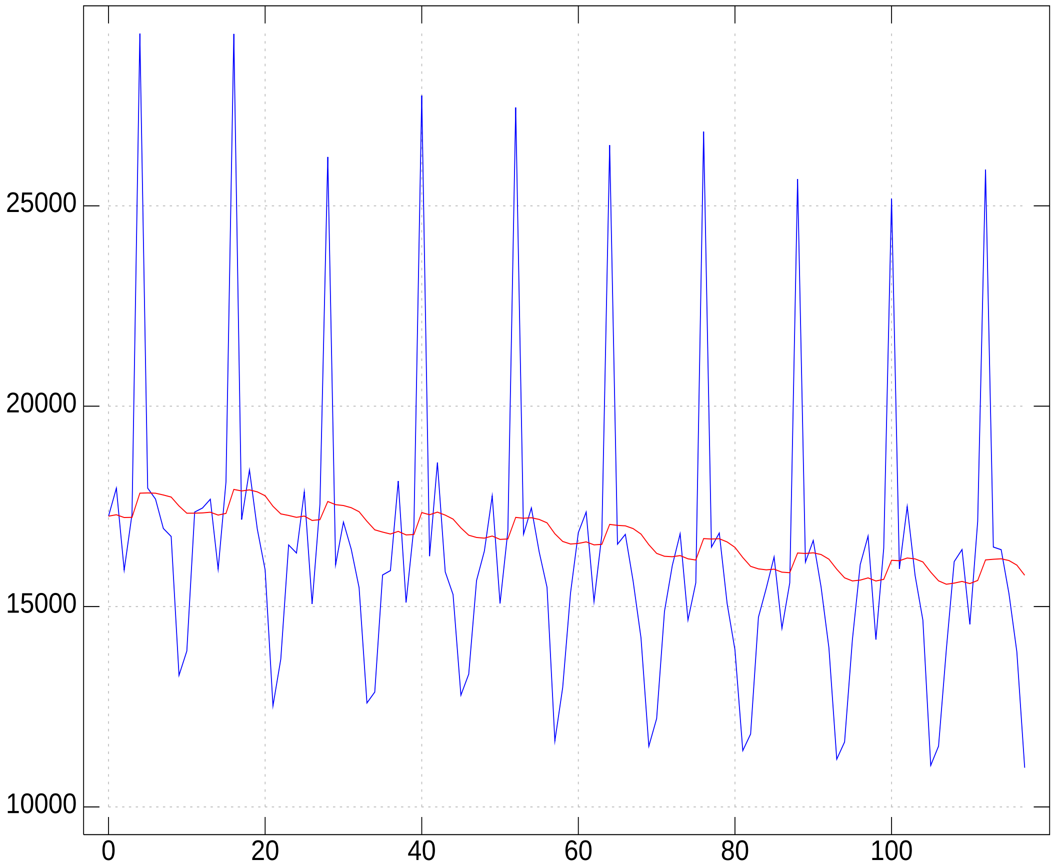

図 2.8　東北道那須付近の交通量の指数平滑化

β = trend smoothing constant b_1 = smoothed trend in period t m = forecast horizon	$(Y_2 - Y_1 + Y_4 - Y_3)/2$

α, β は一意には定まらない。この例では、$\alpha = 0.2\,\beta = 0.6$ が最良であるが、周辺の数字を入れ替えてトライして、残差が最少のものを選択して、モデルとするとされ、２０パターンの α,
beta の組み合わせが紹介されている。ＡＩＣを用いるとよいのではないか。

```
0.2 0.6 hw hdat

(1)   (2)    (3)      (4)
 data ANS_S  ANS_B    Fitted
-----------------------------
63      63     _0.5        0
64     62.8   _0.32     62.5
```

```
 66   63.184   0.1024   62.48
 64 63.4291 0.188032 63.2864
 68 64.4937 0.713974 63.6172
 70 66.1662  1.28905 65.2077
 67 67.3642  1.23443 67.4552
 69 68.6789  1.28259 68.5986
 74 70.7692  1.76722 69.9615
 72 72.4291  1.70285 72.5364
 73 73.9056  1.56702  74.132
```

(2) Smoothed Level

(3) Smoothed Trend

```
   0.2 0.6 hw_res hdat   NB. Residual Etandard Error (RSE)
2.3198

plot |: }: 0 3 sel (0.2 0.6 hw hdat)
plotpdf  'hw_2p.pdf'
```

$$RSE = \sqrt{\sum \frac{e^2}{n-1}}$$

残差分析の自動計算

残差計算が大変そうなので、1に近いパラメーターを与えると、0まで0．05ピッチで0までぐるぐる回し、最小値を選択して、ホルト法を計算するルーチンを作った。

```
0.4 0.8 loop_hw hdat


    10 {. HXX
 0.2  0.8 2.25821
 0.2 0.75 2.26552
0.25 0.65  2.2726
0.25  0.7 2.27268
0.25 0.75 2.27691
 0.2  0.7 2.27745
```

```
0.25  0.6 2.27754
0.25  0.8 2.28468
0.25 0.55  2.2887
```

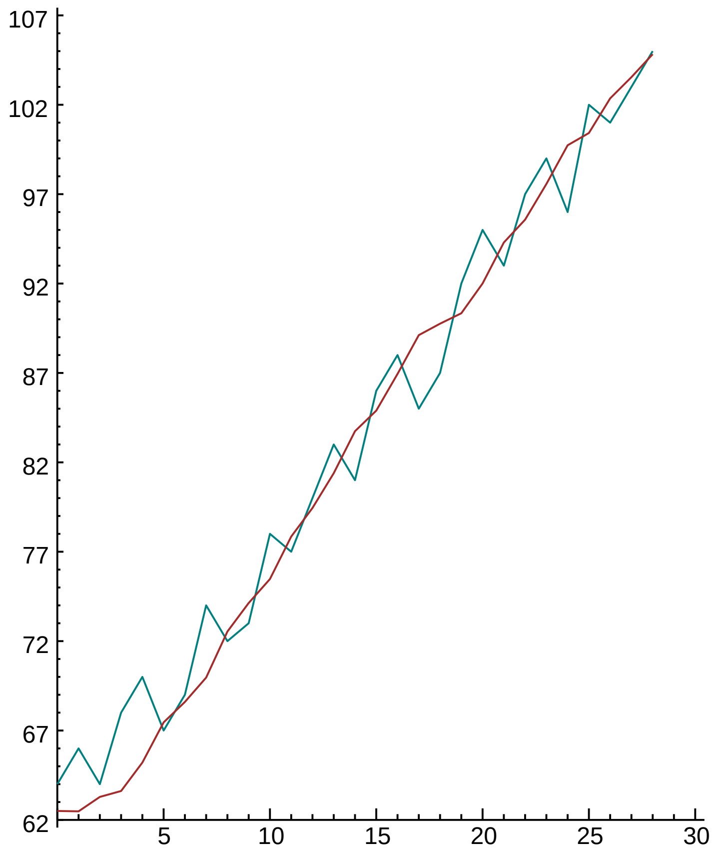

図 2.9　Elastically