6.1.3　季節ダミー

季節性のあるデータ（例として四半期データ）を扱う場合，

$$D_{1i} = \begin{cases} 1, & i\ \text{が第一四半期のとき} \\ 0, & \text{他} \end{cases}$$

$$D_{2i} = \begin{cases} 1, & i\ \text{が第二四半期のとき} \\ 0, & \text{他} \end{cases}$$

$$D_{3i} = \begin{cases} 1, & i\ \text{が第三四半期のとき} \\ 0, & \text{他} \end{cases}$$

という 3 つのダミー変数を作り，

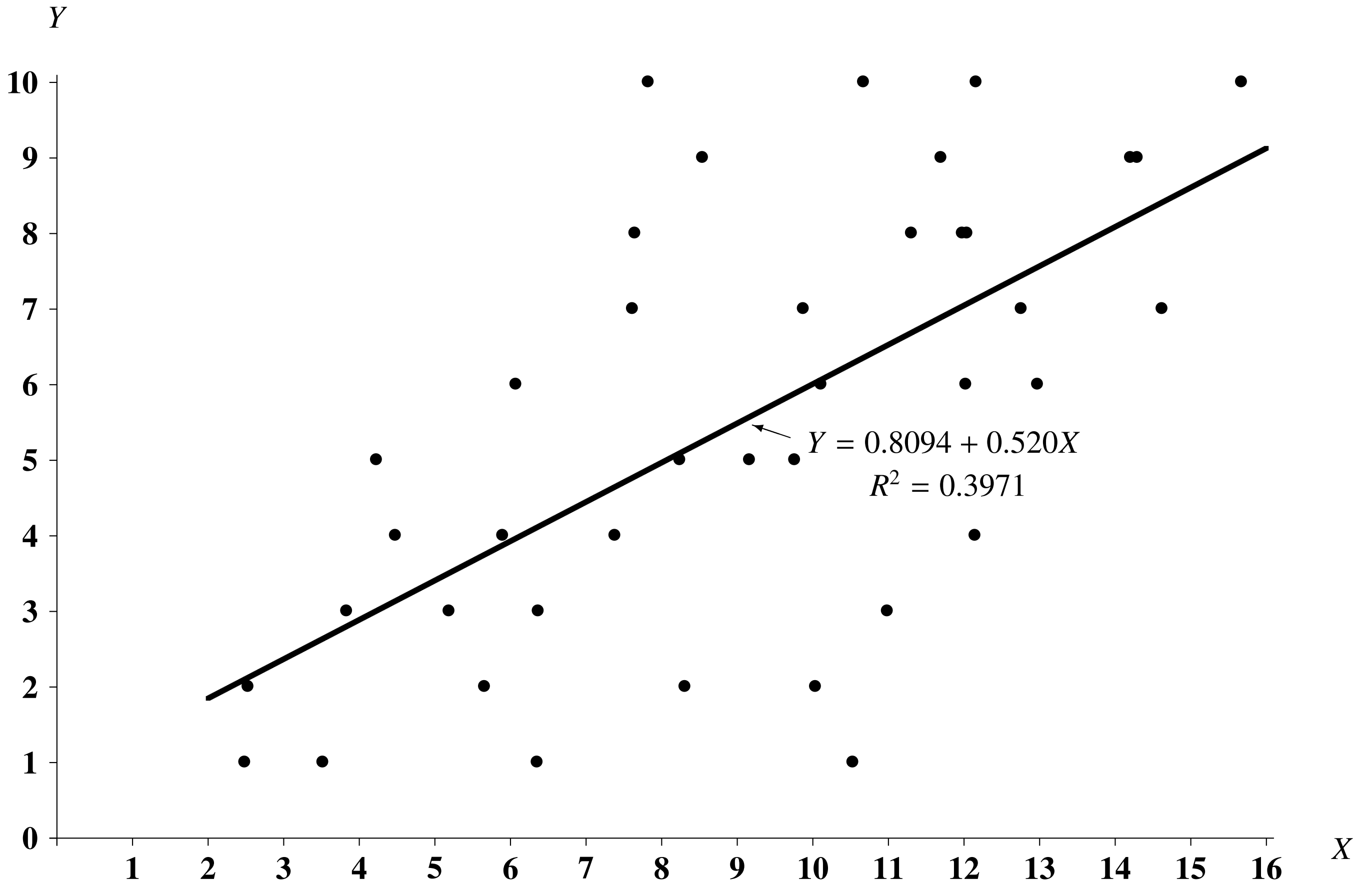
$$Y_i = \alpha + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i$$

を推定することになる（この場合，i は時間を表す）。

まず初めに，$i = 1, 2, \cdots, 40$ として，

$$Y_i = \alpha + \beta X_i + u_i$$

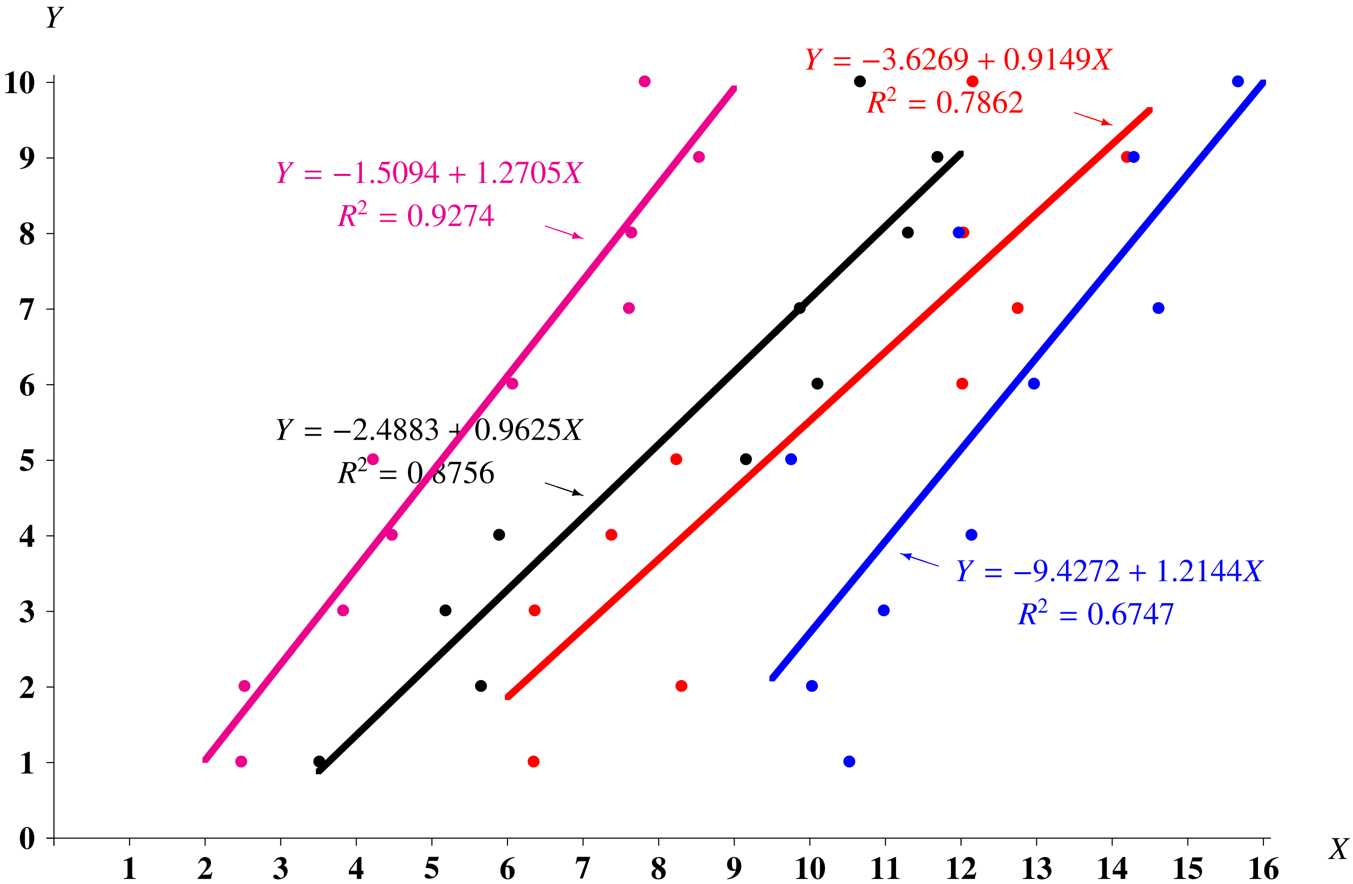
を推定してみる。 $\Longrightarrow$ その結果は次ページ


Y
X
Y = 0.8094 + 0.520X
R2 = 0.3971
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

第1四半期のデータ（黒）だけを集めて，回帰分析を行う。

同様に，第2四半期データ（赤），第3四半期データ（青），第4四半期データ（ピンク）のそれぞれのデータだけを用いて推定してみる。

それぞれの直線は，10個のデータを用いて推定することになる（この場合，自由度は $n-k = 10-2=8$）。

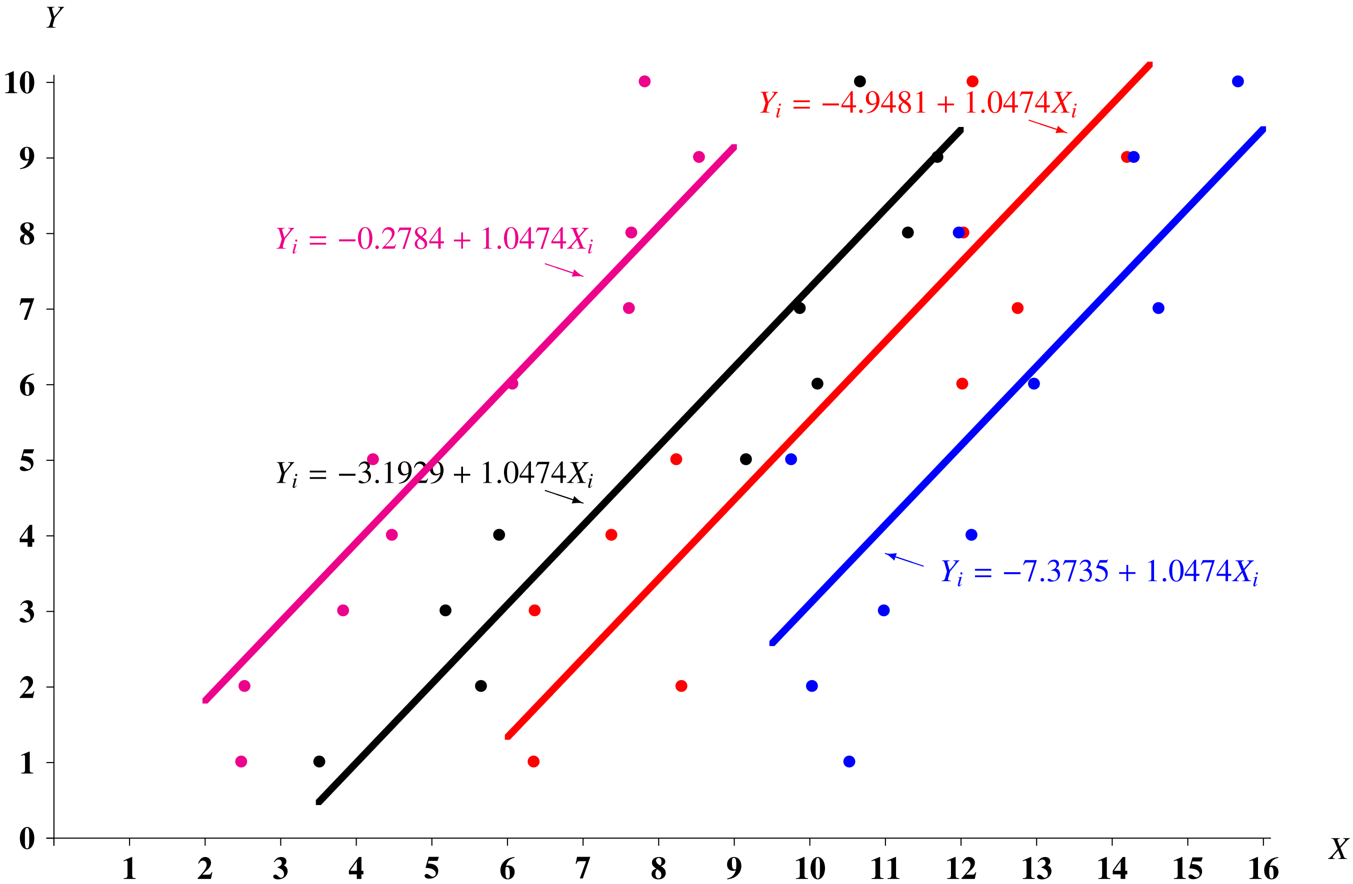

Y
X
$Y = -1.5094 + 1.2705X$
$R^2 = 0.9274$
$Y = -2.4883 + 0.9625X$
$R^2 = 0.8756$
$Y = -3.6269 + 0.9149X$
$R^2 = 0.7862$
$Y = -9.4272 + 1.2144X$
$R^2 = 0.6747$
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

傾きが一定で，季節によって定数項がシフトするというモデルを仮定すると，下記のように推定される。

$$Y_i = -\underset{(0.426)}{\mathbf{0.2784}} - \underset{(4.402)}{\mathbf{2.9145}} D_{1i} - \underset{(6.393)}{\mathbf{4.6697}} D_{2i} - \underset{(8.262)}{\mathbf{7.0951}} D_{3i} + \underset{(11.825)}{\mathbf{1.0474}} X_i$$

$$R^2 = 0.7998, \qquad \overline{R}^2 = 0.7769, \qquad s^2 = 1.3739^2$$

となる（自由度は $n - k = 40 - 5 = 35$）。ただし，係数の推定値の下の括弧内は t 値を表すものとする。


Y
X
$Y_i = -0.2784 + 1.0474X_i$
$Y_i = -3.1929 + 1.0474X_i$
$Y_i = -4.9481 + 1.0474X_i$
$Y_i = -7.3735 + 1.0474X_i$
0
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

6.1.4　地域差ダミー

関西と関東とで賃金格差があるかどうかを調べたい。

w_i を賃金，D_i を地域ダミー（後述），$Year_i$ を勤続年数とすると，例として，下記のような賃金関数を考えることができる。

$$w_i = \alpha + \beta D_i + \gamma Year_i + u_i$$

添え字 i は個人を表すものとする。

D_i を下記のように定義する。

$$D_i = \begin{cases} 1, & i \text{ 番目の人が関東に住んでいるとき} \\ 0, & i \text{ 番目の人が関西に住んでいるとき} \end{cases}$$

β の有意性検定を行うことによって，賃金に格差があるかどうかを検定できる。

6.1.5　男女別ダミー

男女間で賃金格差があるかどうかを調べたい。この場合も，添え字 i は個人を表す。

w_i を賃金，D_i を男女別ダミー（後述），$Year_i$ を勤続年数とすると，例として，下記のような賃金関数を考えることができる。

$$w_i = \alpha + \beta D_i + \gamma Year_i + u_i$$

添え字 i は個人を表すものとする。D_i は，

$$D_i = \begin{cases} 1, & i \text{ 番目の人が女性のとき} \\ 0, & i \text{ 番目の人が男性のとき} \end{cases}$$

となるダミー変数とする。

さらに，学歴ダミーを加えて，下記のような賃金関数を考えることもできる。

$$w_i = \alpha + \beta D_i + \gamma Year_i + \delta Edu_i + u_i$$

添え字 i は個人を表すものとする。Edu_i は，

$$Edu_i = \begin{cases} 1, & i \text{ 番目の人の最終学歴が大学卒以上のとき} \\ 0, & i \text{ 番目の人の最終学歴がそれ以外のとき} \end{cases}$$

となるダミー変数とする。

この場合，定数項で 4 つのケースに分類される。

$$w_i = \begin{cases} \alpha + \gamma Year_i + u_i, & D_i = 0\text{，} Edu_i = 0 \text{ のとき（男性，かつ，大学卒以下）} \\ \alpha + \delta + \gamma Year_i + u_i, & D_i = 0\text{，} Edu_i = 1 \text{ のとき（男性，かつ，大学卒以上）} \\ \alpha + \beta + \gamma Year_i + u_i, & D_i = 1\text{，} Edu_i = 0 \text{ のとき（女性，かつ，大学卒以下）} \\ \alpha + \beta + \delta + \gamma Year_i + u_i, & D_i = 1\text{，} Edu_i = 1 \text{ のとき（女性，かつ，大学卒以上）} \end{cases}$$

4 つのケースを一つの式にまとめて推定することができる。

その他にも，分析者側の問題意識によって，様々なダミー変数が考えられる。

6.2　関数形について

今まで，簡単化のために線形関数を扱った。関数形も未知と考えるのが自然である。現実社会が線形で表されるというのは考えにくいことではあるが，何らかの仮定を置かなければ分析すらできないことになる。一つの考え方としては，現実社会は複雑な非線形関数で，しかも，未知の関数で表されるが，実際に観測されるデータの範囲内では近似的に線形関数を用いることができると考える。

または，非線形関数であっても，変数のある変換によって，線形に表される場合も多い。線形に変換できたとしても，回帰係数の意味が異なってくる。

● 線形：　通常の回帰式：

$$Y_i = \alpha + \beta X_i + u_i,$$

を考える。

この場合，回帰係数 β の意味は，

$$\beta = \frac{\mathbf{d}Y_i}{\mathbf{d}X_i}$$

となる（Y_i を X_i で微分した値，すなわち，微分係数）。

β は，X_i が一単位上昇 (下落) したとき，Y_i は何単位上昇 (下落) するのかを表す。

β は限界係数と呼ばれる。

● 成長率：　被説明変数，または，説明変数に成長率を使いたい場合，例えば，

$$100 \times \frac{Y_i - Y_{i-1}}{Y_{i-1}} = \alpha + \beta X_i + u_i,$$

では，成長率を被説明変数として用いる場合もある。 $100 \times \dfrac{Y_i - Y_{i-1}}{Y_{i-1}}$ **という変数をあらかじめ作っておき，これをこれまでの** Y_i **として扱えばよい。**

この場合， β **意味は，**

$$\beta = \frac{\mathbf{d}\ 100(Y_i - Y_{i-1})/Y_{i-1}}{\mathbf{d}\ X_i}$$

となり， X_i **が 1 単位増加したとき，** Y_i **が** β % **増加するということになる。**

注意：

$Y_i = \alpha + \beta X_i + u_i$ と $100 \times \dfrac{Y_i - Y_{i-1}}{Y_{i-1}} = \alpha + \beta X_i + u_i$ では，得られる決定係数の大きさが全く異なる（被説明変数の値が異なるので）。

単純に，R^2 や $\overline{R}^2$ による比較はこの場合出来ない。

● 対数線形：　$Y_i = aX_i^{\beta}$ という関数を考えることもよくある。

両辺に対数を取って，誤差項 u_i を加えると，

$$\log(Y_i) = \alpha + \beta \log(X_i) + u_i,$$

となる。ただし，$\alpha = \log(a)$ となる。

この場合，

$$\beta = \frac{\mathbf{d}\log(Y_i)}{\mathbf{d}\log(X_i)} = \frac{\dfrac{\mathbf{d}Y_i}{Y_i}}{\dfrac{\mathbf{d}X_i}{X_i}} = \frac{100\dfrac{\mathbf{d}Y_i}{Y_i}}{100\dfrac{\mathbf{d}X_i}{X_i}}$$

となる。

2 つ目の等号では，$\dfrac{\mathbf{d}\log(Y_i)}{\mathbf{d}Y_i} = \dfrac{1}{Y_i}$ が利用される。

3 つ目の等号の分子 $100\dfrac{\mathbf{d}Y_i}{Y_i}$ や分母 $100\dfrac{\mathbf{d}X_i}{X_i}$ は，i が時間であれば，i 時点の瞬間的な上昇率を表す（i が個人であれば，ある時点の瞬間的な上昇率を表す）。

したがって，β は，X_i が 1% 上昇 (下落) したとき，Y_i は何% 上昇 (下落) するのかを表す。β は弾力性と呼ばれる。

例： コブ＝ダグラス型生産関数：

$$Q_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3}$$

ただし，Q_i は生産量，K_i は資本，L_i は労働である。

この場合，両辺に対数を取って，誤差項を加えると，

$$\log(Q_i) = \beta_1' + \beta_2 \log(K_i) + \beta_3 \log(L_i) + u_i$$

として，$\log(Q_i), \log(K_i), \log(L_i)$ のデータをあらかじめ変換しておき，最小二乗法で $\beta_1', \beta_2, \beta_3$ を推定する。

また，生産関数には一次同次の制約 $\beta_2 + \beta_3 = 1$ を置く場合が多い。

この場合は，

$$
\begin{aligned}
\log(Q_i) &= \beta_1' + \beta_2 \log(K_i) + \beta_3 \log(L_i) \\
&= \beta_1' + \beta_2 \log(K_i) + (1 - \beta_2) \log(L_i) + u_i \\
&= \beta_1' + \beta_2\Big(\log(K_i) - \log(L_i)\Big) + \log(L_i) + u_i
\end{aligned}
$$

となるので，

$$
\log(Q_i) - \log(L_i) = \beta_1' + \beta_2\Big(\log(K_i) - \log(L_i)\Big) + u_i
$$

を最小二乗法で推定し，β_1', β_2 を求めることになる。この場合も同様に，各変数をあらかじめ，$\log(Q_i) - \log(L_i)$，$\log(K_i) - \log(L_i)$ としてデータを作っておく必要がある。

● 二次式：　平均費用または限界費用（縦軸）と生産量（横軸）との関係は，下記のように，二次式で表される場合がある。

$$Y_i = \alpha + \beta X_i + \gamma X_i^2 + u_i$$

$X_i^2 = Z_i$ とデータをあらかじめ作っておき，X_i と Z_i を説明変数，Y_i を被説明変数として，回帰分析を行えばよい。

● 逆数：　賃金上昇率または物価上昇率（縦軸）と失業率（横軸）との関係を表す右下がりの曲線は，フィリップス曲線と呼ばれる。下記のような双曲線で表される場合が多い。

$$Y_i = \alpha + \beta \frac{1}{X_i} + u_i$$

$\frac{1}{X_i} = Z_i$ という変数をあらかじめ作っておき，説明変数を Z_i，被説明変数を Y_i として回帰分析を行う。

● 遅れのある変数：　習慣的効果を考慮に入れたモデル：

$$Y_i = \alpha + \beta X_i + \gamma Y_{i-1} + u_i$$

ラグ付き内生変数が説明変数に用いられる。

Excel では，$Y_{i-1} = Z_i$ というデータを作って，X_i と Z_i を説明変数，Y_i を被説明変数として回帰分析を行えばよい。

（注意） 内生変数に対して，外生変数という。被説明変数は内生変数となる。説明変数かつ非確率変数は外生変数となる。

X_i の Y_i への効果は，短期効果，長期効果の 2 つある。

β は短期効果を表す係数である。

長期効果とは，$Y_i = Y_{i-1}$，$X_i = X$（添え字 i に依存しなくなる収束先，または，定常状態とも言う）となるときの，X_i から Y_i への影響を示す効果である。

誤差項 u_i は無視して，

$$Y = \alpha + \beta X + \gamma Y$$

として，Y について解くと，

$$Y = \frac{\alpha}{1-\gamma} + \frac{\beta}{1-\gamma}X$$

となり，$\dfrac{\beta}{1-\gamma}$ が X_i の Y_i への長期効果を表す係数となる。

問題点：

1. 最小二乗法の仮定の一つに，説明変数は確率変数ではないという仮定がある。ラグ付き内生変数を説明変数に加えることによって，この仮定が満たされなくなる。最小二乗推定量は最小分散線形不偏推定量ではなくなる。

2. Y_i と X_i とは，経済理論的に考えると，相関が高いはず。Y_i と Y_{i-1} は相関が高い。当然，Y_{i-1} と X_i も高い相関を示す。

$\Longrightarrow$ 多重共線性の可能性が高い（多重共線性については後述）。

3. ダービン・ワトソン比（DW）統計量は意味をなさない（DW については後述）。

● 遅れのある変数の解釈 (部分調整モデル)： X_i が与えられたときの Y の最適水準を Y_i^* とする。

$$Y_i^* = \alpha + \beta X_i,$$

現実の水準 Y_i は，最適水準 Y_i^* と前期の水準 Y_{i-1} との差の一定割合と前期の水準 Y_{i-1} との和で与えられるとする。調整関数を考える。

$$Y_i - Y_{i-1} = \lambda(Y_i^* - Y_{i-1}) + u_i,$$

ただし，u_i は互いに独立で同一な分布の誤差項，$0 < \lambda < 1$ とする。

よって，

$$Y_i = \lambda\alpha + \lambda\beta X_i + (1 - \lambda)Y_{i-1} + u_i,$$

を得る。

Y_{i-1} と u_i との相関はない。

しかし，Y_{i-1} が説明変数の一つに入っている（説明変数間が確率変数でないという仮定に反する）。

推定量は不偏推定量ではないが，一致推定量である (証明略)。