計量経済学(講義ノート・大学院用— 99年度作成)

谷﨑　久志
神戸大学・経済学部

目 次

1　重回帰の行列表示

1. 回帰モデル

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \ \cdots \ + \beta_k x_{kt} + u_t,$$

ただし，$t = 1, 2, \cdots, T$ とする。

2. 撹乱項 u_t の仮定

   (a) $u_t \sim$ iid $(0, \sigma^2)$

   (b) $x_{1t}, x_{2t}, \cdots, x_{kt}$ と u_t は無相関

   (c) $x_{1t}, x_{2t}, \cdots, x_{kt}$, y_t は定常 (大標本特性を調べる場合に必要)

   (d) β の推定値の仮説検定を行うときには，$u_t \sim$ iid $N(0, \sigma^2)$ の仮定が必要

3. $x_{1t} = 1$ のとき，β_1 は定数項となる。

4. 最小自乗法

$$\begin{aligned} & S(\beta_1, \beta_2, \cdots, \beta_k) \\ =& \sum_{t=1}^{T} (y_t - \beta_1 x_{1t} - \beta_2 x_{2t} - \ \cdots \ - \beta_k x_{kt})^2 \end{aligned}$$

$S(\beta_1, \beta_2, \cdots, \beta_k)$ を最小にする $\beta_1, \beta_2, \cdots, \beta_k$ をその推定値とする。$\Longrightarrow \widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$

それぞれの要素で偏微分する。

$$\begin{aligned} & \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_{1t}} \\ =& -2 \sum_{t=1}^{T} x_{1t} (y_t - \beta_1 x_{1t} - \beta_2 x_{2t} - \ \cdots \ - \beta_k x_{kt}) \\ =& \ 0 \\ & \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_{2t}} \\ =& -2 \sum_{t=1}^{T} x_{2t} (y_t - \beta_1 x_{1t} - \beta_2 x_{2t} - \ \cdots \ - \beta_k x_{kt}) \\ =& \ 0 \\ & \vdots \\ & \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_{kt}} \\ =& -2 \sum_{t=1}^{T} x_{kt} (y_t - \beta_1 x_{1t} - \beta_2 x_{2t} - \ \cdots \ - \beta_k x_{kt}) \\ =& \ 0 \end{aligned}$$

まとめて，

$$\begin{aligned} \sum x_{1t} y_t &= \beta_1 \sum x_{1t}^2 + \beta_2 \sum x_{1t} x_{2t} \\ &\quad + \ \cdots \ + \beta_k \sum x_{1t} x_{kt} \\ \sum x_{2t} y_t &= \beta_1 \sum x_{2t} x_{1t} + \beta_2 \sum x_{2t}^2 \\ &\quad + \ \cdots \ + \beta_k \sum x_{2t} x_{kt} \\ &\quad \vdots \\ \sum x_{kt} y_t &= \beta_1 \sum x_{kt} x_{1t} + \beta_2 \sum x_{kt} x_{2t} \\ &\quad + \ \cdots \ + \beta_k \sum x_{kt}^2 \end{aligned}$$

さらに，行列でまとめて，

$$\begin{pmatrix} \sum x_{1t}^2 & \sum x_{1t} x_{2t} & \cdots & \sum x_{1t} x_{kt} \\ \sum x_{2t} x_{1t} & \sum x_{2t}^2 & \cdots & \sum x_{2t} x_{kt} \\ \vdots & \vdots & \vdots & \cdots \\ \sum x_{kt} x_{1t} & \sum x_{kt} x_{2t} & \cdots & \sum x_{kt}^2 \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \sum x_{1t} y_t \\ \sum x_{2t} y_t \\ \vdots \\ \sum x_{kt} y_t \end{pmatrix}$$

5. 行列表示が簡単

$$\begin{aligned} y_t &= \beta_1 x_{1t} + \beta_2 x_{2t} + \ \cdots \ + \beta_k x_{kt} + u_t \\ &= (x_{1t} \ x_{2t} \ \cdots \ x_{kt}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + u_t \\ &= x_t \beta + u_t \end{aligned}$$

ただし，

$$x_t = (x_{1t} \ x_{2t} \ \cdots \ x_{kt}),$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix},$$

とする。

$$y_1 = x_1 \beta + u_1$$

$$y_2 = x_2\beta + u_2$$
$$\vdots$$
$$y_T = x_T\beta + u_T$$

行列でまとめる。

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$$

$$y = X\beta + u$$

ただし，

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix},$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix},$$

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$$

とする。

6. それぞれの行列の次元は以下のとおり。

   x_t: $1 \times k$,

   β: $k \times 1$,

   y: $T \times 1$,

   X: $T \times k$,

   u: $T \times 1$

2 行列について

$A : T \times T$,
$B : n \times m$,
$C : m \times k$,
$D : k \times n$,

1. $\mathrm{tr}(A) = \sum_{i=1}^{T} a_{ii}$, where $A = [a_{ij}]$.
2. If A is idempotent, $A = A^2 = A'A$.
3. A is idempotent if and only if the eigen values of A consist of 1 and 0.
4. If A is idempotent, $\mathrm{rank}(A) = \mathrm{tr}(A)$.
5. $\mathrm{tr}(BCD) = \mathrm{tr}(CDB)$

2.1 行列の微分

$a, x : T \times 1$,
$y : K \times 1$,
$A : T \times T$,
$B : T \times K$

1. $\dfrac{\partial a'x}{\partial x} = \dfrac{\partial x'a}{\partial x} = a$
   $a'x$, $x'a$ 共にスカラー
2. $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$
3. $\dfrac{\partial^2 x'Ax}{\partial x \partial x'} = (A + A')$
4. $\dfrac{\partial x'By}{\partial B} = xy'$
5. $\dfrac{\partial \mathrm{tr}\ A}{\partial A} = I$
6. $\dfrac{\partial \log|A|}{\partial A} = (A')^{-1}$

2.2 分布関数

$a, x, y, \mu : T \times 1$,
$\Sigma, A, B : T \times T$,
σ : scalar

1. If $x \sim N(\mu, \Sigma)$, then $a'x \sim N(a'\mu, a'\Sigma a)$.

2. If $x \sim N(\mu, \Sigma)$, then $(x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi^2(T)$

3. x: $n \times 1$,

   y: $m \times 1$ とする。$x \sim N(\mu_x, \Sigma_x)$, $y \sim N(\mu_y, \Sigma_y)$, x と y は独立 ($\mathrm{E}\big((x-\mu_x)(y-\mu_y)'\big) = 0$) のとき，
   $\dfrac{(x-\mu_x)'\Sigma_x^{-1}(x-\mu_x)/n}{(y-\mu_y)'\Sigma_y^{-1}(y-\mu_y)/m} \sim F(n, m)$

4. If $x \sim N(0, \sigma^2 I)$ and A is a symmetric idempotent $T \times T$ matrix of rank G, then $x'Ax/\sigma^2 \sim \chi^2(G)$.

5. If $x \sim N(0, \sigma^2 I)$, $y \sim N(0, \sigma^2 I)$, A and B are symmetric idempotent $T \times T$ matrices of rank G and K, and $AB = 0$, then

$$\frac{y'Ay}{G\sigma^2} \Big/ \frac{x'Bx}{K\sigma^2} = \frac{y'Ay/G}{x'Bx/K} \sim F(G, K).$$

3　多重回帰： 再考

$y : T \times 1$,
$X : T \times k$,
$\beta : k \times 1$,
$u : T \times 1$,

1. $u = (u_1, u_2, \cdots, u_t)'$ とする。

$$\sum_{t=1}^{T} u_t^2 = u'u$$

2. Regression model: $y = X\beta + u$, $u \sim (0, \sigma^2 I)$

   最小二乗推定量

$$\min_{\beta} \ (y - X\beta)'(y - X\beta)$$

$$\begin{aligned} u'u &= (y - X\beta)'(y - X\beta) \\ &= y'y - 2\beta'X'y + \beta'X'X\beta \end{aligned}$$

   最小化のためには

$$\frac{\partial u'u}{\partial \beta} = -2X'y + 2X'X\beta = 0$$

   で，これを満たす β を最小二乗推定量と呼び，$\widehat{\beta}$ で表す。

$$\widehat{\beta} = (X'X)^{-1}X'y$$

3. 最小化のための 2 階の条件：

$$\frac{\partial^2 u'u}{\partial \beta \partial \beta'} = 2X'X$$

   は正値定符号行列

   $c = Xd$ とすると，任意の $d \neq 0$ について

   $c'c = d'X'Xd > 0$ となる。

   (*) 正値定符号行列，負値定符号行列について

   (a) 正値定符号行列：
   任意のベクトル $x \neq 0$ について，$x'Ax > 0$ が成り立つとき，A は正値定符号行列 (positive definite matrix) という。
   正値定符号行列の固有値はすべて正である。

   (b) 負値定符号行列：
   任意のベクトル $x \neq 0$ について，$x'Ax < 0$ が成り立つとき，A は負値定符号行列 (positive definite matrix) という。
   負値定符号行列の固有値はすべて負である。

4. $\widehat{\beta} = (X'X)^{-1}X'y$
   $= (X'X)^{-1}X'(X\beta + u)$
   $= \beta + (X'X)^{-1}X'u$

5. 最小二乗推定量 $\widehat{\beta} = (X'X)^{-1}X'y$ は y に関して線形
   $\Longrightarrow$ 線形推定量

6. $\mathrm{E}(\widehat{\beta}) = \beta + (X'X)^{-1}X'\mathrm{E}(u) = \beta$
   $\Longrightarrow$ 線形不偏推定量

7. $\mathrm{V}(\widehat{\beta}) = \mathrm{E}\big((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'\big)$
   $= \mathrm{E}\Big(\big((X'X)^{-1}X'u\big)\big((X'X)^{-1}X'u\big)'\Big)$
   $= (X'X)^{-1}X'\mathrm{E}(uu')X(X'X)^{-1} = \sigma^2(X'X)^{-1}$

8. $u \sim N(0, \sigma^2 I_T)$ のとき，$\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$
   さらに，$\dfrac{(\widehat{\beta} - \beta)'X'X(\widehat{\beta} - \beta)}{\sigma^2} \sim \chi^2(k)$ を得る。
   (*) x を $k \times 1$ の確率変数とする。
   $x \sim N(\mu, \Sigma) \Longrightarrow (x - \mu)'\Sigma^{-1}(x - \mu) \sim \chi^2(k)$

9. Properties of $\widehat{\beta}$: BLUE (best linear unbiased estimator), i.e., Unbiased and efficient estimator in linear class (Gauss-Markov theorem)

証明：

適当な他の線形不偏推定量を $\widetilde{\beta} = Cy$ とする。

$$\begin{aligned}\widetilde{\beta} &= Cy \\ &= C(X\beta + u) \\ &= CX\beta + Cu\end{aligned}$$

なので，

$$\begin{aligned}\mathrm{E}(\widetilde{\beta}) &= CX\beta + C\mathrm{E}(u) \\ &= CX\beta\end{aligned}$$

を得る。

$\widetilde{\beta} = Cy$ を線形不偏推定量と仮定したので，$\mathrm{E}(\widetilde{\beta}) = \beta$，すなわち，$CX = I$ を得る。

次に，$\widetilde{\beta} = Cy$ の分散を求める。

$$\begin{aligned}\widetilde{\beta} &= C(X\beta + u) \\ &= \beta + Cu\end{aligned}$$

なので，

$$\begin{aligned}\mathrm{V}(\widetilde{\beta}) &= \mathrm{E}(\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)' \\ &= \mathrm{E}(Cuu'C') \\ &= \sigma^2 CC'\end{aligned}$$

を得る。

$D = C - (X'X)^{-1}X'$ を定義する。

$$\begin{aligned}\mathrm{V}(\widetilde{\beta}) &= \sigma^2 CC' \\ &= \sigma^2 \big(D + (X'X)^{-1}X'\big)\big(D + (X'X)^{-1}X'\big)'\end{aligned}$$

さらに，

$$CX = I = \big(D + (X'X)^{-1}X'\big)X = DX + I$$

により，

$$DX = 0$$

を得る。したがって，

$$\begin{aligned}&\mathrm{V}(\widetilde{\beta}) \\ &= \sigma^2 CC' \\ &= \sigma^2 \big(D + (X'X)^{-1}X'\big)\big(D + (X'X)^{-1}X'\big)' \\ &= \sigma^2 (X'X)^{-1} + \sigma^2 DD' \\ &= \mathrm{V}(\widehat{\beta}) + \sigma^2 DD'\end{aligned}$$

よって，$\mathrm{V}(\widetilde{\beta}) - \mathrm{V}(\widehat{\beta})$ は正値定符号行列

$\Longrightarrow \widehat{\beta}$ は最小分散不偏推定量

10. σ^2 の推定量を s^2 とする。

$s^2 = \dfrac{1}{T-k}(y - X\widehat{\beta})'(y - X\widehat{\beta})$,

because

$$\begin{aligned}y - X\widehat{\beta} &= y - X(X'X)^{-1}X'y \\ &= (I_T - X(X'X)^{-1}X')y \\ &= (I_T - X(X'X)^{-1}X')(X\beta + u) \\ &= (I_T - X(X'X)^{-1}X')u\end{aligned}$$

and

$$\begin{aligned}&\mathrm{E}\Big(\frac{1}{T-k}(y - X\widehat{\beta})'(y - X\widehat{\beta})\Big) \\ &= \frac{1}{T-k}\mathrm{E}\big(((I_T - X(X'X)^{-1}X')u)' \\ &\qquad ((I_T - X(X'X)^{-1}X')u)\ \big) \\ &= \frac{1}{T-k}\mathrm{E}\big(u'(I_T - X(X'X)^{-1}X')' \\ &\qquad (I_T - X(X'X)^{-1}X')u\ \big) \\ &= \frac{1}{T-k}\mathrm{E}\big(u'(I_T - X(X'X)^{-1}X')u\big) \\ &= \frac{1}{T-k}\mathrm{E}\big(\mathrm{tr}u'(I_T - X(X'X)^{-1}X')u\big) \\ &= \frac{1}{T-k}\mathrm{tr}\big((I_T - X(X'X)^{-1}X')\mathrm{E}(uu')\big) \\ &= \frac{1}{T-k}\sigma^2\mathrm{tr}\big((I_T - X(X'X)^{-1}X')\big) \\ &= \frac{1}{T-k}\sigma^2(T-k) \\ &= \sigma^2\end{aligned}$$

トレースについて：

1. A: $k \times k$

   $\mathrm{tr}(A) = \sum_{i=1}^{k} a_{ii}$

   a_{ii} は A の第 i 行，第 j 列の要素

2. a: スカラー (1×1)

   $\mathrm{tr}(a) = a$

3. A: $T \times k$, B: $k \times T$ のとき，

   $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

4. $\mathrm{tr}[X(X'X)^{-1}X'] = \mathrm{tr}[(X'X)^{-1}X'X] = \mathrm{tr}(I_k)$

5. X: 確率変数行列のとき，

   $\mathrm{E}[\mathrm{tr}(X)] = \mathrm{tr}[\mathrm{E}(X)]$

***F* 分布 (H_0： $\beta = 0$) について：**

1. If $u \sim N(0, \sigma^2 I_T)$, then $\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$.

   Therefore,
   $$\frac{(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)}{\sigma^2} \sim \chi^2(k).$$

   (a) $\dfrac{(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)}{\sigma^2} \sim \chi^2(k)$ の証明：

   $$\widehat{\beta} - \beta = (X'X)^{-1}X'u$$

   なので，

   $$\begin{aligned}
   &(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta) \\
   &= \big((X'X)^{-1}X'u\big)'X'X(X'X)^{-1}X'u \\
   &= u'X(X'X)^{-1}X'X(X'X)^{-1}X'u \\
   &= u'X(X'X)^{-1}X'u
   \end{aligned}$$

   を得る。

   $X(X'X)^{-1}X'$ はべき等行列 (idempotent, i.e., $A'A = A$) であることに注意すると，

   $$\begin{aligned}
   &\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \\
   &\sim \chi^2\Big(\mathrm{tr}\big(X(X'X)^{-1}X'\big)\Big)
   \end{aligned}$$

   しかも，

   $$\begin{aligned}
   &\mathrm{tr}\big(X(X'X)^{-1}X'\big) \\
   &= \mathrm{tr}\big((X'X)^{-1}X'X\big) \\
   &= \mathrm{tr}(I_k) \\
   &= k
   \end{aligned}$$

   なので，

   $$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(k)$$

   (b) (*) 公式：

   確率変数 $x \sim N(0, I_k)$ とする。このとき，

   $$x'Ax \sim \chi^2\big(\mathrm{Rank}(A)\big)$$

   となる。

   また，A がべき等行列 (idempotent, i.e., $A'A = A$) のとき，

   $\mathrm{Rank}(A) = \mathrm{tr}(A)$

   なので，

   $$x'Ax \sim \chi^2\big(\mathrm{tr}(A)\big)$$

   となる。

   ここでは，

   $$u \sim N(0, \sigma^2 I_T)$$

   なので，

   $$\frac{1}{\sigma}u \sim N(0, I_T)$$

   を当てはめればよい。

   行列の階数 (Rank) について：

   一次独立な行ベクトル (または，列ベクトル) の最大個数

   (行ベクトルが張る空間の次元)

2. 残差平方和について，

   $$\begin{aligned}
   \widehat{u} &= y - X\widehat{\beta} \\
   &= y - X(X'X)^{-1}X'y \\
   &= \big(I - X(X'X)^{-1}X'\big)y \\
   &= \big(I - X(X'X)^{-1}X'\big)(X\beta + u) \\
   &= \big(I - X(X'X)^{-1}X'\big)X\beta
   \end{aligned}$$

$$
\begin{aligned}
&+\big(I-X(X'X)^{-1}X'\big)u\\
&=\big(X-X(X'X)^{-1}X'X\big)\beta\\
&+\big(I-X(X'X)^{-1}X'\big)u\\
&=\big(I-X(X'X)^{-1}X'\big)u
\end{aligned}
$$

なので，

$$
\begin{aligned}
\widehat{u}'\widehat{u}&=\Big(\big(I-X(X'X)^{-1}X'\big)u\Big)'\\
&\quad\big(I-X(X'X)^{-1}X'\big)u\\
&=u'\big(I-X(X'X)^{-1}X'\big)\\
&\quad\big(I-X(X'X)^{-1}X'\big)u\\
&=u'\big(I-X(X'X)^{-1}X'\big)u
\end{aligned}
$$

と変形され，

$$
\begin{aligned}
\frac{\widehat{u}'\widehat{u}}{\sigma^2}&=\frac{u'\big(I-X(X'X)^{-1}X'\big)u}{\sigma^2}\\
&\sim\chi^2\Big(\mathrm{tr}\big(I-X(X'X)^{-1}X'\big)\Big)
\end{aligned}
$$

を得る。

$$
\begin{aligned}
&\mathrm{tr}\big(I-X(X'X)^{-1}X'\big)\\
&=\mathrm{tr}(I_T)-\mathrm{tr}\big(X(X'X)^{-1}X'\big)\\
&=\mathrm{tr}(I_T)-\mathrm{tr}\big((X'X)^{-1}X'X\big)\\
&=\mathrm{tr}(I_T)-\mathrm{tr}(I_k)\\
&=T-k
\end{aligned}
$$

なので，

$$
\begin{aligned}
\frac{\widehat{u}'\widehat{u}}{\sigma^2}&=\frac{(T-k)s^2}{\sigma^2}\\
&\sim\chi^2(T-k)
\end{aligned}
$$

途中で以下の式が使われている。

$$
s^2=\frac{1}{T-k}\widehat{u}'\widehat{u}
$$

3. さらに，$\widehat{\beta}$ と $\widehat{u}$ は独立になる。

   [証明]

   $u\sim N(0,\sigma^2 I)$ なので，$\mathrm{Cov}(\widehat{u},\widehat{\beta})=0$ を示せばよい。

$$
\begin{aligned}
&\mathrm{Cov}(\widehat{u},\widehat{\beta})\\
&=\mathrm{E}\big(\widehat{u}(\widehat{\beta}-\beta)'\big)\\
&=\mathrm{E}\Big(\big(I-X(X'X)^{-1}X'\big)u\big((X'X)^{-1}X'u\big)'\Big)\\
&=\mathrm{E}\Big(\big(I-X(X'X)^{-1}X'\big)uu'X(X'X)^{-1}\Big)\\
&=\big(I-X(X'X)^{-1}X'\big)\mathrm{E}(uu')X(X'X)^{-1}\\
&=\big(I-X(X'X)^{-1}X'\big)(\sigma^2 I)X(X'X)^{-1}\\
&=\sigma^2\big(I-X(X'X)^{-1}X'\big)X(X'X)^{-1}\\
&=\sigma^2\big(X(X'X)^{-1}-X(X'X)^{-1}X'X(X'X)^{-1}\big)\\
&=\sigma^2\big(X(X'X)^{-1}-X(X'X)^{-1}\big)\\
&=0
\end{aligned}
$$

   よって，$\widehat{\beta}$ と $\widehat{u}$ は独立となる。

4. したがって，以下の分布を得る。

$$
\frac{(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)}{\sigma^2}\sim\chi^2(k),
$$

$$
\frac{\widehat{u}'\widehat{u}}{\sigma^2}\sim\chi^2(T-k)
$$

   しかも，$(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)$ と $\widehat{u}'\widehat{u}$ とは独立となる。

   よって，

$$
\frac{\dfrac{(\widehat{\beta}-\beta)'X'X(\widehat{\beta}-\beta)}{\sigma^2}\Big/k}{\dfrac{\widehat{u}'\widehat{u}}{\sigma^2}\Big/(T-k)}\sim F(k,T-k)
$$

   を得る。

決定係数 R^2 について：

1. 決定係数 R^2 の定義：

   $R^2=1-\dfrac{\sum_{t=1}^T\widehat{u}_t^2}{\sum_{t=1}^T(y_t-\overline{y})^2}$

2. 分子：

   $\sum_{t=1}^T\widehat{u}_t^2=\widehat{u}'\widehat{u}$

3. 分母：

$$
\begin{aligned}
&\textstyle\sum_{t=1}^T(y_t-\overline{y})^2\\
&=y'(I_T-\tfrac{1}{T}ii')'(I_T-\tfrac{1}{T}ii')y\\
&=y'(I_T-\tfrac{1}{T}ii')y
\end{aligned}
$$

(a) (注)

$$\begin{pmatrix} y_1-\overline{y} \\ y_2-\overline{y} \\ \vdots \\ y_T-\overline{y} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} - \begin{pmatrix} \overline{y} \\ \overline{y} \\ \vdots \\ \overline{y} \end{pmatrix} = y-\frac{1}{T}ii'y = (I_T-\frac{1}{T}ii')y$$

ただし，$i=(1,1,\cdots,1)'$ とする。

4. 行列表示で書き直すと，

$$R^2 = 1-\frac{\widehat{u}'\widehat{u}}{y'(I_T-\frac{1}{T}ii')y}$$

F 分布 (H_0： $\beta^*=0$) と決定係数 R^2 について：

1. 2 つの回帰式を考える。

$y_t = x_t\beta+u_t \Longrightarrow y_t=\beta_1+x_t^*\beta^*+u_t$

$y_t-\overline{y}=(x_t^*-\overline{x}^*)\beta^*+u_t^*$

ただし，

$$x_t=(x_{1t},x_{2t},\cdots,x_{kt})=(x_{1t},x_t^*),\quad x_{1t}=1,$$

$$\overline{x}^*=(\overline{x}_{2t},\cdots,\overline{x}_{kt}),$$

$$\beta=\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta^* \end{pmatrix},$$

とする。行列表示により，

$y_t=x_t\beta+u_t \Longrightarrow y=X\beta+u$

$y_t-\overline{y}=(x_t^*-\overline{x}^*)\beta^*+u_t^* \Longrightarrow y^*=X^*\beta^*+u^*$

2. $\widehat{u}_t=\widehat{u}_t^*$

[証明]

正規方程式から，

$\sum_{t=1}^T \widehat{u}_t=0,$

すなわち，

$\sum_{t=1}^T(y_t-\widehat{\beta}_1-\widehat{\beta}_2x_{2t}-\cdots-\widehat{\beta}_kx_{kt})=0,$

$\sum_{t=1}^T(y_t-\widehat{\beta}_1-x_t^*\widehat{\beta}^*)=0,$

$\overline{y}=\widehat{\beta}_1+\overline{x}^*\widehat{\beta}^*,$

を得る。

$y_t=\widehat{\beta}_1+x_t^*\widehat{\beta}^*+\widehat{u}_t$ を使うと，

$y_t-\overline{y}=(x_t^*-\overline{x}^*)\widehat{\beta}^*+\widehat{u}_t$

を得る。

3. 回帰モデル：

$y_t-\overline{y}=(x_t^*-\overline{x}^*)\beta^*+u_t^*$

について，

H_0： $\beta^*=0$ の検定は，

行列表示で，

$y^*=X^*\beta^*+u^*$

を用いると，

$$\frac{\widehat{\beta}^{*\prime}X^{*\prime}X^*\widehat{\beta}^*/(k-1)}{\widehat{u}^{*\prime}\widehat{u}^*/(T-k)} \sim F(k-1,T-k)$$

を得る。

4. さらに，

$$y^{*\prime}y^* = \widehat{\beta}^{*\prime}X^{*\prime}X^*\widehat{\beta}^*+\widehat{u}^{*\prime}\widehat{u}^*$$
$$y^*=(I_T-\frac{1}{T}ii')y$$
$$\widehat{u}^*=\widehat{u}$$

から，

$$\frac{\widehat{\beta}^{*\prime}X^{*\prime}X^*\widehat{\beta}^*/(k-1)}{\widehat{u}^{*\prime}\widehat{u}^*/(T-k)} = \frac{\Big(y'(I_T-\frac{1}{T}ii')y-\widehat{u}'\widehat{u}\Big)/(k-1)}{\widehat{u}'\widehat{u}/(T-k)} \sim F(k-1,T-k)$$

と変形される。

5. 一方，回帰モデル：

$y=X\beta+u$

のもとでの決定係数 R^2 は，

$$R^2=1-\frac{\widehat{u}'\widehat{u}}{y'(I_T-\frac{1}{T}ii')y}$$

6. したがって，F 分布は，さらに，

$$\frac{\widehat{\beta}^{*\prime}X^{*\prime}X^{*}\widehat{\beta}^{*}/(k-1)}{\widehat{u}^{*\prime}\widehat{u}^{*}/(T-k)}$$
$$=\frac{\Big(y'(I_T-\frac{1}{T}ii')y-\widehat{u}'\widehat{u}\Big)/(k-1)}{\widehat{u}'\widehat{u}/(T-k)}$$
$$=\frac{R^2/(k-1)}{(1-R^2)/(T-k)}\sim F(k-1,T-k)$$

と変形される。

回帰モデル：$y=X\beta+u$ の決定係数から H_0： $\beta^*=0$ の検定 (定数項を除く説明変数が有意かどうかの同時検定) が可能。

制約の検定 (F 検定)：

1. If $u\sim N(0,\sigma^2 I_T)$, then $\widehat{\beta}\sim N(\beta,\sigma^2(X'X)^{-1})$.

   Consider testing the hypothesis $H_0: R\beta=r$.

   R： $G\times k$, $\mathrm{rank}(R)=G\le k$.

   $R\widehat{\beta}\sim N(R\beta,\sigma^2R(X'X)^{-1}R')$.

   Therefore,
   $$\frac{(R\widehat{\beta}-r)'\big(R(X'X)^{-1}R'\big)^{-1}(R\widehat{\beta}-r)}{\sigma^2}\sim\chi^2(G).$$
   Note that $R\beta=r$.

   (a) $\widehat{\beta}\sim N(\beta,\sigma^2(X'X)^{-1})$ のとき，
   $\mathrm{E}(R\widehat{\beta})=R\mathrm{E}(\widehat{\beta})=R\beta$

   (b) $\widehat{\beta}\sim N(\beta,\sigma^2(X'X)^{-1})$ のとき，
   $$\begin{aligned}\mathrm{V}(R\widehat{\beta})&=\mathrm{E}\big((R\widehat{\beta}-R\beta)(R\widehat{\beta}-R\beta)'\big)\\&=\mathrm{E}\big(R(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'R'\big)\\&=R\mathrm{E}\big((\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'\big)R'\\&=R\mathrm{V}(\widehat{\beta})R'\\&=\sigma^2R(X'X)^{-1}R'\end{aligned}$$

2. We have the following:
$$\frac{\dfrac{(R\widehat{\beta}-r)'\big(R(X'X)^{-1}R'\big)^{-1}(R\widehat{\beta}-r)}{G}}{\dfrac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T-k}}\sim F(G,T-k)$$

3. いくつかの例：

   (a) t 検定：

   $G=1,\ r=0,\ R=(0,\cdots,1,\cdots,0)$ の場合 (R の第 i 番目の要素が 1 で，それ以外は 0)：

   すなわち，$\beta_i=0$ の検定：

   $s^2=\dfrac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T-k}$ とすると，
   $$\frac{\dfrac{(R\widehat{\beta}-r)'\big(R(X'X)^{-1}R'\big)^{-1}(R\widehat{\beta}-r)}{G}}{s^2}$$
   $$=\frac{\widehat{\beta}_i^2}{s^2a_{ii}}\sim F(1,T-k)$$
   ただし，

   $R\widehat{\beta}=\widehat{\beta}_i$

   $a_{ii}=$ $(X'X)^{-1}$ の i 行 i 列目の要素

   とする。

   したがって，$\beta_i=0$ の検定は，
   $$\frac{\widehat{\beta}_i}{s\sqrt{a_{ii}}}\sim t(T-k)$$
   によって行われる。

   *) Recall that $\big(t(T-k)\big)^2=F(1,T-k)$

   (b) 構造変化の検定 (その 1)：
   $$y_t=\begin{cases}x_t\beta_1+u_t, & t=1,2,\cdots,n\\ x_t\beta_2+u_t, & t=n+1,n+2,\cdots,T\end{cases}$$
   $u_i\sim N(0,\sigma^2)$ を仮定する。

   行列表示：
   $$\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\\y_{n+1}\\y_{n+2}\\\vdots\\y_T\end{pmatrix}=\begin{pmatrix}x_1&0\\x_2&0\\\vdots&\vdots\\x_n&0\\0&x_{n+1}\\0&x_{n+2}\\\vdots&\vdots\\0&x_T\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\\u_{n+1}\\u_{n+2}\\\vdots\\u_T\end{pmatrix}$$
   さらに行列表示：
   $$\begin{pmatrix}Y_1\\Y_2\end{pmatrix}=\begin{pmatrix}X_1&0\\0&X_2\end{pmatrix}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+u$$
   さらに行列表示：
   $$Y=X\beta+u$$

$H_0 : \beta_1 = \beta_2$ の検定：

$R = (I \ \ -I)$, $r = 0$ として，F 検定に当てはめる。

この場合，$G = \mathrm{rank}(R) = k$, β は $2k \times 1$ ベクトル。

分布は $F(k, T-2k)$

(c) 1 番目の係数と 2 番目の係数を足すと 1 となるという仮説：

$R = (1, 1, 0, \cdots, 0)$, $r = 1$

この場合，$G = \mathrm{rank}(R) = 1$

分布は $F(1, T-k)$

(d) 季節性があるかないかの検定：

モデル： 四半期データの場合

$$y = \alpha + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + X\beta_0 + u$$

「$D_i =$ 第 i 四半期に 1，その他は 0」というダミー変数

季節性の検定 $\Longrightarrow \alpha_1 = \alpha_2 = \alpha_3 = 0$ の検定

$\beta = (\alpha, \alpha_1, \alpha_2, \alpha_3, \beta_0')'$ の次元を k

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

$$r = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

この場合，$G = \mathrm{rank}(R) = 3$, β は $k \times 1$ ベクトル。

分布は $F(3, T-k)$

(e) コブ＝ダグラス型生産関数：

Q_t は生産量，K_t は資本，L_t は労働とする。生産関数を推定する。

$$\log(Q_t) = \beta_1' + \beta_2 \log(K_t) + \beta_3 \log(L_t) + u_t,$$

において，一次同時の制約 $\beta_2 + \beta_3 = 1$ を検定したい。すなわち，帰無仮説，対立仮説は以下のように表される。

帰無仮説 H_0： $\beta_2 + \beta_3 = 1$,
対立仮説 H_1： $\beta_2 + \beta_3 \neq 1$,

このとき，

$$R = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}, \quad r = 1$$

(f) 構造変化の検定 (その 2)：

n 期以前と $n+1$ 期以降とで経済構造が変化したと考えて推定を行う。しかも，定数項，傾き共に変化したと想定した場合，回帰式は以下のようになる。

$$Y_t = \alpha + \beta X_t + \gamma d_t + \delta d_t X_t + u_t,$$

ただし，

$$d_t = \begin{cases} 0, & t = 1, 2, \cdots, n \text{ のとき}, \\ 1, & t = n+1, n+2, \cdots, T \text{ のとき}, \end{cases}$$

とする。構造変化が $n+1$ 期で起こったかどうかを検定したい。すなわち，帰無仮説，対立仮説は以下のように表される。

帰無仮説 H_0： $\gamma = \delta = 0$,
対立仮説 H_1： $\gamma \neq 0$，または，$\delta \neq 0$,

このとき，

$$R = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(g) 多重回帰モデルの係数の同時検定：

2 つの説明変数が含まれる場合を考える。

$$Y_t = \alpha + \beta X_t + \gamma Z_t + u_t,$$

のモデルにおいて，X_t と Z_t のどちらも，Y_t に影響を与えていないという仮説を検定したい。この場合，帰無仮説，対立仮説は以下のように表される。

帰無仮説 H_0： $\beta = \gamma = 0$,
対立仮説 H_1： $\beta \neq 0$，または，$\gamma \neq 0$,

このとき，

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

その他：

1. Define $\widehat{u}$ as $\widehat{u} = y - X\widehat{\beta}$. The coefficient of determinant, R^2, is

$$R^2 = 1 - \frac{\widehat{u}'\widehat{u}}{y'My},$$

where $M = I - \dfrac{1}{T} i i'$, I is a $T \times T$ identity matrix and i is a $T \times 1$ vector consisting of 1, i.e., $i = (1, 1, \cdots, 1)'$. Note that

$$\begin{pmatrix} y_1 - \overline{y} \\ y_2 - \overline{y} \\ \vdots \\ y_T - \overline{y} \end{pmatrix} = My$$

M is idempotent.

2. Durbin=Watson ratio, DW, is defined as

$$DW = \frac{\sum_{t=2}^{T} (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T} \widehat{u}_t^2} = \frac{\widehat{u}' A \widehat{u}}{\widehat{u}' \widehat{u}},$$

where $\widehat{u} = (\widehat{u}_1, \widehat{u}_2, \cdots, \widehat{u}_T)'$.

$$A = \begin{pmatrix} 1 & -1 & 0 & & \cdots & & 0 \\ -1 & 2 & -1 & & \cdots & & 0 \\ 0 & -1 & 2 & -1 & \cdots & & 0 \\ \vdots & & & & \ddots & & \\ \vdots & & & & & \ddots & \\ 0 & & & & -1 & 2 & -1 \end{pmatrix}$$

4 制約付き最小自乗法

1. 制約 $R\beta = r$ のもとで $(y - X\beta)'(y - X\beta)$ の最小化問題，L をラグランジェ関数とする。

$$L = (y - X\beta)'(y - X\beta) - 2\lambda'(R\beta - r)$$

を最小にする β, λ の解を β^*, λ^* とする。

$$\frac{\partial L}{\partial \beta} = -2X'(y - X\beta^*) - 2R'\lambda^* = 0$$

$$\frac{\partial L}{\partial \lambda} = -2(R\beta^* - r) = 0$$

$\partial L / \partial \beta = 0$ から,

$$\begin{aligned} \beta^* &= (X'X)^{-1} X'y + (X'X)^{-1} R' \lambda^* \\ &= \widehat{\beta} + (X'X)^{-1} R' \lambda^* \end{aligned}$$

さらに,

$$R\beta^* = R\widehat{\beta} + R(X'X)^{-1} R' \lambda^*$$

$$r = R\widehat{\beta} + R(X'X)^{-1} R' \lambda^*$$

したがって,

$$\lambda^* = \big(R(X'X)^{-1} R' \big)^{-1} (r - R\widehat{\beta})$$

代入して，制約付き最小自乗法の推定値は,

$$\beta^* = \widehat{\beta} + (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} (r - R\widehat{\beta})$$

となる。

(a) 期待値は,

$$\begin{aligned} & \mathrm{E}(\beta^*) \\ &= \mathrm{E}(\widehat{\beta}) \\ &+ (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} \big(r - R\mathrm{E}(\widehat{\beta}) \big) \\ &= \beta \\ &+ (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} \big(r - R\beta \big) \\ &= \beta \end{aligned}$$

となる。

(b) 分散は,

$$\begin{aligned} & (\beta^* - \beta) \\ &= (\widehat{\beta} - \beta) \\ &+ (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} (R\beta - R\widehat{\beta}) \\ &= (\widehat{\beta} - \beta) \\ &- (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} (R\widehat{\beta} - R\beta) \\ &= (\widehat{\beta} - \beta) \\ &- (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} R(\widehat{\beta} - \beta) \\ &= [I - (X'X)^{-1} R' \big(R(X'X)^{-1} R' \big)^{-1} R](\widehat{\beta} - \beta) \\ &= W(\widehat{\beta} - \beta) \end{aligned}$$

と変形されるので,

$$\begin{aligned} & \mathrm{V}(\beta^*) \\ &\equiv \mathrm{E}\big((\beta^* - \beta)(\beta^* - \beta)' \big) \\ &= \mathrm{E}\big(W(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' W' \big) \\ &= W\mathrm{E}\big((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)' \big) W' \\ &= W\mathrm{V}(\widehat{\beta}) W' \end{aligned}$$

$$
\begin{aligned}
&= \sigma^2 W(X'X)^{-1}W' \\
&= \sigma^2 \Big(I - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R\Big) \\
&\qquad (X'X)^{-1} \\
&\qquad \Big(I - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R\Big)' \\
&= \sigma^2 \Big(I - (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R\Big) \\
&\qquad (X'X)^{-1} \\
&\qquad \Big(I - R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}\Big) \\
&= \sigma^2 \Big((X'X)^{-1} \\
&\qquad -(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}\Big) \\
&\qquad \Big(I - R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}\Big) \\
&= \sigma^2 \Big((X'X)^{-1} \\
&\qquad -(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1} \\
&\qquad -(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1} \\
&\qquad +(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}R' \\
&\qquad\quad \left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}\Big) \\
&= \sigma^2 \Big((X'X)^{-1} \\
&\qquad -(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}\Big) \\
&= \sigma^2(X'X)^{-1} \\
&\quad - \sigma^2(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1} \\
&= \mathrm{V}(\widehat{\beta}) - \sigma^2(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}R(X'X)^{-1}
\end{aligned}
$$

となる。

2. 別解：

(a) もう一度書くと,

$$
\frac{\partial L}{\partial \beta} = -2X'(y - X\beta^*) - 2R'\lambda^* = 0
$$

$$
\frac{\partial L}{\partial \lambda} = -2(R\beta^* - r) = 0
$$

から,

$$
X'X\beta^* - R'\lambda^* = X'y
$$

$$
R\beta^* = r
$$

行列表示して,

$$
\begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}\begin{pmatrix} \beta^* \\ -\lambda^* \end{pmatrix} = \begin{pmatrix} X'y \\ r \end{pmatrix}
$$

$$
\begin{pmatrix} \beta^* \\ -\lambda^* \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1}\begin{pmatrix} X'y \\ r \end{pmatrix}
$$

(b) 逆行列の公式

$$
\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} E & F \\ F' & G \end{pmatrix}
$$

E, F, G は次に与えられる。

$$
\begin{aligned}
E &= (A - BD^{-1}B')^{-1} \\
&= A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A^{-1}
\end{aligned}
$$

$$
\begin{aligned}
F &= -(A - BD^{-1}B')^{-1}BD^{-1} \\
&= -A^{-1}B(D - B'A^{-1}B)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
G &= (D - B'A^{-1}B)^{-1} \\
&= D^{-1} + D^{-1}B'(A - BD^{-1}B')^{-1}BD^{-1}
\end{aligned}
$$

(c) E, F は,

$$
\begin{aligned}
E &= (X'X)^{-1} \\
&\quad - (X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)R(X'X)^{-1} \\
F &= (X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)
\end{aligned}
$$

なので,

$$
\beta^* = EX'y + Fr
$$

よって,

$$
\beta^* = \widehat{\beta} + (X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(r - R\widehat{\beta})
$$

となる。

(d) 分散は,

$$
\mathrm{V}\begin{pmatrix} \beta^* \\ -\lambda^* \end{pmatrix} = \sigma^2\begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1}
$$

したがって,

$$
\mathrm{E}(\beta^*) = \sigma^2 E
$$

すなわち,

$$
\begin{aligned}
&\mathrm{V}(\beta^*) \\
&= \sigma^2\Big((X'X)^{-1} \\
&\qquad -(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)R(X'X)^{-1}\Big)
\end{aligned}
$$

(e) $R\beta = r$ が正しいもとでは,

$$
\begin{aligned}
&\mathrm{V}(\widehat{\beta}) - \mathrm{V}(\beta^*) \\
&= \sigma^2(X'X)^{-1}R'\big(R(X'X)^{-1}R'\big)R(X'X)^{-1}
\end{aligned}
$$

は正値定符号行列になる。

5 F 分布 (制約付き最小自乗法と制約なし最小自乗法との関係)

1. 前述の通り，

$$\frac{\dfrac{(R\widehat{\beta}-r)'\big(R(X'X)^{-1}R'\big)^{-1}(R\widehat{\beta}-r)}{G}}{\dfrac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T-k}} \sim F(G,T-k)$$

分子は，以下のように書き換えられる。

$$\begin{aligned}&(R\widehat{\beta}-r)'\big(R(X'X)^{-1}R'\big)^{-1}(R\widehat{\beta}-r)\\=&(\widehat{\beta}-\beta^*)'(X'X)(\widehat{\beta}-\beta^*)\end{aligned}$$

なぜなら，

$$\begin{aligned}\beta^* =& \widehat{\beta}\\&+(X'X)^{-1}R'\left(R(X'X)^{-1}R'\right)^{-1}(r-R\widehat{\beta})\end{aligned}$$

を思い起こせ。

さらに，

$$\begin{aligned}&(y-X\beta^*)'(y-X\beta^*)\\=&\big(y-X(\beta^*-\widehat{\beta})-X\widehat{\beta}\big)'\\&\quad\times\big(y-X(\beta^*-\widehat{\beta})-X\widehat{\beta}\big)\\=&(y-X\widehat{\beta})'(y-X\widehat{\beta})\\&+(\beta^*-\widehat{\beta})'X'X(\beta^*-\widehat{\beta})\\&-(y-X\widehat{\beta})'X(\beta^*-\widehat{\beta})\\&-(\beta^*-\widehat{\beta})'X'(y-X\widehat{\beta})\\=&(y-X\widehat{\beta})'(y-X\widehat{\beta})\\&+(\beta^*-\widehat{\beta})'X'X(\beta^*-\widehat{\beta})\end{aligned}$$

($X'\widehat{u}=0$ が用いられる。)

したがって，

$$\begin{aligned}&(\beta^*-\widehat{\beta})'X'X(\beta^*-\widehat{\beta})\\=&(y-X\beta^*)'(y-X\beta^*)-(y-X\widehat{\beta})'(y-X\widehat{\beta})\end{aligned}$$

を得る。

$\widehat{u}$, u^* をそれぞれ制約付き残差，制約なし残差として，以下のように定義する。

$$\widehat{u}=y-X\widehat{\beta}$$

$$u^*=y-X\beta^*$$

もとの，検定統計量に代入して，

$$\begin{aligned}&\frac{\dfrac{(R\widehat{\beta}-r)'\big(R(X'X)^{-1}R'\big)^{-1}(R\widehat{\beta}-r)}{G}}{\dfrac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T-k}}\\=&\frac{\dfrac{(\beta^*-\widehat{\beta})'X'X(\beta^*-\widehat{\beta})}{G}}{\dfrac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T-k}}\\=&\frac{\dfrac{(y-X\beta^*)'(y-X\beta^*)-(y-X\widehat{\beta})'(y-X\widehat{\beta})}{G}}{\dfrac{(y-X\widehat{\beta})'(y-X\widehat{\beta})}{T-k}}\\=&\frac{(u^{*\prime}u^*-\widehat{u}'\widehat{u})/G}{\widehat{u}'\widehat{u}/(T-k)}\end{aligned}$$

6 例： F 分布 (制約付き最小自乗法と制約なし最小自乗法との関係)

データ・ファイル (cons99.txt)

```
1955   5430.1   6135.0  18.1
1956   5974.2   6828.4  18.3
1957   6686.3   7619.5  19.0
1958   7169.7   8153.3  19.1
1959   8019.3   9274.3  19.7
1960   9234.9  10776.5  20.5
1961  10836.2  12869.4  21.8
1962  12430.8  14701.4  23.2
1963  14506.6  17042.7  24.9
1964  16674.9  19709.9  26.0
1965  18820.5  22337.4  27.8
1966  21680.6  25514.5  29.0
1967  24914.0  29012.6  30.1
1968  28452.7  34233.6  31.6
1969  32705.2  39486.3  32.9
1970  37784.1  45913.2  35.2
1971  42571.6  51944.3  37.5
1972  49124.1  60245.4  39.7
1973  59366.1  74924.8  44.1
1974  71782.1  93833.2  53.3
1975  83591.1 108712.8  59.4
1976  94443.7 123540.9  65.2
1977 105397.8 135318.4  70.1
1978 115960.3 147244.2  73.5
```

```
1979 127600.9 157071.1  76.0
1980 138585.0 169931.5  81.6
1981 147103.4 181349.2  85.4
1982 157994.0 190611.5  87.7
1983 166631.6 199587.8  89.5
1984 175383.4 209451.9  91.8
1985 185335.1 220655.6  93.9
1986 193069.6 229938.8  94.8
1987 202072.8 235924.0  95.3
1988 212939.9 247159.7  95.8
1989 227122.2 263940.5  97.7
1990 243035.7 280133.0 100.0
1991 255531.8 297512.9 102.5
1992 265701.6 309256.6 104.5
1993 272075.3 317021.6 105.9
1994 279538.7 325655.7 106.7
1995 283245.4 331967.5 106.2
1996 291458.5 340619.1 106.0
1997 298475.2 345522.7 107.3
```

左から，年，名目家計最終消費支出 (10 億円)，家計可処分所得 (10 億円)，家計最終消費支出デフレータ (1990 年 =100)

```
         PROGRAM
LINE  ************************************************
|       freq a;
|       smpl 1955 1997;
|       read(file='cons99.txt') year cons yd price;
|       rcons=cons/(price/100);
|       ryd=yd/(price/100);
|       d1=0.0;
|       smpl 1974 1997;
|       d1=1.0;
|       smpl 1956 1997;
|       d1ryd=d1*ryd;
|       dcons=rcons-rcons(-1);
|       olsq rcons c ryd;
|       olsq rcons c d1 ryd d1ryd;
|       olsq rcons c ryd rcons(-1);
|       olsq dcons c;
|       end;
         EXECUTION
******************************************************


                 Equation   1
                 ============

  Method of estimation = Ordinary Least Squares


Dependent variable: RCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 149038.
Std. dev. of dependent var. = 78147.9
   Sum of squared residuals = .127951E+10
      Variance of residuals = .319878E+08
   Std. error of regression = 5655.77
                  R-squared = .994890
         Adjusted R-squared = .994762
    Durbin-Watson statistic = .116873
  F-statistic (zero slopes) = 7787.70
 Schwarz Bayes. Info. Crit. = 17.4101
 Log of likelihood function = -421.469

          Estimated    Standard
Variable  Coefficient    Error       t-statistic
C         -3317.80      1934.49      -1.71508
RYD       .854577       .968382E-02  88.2480


                 Equation   2
                 ============

  Method of estimation = Ordinary Least Squares


Dependent variable: RCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 149038.
Std. dev. of dependent var. = 78147.9
   Sum of squared residuals = .244501E+09
      Variance of residuals = .643423E+07
   Std. error of regression = 2536.58
                  R-squared = .999024
         Adjusted R-squared = .998946
    Durbin-Watson statistic = .420979
  F-statistic (zero slopes) = 12959.1
 Schwarz Bayes. Info. Crit. = 15.9330
 Log of likelihood function = -386.714

          Estimated    Standard
Variable  Coefficient    Error       t-statistic
C         4204.11       1440.45      2.91861
D1        -39915.3      3154.24      -12.6545
RYD       .786609       .015024      52.3561
D1RYD     .194495       .018731      10.3839


                 Equation   3
                 ============

  Method of estimation = Ordinary Least Squares


Dependent variable: RCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 149038.
Std. dev. of dependent var. = 78147.9
   Sum of squared residuals = .246205E+09
      Variance of residuals = .631296E+07
   Std. error of regression = 2512.56
                  R-squared = .999017
         Adjusted R-squared = .998966
    Durbin-Watson statistic = 1.25472
                 Durbin's h = 2.62625
     Durbin's h alternative = 2.44578
  F-statistic (zero slopes) = 19812.0
 Schwarz Bayes. Info. Crit. = 15.8510
 Log of likelihood function = -386.860

           Estimated    Standard
Variable   Coefficient    Error       t-statistic
C          3281.37       1002.31      3.27383
RYD        .150357       .055212      2.72328
```

```
RCONS(-1)  .831071       .064959       12.7938


                 Equation   4
                 ============

  Method of estimation = Ordinary Least Squares


Dependent variable: DCONS
Current sample:  1956 to 1997
Number of observations:  42

 Mean of dependent variable = 5908.77
Std. dev. of dependent var. = 2734.81
   Sum of squared residuals = .306647E+09
      Variance of residuals = .747919E+07
   Std. error of regression = 2734.81
                  R-squared = .136129E-49
         Adjusted R-squared = 0.
    Durbin-Watson statistic = 1.30871
 Schwarz Bayes. Info. Crit. = 15.8925
 Log of likelihood function = -391.470

           Estimated    Standard
Variable  Coefficient     Error       t-statistic
C         5908.77       421.991       14.0021

******************************************************
```

1. `Equation 1` vs. `Equation 2`

   構造変化の検定 (1974 年以降と以前とでは経済構造は変化したかどうか)：

   `Equation 2` は

   $$\texttt{RCONS} = \beta_1 + \beta_2\texttt{D1} + \beta_3\texttt{RYD} + \beta_4\texttt{RYD} \times \texttt{D1}$$

   $H_0:\ \beta_2 = \beta_4 = 0$

   制約付き OLS $\Longrightarrow$ `Equation 1`

   制約なし OLS $\Longrightarrow$ `Equation 2`

   $$\begin{aligned}&\frac{(u^{*\prime}u^* - \widehat{u}'\widehat{u})/G}{\widehat{u}'\widehat{u}/(T-k)}\\ =&\frac{(.127951\mathrm{E}+10 - .244501\mathrm{E}+09)/2}{.244501\mathrm{E}+09/(42-4)}\\ =&\,80.43\\ &\sim F(2,38)\end{aligned}$$

   $F(2,38)$ の上側 1% 点 $= 5.211 < 80.43$ なので，

   $H_0:\ \beta_2 = \beta_4 = 0$ を棄却

   $\Longrightarrow$ 1974 年で経済構造は変化したと言える。

2. `Equation 1` vs. `Equation 3`

   有意性の検定：

   `Equation 3` は

   $$\texttt{RCONS} = \beta_1 + \beta_2\texttt{RYD} + \beta_3\texttt{RCONS}(-1)$$

   $H_0:\ \beta_3 = 0$

   制約付き OLS $\Longrightarrow$ `Equation 1`

   制約なし OLS $\Longrightarrow$ `Equation 3`

   $$\begin{aligned}&\frac{(u^{*\prime}u^* - \widehat{u}'\widehat{u})/G}{\widehat{u}'\widehat{u}/(T-k)}\\ =&\frac{(.127951\mathrm{E}+10 - .246205\mathrm{E}+09)/1}{.246205\mathrm{E}+09/(42-3)}\\ =&\,163.68\\ &\sim F(1,39)\end{aligned}$$

   $F(1,39)$ の上側 1% 点 $= 7.333 < 163.68$ なので，

   $H_0:\ \beta_3 = 0$ を棄却

   $\Longrightarrow$ `RCONS(-1)` は `RCONS` に有意に影響を与えると言える。

   $\sqrt{163.68} = 12.7938$ となり，これは `RCONS(-1)` の `t-statistic` に等しい。

3. `Equation 3` vs. `Equation 4`

   有意性の同時検定：

   `Equation 3` は

   $$\texttt{RCONS} = \beta_1 + \beta_2\texttt{RYD} + \beta_3\texttt{RCONS}(-1)$$

   $H_0:\ \beta_2 = 0$ and $\beta_3 = 1$

   制約付き OLS $\Longrightarrow$ `Equation 4`

   制約なし OLS $\Longrightarrow$ `Equation 3`

   $$\begin{aligned}&\frac{(u^{*\prime}u^* - \widehat{u}'\widehat{u})/G}{\widehat{u}'\widehat{u}/(T-k)}\\ =&\frac{(.306647\mathrm{E}+09 - .246205\mathrm{E}+09)/2}{.246205\mathrm{E}+09/(42-3)}\\ =&\,4.910\\ &\sim F(2,39)\end{aligned}$$

   $F(2,39)$ の上側 1% 点 $= 5.194 > 4.910$ なので，

   $H_0:\ \beta_2 = 0$ and $\beta_3 = 1$

   を棄却できない。

7 一般化最小自乗法 (GLS)

1. Regression model: $y = X\beta + u$, $u \sim (0, \sigma^2\Omega)$

2. 不等分散 (Heteroscedasticity)

$$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_T^2 \end{pmatrix}$$

一階の自己相関 (First-Order Autocorrelation)

$u_t = \rho u_{t-1} + \epsilon_t$, $\epsilon_t \sim$ iid $N(0, \sigma_\epsilon^2)$

$$\sigma^2\Omega = \frac{\sigma_\epsilon^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}$$

$$\mathrm{V}(u_t) = \sigma^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$$

3. β の GLS 推定値 b は以下の問題を解くことに等しい。

$$\min_\beta \ (y - X\beta)'\Omega^{-1}(y - X\beta)$$

GLSE of β is b .

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

4. 一般的に，Ω が対称行列のとき，Ω は以下のように分解される。

$$\Omega = A'\Lambda A$$

Λ は対角要素が固有値となる対角行列，A は固有ベクトルから成る行列とする。

Ω が正値定符号行列のとき，Λ の対角要素はすべて正となる。(正値定符号行列とは，任意のベクトル x について，$x'\Omega x > 0$ となる行列 Ω である)

5. There exists P shuch that $\Omega = PP'$

(take $P = \Lambda^{1/2}A$).

Multiply P^{-1} on both sides of $y = X\beta + u$.

We have:

$y^\star = X^\star\beta + u^\star$,

where

$y^\star = P^{-1}y$,

$X^\star = P^{-1}X$, and

$u^\star = P^{-1}u$.

Note that

$\mathrm{V}(u^\star)$

$= \mathrm{V}(P^{-1}u)$

$= P^{-1}\mathrm{V}(u)P'^{-1}$

$= \sigma^2 P^{-1}\Omega P'^{-1} = I$,

because $\Omega = PP'$.

Accordingly, the regression model is rewritten as:

$$y^\star = X^\star\beta + u^\star, \qquad u^\star \sim (0, \sigma^2 I)$$

Apply OLS to the above model. That is,

$$\min_\beta \ (y^\star - X^\star\beta)'(y^\star - X^\star\beta)$$

is equivalent to:

$$\min_\beta \ (y - X\beta)'\Omega^{-1}(y - X\beta)$$

$$\begin{aligned} b &= (X^{\star\prime}X^\star)^{-1}X^{\star\prime}y^\star \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \end{aligned}$$

$$\begin{aligned} b &= \beta + (X^{\star\prime}X^\star)^{-1}X^{\star\prime}u^\star \\ &= \beta + (X^{\star\prime}X^\star)^{-1}X'\Omega^{-1}u \end{aligned}$$

$$\mathrm{E}(b) = \beta$$

$$\begin{aligned} \mathrm{V}(b) &= \sigma^2(X^{\star\prime}X^\star)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1} \end{aligned}$$

6. 回帰モデルが

$$y = X\beta + u, \qquad u \sim N(0, \sigma^2\Omega),$$

のとき，最小自乗法を適用する。

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'u\end{aligned}$$

$$\mathrm{V}(\widehat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

GLS と OLS との比較

(a) 期待値について

$$\mathrm{E}(\widehat{\beta}) = \beta$$

$$\mathrm{E}(b) = \beta$$

となり，$\widehat{\beta}$ も b も共に不偏推定値

(b) 分散について

$$\mathrm{V}(\widehat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

$$\mathrm{V}(b) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

となり，どちらが有効推定値か？

$$\begin{aligned}&\mathrm{V}(\widehat{\beta}) - \mathrm{V}(b) \\ =\ &\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &-\sigma^2(X'\Omega^{-1}X)^{-1} \\ =\ &\sigma^2\Big((X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Big) \\ &\times\Omega \\ &\times\Big((X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Big)' \\ =\ &\sigma^2 A\Omega A'\end{aligned}$$

Ω は u の分散共分散行列により正値定符号行列。よって，$\Omega = I_T$ でなければ， $A\Omega A'$ も正値定符号行列。b が $\widehat{\beta}$ より有効。

7. If $u \sim N(0, \sigma^2\Omega)$, then $b \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$.

   Consider testing the hypothesis $H_0 : R\beta = r$.

   $R : G \times k$, $\mathrm{rank}(R) = G \leq k$.

   $R\widehat{\beta} \sim N(R\beta, \sigma^2 R(X'\Omega^{-1}X)^{-1}R')$.

   Therefore,

$$\frac{(R\widehat{\beta} - r)'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}(R\widehat{\beta} - r)}{\sigma^2} \sim \chi^2(G)$$

8. $$\frac{(y - Xb)'\Omega^{-1}(y - Xb)}{\sigma^2} \sim \chi^2(T - k)$$

9. We have:

$$\frac{\dfrac{(R\widehat{\beta} - r)'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}(R\widehat{\beta} - r)}{G}}{\dfrac{(y - Xb)'\Omega^{-1}(y - Xb)}{T - k}} \sim F(G, T - k)$$

b を制約なし一般化最小自乗推定量，b^* を制約付き一般化最小自乗推定量とし，それに対応する残差をそれぞれ e，e^* とする。

$$e = y - Xb, \qquad e^* = y - Xb^*$$

このとき，F 検定統計量は以下のように書き換えられる。

$$\frac{(e^{*\prime}\Omega^{-1}e^* - e'\Omega^{-1}e)/G}{e'\Omega^{-1}e/(T - k)} \sim F(G, T - k)$$

7.1 例： 混合推定 (Theil and Goldberger Model)

制約付き最小二乗推定量の一般化：
$\Longrightarrow$ 確率的線形制約

$$r = R\beta + v, \quad v \sim (0, \Psi)$$

$$y = X\beta + u, \quad u \sim (0, \Omega)$$

２つを行列表示

$$\begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix}\beta + \begin{pmatrix} u \\ v \end{pmatrix}$$

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim (0, \begin{pmatrix} \Omega & 0 \\ 0 & \Psi \end{pmatrix})$$

一般化最小二乗法を適用

$$\begin{aligned}\widehat{\beta}^* &= \left(\begin{pmatrix} X' & R' \end{pmatrix}\begin{pmatrix} \Omega & 0 \\ 0 & \Psi \end{pmatrix}^{-1}\begin{pmatrix} X \\ R \end{pmatrix}\right)^{-1} \\ &\times\left(\begin{pmatrix} X' & R' \end{pmatrix}\begin{pmatrix} \Omega & 0 \\ 0 & \Psi \end{pmatrix}^{-1}\begin{pmatrix} y \\ r \end{pmatrix}\right) \\ &= (X'\Omega^{-1}X + R'\Psi^{-1}R)^{-1}(X'\Omega^{-1}y + R'\Psi^{-1}r)\end{aligned}$$

$$
\begin{aligned}
\mathrm{V}(\widehat{\beta}^*) &= \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} \Omega & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \\
&= (X'\Omega^{-1}X + R'\Psi^{-1}R)^{-1}
\end{aligned}
$$

8 最尤法 (MLE)

最尤法 = Maximum Likelihood Estimation (MLE)

1. The distribution function of $\{x_i\}_{i=1}^T$ is $f(x;\theta)$, where $x = (x_1, \cdots, x_T)$ and $\theta = (\mu, \Sigma)$. Likelihood function $L(\cdot)$ is defined as $L(\theta;x) = f(x;\theta)$. Maximum likelihood estimate (MLE) of θ is θ such that:

$$\max_{\theta} \ L(\theta;x).$$

MLE satisfies the following:

(a) $$\frac{\partial \log L(\theta;x)}{\partial\theta} = 0.$$

(b) $$\frac{\partial^2 \log L(\theta;x)}{\partial\theta\partial\theta'}$$
is a negative definite matrix.

2. Fisher's information matrix is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta;x)}{\partial\theta\partial\theta'}\Big)$$

Note as follows:

$$
\begin{aligned}
&-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta;x)}{\partial\theta\partial\theta'}\Big) \\
&= \mathrm{E}\Big(\frac{\partial \log L(\theta;x)}{\partial\theta}\frac{\partial \log L(\theta;x)}{\partial\theta'}\Big) \\
&= \mathrm{V}\Big(\frac{\partial \log L(\theta;x)}{\partial\theta}\Big)
\end{aligned}
$$

証明：

$$\int L(\theta;x)\mathrm{d}x = 1$$

θ について微分

$$\int \frac{\partial L(\theta;x)}{\partial\theta}\mathrm{d}x = 0$$

(x の範囲は θ に依存しないもの，微分 $\partial L/\partial\theta$ が存在するものと仮定される)

上式の変形により

$$\int \frac{\partial \log L(\theta;x)}{\partial\theta} L(\theta;x)\mathrm{d}x = 0$$

すなわち，

$$\mathrm{E}\left(\frac{\partial \log L(\theta;x)}{\partial\theta}\right) = 0$$

さらに，θ について微分

$$
\begin{aligned}
&\int \frac{\partial^2 \log L(\theta;x)}{\partial\theta\partial\theta'} L(\theta;x)\mathrm{d}x \\
&+ \int \frac{\partial \log L(\theta;x)}{\partial\theta}\frac{\partial L(\theta;x)}{\partial'\theta}\mathrm{d}x \\
&= \int \frac{\partial^2 \log L(\theta;x)}{\partial\theta\partial\theta'} L(\theta;x)\mathrm{d}x \\
&+ \int \frac{\partial \log L(\theta;x)}{\partial\theta}\frac{\partial \log L(\theta;x)}{\partial\theta'} L(\theta;x)\mathrm{d}x
\end{aligned}
$$

よって，

$$
\begin{aligned}
&-\mathrm{E}\left(\frac{\partial^2 \log L(\theta;x)}{\partial\theta\partial\theta'}\right) \\
&= \mathrm{E}\left(\frac{\partial \log L(\theta;x)}{\partial\theta}\frac{\partial \log L(\theta;x)}{\partial\theta'}\right) \\
&= \mathrm{V}\left(\frac{\partial \log L(\theta;x)}{\partial\theta}\right)
\end{aligned}
$$

を得る。

3. Cramer-Rao の下限 $I(\theta)$:

今，θ の推定量を $s(x)$ とおく。

$$\mathrm{E}(s(x)) = \int s(x)L(\theta;x)\mathrm{d}x$$

θ について微分

$$
\begin{aligned}
\frac{\partial \mathrm{E}(s(x))}{\partial\theta} &= \int s(x)\frac{\partial L(\theta;x)}{\partial\theta}\mathrm{d}x \\
&= \int s(x)\frac{\partial \log L(\theta;x)}{\partial\theta} L(\theta;x)\mathrm{d}x \\
&= \mathrm{Cov}\left(s(x), \frac{\partial \log L(\theta;x)}{\partial\theta}\right)
\end{aligned}
$$

簡単化のために，$s(x)$，θ をスカラーとする。このとき，

$$\left(\frac{\partial \mathrm{E}(s(x))}{\partial\theta}\right)^2$$

$$
\begin{aligned}
&= \left(\mathrm{Cov}\left(s(x), \frac{\partial \log L(\theta;x)}{\partial \theta}\right)\right)^2 \\
&= \rho^2 \mathrm{V}\left(s(x)\right) \mathrm{V}\left(\frac{\partial \log L(\theta;x)}{\partial \theta}\right) \\
&\le \mathrm{V}\left(s(x)\right) \mathrm{V}\left(\frac{\partial \log L(\theta;x)}{\partial \theta}\right)
\end{aligned}
$$

ただし，ρ は $s(x)$ と $\dfrac{\partial \log L(\theta;x)}{\partial \theta}$ との相関係数とする。すなわち，$-1 \le \rho \le 1$

$$
\rho = \frac{\mathrm{Cov}\left(s(x), \dfrac{\partial \log L(\theta;x)}{\partial \theta}\right)}{\sqrt{\mathrm{V}\left(s(x)\right)}\sqrt{\mathrm{V}\left(\dfrac{\partial \log L(\theta;x)}{\partial \theta}\right)}}
$$

よって，

$$
\left(\frac{\partial \mathrm{E}(s(x))}{\partial \theta}\right)^2 \le \mathrm{V}(s(x))\ \mathrm{V}\left(\frac{\partial \log L(\theta;x)}{\partial \theta}\right)
$$

すなわち，

$$
\mathrm{V}(s(x)) \ge \frac{\left(\dfrac{\partial \mathrm{E}(s(x))}{\partial \theta}\right)^2}{\mathrm{V}\left(\dfrac{\partial \log L(\theta;x)}{\partial \theta}\right)}
$$

$\mathrm{E}(s(x)) = \theta$ のとき，

$$
\mathrm{V}(s(x)) \ge \frac{1}{-\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta;x)}{\partial \theta^2}\right)} = (I(\theta))^{-1}
$$

$s(x)$ がベクトルの場合でも以下の式が成り立つ。

$$
\mathrm{V}(s(x)) \ge (I(\theta))^{-1}
$$

4. 最尤推定値の漸近的正規性：

T が大きくなるにつれて，

$$
\sqrt{T}(\widetilde{\theta} - \theta) \ \longrightarrow \ N\left(0, \lim_{T\to\infty}\left(\frac{I(\theta)}{T}\right)^{-1}\right)
$$

が成り立つ。

すなわち，T が大きいとき，

$$
\widetilde{\theta} \sim N\left(\theta, \big(I(\theta)\big)^{-1}\right)
$$

とすればよい。

$s(x) = \widetilde{\theta}$ とする。このとき，T が大きいとき，$\mathrm{V}\big(s(x)\big)$ は $\big(I(\theta)\big)^{-1}$ に近づく。

5. 最適化方法：

$$
\begin{aligned}
\frac{\partial \log L(\theta;x)}{\partial \theta} &= 0 \\
&= \frac{\partial \log L(\theta^*;x)}{\partial \theta} \\
&\quad + \frac{\partial^2 \log L(\theta^*;x)}{\partial \theta \partial \theta'}(\theta - \theta^*)
\end{aligned}
$$

$$
\theta = \theta^* - \left(\frac{\partial^2 \log L(\theta^*;x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log L(\theta^*;x)}{\partial \theta}
$$

Replace the variables as follows:

$$
\begin{aligned}
\theta &\longrightarrow \theta^{(i+1)} \\
\theta^* &\longrightarrow \theta^{(i)}
\end{aligned}
$$

Then, we have:

$$
\begin{aligned}
\theta^{(i+1)} &= \theta^{(i)} \\
&\quad - \left(\frac{\partial^2 \log L(\theta^{(i)};x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log L(\theta^{(i)};x)}{\partial \theta}
\end{aligned}
$$

$\Longrightarrow$ ニュートンラプソン法

また，

$$
\frac{\partial^2 \log L(\theta^{(i)};x)}{\partial \theta \partial \theta'} \longrightarrow \mathrm{E}\left(\frac{\partial^2 \log L(\theta^{(i)};x)}{\partial \theta \partial \theta'}\right)
$$

として置き代えると，

$$
\begin{aligned}
\theta^{(i+1)} &= \theta^{(i)} \\
&\quad - \left(\mathrm{E}\left(\frac{\partial^2 \log L(\theta^{(i)};x)}{\partial \theta \partial \theta'}\right)\right)^{-1} \frac{\partial \log L(\theta^{(i)};x)}{\partial \theta} \\
&= \theta^{(i)} - \Big(I(\theta^{(i)})\Big)^{-1} \frac{\partial \log L(\theta^{(i)};x)}{\partial \theta}
\end{aligned}
$$

$\Longrightarrow$ スコア法

8.1 回帰モデルの最尤法： 2 変数の場合

回帰モデル：

$$y_t = \alpha + \beta x_t + u_t,$$

について，

1. u_t に正規分布を仮定する。すなわち， $u_t \sim N(0,\sigma^2)$ とする。

2. u_t の分布関数は

$$f(u_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}u_t^2\right)$$

となる。$u_1, u_2, \cdots, u_T$ の結合分布は，それぞれが独立に分布していることに注意すると，次のように書き直される。

$$\begin{aligned} & f(u_1, u_2, \cdots, u_T) \\ &= f(u_1)f(u_2)\cdots f(u_T) \\ &= \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T} u_t^2\right) \end{aligned}$$

3. $y_1, \cdots, y_T$ の結合分布は，変数変換によって（ $u_t = y_t - \alpha - \beta x_t$ ），

$$\begin{aligned} & f(y_1, y_2, \cdots, y_T) \\ &= \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2\right) \\ &\equiv L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T) \end{aligned}$$

となる。$L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)$ を尤度関数と呼び，$\log L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)$ を対数尤度関数と呼ぶ。

$$\begin{aligned} & \log L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T) \\ &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2 \end{aligned}$$

4. 変数変換について：

確率変数 x は $f_x(x)$ の分布に従う。このとき， $x = g(z)$ となるような z の分布関数 $f_z(z)$ は

$$f_z(z) = f_x\big(g(z)\big)\left|\frac{\mathrm{d}g(z)}{\mathrm{d}z}\right|$$

となる。

例：$x \sim U(0,1)$ のとき，$z = -\log(x)$ の分布を求める。

$$f_x(x) = 1$$

$x = \exp(-z)$ となる。

$$\begin{aligned} f_z(z) &= \left|\frac{\mathrm{d}x}{\mathrm{d}z}\right| f_x\big(g(z)\big) \\ &= |-\exp(-z)| \\ &= \exp(-z) \end{aligned}$$

5. 最尤法とは，データ $y_1, y_2, \cdots, y_T$ を与えたもとで，尤度関数 $L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)$，または，対数尤度関数 $\log L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)$ を最大にするような (α, β, σ^2) を推定値とする考え方である。(α, β, σ^2) の最尤推定値を ($\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\sigma}^2$) とすると，

$$\begin{aligned} & \frac{\partial \log L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)}{\partial \alpha} \\ &= \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t) = 0 \end{aligned}$$

$$\begin{aligned} & \frac{\partial \log L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)}{\partial \beta} \\ &= \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)x_t = 0 \end{aligned}$$

$$\begin{aligned} & \frac{\partial \log L(\alpha, \beta, \sigma^2 | y_1, y_2, \cdots, y_T)}{\partial \sigma^2} \\ &= -\frac{T}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2 = 0 \end{aligned}$$

を解くことになる。よって，最尤推定値は

$$\begin{aligned} \widetilde{\beta} &= \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{T}(x_t - \overline{x})^2} \\ \widetilde{\alpha} &= \overline{y} - \widetilde{\beta}\overline{x} \\ \widetilde{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \widetilde{\alpha} - \widetilde{\beta}x_t)^2 \end{aligned}$$

として与えられる。最小自乗法 (OLS) と最尤法 (ML) との違いは σ^2 の推定値にある。

8.2 回帰モデルの最尤法： 多変数の場合 I

1. Regression model: $y = X\beta + u$, $u \sim N(0, \sigma^2 I)$

   Log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta),$$

   where $\theta = (\beta, \sigma^2)$.

2. $\max_{\theta} \ \log L(\theta; y, X)$.

$$\frac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$$

   We obtain MLE of β and σ^2 :

$$\widetilde{\beta} = (X'X)^{-1}X'y,$$
$$\widetilde{\sigma}^2 = \frac{(y - X\widetilde{\beta})'(y - X\widetilde{\beta})}{T}$$

3. Fisher's information matrix is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; y, X)}{\partial\theta\partial\theta'}\Big)$$

   The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of θ .

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{T} \end{pmatrix}$$

8.3 回帰モデルの最尤法： 多変数の場合 II

1. Regression model: $y = X\beta + u$, $u \sim N(0, \sigma^2\Omega)$

   Log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta),$$

   where $\theta = (\beta, \sigma^2)$.

2. $\max_{\theta} \ \log L(\theta; y, X)$.

$$\frac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$$

   We obtain MLE of β and σ^2 :

$$\widetilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$
$$\widetilde{\sigma}^2 = \frac{(y - X\widetilde{\beta})'\Omega^{-1}(y - X\widetilde{\beta})}{T}$$

3. Fisher's information matrix is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; y, X)}{\partial\theta\partial\theta'}\Big)$$

   The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of θ .

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'\Omega^{-1}X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{T} \end{pmatrix}$$

8.4 AR(1) モデルの最尤法

AR(1) モデル： $t = 2, 3, \cdots, T$, $|\phi_1| < 1$ を仮定する。

$$y_t = \phi_1 y_{t-1} + u_t, \quad u_t \sim N(0, \sigma^2)$$

$y_1, y_2, \cdots, y_T$ の結合密度関数 $f(y_T, y_{T-1}, \cdots, y_1)$ を求める。

$$f(y_T, y_{T-1}, \cdots, y_1) = f(y_1)\prod_{t=2}^{T} f(y_t|y_{t-1}, \cdots, y_1)$$

となる。
条件付き分布 $f(y_t|y_{t-1}, \cdots, y_1)$ は，$y_t = \phi_1 y_{t-1} + u_t$ から，
$\mathrm{E}(y_t|y_{t-1}, \cdots, y_1) = \phi_1 y_{t-1}$,
$\mathrm{V}(y_t|y_{t-1}, \cdots, y_1) = \sigma^2$
となるので，

$$f(y_t|y_{t-1}, \cdots, y_1) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1})^2\right)$$

を得る。
このように，$f(y_t|y_{t-1}, \cdots, y_1)$ について，

$\mathrm{E}(y_t|y_{t-1},\cdots,y_1)$,
$\mathrm{V}(y_t|y_{t-1},\cdots,y_1)$
を求めれば $f(y_t|y_{t-1},\cdots,y_1)$ が得られる。

条件なしの分布 $f(y_t)$ は,

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + u_t \\
&= \phi_1^2 y_{t-2} + u_t + \phi_1 u_{t-1} \\
&\ \ \vdots \\
&= \phi_1^t y_0 + u_t + \phi_1 u_{t-1} + \cdots + \phi_1^{t-1} u_1 \\
&\ \ \vdots \\
&= u_t + \phi_1 u_{t-1} + \phi_1^2 u_{t-2} + \cdots
\end{aligned}
$$

なので,
$\mathrm{E}(y_t) = 0$,
$\mathrm{V}(y_t) = \sigma^2(1+\phi_1^2+\phi_1^4+\cdots) = \dfrac{\sigma^2}{1-\phi_1^2}$,
となる。よって,

$$
f(y_t) = \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}} \exp\left(-\frac{1}{2\sigma^2/(1-\phi_1^2)} y_t^2\right)
$$

を得る。

したがって,
$y_1, y_2, \cdots, y_T$ の結合密度関数 $f(y_T, y_{T-1}, \cdots, y_1)$ は,

$$
\begin{aligned}
& f(y_T, y_{T-1}, \cdots, y_1) \\
&= f(y_1) \prod_{t=2}^{T} f(y_t|y_{t-1},\cdots,y_1) \\
&= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}} \exp\left(-\frac{1}{2\sigma^2/(1-\phi_1^2)} y_1^2\right) \\
&\times \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1})^2\right)
\end{aligned}
$$

となる。
対数尤度関数は,

$$
\begin{aligned}
& L(\phi_1, \sigma^2; y_T, y_{T-1}, \cdots, y_1) \\
&= -\frac{1}{2}\log(2\pi\sigma^2/(1-\phi_1^2)) - \frac{1}{2\sigma^2/(1-\phi_1^2)} y_1^2 \\
&\quad -\frac{T-1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi_1 y_{t-1})^2
\end{aligned}
$$

となる。
ニュートン・ラプソン法，スコア法による最大化
$-1 < \rho < 1$ の範囲で，例えば 0.01 刻みで，探査 (grid search) 法を行う・

8.5 回帰モデルの最尤法： 一階の自己相関のケース

攪乱項が一階の自己相関に従う回帰モデルは,

$$
\begin{aligned}
y_t &= x_t\beta + u_t \\
u_t &= \rho u_{t-1} + \epsilon_t \\
\epsilon_t &\sim \ \text{iid } N(0, \sigma_\epsilon^2)
\end{aligned}
$$

となる。
$u_T, u_{T-1}, \cdots, u_1$ の結合分布 $f_u(\cdot;\cdot)$ の対数は，前節を参考にすると,

$$
\begin{aligned}
& \log f_u(u_T, u_{T-1}, \cdots, u_1; \rho, \sigma_\epsilon^2) \\
&= \log f(u_1; \rho, \sigma_\epsilon^2) + \sum_{t=2}^{T} \log f(u_t|u_{t-1},\cdots,u_1;\rho,\sigma_\epsilon^2) \\
&= -\frac{1}{2}\log(2\pi\sigma_\epsilon^2/(1-\rho^2)) - \frac{1}{2\sigma_\epsilon^2/(1-\rho^2)} u_1^2 \\
&\quad -\frac{T-1}{2}\log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}(u_t - \rho u_{t-1})^2
\end{aligned}
$$

となる。
$u_T, u_{T-1}, \cdots, u_1$ から $y_T, y_{T-1}, \cdots, y_1$ へ変数変換を行い，対数尤度関数は,

$$
\begin{aligned}
& L(\rho, \sigma_\epsilon^2, \beta; y_T, y_{T-1}, \cdots, y_1) \\
&= \log f_y(y_T, y_{T-1}, \cdots, y_1; \rho, \sigma_\epsilon^2, \beta) \\
&= \log f_u(y_T - x_T\beta, y_{T-1} - x_{T-1}\beta, \\
&\qquad\qquad \cdots, y_1 - x_1\beta; \rho, \sigma_\epsilon^2) \left|\frac{\partial u}{\partial y}\right| \\
&= -\frac{1}{2}\log\bigl(2\pi\sigma_\epsilon^2/(1-\rho^2)\bigr) \\
&\quad -\frac{1}{2\sigma_\epsilon^2/(1-\rho^2)}(y_1 - x_1\beta)^2 \\
&\quad -\frac{T-1}{2}\log(2\pi\sigma_\epsilon^2) \\
&\quad -\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}\Bigl((y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})\beta\Bigr)^2 \\
&= -\frac{1}{2}\log(2\pi\sigma_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{1}{2\sigma_\epsilon^2}(\sqrt{1-\rho^2}y_1 - \sqrt{1-\rho^2}x_1\beta)^2 \\
&\quad -\frac{T-1}{2}\log(2\pi\sigma_\epsilon^2) \\
&\quad -\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}\Bigl((y_t - \rho y_{t-1}) - (x_t - \rho x_{t-1})\beta\Bigr)^2
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{T}{2}\log(2\pi\sigma_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{1}{2\sigma_\epsilon^2}(y_1^* - x_1^*\beta)^2 \\
&\quad -\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{T}(y_t^* - x_t^*\beta)^2 \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}(y_t^* - x_t^*\beta)^2
\end{aligned}
$$

ただし，y_t^*, x_t^* は次のように表される。

$$
y_t^* = \begin{cases} \sqrt{1-\rho^2}y_t, & \text{for } t=1, \\ y_t - \rho y_{t-1}, & \text{for } t = 2,3,\cdots,T, \end{cases}
$$
$$
x_t^* = \begin{cases} \sqrt{1-\rho^2}x_t, & \text{for } t=1, \\ x_t - \rho x_{t-1}, & \text{for } t = 2,3,\cdots,T, \end{cases}
$$

◎ β に関して，$L(\rho, \sigma_\epsilon^2, \beta; y_T, y_{T-1}, \cdots, y_1)$ の最大化：

$$
\begin{aligned}
\widetilde{\beta} &= (\sum_{t=1}^{T} x_t^{*\prime} x_t^*)^{-1}(\sum_{t=1}^{T} x_t^{*\prime} y_t^*) \\
&= (X^{*\prime}X^*)^{-1}X^{*\prime}y^*
\end{aligned}
$$

$\Longrightarrow$ 回帰モデル $y^* = X^*\beta + \epsilon$, $\epsilon \sim N(0, \sigma_\epsilon^2 I_T)$ の最小二乗推定量に一致する。

◎ σ_ϵ^2 に関して，$L(\rho, \sigma_\epsilon^2, \beta; y_T, y_{T-1}, \cdots, y_1)$ の最大化：

$$
\begin{aligned}
\widetilde{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_t^* - x_t^*\beta)^2 \\
&= \frac{1}{T}(y^* - X^*\beta)'(y^* - X^*\beta)
\end{aligned}
$$

ただし，

$$
y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_T^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_T - \rho y_{T-1} \end{pmatrix}
$$
$$
x^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_T^* \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}x_1 \\ x_2 - \rho x_1 \\ \vdots \\ x_T - \rho x_{T-1} \end{pmatrix}
$$

とする。

◎ ρ に関して，$L(\rho, \sigma_\epsilon^2, \beta; y_T, y_{T-1}, \cdots, y_1)$ の最大化：

$\max\limits_{\beta, \sigma_\epsilon^2, \rho} L(\rho, \sigma_\epsilon^2, \beta; y)$

と

$\max\limits_{\rho} L(\rho, \widetilde{\sigma}_\epsilon^2, \widetilde{\beta}; y)$

とは同値。

$L(\rho, \widetilde{\sigma}_\epsilon^2, \widetilde{\beta}; y)$ は，集約対数尤度関数 (concentrated log-likelihood function) と呼ばれる。

よって，

$$
\begin{aligned}
&L(\rho, \widetilde{\sigma}_\epsilon^2, \widetilde{\beta}; y) \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\widetilde{\sigma}_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{T}{2} \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2} \\
&\quad -\frac{T}{2}\log\Big(\widetilde{\sigma}_\epsilon^2(\rho)\Big) + \frac{1}{2}\log(1-\rho^2)
\end{aligned}
$$

を ρ について最大化する。

$\Longrightarrow$ 探査法，ニュートン・ラプソン法，スコア法

$\widetilde{\sigma}_\epsilon^2 = \widetilde{\sigma}_\epsilon^2(\rho)$ に注意

8.6 回帰モデルの最尤法：不均一分散のケース

攪乱項が他の外生変数に依存する回帰モデルは，

$$
\begin{aligned}
&y_t = x_t\beta + u_t \\
&u_t \sim \text{ iid } N(0, \sigma_t^2) \\
&\sigma_t^2 = (z_t\alpha)^2
\end{aligned}
$$

となる。

$u_T, u_{T-1}, \cdots, u_1$ の結合分布 $f_u(\cdot;\cdot)$ は，

$$
\begin{aligned}
&\log f_u(u_T, u_{T-1}, \cdots, u_1; \sigma_t^2) \\
&= \sum_{t=1}^{T}\log f_u(u_t; \sigma_t^2) \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_t^2) - \frac{1}{2}\sum_{t=2}^{T}\left(\frac{u_t}{\sigma_t}\right)^2 \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(z_t\alpha)^2 - \frac{1}{2}\sum_{t=2}^{T}\left(\frac{u_t}{z_t\alpha}\right)^2
\end{aligned}
$$

となる。

$u_T, u_{T-1}, \cdots, u_1$ から $y_T, y_{T-1}, \cdots, y_1$ へ変数変換を行い，対数尤度関数は，

$$
L(\sigma_t^2, \beta; y_T, y_{T-1}, \cdots, y_1)
$$

$$
\begin{aligned}
&= \log f_y(y_T, y_{T-1}, \cdots, y_1; \sigma_t, \beta) \\
&= \log f_u(y_T - x_T\beta, y_{T-1} - x_{T-1}\beta, \\
&\qquad\qquad \cdots, y_1 - x_1\beta; \sigma_t^2) \left|\frac{\partial u}{\partial y}\right| \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(z_t\alpha)^2 - \frac{1}{2}\sum_{t=2}^{T}\left(\frac{y_t - x_t\beta}{z_t\alpha}\right)^2
\end{aligned}
$$

$\Longrightarrow$ β, α に関して最大化

9 漸近理論

1. **定義: 分布収束**

   確率変数の系列 x_1, x_2, $\cdots$ は分布関数 F_1, F_2, $\cdots$ をそれぞれ持つ。もし，

   $$\lim_{t\to\infty} F_t = F$$

   となるとき，確率変数の系列 x_1, x_2, $\cdots$ は F に分布収束 (convergence in distribution) するという。

2. **一致性について**

   (a) 確率収束の定義

   $\{z_T:\ T = 1, 2, \cdots\}$ を確率変数の系列であるとする。このとき，以下が成り立つならば，z_T は θ に確率収束 (convergence in probability) するという。

   $$\lim_{T\to\infty} \mathrm{Prob}(|z_T - \theta| < \epsilon) = 1,$$

   ただし，ϵ は任意の定数とする。

   さらに，定数 θ を z_T の確率極限 (probability limit) と呼ぶ。

   $$\mathrm{plim}\ z_T = \theta,$$

   として表される。

   (b) $\widehat{\theta}_T$ をパラメータ θ の推定値とする。このとき，$\widehat{\theta}_T$ が θ に確率収束すれば，$\widehat{\theta}_T$ は θ の一致推定量 (consistent estimator) であると言われる。

3. Chebyshev の不等式：

   $g(X) \geq 0$ について，

   $$\mathrm{Prob}\big(g(X) \geq k\big) \leq \frac{\mathrm{E}\big(g(X)\big)}{k}$$

   となる。ただし，k は正の定数とする。

   証明：

   $g(X) \geq k$ のとき $U = 1$，$g(X) < k$ のとき $U = 0$ となる確率変数 U を導入する。常に，以下の式が成り立つ。

   $$g(X) \geq kU$$

   よって，両辺に期待値を取ると，

   $$\mathrm{E}\big(g(X)\big) \geq k\mathrm{E}(U)$$

   となる。$\mathrm{E}(U)$ を求める。

   $$
   \begin{aligned}
   \mathrm{E}(U) &= 1 \times \mathrm{Prob}\big(g(X) \geq k\big) + 0 \times \mathrm{Prob}\big(g(X) < k\big) \\
   &= \mathrm{P}\big(g(X) \geq k\big)
   \end{aligned}
   $$

   したがって，

   $$\mathrm{E}\big(g(X)\big) \geq k\mathrm{Prob}\big(g(X) \geq k\big)$$

   から，

   $$\mathrm{Prob}\big(g(X) \geq k\big) \leq \frac{\mathrm{E}\big(g(X)\big)}{k}$$

   を得る。

4. ある確率変数 X について，$g(X) = (X-\mu)'(X-\mu)$，$\mathrm{E}(X) = \mu$，$\mathrm{Var}(X) = \Sigma$ とするとき，

   $$\mathrm{Prob}\big((X-\mu)'(X-\mu) \geq k\big) \leq \frac{\mathrm{tr}(\Sigma)}{k}$$

   を得る。なぜなら，

   $$
   \begin{aligned}
   &\mathrm{E}\big((X-\mu)'(X-\mu)\big) \\
   &= \mathrm{E}\Big(\mathrm{tr}\big((X-\mu)'(X-\mu)\big)\Big) \\
   &= \mathrm{E}\Big(\mathrm{tr}\big((x-\mu)(x-\mu)'\big)\Big) \\
   &= \mathrm{tr}\Big(\mathrm{E}\big((x-\mu)(x-\mu)'\big)\Big) \\
   &= \mathrm{tr}(\Sigma)
   \end{aligned}
   $$

   となることに注意。

5. **例：**

確率変数 $X_i \sim (\mu, \sigma^2)$, $i = 1, 2, \cdots, T$

標本平均 $\overline{X}$ は μ の一致推定量である。

証明：

Chebyshev の不等式を当てはめる。

$g(\overline{X}) = (\overline{X} - \mu)^2$, $\epsilon^2 = k$

$\mathrm{E}\big(g(\overline{X})\big) = \mathrm{V}(\overline{X}) = \dfrac{\sigma^2}{T}$

に注意すると，$T \longrightarrow \infty$ のとき，

$$P\big(|\overline{X} - \mu| \geq \epsilon\big) \leq \frac{\sigma^2}{T\epsilon^2} \longrightarrow 0$$

すなわち，任意の ϵ について，

$$\lim_{T\to\infty} P\big(|\overline{X} - \mu| < \epsilon\big) = 1$$

6. **公式：**

x_T と y_T は確率変数で，

$\mathrm{plim}\ x_T = c$, $\mathrm{plim}\ y_T = d$ とする。

このとき，

(a) $\mathrm{plim}\ (x_T + y_T) = c + d$

(b) $\mathrm{plim}\ x_T y_T = cd$

(c) $\mathrm{plim}\ x_T / y_T = c/d$, ただし，$d \neq 0$

(d) $\mathrm{plim}\ g(x_T) = g(c)$, ただし，$g(\cdot)$ はある関数とする。

$\Longrightarrow$ Slutsky's Theorem

7. **Lidberg-Levy Central Limit Theorem**

x_1, x_2, $\cdots$, x_T は互いに独立に，$x_t \sim (\mu,\Sigma)$ の同一の分布をする (independent and identically distributed, iid)。このとき，

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu) \ \longrightarrow\ N(0, \Sigma)$$

8. **Central Limit Theorem** (Greenberg and Webster (1983))

一般化：

x_1, x_2, $\cdots$, x_T は互いに独立に，$x_t \sim (\mu,\Sigma_t)$ の分布をする。このとき，

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mu) \ \longrightarrow\ N(0, \Sigma)$$

ただし，

$$\Sigma = \lim_{T\to\infty} \left(\frac{1}{T} \sum_{t=1}^{T} \Sigma_t \right)$$

とする。

9. **定義:** $\widehat{\theta}_T$ を θ の一致推定量とする。$\sqrt{T}(\widehat{\theta}_T - \theta)$ は $N(0, \Sigma)$ に分布収束すると仮定する。このとき，$\widehat{\theta}_T$ は漸近分布 $N(\theta, \Sigma/T)$ を持つという。

10. **定義:** 次の 3 つの条件が満たされるとき，$\widehat{\theta}_T$ は consistent uniformly asymptotically normal であると言われる。

(a) $\widehat{\theta}_T$ は一致性がある

(b) $\sqrt{T}(\widehat{\theta}_T - \theta)$ は $N(0, \Sigma)$ に分布収束する

(c) 一様収束である

11. **定義:** $\widehat{\theta}_T$ と $\widetilde{\theta}_T$ は consistent, uniformly, asymptotically normal で，その漸近分散は Σ/T, Ω/T とする。このとき，$\Omega - \Sigma$ が非負値定符号行列 (positive semidefinite) であれば，$\widehat{\theta}_T$ は，$\widetilde{\theta}_T$ に比べて，漸近的に有効 (asymptotically efficient) である。

12. **定義:** もし，他のどの consistent, uniformly, asymptotically normal estimator と比べても asymptotically efficient であれば，その consistent, uniformly, asymptotically normal estimator は漸近的有効 (asymptotically efficient) であると言われる。

13. asymptotically efficient であり，しかも，

consistent, uniformly, asymptotically normal estimator となるための十分条件は，その推定値の漸近分散がクラメール・ラオの下限に一致することである。

14. $x_1, x_2, \cdots, x_T$ は密度関数 $f(x;\theta)$ からの確率変数である。$\widehat{\theta}_T$ を θ の最尤推定量とする。このとき，ある条件 (regularity conditions) のもとで，$\widehat{\theta}_T$ は θ の一致推定量であり，$\sqrt{T}(\widehat{\theta} - \theta)$ の漸近分布は $N\left(0, \lim\left(\dfrac{I(\theta)}{T}\right)^{-1}\right)$ となる。

15. Regularity Conditions について

   (a) x_t の範囲は θ に依存しない

   (b) $f(x;\theta)$ は，θ について，少なくとも 3 階微分可能であり，それらの微分は有界である。

16. このように，最尤推定量は,

   (i) consistent,

   (ii) asymptotically normal,

   (iii) asymptotically efficient,

   である。

17. **Slutsky's Theorem**

   $\widehat{\theta}$ は θ の一致推定量であるとする。このとき，$g(\widehat{\theta})$ は $g(\theta)$ の一致推定量となる。ただし，g は well-defined continuous function である。

18. **最尤法の不変性 (Invariance of Maximum Likelihood Estimation)**

   $\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k$ を $\theta_1, \theta_2, \cdots, \theta_k$ の最尤推定量であるとする。

   $\alpha_1 = \alpha_1(\theta_1, \theta_2, \cdots, \theta_k)$, $\alpha_2 = \alpha_2(\theta_1, \theta_2, \cdots, \theta_k)$, $\cdots$, $\alpha_k = \alpha_k(\theta_1, \theta_2, \cdots, \theta_k)$ のような 1 対 1 変換を考える。

   このとき，α_1，α_2，$\cdots$，α_k の最尤推定量は $\widehat{\alpha}_1 = \alpha_1(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k)$, $\widehat{\alpha}_2 = \alpha_2(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k)$, $\cdots$, $\widehat{\alpha}_k = \alpha_k(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_k)$ である。

10 最小自乗推定値の一致性と漸近的正規性

1. T 組の標本を用いた最小自乗推定値を便宜上 $\widehat{\beta}_T = (X'X)^{-1}X'y$ と書く。一致性とは,「 T が大きくなるにつれて，$\widehat{\beta}_T$ は β に近づく」というものである。

2. データの定常性の仮定：

$$\frac{1}{T}X'X \longrightarrow M_{xx}$$

   を仮定する。

   このとき,

$$\frac{1}{T}X'u \longrightarrow 0$$

   が成り立つ。

   証明：

   チェビシェフの不等式 (Chebyshev's inequality) によると，$g(x) \geq 0$ について,

$$\mathrm{Prob}\big(g(x) \geq k\big) \leq \frac{\mathrm{E}\big(g(x)\big)}{k}$$

   となる。ただし，k は正の定数とする。

   ここで，$x \Longrightarrow \dfrac{1}{T}X'u$,

   かつ，$g(x) = x'x$ として，チェビシェフの不等式を当てはめる。

$$\begin{aligned}
&\mathrm{E}\Big((\frac{1}{T}X'u)'\frac{1}{T}X'u\Big) \\
&= \frac{1}{T^2}\mathrm{E}\Big(u'XX'u\Big) \\
&= \frac{1}{T^2}\mathrm{E}\Big(\mathrm{tr}\big(u'XX'u\big)\Big) \\
&= \frac{1}{T^2}\mathrm{E}\Big(\mathrm{tr}\big(XX'uu'\big)\Big) \\
&= \frac{1}{T^2}\mathrm{tr}\Big(XX'\mathrm{E}(uu')\Big) \\
&= \frac{\sigma^2}{T^2}\mathrm{tr}(XX') \\
&= \frac{\sigma^2}{T^2}\mathrm{tr}(X'X) \\
&= \frac{\sigma^2}{T}\mathrm{tr}(\frac{1}{T}X'X)
\end{aligned}$$

   に注意すると,

$$\begin{aligned}
\mathrm{Prob}\Big((\frac{1}{T}X'u)'\frac{1}{T}X'u \geq k\Big) &\leq \frac{\sigma^2}{Tk}\mathrm{tr}(\frac{1}{T}X'X) \\
&\longrightarrow 0 \times \mathrm{tr}(M_{xx})
\end{aligned}$$

   が得られる。

   最後の行では，仮定により,

$$\frac{1}{T}X'X \longrightarrow M_{xx}$$

   が成り立つことを利用している。

   よって,

$$(\frac{1}{T}X'u)'\frac{1}{T}X'u \longrightarrow 0$$

   が成り立ち，2 次形式なので，これは,

$$\frac{1}{T}X'u \longrightarrow 0$$

   を意味することになる。

3. $$\frac{1}{T}X'X \ \longrightarrow\ M_{xx}$$

ならば

$$(\frac{1}{T}X'X)^{-1} \ \longrightarrow\ M_{xx}^{-1}$$

となる。$\Longrightarrow$ Slutsky's Theorem

(*) **Slutsky's Theorem**

$\widehat{\theta} \longrightarrow \theta$ のとき，$g(\widehat{\theta}) \longrightarrow g(\theta)$ となる。

4. 最小自乗推定値は

$$\begin{aligned}\widehat{\beta}_T &= \beta + (X'X)^{-1}X'u \\ &= \beta + (\frac{1}{T}X'X)^{-1}(\frac{1}{T}X'u)\end{aligned}$$

と表されるので，

$$\begin{aligned}\widehat{\beta}_T &\longrightarrow \beta + M_{xx}^{-1} \times 0 \\ &\longrightarrow \beta\end{aligned}$$

5. 以上から，最小自乗推定値は (i) 漸近的に不偏性，(ii) 漸近的に有効性，(iii) 一致性 を持つ。

6. Asymptotic Normality of OLSE

(a) **Central Limit Theorem:** Greenberg and Webster (1983)

$z_1, z_2, \cdots, z_T$ は互いに独立に，平均 μ，分散 Σ_t の分布をする。このとき，

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_t - \mu) \ \longrightarrow\ N(0, \Sigma)$$

ただし，

$$\Sigma = \lim_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T}\Sigma_t\right)$$

とする。

z_t の分布の形状は仮定されないことに注意。

(b) $z_t = x_t u_t$ とすると，$\Sigma_t = \sigma^2 x_t' x_t$ となる。

(c) さらに，Σ を以下のように定義する。

$$\begin{aligned}\Sigma &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sigma^2 x_t' x_t \\ &= \sigma^2 \lim_{T\to\infty}\frac{1}{T}X'X \\ &= \sigma^2 M_{xx}\end{aligned}$$

(d) Central Limit Theorem (Greenberg and Webster (1983) を適用すると，

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_t' u_t &= \frac{1}{\sqrt{T}}X'u \\ &\longrightarrow N(0, \sigma^2 M_{xx})\end{aligned}$$

を得る。

$$\sqrt{T}(\widehat{\beta} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\frac{1}{\sqrt{T}}X'u$$

なので，

$$\sqrt{T}(\widehat{\beta} - \beta) \ \longrightarrow\ N(0, \sigma^2 M_{xx}^{-1})$$

を得る。$\Longrightarrow$ 最小二乗推定量の漸近的正規性

u_t の正規分布を仮定する必要なし

11 操作変数法

11.1 測定誤差

Errors in Variables

1. 真の回帰モデル

$$y = \widetilde{X}\beta + u$$

2. 観測される外生変数

$$X = \widetilde{X} + V$$

V: 測定誤差，観測誤差 (Measurement Errors)

3. X の中で観測誤差を含まない変数では，V の対応する列はすべてゼロとなる。

4. 観測される変数との関係

$$y = X\beta + (u - V\beta)$$

β の最小二乗推定量は

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'(u - V\beta)\end{aligned}$$

5. 仮定：

(a) X の観測誤差と真の値 $\widetilde{X}$ とは，極限では無相関を仮定

$$\mathrm{plim}\left(\frac{1}{T}\widetilde{X}'V\right)=0$$

よって，

$$\begin{aligned}&\mathrm{plim}\left(\frac{1}{T}X'X\right)\\ &=\mathrm{plim}\left(\frac{1}{T}\widetilde{X}'\widetilde{X}\right)+\mathrm{plim}\left(\frac{1}{T}V'V\right)\\ &=\Sigma+\Omega\end{aligned}$$

(b) 攪乱項 u と X の観測誤差とは無相関，攪乱項 u と真の値 $\widetilde{X}$ とは無相関と仮定

$$\mathrm{plim}\left(\frac{1}{T}V'u\right)=0$$
$$\mathrm{plim}\left(\frac{1}{T}\widetilde{X}'u\right)=0$$

6. β の最小二乗推定量は

$$\begin{aligned}\widehat{\beta}&=\beta+(X'X)^{-1}X'(u-V\beta)\\ &=\beta+(X'X)^{-1}(\widetilde{X}+V)'(u-V\beta)\end{aligned}$$

なので，

$$\mathrm{plim}\ \widehat{\beta}=\beta-(\Sigma+\Omega)^{-1}\Omega\beta$$

7. 例： 2 変数の場合

$$\begin{aligned}y_t&=\alpha+\beta\widetilde{x}_t+u_t\\ x_t&=\widetilde{x}_t+v_t\end{aligned}$$

以上のモデルでは，

$$\begin{aligned}\Sigma&=\mathrm{plim}\left(\frac{1}{T}\widetilde{X}'\widetilde{X}\right)\\ &=\mathrm{plim}\begin{pmatrix}1 & \frac{1}{T}\sum\widetilde{x}_t\\ \frac{1}{T}\sum\widetilde{x}_t & \frac{1}{T}\sum\widetilde{x}_t^2\end{pmatrix}\\ &=\begin{pmatrix}1 & \mu\\ \mu & \mu^2+\sigma^2\end{pmatrix}\end{aligned}$$

ただし，μ, σ^2 は $\widetilde{x}_t$ の平均，分散を表す。

$$\begin{aligned}\Omega&=\mathrm{plim}\left(\frac{1}{T}V'V\right)\\ &=\mathrm{plim}\begin{pmatrix}0 & 0\\ 0 & \frac{1}{T}\sum v_t^2\end{pmatrix}\\ &=\begin{pmatrix}0 & 0\\ 0 & \sigma_v^2\end{pmatrix}\end{aligned}$$

したがって，

$$\begin{aligned}&\mathrm{plim}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}\\ &=\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\\ &-\left(\begin{pmatrix}1 & \mu\\ \mu & \mu^2+\sigma^2\end{pmatrix}+\begin{pmatrix}0 & 0\\ 0 & \sigma_v^2\end{pmatrix}\right)^{-1}\\ &\times\begin{pmatrix}0 & 0\\ 0 & \sigma_v^2\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\\ &=\begin{pmatrix}\alpha\\ \beta\end{pmatrix}-\frac{1}{\sigma^2+\sigma_v^2}\begin{pmatrix}-\mu\sigma_v^2\beta\\ \sigma_v^2\beta\end{pmatrix}\end{aligned}$$

β について，

$$\begin{aligned}\mathrm{plim}(\widehat{\beta})&=\beta-\frac{\sigma_v^2\beta}{\sigma^2+\sigma_v^2}\\ &=\frac{\beta}{1+\sigma_v^2/\sigma^2}\end{aligned}$$

11.2　操作変数法

Instrumental Variable (IV)

1. 回帰モデル：$y=X\beta+u$, $u\sim N(0,\sigma^2I)$ において，$\mathrm{E}(X'u)\neq 0$ の場合，最小自乗推定量は一致性を持たない。

2. 最小自乗推定量は一致性を持たないことの証明

$$\begin{aligned}\widehat{\beta}&=\beta+(\frac{1}{T}X'X)^{-1}\frac{1}{T}X'u\\ &\longrightarrow\beta+M_{xx}^{-1}M_{xu}\end{aligned}$$

$$\frac{1}{T}X'X\longrightarrow M_{xx}$$
$$\frac{1}{T}X'u\longrightarrow M_{xu}\neq 0$$

3. $\frac{1}{T}Z'u \;\longrightarrow\; M_{zu}=0$ となるような X とは別の変数 Z を見つける。

両辺に Z' を掛けて，

$$Z'y = Z'X\beta + Z'u$$

さらに，両辺を T で割って，plim をとる。

$$\begin{aligned}\operatorname{plim}\left(\frac{1}{T}Z'y\right) &= \operatorname{plim}\left(\frac{1}{T}Z'X\right)\beta \\ &\quad + \operatorname{plim}\left(\frac{1}{T}Z'u\right) \\ &= \operatorname{plim}\left(\frac{1}{T}Z'X\right)\beta\end{aligned}$$

したがって，

$$\beta = \left(\operatorname{plim}\left(\frac{1}{T}Z'X\right)\right)^{-1}\operatorname{plim}\left(\frac{1}{T}Z'y\right)$$

となるので，

$$\beta_{IV} = (Z'X)^{-1}Z'y$$

を β の推定量とする。

$\Longrightarrow$ 操作変数法

4. 次を定義する。

$$\begin{aligned}\frac{1}{T}Z'X &\longrightarrow M_{zx} \\ \frac{1}{T}Z'Z &\longrightarrow M_{zz} \\ \frac{1}{T}Z'u &\longrightarrow 0\end{aligned}$$

5. β_{IV} の分散

$$\begin{aligned}&\beta_{IV} \\ &= (Z'X)^{-1}Z'y \\ &= (Z'X)^{-1}Z'(X\beta+u) \\ &= \beta + (Z'X)^{-1}Z'u\end{aligned}$$

$$\begin{aligned}&\sqrt{T}(\beta_{IV}-\beta) \\ &= \left(\frac{1}{T}Z'X\right)^{-1}\left(\frac{1}{\sqrt{T}}Z'u\right)\end{aligned}$$

Central Limit Theorem を適用する。

$$\frac{1}{\sqrt{T}}Z'u \;\longrightarrow\; N(0,\sigma^2M_{zz})$$

なので，

$$\begin{aligned}\beta_{IV} &= \left(\frac{1}{T}Z'X\right)^{-1}\left(\frac{1}{\sqrt{T}}Z'u\right) \\ &\longrightarrow N(0,\sigma^2M_{zx}^{-1}M_{zz}{M'_{zx}}^{-1})\end{aligned}$$

を得る。

$\Longrightarrow$ 一致性と漸近的正規性

6. **Central Limit Theorem:** Greenberg and Webster (1983)

z_1, z_2, $\cdots$, z_T は互いに独立に，平均 μ，分散 Σ_t の分布をする。このとき，

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_t-\mu) \;\longrightarrow\; N(0,\Sigma)$$

ただし，

$$\Sigma = \lim_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T}\Sigma_t\right)$$

とする。

分布の形状は仮定されないことに注意。

7. β_{IV} の分散は次のように推定される。

$$\mathrm{V}(\beta_{IV}) = s^2(Z'X)^{-1}Z'Z(X'Z)^{-1}$$

ただし，

$$s^2 = \frac{(y-X\beta_{IV})'(y-X\beta_{IV})}{T-k}$$

とする。

11.3　2段階最小二乗法

1. 回帰モデル：

$y = X\beta+u$, $u \sim N(0,\sigma^2I)$

において，

$\mathrm{E}(X'u)\neq 0$

の場合，最小自乗推定量は一致性を持たない。

2. $\dfrac{1}{T}Z'u \ \longrightarrow\ M_{zu}=0$ となるような X とは別の変数 Z を見つける。

3. $Z=\widehat{X}$ を操作変数に使う。ただし，$\widehat{X}$ は X を他の外生変数 (仮に，W) に回帰させたときの予測値とする。

   すなわち，

$$X = WB + V$$

   から，B を推定し，

$$\widehat{X} = W\widehat{B}$$

   を得る。ただし，$\widehat{B}=(W'W)^{-1}W'X$ である。

   すなわち，

$$\widehat{X} = W(W'W)^{-1}W'X$$

   を操作変数 Z として用いる。

4. 操作変数法は

$$\begin{aligned}\beta_{IV} &= (\widehat{X}'X)^{-1}\widehat{X}'y \\ &= \big(X'W(W'W)^{-1}W'X\big)^{-1}X'W(W'W)^{-1}W'y\end{aligned}$$

   変形する。

$$\begin{aligned}&\beta_{IV} \\ &= \beta + \big(X'W(W'W)^{-1}W'X\big)^{-1}X'W(W'W)^{-1}W'u\end{aligned}$$

   なので，

$$\begin{aligned}&\sqrt{T}(\beta_{IV}-\beta) \\ &= \left(\left(\frac{1}{T}X'W\right)\left(\frac{1}{T}W'W\right)^{-1}\left(\frac{1}{T}XW'\right)'\right)^{-1} \\ &\times\left(\frac{1}{T}X'W\right)\left(\frac{1}{T}W'W\right)^{-1}\left(\frac{1}{\sqrt{T}}W'u\right) \\ &\longrightarrow N\big(0,(M_{xz}M_{zz}^{-1}M_{xz}')^{-1}\big)\end{aligned}$$

5. 明らかに，W と u との相関はゼロ

$$\mathrm{plim}\left(\frac{1}{T}W'u\right)=0$$

6. 注意

$$\begin{aligned}\widehat{X}'X &= X'W(W'W)^{-1}W'X \\ &= X'W(W'W)^{-1}W'W(W'W)^{-1}W'X \\ &= \widehat{X}'\widehat{X}\end{aligned}$$

   となるので，

$$\begin{aligned}\beta_{IV} &= (\widehat{X}'X)^{-1}\widehat{X}'y \\ &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y\end{aligned}$$

12 大標本検定

12.1 Wald, LM, LR テスト

$\theta : K\times 1$
$h(\theta) : G\times 1$ vector function, $G\le K$
$\theta : K\times 1$
The null hypothesis $H_0 : h(\theta)=0$
$\theta^* : k\times 1$, restricted maximum likelihood estimate
$\widehat{\theta} : k\times 1$, unrestricted maximum likelihood estimate
$I(\theta) : K\times K$, information matrix

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2\log L(\theta)}{\partial\theta\partial\theta'}\Big)$$

$\log L(\theta)$: log-likelihood function
$R_\theta = \dfrac{\partial h(\theta)}{\partial\theta'} : G\times K$
$F_\theta = \dfrac{\partial \log L(\theta)}{\partial\theta} : K\times 1$

1. Wald 検定: $W = h(\widehat{\theta})'\Big(R_{\widehat{\theta}}\Big(I(\widehat{\theta})\Big)^{-1}R'_{\widehat{\theta}}\Big)^{-1}h(\widehat{\theta})$

   (a) $h(\theta)\approx h(\widehat{\theta}) + \dfrac{\partial h(\widehat{\theta})}{\partial\theta'}(\theta-\widehat{\theta})$
   により，

$$\begin{aligned}h(\widehat{\theta}) &\approx \frac{\partial h(\widehat{\theta})}{\partial\theta'}(\widehat{\theta}-\theta) \\ &= R_{\widehat{\theta}}(\widehat{\theta}-\theta)\end{aligned}$$

   (b) $\widehat{\theta}$ は最尤推定値である。最尤法の性質により，

$$\sqrt{T}(\widehat{\theta}-\theta)\ \longrightarrow\ N\left(0,\lim\left(\frac{I(\theta)}{T}\right)^{-1}\right)$$

$$(\widehat{\theta}-\theta)\ \longrightarrow\ N\Big(0,\big(I(\theta)\big)^{-1}\Big)$$

(c) $h(\widehat{\theta})$ の分布は,

$$h(\widehat{\theta}) \longrightarrow N\Big(0, R_{\widehat{\theta}}\big(I(\theta)\big)^{-1}R'_{\widehat{\theta}}\Big)$$

(d) χ^2 分布を導出する。

$$h(\widehat{\theta})\Big(R_{\widehat{\theta}}\big(I(\theta)\big)^{-1}R'_{\widehat{\theta}}\Big)^{-1}h(\widehat{\theta})' \longrightarrow \chi^2(G)$$

さらに，$I(\widehat{\theta}) \longrightarrow I(\theta)$ により (すなわち，確率収束，convergence in probability),

$$h(\widehat{\theta})\Big(R_{\widehat{\theta}}\big(I(\widehat{\theta})\big)^{-1}R'_{\widehat{\theta}}\Big)^{-1}h(\widehat{\theta})' \longrightarrow \chi^2(G)$$

を得る。

2. Lagrange Multiplier 検定: $LM = F'_{\theta^*}\left(I(\theta^*)\right)^{-1}F_{\theta^*}$

(a) $h(\theta)=0$ の制約付き最尤法

$$\max_{\theta}\ \log L(\theta), \quad \text{subject to} \quad h(\theta)=0$$

ラグランジェ関数

$$L = \log L(\theta) + \lambda h(\theta)$$

(b) 最大化のために,

$$\frac{\partial L}{\partial \theta} = \frac{\partial \log L(\theta)}{\partial \theta} + \lambda \frac{\partial h(\theta)}{\partial \theta} = 0$$

$$\frac{\partial L}{\partial \lambda} = h(\theta) = 0$$

(c) $\dfrac{\partial \log L(\theta)}{\partial \theta}$ の平均，分散は

$$\mathrm{E}\left(\frac{\partial \log L(\theta)}{\partial \theta}\right) = 0,$$

$$\mathrm{V}\left(\frac{\partial \log L(\theta)}{\partial \theta}\right) = -\mathrm{E}\left(\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right) = I(\theta)$$

である。

(d) したがって,

$$\frac{\partial \log L(\theta)}{\partial \theta} \longrightarrow N\big(0, I(\theta)\big)$$

(e) よって,

$$\frac{\partial \log L(\theta)}{\partial \theta}\big(I(\theta)\big)^{-1}\frac{\partial \log L(\theta)}{\partial \theta'} \longrightarrow \chi^2(G)$$

最尤法の推定値は一致性を持つので，$\theta^* \longrightarrow \theta$ となり,

$$F'_{\theta^*}\big(I(\theta^*)\big)^{-1}F_{\theta^*} \longrightarrow \chi^2(G)$$

を得る。

3. Likelihood Ratio 検定: $LR = -2\log\lambda \longrightarrow \chi^2(G)$

$$\lambda = \frac{L(\theta^*)}{L(\widehat{\theta})}$$

For proof, see Theil (1971, p.396).

4. All of W, LM and LR are asymptotically distributed as $\chi^2(G)$ random variables under the null hypothesis $H_0:\ h(\theta)=0$.

5. Under some comditions, we have $W \ge LR \ge LM$. See Engle (1981) "Wald, Likelihood and Lagrange Multiplier Tests in Econometrics," Chap. 13 in *Handbook of Econometrics*, Vol.2, Grilliches and Intriligator eds, North-Holland.

12.2 尤度比検定の使用例

データ・ファイル (cons99.txt)

```
1955   5430.1   6135.0  18.1
1956   5974.2   6828.4  18.3
1957   6686.3   7619.5  19.0
1958   7169.7   8153.3  19.1
1959   8019.3   9274.3  19.7
1960   9234.9  10776.5  20.5
1961  10836.2  12869.4  21.8
1962  12430.8  14701.4  23.2
1963  14506.6  17042.7  24.9
1964  16674.9  19709.9  26.0
1965  18820.5  22337.4  27.8
1966  21680.6  25514.5  29.0
1967  24914.0  29012.6  30.1
1968  28452.7  34233.6  31.6
1969  32705.2  39486.3  32.9
1970  37784.1  45913.2  35.2
1971  42571.6  51944.3  37.5
1972  49124.1  60245.4  39.7
1973  59366.1  74924.8  44.1
1974  71782.1  93833.2  53.3
1975  83591.1 108712.8  59.4
1976  94443.7 123540.9  65.2
1977 105397.8 135318.4  70.1
1978 115960.3 147244.2  73.5
1979 127600.9 157071.1  76.0
1980 138585.0 169931.5  81.6
1981 147103.4 181349.2  85.4
1982 157994.0 190611.5  87.7
1983 166631.6 199587.8  89.5
1984 175383.4 209451.9  91.8
1985 185335.1 220655.6  93.9
1986 193069.6 229938.8  94.8
1987 202072.8 235924.0  95.3
1988 212939.9 247159.7  95.8
1989 227122.2 263940.5  97.7
1990 243035.7 280133.0 100.0
1991 255531.8 297512.9 102.5
```

```
1992 265701.6 309256.6 104.5
1993 272075.3 317021.6 105.9
1994 279538.7 325655.7 106.7
1995 283245.4 331967.5 106.2
1996 291458.5 340619.1 106.0
1997 298475.2 345522.7 107.3
```

左から，年，名目家計最終消費支出 (10 億円)，家計可処分所得 (10 億円)，家計最終消費支出デフレータ (1990 年 =100)

```
          PROGRAM
 LINE  ***********************************************
 |      1  freq a;
 |      2  smpl 1955 1997;
 |      3  read(file='cons99.txt') year cons yd price;
 |      4  rcons=cons/(price/100);
 |      5  ryd=yd/(price/100);
 |      6  lyd=log(ryd);
 |      7  olsq rcons c ryd;
 |      8  ar1  rcons c ryd;
 |      9  olsq rcons c lyd;
 |     10  param a1 0 a2 0 a3 1;
 |     11  frml eq rcons=a1+a2*((ryd**a3)-1.)/a3;
 |     12  lsq(tol=0.00001,maxit=100) eq;
 |     13  a3=1.15;
 |     14  rryd=((ryd**a3)-1.)/a3;
 |     15  ar1  rcons c rryd;
 |     16  end;
          EXECUTION
*****************************************************

                Equation   1
                ============

  Method of estimation = Ordinary Least Squares

Dependent variable: RCONS
Current sample:  1955 to 1997
Number of observations:  43

  Mean of dependent variable = 146270.
 Std. dev. of dependent var. = 79317.2
    Sum of squared residuals = .129697E+10
       Variance of residuals = .316335E+08
    Std. error of regression = 5624.36
                   R-squared = .995092
          Adjusted R-squared = .994972
     Durbin-Watson statistic = .115101
  F-statistic (zero slopes) = 8311.90
  Schwarz Bayes. Info. Crit. = 17.3970
  Log of likelihood function = -431.289

           Estimated    Standard
Variable  Coefficient     Error       t-statistic
C         -2919.54      1847.55       -1.58022
RYD       .852879       .935486E-02   91.1696


                    Equation   2
                    ============

   FIRST-ORDER SERIAL CORRELATION OF THE ERROR

   MAXIMUM LIKELIHOOD ITERATIVE TECHNIQUE

   CONVERGENCE ACHIEVED AFTER    7 ITERATIONS

Dependent variable: RCONS
Current sample:  1955 to 1997
Number of observations:  43

(Statistics based on transformed data)
 Mean of dependent variable = 13685.2
Std. dev. of dependent var. = 5495.11
   Sum of squared residuals = .145807E+09
      Variance of residuals = .355627E+07
   Std. error of regression = 1885.81
                  R-squared = .885040
         Adjusted R-squared = .882236
    Durbin-Watson statistic = 1.38750
Rho (autocorrelation coef.) = .945024
      Standard error of rho = .040839
        t-statistic for rho = 23.1405
  F-statistic (zero slopes) = 315.621
 Log of likelihood function = -385.419
(Statistics based on original data)
 Mean of dependent variable = 146270.
Std. dev. of dependent var. = 79317.2
   Sum of squared residuals = .145826E+09
      Variance of residuals = .355672E+07
   Std. error of regression = 1885.93
                  R-squared = .999480
         Adjusted R-squared = .999467
    Durbin-Watson statistic = 1.38714

           Estimated    Standard
Variable  Coefficient     Error       t-statistic
C         1672.37       5919.24       .282531
RYD       .840011       .025263       33.2501


                 Equation   3
                 ============

  Method of estimation = Ordinary Least Squares

Dependent variable: RCONS
Current sample:  1955 to 1997
Number of observations:  43

 Mean of dependent variable = 146270.
Std. dev. of dependent var. = 79317.2
   Sum of squared residuals = .256040E+11
      Variance of residuals = .624487E+09
   Std. error of regression = 24989.7
                  R-squared = .903100
         Adjusted R-squared = .900737
    Durbin-Watson statistic = .029725
  F-statistic (zero slopes) = 382.117
 Schwarz Bayes. Info. Crit. = 20.3798
 Log of likelihood function = -495.418

           Estimated    Standard
Variable  Coefficient     Error       t-statistic
C         -.115228E+07  66538.5       -17.3175
LYD       109305.       5591.69       19.5478


            NONLINEAR LEAST SQUARES
            =======================
```

```
EQUATIONS: EQ

CONVERGENCE ACHIEVED AFTER  72 ITERATIONS

        Log of Likelihood Function =    -414.362
            Number of Observations =     43

                         Standard
Parameter  Estimate        Error        t-statistic
A1         16544.5       2615.60        6.32530
A2         .063304       .024133        2.62307
A3         1.21694       .031705        38.3839

Standard Errors computed from quadratic form of
analytic first derivatives (Gauss)

Dependent variable: RCONS

 Mean of dependent variable = 146270.
Std. dev. of dependent var. = 79317.2
   Sum of squared residuals = .590213E+09
      Variance of residuals = .147553E+08
   Std. error of regression = 3841.27
                  R-squared = .997766
         Adjusted R-squared = .997655
    Durbin-Watson statistic = .253234


                      Equation   4
                      ============

   FIRST-ORDER SERIAL CORRELATION OF THE ERROR

   MAXIMUM LIKELIHOOD ITERATIVE TECHNIQUE

   CONVERGENCE ACHIEVED AFTER    5 ITERATIONS

Dependent variable: RCONS
Current sample:  1955 to 1997
Number of observations:  43

(Statistics based on transformed data)
 Mean of dependent variable = 23312.9
Std. dev. of dependent var. = 10432.7
   Sum of squared residuals = .137718E+09
      Variance of residuals = .335899E+07
   Std. error of regression = 1832.75
                  R-squared = .970084
         Adjusted R-squared = .969354
    Durbin-Watson statistic = 1.44365
Rho (autocorrelation coef.) = .876923
      Standard error of rho = .066300
        t-statistic for rho = 13.2266
  F-statistic (zero slopes) = 1319.94
 Log of likelihood function = -383.807
(Statistics based on original data)
 Mean of dependent variable = 146270.
Std. dev. of dependent var. = 79317.2
   Sum of squared residuals = .140391E+09
      Variance of residuals = .342417E+07
   Std. error of regression = 1850.45
                  R-squared = .999470
         Adjusted R-squared = .999457
    Durbin-Watson statistic = 1.43657

           Estimated    Standard
Variable  Coefficient     Error       t-statistic
C         12034.8       3315.11       3.63029
RRYD      .140723       .275670E-02   51.0476

*****************************************************
```

1. `Equation 1` vs. `Equation 2`

   系列相関の検定：

   `Equation 2` は

$$\mathtt{RCONS}_t = \beta_1 + \beta_2 \mathtt{RYD}_t + u_t,$$
$$u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$$

   $H_0: \ \rho = 0$

   制約付き MLE $\Longrightarrow$ `Equation 1`

   制約なし MLE $\Longrightarrow$ `Equation 2`

   尤度関数は,

$$\begin{aligned}
&\log L(\beta, \sigma_\epsilon^2, \rho) \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma_\epsilon^2) + \frac{1}{2}\log(1-\rho^2) \\
&\quad -\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}(\mathtt{RCONS}_t^* - \beta_1 \mathtt{CONST}_t^* - \beta_2 \mathtt{RYD}_t^*)^2
\end{aligned}$$

   となる。ただし,

$$\mathtt{RCONS}_t^* = \begin{cases} \sqrt{1-\rho^2}\,\mathtt{RCONS}_t, & t=1 \text{ のとき}, \\ \mathtt{RCONS}_t - \rho \mathtt{RCONS}_{t-1}, & t = 2, 3, \cdots, T \end{cases}$$

$$\mathtt{CONST}_t^* = \begin{cases} \sqrt{1-\rho^2}, & t=1 \text{ のとき}, \\ 1-\rho, & t = 2, 3, \cdots, T \text{ のとき}, \end{cases}$$

$$\mathtt{RYD}_t^* = \begin{cases} \sqrt{1-\rho^2}\,\mathtt{RYD}_t, & t=1 \text{ のとき}, \\ \mathtt{RYD}_t - \rho \mathtt{RYD}_{t-1}, & t = 2, 3, \cdots, T \text{ のとき}, \end{cases}$$

   とする。

   - $\rho = 0$ の制約付き MLE (`Equation 1`)

$$\max_{\beta, \sigma_\epsilon^2} \log L(\beta, \sigma_\epsilon^2, 0)$$

     制約付き最尤推定量 $\Longrightarrow \widetilde{\beta}, \widetilde{\sigma}_\epsilon^2$

     `Log of likelihood function = -431.289`

- $\rho = 0$ の制約なし MLE (`Equation 2`)

$$\max_{\beta, \sigma_\epsilon^2, \rho} \log L(\beta, \sigma_\epsilon^2, \rho)$$

制約なし最尤推定量 $\Longrightarrow \widehat{\beta}, \widehat{\sigma}_\epsilon^2, \widehat{\rho}$

`Log of likelihood function = -385.419`

尤度比検定統計量は,

$$\begin{aligned}
&-2\log(\lambda) \\
&= -2\log\left(\frac{L(\widetilde{\beta}, \widetilde{\sigma}_\epsilon^2, 0)}{L(\widehat{\beta}, \widehat{\sigma}_\epsilon^2, \widehat{\rho})}\right) \\
&= -2\left(\log L(\widetilde{\beta}, \widetilde{\sigma}_\epsilon^2, 0) - \log L(\widehat{\beta}, \widehat{\sigma}_\epsilon^2, \widehat{\rho})\right) \\
&= -2\Big(-431.289 - (-385.419)\Big) \\
&= 91.74
\end{aligned}$$

となり，そのときの漸近分布は,

$$-2\log(\lambda) \sim \chi^2(G)$$

となる。G は制約数とする。

この場合，$G = 1$ なので，$-2\log(\lambda) \sim \chi^2(1)$ となる。$chi^2(1)$ の上側 1 % 点は 6.635 なので,

$$91.74 > 6.635$$

となり，$\rho = 0$ の仮説を棄却する。

したがって，誤差項に系列相関があると判定される。

2. `Equation 1` vs. `NONLINEAR LEAST SQUARES`

関数型の選択 (線型かどうか)：

`NONLINEAR LEAST SQUARES` で,

$$\mathtt{RCONS}_t = a1 + a2\frac{\mathtt{RYD}_t^{a3} - 1}{a3} + u_t$$

を推定する。

$a3 = 1$ ならば,

$$\mathtt{RCONS}_t = (a1 - a2) + a2\mathtt{RYD}_t + u_t$$

となり，`Equation 1` を推定するのと同じ。

H_0： $a3 = 1$ とする。($G = 1$)

- $a3 = 1$ の制約付き MLE (`Equation 1`)

`Log of likelihood function = -431.289`

- $a3 = 1$ の制約なし MLE (`NONLINEAR LEAST SQUARES`)

`Log of likelihood function = -414.362`

尤度比検定統計量は,

$$\begin{aligned}
&-2\log(\lambda) \\
&= -2\Big(-431.289 - (-414.362)\Big) \\
&= 33.854
\end{aligned}$$

となり，$-2\log(\lambda) \sim \chi^2(1)$ となる。$chi^2(1)$ の上側 1 % 点は 6.635 なので,

$$33.854 > 6.635$$

となり，$a3 = 1$ の仮説を棄却する。

したがって，関数型は線型ではない。

3. `Equation 3` vs. `NONLINEAR LEAST SQUARES`

関数型の選択 (対数線型かどうか)：

`NONLINEAR LEAST SQUARES` の

$$\mathtt{RCONS}_t = a1 + a2\frac{\mathtt{RYD}_t^{a3} - 1}{a3} + u_t$$

について，$a3 = 0$ ならば,

$$\mathtt{RCONS}_t = a1 + a2\log(\mathtt{RYD}_t) + u_t$$

となり，`Equation 3` を推定するのと同じ。

H_0： $a3 = 0$ とする。($G = 1$)

- $a3 = 0$ の制約付き MLE (`Equation 3`)

`Log of likelihood function = -495.418`

- $a3 = 0$ の制約なし MLE (`NONLINEAR LEAST SQUARES`)

`Log of likelihood function = -414.362`

尤度比検定統計量は,

$$\begin{aligned}
&-2\log(\lambda) \\
&= -2\Big(-495.418 - (-414.362)\Big) \\
&= 162.112
\end{aligned}$$

となり，$-2\log(\lambda) \sim \chi^2(1)$ となる。$chi^2(1)$ の上側 1 % 点は 6.635 なので，

$$162.112 > 6.635$$

となり，$a3 = 0$ の仮説を棄却する。

したがって，関数型は対数線型でもない。

4. `Equation 1` vs. `Equation 4`

系列相関と関数型 (線型) の同時検定：

`Equation 4` は

$$\begin{aligned} \texttt{RCONS}_t &= a1 + a2\frac{\texttt{RYD}_t^{a3} - 1}{a3} + u_t, \\ u_t &= \rho u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2) \end{aligned}$$

H_0： $a3 = 1,\ \rho = 0$

制約付き MLE $\Longrightarrow$ `Equation 1`

制約なし MLE $\Longrightarrow$ `Equation 4`

注)

`PROGRAM` の 13〜15 行について，$a3$ の値を 0.01 刻みで変えて，推定を行った結果，$a3 = 1.15$ のときが，最も尤度関数が大きくなった。

尤度比検定統計量は，

$$\begin{aligned} &-2\log(\lambda) \\ &= -2\Big(-431.289 - (-383.807)\Big) \\ &= 94.964 \end{aligned}$$

となり，$-2\log(\lambda) \sim \chi^2(2)$ となる。$chi^2(2)$ の上側 1 % 点は 9.210 なので，

$$94.964 > 9.210$$

となり，$a3 = 1$ かつ $\rho = 0$ の仮説を棄却する。

したがって，系列相関を考慮に入れても線型ではない。

13　不均一分散

回帰式が

$$y_t = \alpha + \beta x_t + u_t$$

の場合を考える。x_t が外生変数，y_t は内生変数，u_t は互いに独立な同一の分布を持つ攪乱項 (最小二乗法に必要な仮定) とする。「独立な同一の分布」の意味は「攪乱項 u_1, u_2, $\cdots$, u_T はそれぞれ独立に平均ゼロ，分散 σ^2 の分布する」である。

分散が時点に依存する場合，代表的には，分散が他の変数 (例えば，z_t) に依存する場合，すなわち，u_t の平均はゼロ，分散は $\sigma_*^2 z_t^2$ の場合は，最小二乗法の仮定に反する。そのため，単純には，$y_t = \alpha + \beta x_t + u_t$ に最小二乗法を適用できない。以下のような修正が必要となる。

$$\begin{aligned} \frac{y_t}{z_t} &= \alpha\frac{1}{z_t} + \beta\frac{x_t}{z_t} + \frac{u_t}{z_t} \\ &= \alpha\frac{1}{z_t} + \beta\frac{x_t}{z_t} + u_t^* \end{aligned}$$

このとき，新たな攪乱項 u_t^* は平均ゼロ，分散 σ_*^2 の分布となる (すなわち,「同一の」分布)。

$$\mathrm{E}(u_t^*) = \mathrm{E}\left(\frac{u_t}{z_t}\right) = \left(\frac{1}{z_t}\right)\mathrm{E}(u_t) = 0$$

u_t の仮定 $\mathrm{E}(u_t) = 0$ が使われている。

$$\mathrm{V}(u_t^*) = \mathrm{V}\left(\frac{u_t}{z_t}\right) = \left(\frac{1}{z_t}\right)^2\mathrm{V}(u_t) = \sigma_*^2$$

u_t の仮定 $\mathrm{V}(u_t) = \sigma_*^2 z_t^2$ が最後に使われている。

よって，$\dfrac{y_t}{z_t}$, $\dfrac{1}{z_t}$, $\dfrac{x_t}{z_t}$ を新たな変数として，最小二乗法を適用することができる。

不均一分散の検定について

$$\widehat{u}_t^2 = \gamma z_t + \epsilon_t$$

を推定し，γ の推定値 $\widehat{\gamma}$ の有意性の検定を行う (通常の t 検定)。

z_t は回帰式に含まれる変数でもよい。例えば，u_t の平均はゼロ，分散は $\sigma_*^2 x_t^2$ の場合，各変数を x_t で割って，

$$\begin{aligned} \frac{y_t}{x_t} &= \alpha\frac{1}{x_t} + \beta + \frac{u_t}{x_t} \\ &= \alpha\frac{1}{x_t} + \beta + u_t^* \end{aligned}$$

を推定すればよい。β は定数項として推定されるが，意味は限界係数 (すなわち，傾き) と同じなので注意すること。

1. $$y_t = x_t\beta + u_t, \qquad u_t \sim N(0, \sigma_t^2)$$

$$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_T^2 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_T^2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_T} \end{pmatrix}' \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_T} \end{pmatrix}$$

2. $$y_t = x_t\beta + u_t, \qquad u_t \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = (z_t\alpha)^2$$

最尤法により推定

$$\begin{aligned} L = &-\frac{T}{2}\log(2\pi) - \log(z_t\alpha) \\ &-\frac{1}{2}\sum_{t=1}^T \left(\frac{y_t - x_t\beta}{z_t\alpha}\right)^2 \end{aligned}$$

3. $$y_t = x_t\beta + u_t, \qquad u_t \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = \sigma^2 (x_t\beta)^p$$

4. $$y_t = x_t\beta + u_t, \qquad u_t \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = \exp(z_t\alpha)$$

5. ARCH (autoregressive conditional heteroscedasticity) モデル

$$y_t = x_t\beta + u_t, \qquad u_t \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

ただし，$\alpha_0 > 0,\ \alpha_1 \ge 0$

14 自己相関

***DW* について：** 最小自乗法の仮定の一つに,「撹乱項 u_1, u_2, $\cdots$, u_T はそれぞれ独立に分布する」というものがあった。ダービン・ワトソン比 (DW) とは，誤差項の系列相関，すなわち，u_t と u_{t-1} との間の相関の有無を検定するために考案された。

$\Longrightarrow$ 時系列データのときのみ有効

$u_1, u_2, \cdots, u_T$ の系列について，それぞれの符号が，+ + + - - - - + + - - - + + のように，プラスが連続で続いた後で，マイナスが連続で続くというような場合，$u_1, u_2, \cdots$, u_T は正の系列相関があると言う。また，+ - + - + - + - + のように交互にプラス，マイナスになる場合，$u_1, u_2, \cdots$, u_T 負の系列相関があると言う。

特徴：$u_1, u_2, \cdots, u_t$ から u_{t+1} の符号が予想できる。$\Longrightarrow$「$u_1, u_2, \cdots, u_T$ はそれぞれ独立に分布する」という仮定に反する。

すなわち，ダービン・ワトソン比とは，回帰式が

$$\begin{aligned} y_t &= \alpha + \beta x_t + u_t, \\ u_t &= \rho u_{t-1} + \epsilon_t, \end{aligned}$$

のときに，$H_0:\ \rho = 0$, $H_1:\ \rho \neq 0$ の検定である。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ は互いに独立とする。

ダービン・ワトソン比の定義は次の通りである。

$$DW = \frac{\sum_{t=2}^T (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^T \widehat{u}_t^2}$$

DW は近似的に，次のように表される。

$$\begin{aligned} DW &= \frac{\sum_{t=2}^T (\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^T \widehat{u}_t^2} \\ &= \frac{\sum_{t=2}^T \widehat{u}_t^2 - 2\sum_{t=2}^T \widehat{u}_t\widehat{u}_{t-1} + \sum_{t=2}^T \widehat{u}_{t-1}^2}{\sum_{t=1}^T \widehat{u}_t^2} \\ &= \frac{2\sum_{t=1}^T \widehat{u}_t^2 - (\widehat{u}_1^2 + \widehat{u}_T^2)}{\sum_{t=1}^T \widehat{u}_t^2} - 2\frac{\sum_{t=2}^T \widehat{u}_t\widehat{u}_{t-1}}{\sum_{t=1}^T \widehat{u}_t^2} \\ &\approx 2(1 - \widehat{\rho}), \end{aligned}$$

以下の 2 つの近似が用いられる。

$$\frac{\widehat{u}_1^2 + \widehat{u}_T^2}{\sum_{t=1}^T \widehat{u}_t^2} \approx 0,$$

$$\frac{\sum_{t=2}^T \widehat{u}_t\widehat{u}_{t-1}}{\sum_{t=1}^T \widehat{u}_t^2} = \frac{\sum_{t=2}^T \widehat{u}_t\widehat{u}_{t-1}}{\sum_{t=2}^T \widehat{u}_{t-1}^2 + \widehat{u}_T^2}$$

$$\approx \frac{\sum_{t=2}^{T} \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^{T} \widehat{u}_{t-1}^2}$$
$$= \widehat{\rho},$$

すなわち，$\widehat{\rho}$ は $\widehat{u}_t$ と $\widehat{u}_{t-1}$ の回帰係数である。$u_t = \rho u_{t-1} + \epsilon_t$ において，u_t, u_{t-1} の代わりに $\widehat{u}_t$, $\widehat{u}_{t-1}$ に置き換えて，ρ の推定値 $\widehat{\rho}$ を求める。

1. DW の値が 2 前後のとき，系列相関なし ($\widehat{\rho} = 0$ のとき，$DW \approx 2$)。

2. DW が 2 より十分に小さいとき，正の系列相関と判定される。

3. DW が 2 より十分に大きいとき，負の系列相関と判定される。

正確な判定には，データ数 T とパラメータ数 k に依存する。表 1 と表 2 を参照せよ。

表 1 と表 2 で，k' は定数項を除く説明変数の数を表すものとする。

図 4： 正の系列相関

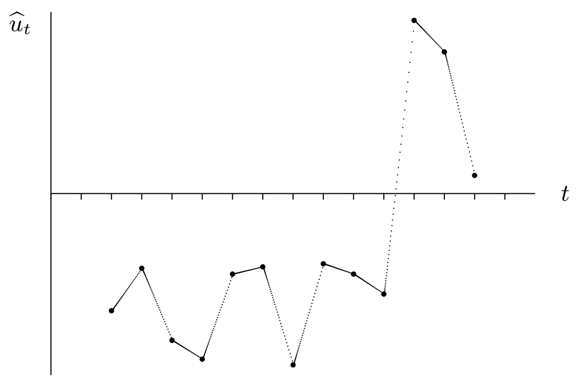
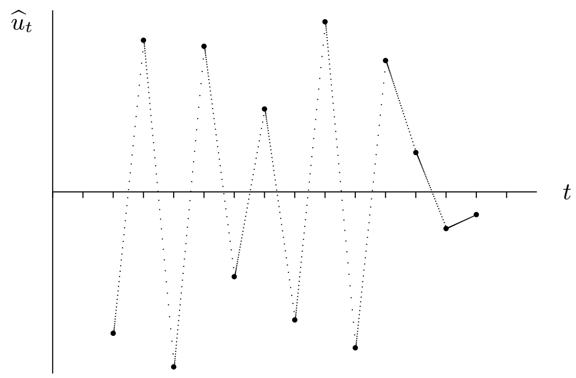


図 5： 負の系列相関



系列相関のもとで回帰式の推定： 回帰式が

$$Y_t = \alpha + \beta X_t + u_t,$$
$$u_t = \rho u_{t-1} + \epsilon_t,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ は互いに独立とする。
u_t を消去すると，

$$(Y_t - \rho Y_{t-1}) = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \epsilon_t,$$

となり，
$Y_t^* = (Y_t - \rho Y_{t-1})$,
$X_t^* = (X_t - \rho X_{t-1})$
を新たな変数として，

$$Y_t^* = \alpha' + \beta X_t^* + \epsilon_t,$$

に最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ は互いに独立とするなので，最小二乗法を適用が可能となる。ただし，$\alpha' = \alpha(1-\rho)$ の関係が成り立つことに注意。
より一般的に，回帰式が

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t,$$
$$u_t = \rho u_{t-1} + \epsilon_t,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ は互いに独立とする。

表 1: ダービン・ワトソン統計量の 5 % 点の上限と下限

(1) $k'=1$

	A		B		C		D		E	
T	下限	上限	下限	上限	下限	上限	下限	上限	下限	上限
	0	dl	dl	du	du	$4-du$	$4-du$	$4-dl$	$4-dl$	4
15	0	1.08	1.08	1.36	1.36	2.64	2.64	2.92	2.92	4
20	0	1.20	1.20	1.41	1.41	2.59	2.59	2.80	2.80	4
25	0	1.29	1.29	1.45	1.45	2.55	2.55	2.71	2.71	4
30	0	1.35	1.35	1.49	1.49	2.51	2.51	2.65	2.65	4

(2) $k'=2$

	A		B		C		D		E	
T	下限	上限	下限	上限	下限	上限	下限	上限	下限	上限
	0	dl	dl	du	du	$4-du$	$4-du$	$4-dl$	$4-dl$	4
15	0	0.95	0.95	1.54	1.54	2.46	2.46	3.05	3.05	4
20	0	1.10	1.10	1.54	1.54	2.46	2.46	2.90	2.90	4
25	0	1.21	1.21	1.55	1.55	2.45	2.45	2.79	2.79	4
30	0	1.28	1.28	1.57	1.57	2.43	2.43	2.72	2.72	4

(3) $k'=3$

	A		B		C		D		E	
T	下限	上限	下限	上限	下限	上限	下限	上限	下限	上限
	0	dl	dl	du	du	$4-du$	$4-du$	$4-dl$	$4-dl$	4
15	0	0.82	0.82	1.75	1.75	2.25	2.25	2.25	3.18	4
20	0	1.00	1.00	1.68	1.68	2.32	2.32	2.32	3.00	4
25	0	1.12	1.12	1.66	1.66	2.34	2.34	2.34	2.88	4
30	0	1.21	1.21	1.65	1.65	2.35	2.35	2.35	2.79	4

(4) $k'=4$

	A		B		C		D		E	
T	下限	上限	下限	上限	下限	上限	下限	上限	下限	上限
	0	dl	dl	du	du	$4-du$	$4-du$	$4-dl$	$4-dl$	4
15	0	0.69	0.69	1.97	1.97	2.03	2.03	3.31	3.31	4
20	0	0.90	0.90	1.83	1.83	2.17	2.17	3.10	3.10	4
25	0	1.04	1.04	1.77	1.77	2.23	2.23	2.96	2.96	4
30	0	1.14	1.14	1.74	1.74	2.26	2.26	2.86	2.86	4

(5) $k'=5$

	A		B		C		D		E	
T	下限	上限	下限	上限	下限	上限	下限	上限	下限	上限
	0	dl	dl	du	du	$4-du$	$4-du$	$4-dl$	$4-dl$	4
15	0	0.56	0.56	2.21	—	—	2.21	3.44	3.44	4
20	0	0.79	0.79	1.99	1.99	2.01	2.01	3.21	3.21	4
25	0	0.95	0.95	1.89	1.89	2.11	2.11	3.05	3.05	4
30	0	1.07	1.07	1.83	1.83	2.17	2.17	2.93	2.93	4

A: 正の系列相関あり
B: 系列相関の有無を判定不能
C: 系列相関なし
D: 系列相関の有無を判定不能
E: 負の系列相関あり

表 2: ダービン・ワトソン統計量の 5 % 点の上限と下限

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
T	dl	du	dl	du	dl	du	dl	du	dl	du
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

u_t を消去すると，

$$\begin{aligned}(Y_t - \rho Y_{t-1}) = &\beta_1(X_{1t} - \rho X_{1,t-1}) \\ &+ \beta_2(X_{1t} - \rho X_{2,t-1}) \\ &+ \cdots \\ &+ \beta_k(X_{1t} - \rho X_{k,t-1}) + \epsilon_t,\end{aligned}$$

となり，
$Y_t^* = (Y_t - \rho Y_{t-1})$,
$X_{1t}^* = (X_{1t} - \rho X_{1,t-1})$,
$X_{2t}^* = (X_{2t} - \rho X_{2,t-1})$,
$\cdots$,
$X_{kt}^* = (X_{kt} - \rho X_{k,t-1})$
を新たな変数として，

$$Y_t^* = \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \cdots + \beta_k X_{kt}^* + \epsilon_t$$

最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ は互いに独立とするなので，最小二乗法を適用が可能となる。
ρ の求め方について

1. ダービン・ワトソン比がゼロに近い場合，$\rho = 1$ と近似して，

$$Y_t^* = \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \cdots + \beta_k X_{kt}^* + \epsilon_t,$$

として推定する。ただし，$Y_t^* = Y_t - Y_{t-1}$，$X_{1t}^* = X_{1t} - X_{1,t-1}, \cdots, X_{kt}^* = X_{kt} - X_{k,t-1}$ とする。

2. DW は近似的に $DW \approx 2(1 - \widehat{\rho})$ と表されるので，DW から ρ の推定値 $\widehat{\rho}$ を逆算して，
$Y_t^* = (Y_t - \widehat{\rho} Y_{t-1})$,
$X_{1t}^* = (X_{1t} - \widehat{\rho} X_{1,t-1})$,
$X_{2t}^* = (X_{2t} - \widehat{\rho} X_{2,t-1})$,
$\cdots$,
$X_{kt}^* = (X_{kt} - \widehat{\rho} X_{k,t-1})$
を新たな変数として，

$$Y_t^* = \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \cdots + \beta_k X_{kt}^* + \epsilon_t,$$

に最小二乗法を適用する。

3. 収束計算によって求める。

(i) まず，

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t,$$

に最小二乗法を適用し，$\widehat{u}_1, \widehat{u}_2, \cdots, \widehat{u}_T$ を求める。

(ii) 次に，

$$\widehat{u}_t = \rho \widehat{u}_{t-1} + \epsilon_t,$$

に最小二乗法を適用し，$\widehat{\rho}$ を求める。

(iii) データを変換する。すなわち，
$Y_t^* = (Y_t - \widehat{\rho} Y_{t-1})$,
$X_{1t}^* = (X_{1t} - \widehat{\rho} X_{1,t-1})$,
$X_{2t}^* = (X_{2t} - \widehat{\rho} X_{2,t-1})$,
$\cdots$,
$X_{kt}^* = (X_{kt} - \widehat{\rho} X_{k,t-1})$
を新たな変数として，

$$Y_t^* = \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \cdots + \beta_k X_{kt}^* + \epsilon_t,$$

を計算して，$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ を求める。

(iv) さらに，

$$\widehat{u}_t = Y_t - \widehat{\beta}_1 X_{1t} - \widehat{\beta}_2 X_{2t} - \cdots - \widehat{\beta}_k X_{kt},$$

から $\widehat{u}_t$ を再計算して，

$$\widehat{u}_t = \rho \widehat{u}_{t-1} + \epsilon_t,$$

から，最小二乗法により $\widehat{\rho}$ を求める。

(v) (iii) と (iv) を収束するまで繰り返す。

$\Longrightarrow$ コクラン・オーカット法

(iv) で得られた $\widehat{\rho}$ の有意性から，H_0 : $\rho = 0$ の検定を行う方法もある。$\Longrightarrow$ Wald 検定

1. 攪乱項に一階の自己相関があるモデル

$$y_t = x_t \beta + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t,$$

$$\epsilon_t \sim \text{iid } N(0, \sigma_\epsilon^2)$$

2. u_t の分散共分散行列

$$\sigma^2 \Omega = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho \\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1 \end{pmatrix}$$

$$
\begin{pmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\
\rho & 1 & \rho & \cdots & \rho^{T-2} \\
\rho^2 & \rho & 1 & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \rho \\
\rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1
\end{pmatrix}^{-1}
$$

$$
= \begin{pmatrix}
1 & -\rho & 0 & \cdots & 0 \\
-\rho & 1+\rho^2 & -\rho & & \vdots \\
0 & -\rho & \ddots & \ddots & 0 \\
\vdots & & \ddots & \ddots & -\rho \\
0 & \cdots & 0 & -\rho & 1
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\
-\rho & 1 & 0 & \cdots & 0 \\
0 & -\rho & 1 & & \vdots \\
\vdots & & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -\rho & 1
\end{pmatrix}'
\times \begin{pmatrix}
\sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\
-\rho & 1 & 0 & \cdots & 0 \\
0 & -\rho & 1 & & \vdots \\
\vdots & & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -\rho & 1
\end{pmatrix}
$$

$$
\sigma^2 = \frac{\sigma_\epsilon}{1-\rho^2}
$$

3. u と ϵ との関係

$$
\begin{pmatrix}
\sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\
-\rho & 1 & 0 & \cdots & 0 \\
0 & -\rho & 1 & & \vdots \\
\vdots & & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -\rho & 1
\end{pmatrix}
\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}
= \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}
$$

4. 系列相関があるかないかの検定

   (a) ダービン・ワトソン比 (遅れのある変数が説明変数に含まれているとき，h 統計量)

   (b) H_0： $\rho = 0$ の検定 (LM, LR, W)

遅れのある変数： 習慣的効果を考慮に入れたモデル：

$$
y_t = \alpha + \beta x_t + \gamma y_{t-1} + u_t,
$$

ラグ付き内生変数が説明変数に用いられる。

(注)

1. x_t の y_t 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，$y_t = y_{t-1}$ となるときの，x_t から y_t への影響を示す効果である。

   $$
   y_t = \alpha + \beta x_t + \gamma y_t + u_t,
   $$

   として，y_t について解くと，

   $$
   y_t = \frac{\alpha}{1-\gamma} + \frac{\beta}{1-\gamma} + \frac{1}{1-\gamma} u_t,
   $$

   となり，$\dfrac{\beta}{1-\gamma}$ が x_t の y_t 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型不偏推定量ではなくなる。

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

   $\Longrightarrow$ 多重共線性の可能性が高い。

3. DW 統計量は意味をなさない (詳細略)。実際には誤差項に系列相関があるにもかかわらず，標本数 (データ数) が増えるにつれて，DW は 2 に近づいてしまう。すなわち，系列相関なしと判定されてしまう。

   $\Longrightarrow$ 代わりに，h 統計量を使う。

   h 統計量は次のように求められる。

   $$
   h = \widehat{\rho}\sqrt{\frac{T}{1 - T s_{\widehat{\gamma}}^2}},
   $$

   T は標本数 (データ数)，$s_{\widehat{\gamma}}$ は $\widehat{\gamma}$ の標準誤差，$\widehat{\rho}$ は $\widehat{\rho} = \dfrac{\sum_{t=2}^T \widehat{u}_t \widehat{u}_{t-1}}{\sum_{t=2}^T \widehat{u}_{t-1}^2}$ で，$\widehat{u}_t$ はもとのモデルの最小二乗法による残差である。

   帰無仮説は，$u_t = \rho u_{t-1} + \epsilon_t$ のモデルにおいて，H_0 : $\rho = 0$ であり，対立仮説は H_1 ： $\rho \neq 0$ である。帰無

仮説 H_0： $\rho = 0$ のもとで，標本数 T が増加するにつれて，h は標準正規分布に近づくことが知られている。よって，

(a) $|h| > z_{\alpha/2}$ のとき，有意水準 100 % で H_0 を棄却する。

(b) $|h| < z_{\alpha/2}$ のとき，有意水準 100 % で H_0 を採択する。

ただし，$z_{\alpha/2}$ は標準正規分布表から得られた $100 \times \dfrac{\alpha}{2}$ % 点の値である。

15 特定化誤差

適切な変数がない場合 (omitting relevant variables) と不適切な変数を含めた場合 (including irrelevant variable)

$$y = X_1\beta_1 + X_2\beta_2 + u,$$

$$\beta_1 : \ k_1 \times 1, \quad \beta_2 : \ k_2 \times 1$$

からの推定値を $\widehat{\beta}_1$, $\widehat{\beta}_2$, s^2 とする。

$$y = X_1\beta_1 + u$$

からの推定値を β_1^*, σ^{*2} とする。

1. 適切な変数がない場合 (omitting relevant variables) 真のモデル

$$y = X_1\beta_1 + X_2\beta_2 + u$$

推定モデル

$$y = X_1\beta_1 + u$$

(a) β_1 の OLSE は

$$\begin{aligned}\beta_1^* &= (X_1'X_1)^{-1}X_1'y \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + u) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'u\end{aligned}$$

$$\begin{aligned}\mathrm{E}(\beta_1^*) &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \\ &\neq \beta_1\end{aligned}$$

2. 不適切な変数を含めた場合 (including irrelevant variable) 真のモデル

$$y = X_1\beta_1 + u$$

推定モデル

$$\begin{aligned}y &= X_1\beta_1 + X_2\beta_2 + u \\ &= (X_1 \ X_2)\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix} + u \\ &= X\beta + u\end{aligned}$$

(a) $\beta = (\beta_1', \beta_2')'$ の OLSE は

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X_1\beta_1 + u) \\ &= (X'X)^{-1}X'(X\begin{pmatrix}\beta_1 \\ 0\end{pmatrix} + u) \\ &= \begin{pmatrix}\beta_1 \\ 0\end{pmatrix} + (X'X)^{-1}X'u\end{aligned}$$

(b) $\widehat{\beta}_1$ は以下の通り不偏

$$\mathrm{E}(\widehat{\beta}) = \begin{pmatrix}\beta_1 \\ 0\end{pmatrix}$$

(c) しかし，分散は

$$\mathrm{V}(\widehat{\beta}) = \sigma^2(X'X)^2$$

$$\mathrm{V}\begin{pmatrix}\widehat{\beta}_1 \\ \widehat{\beta}_2\end{pmatrix} = \sigma^2\begin{pmatrix}X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2\end{pmatrix}^{-1}$$

(d) 逆行列の公式

$$\begin{pmatrix}A & B \\ B' & D\end{pmatrix}^{-1} = \begin{pmatrix}X & Y \\ Y' & Z\end{pmatrix}$$

X, Y, Z は次に与えられる。

$$\begin{aligned}X &= (A - BD^{-1}B')^{-1} \\ &= A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A^{-1}\end{aligned}$$

$$\begin{aligned}Y &= -(A - BD^{-1}B')^{-1}BD^{-1} \\ &= -A^{-1}B(D - B'A^{-1}B)^{-1}\end{aligned}$$

$$\begin{aligned}Z &= (D - B'A^{-1}B)^{-1} \\ &= D^{-1} + D^{-1}B'(A - BD^{-1}B')^{-1}BD^{-1}\end{aligned}$$

		真のモデル	
		$\beta_2 = 0$	$\beta_2 \neq 0$
採用されたモデル	$\beta_2 = 0$	β_1^*, σ^{*2}, efficient and unbiased	β_1^*, σ^{*2}, biased
	$\beta_2 \neq 0$	$\widehat{\beta}_1$, $\widehat{\beta}_2$, s^2, unbiased but inefficient	$\widehat{\beta}_1$, $\widehat{\beta}_2$, s^2, efficient and unbiased

(e) $\widehat{\beta}_1$ の分散は,

$$\begin{aligned} \mathrm{V}(\widehat{\beta}_1) &= \sigma^2(X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1')^{-1} \\ &\neq \sigma^2(X_1'X_1)^{-1} \end{aligned}$$

$\mathrm{V}(\widehat{\beta}_1) - \mathrm{V}(\beta_1^*)$ は正値定符号行列

(f) $\widehat{\beta}_1$ は不偏であるが，有効ではない。

16 多重共線性

1. 回帰モデル $y = X\beta + u$, $u \sim N(0, \sigma^2 I)$ について

説明変数間に線形関係が存在する場合，すなわち，説明変数を

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{pmatrix}$$

としたとき,

$$a_1 \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{T1} \end{pmatrix} + a_2 \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{T2} \end{pmatrix} + \cdots + a_k \begin{pmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{Tk} \end{pmatrix} = 0$$

を満たす a_1, a_2, $\cdots$, a_k が存在する場合，$\widehat{\beta} = (X'X)^{-1}X'y$ の $(X'X)^{-1}$ が無限大になる (なぜなら，$X'X$ の行列式の値がゼロになるためである)。

2. 実際上の問題としては，完全な共線性はないが，説明変数間に高い相関がある場合がある。この場合，多重共線性という問題が起こる。

3. 2 変数の場合の例

$$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + u_t, \quad u_t \sim \mathrm{iid}N(0, \sigma^2)$$

x_{t1} と x_{t2} との相関係数を r とする。

$$\begin{aligned} r^2 &= \frac{\left(\sum_t (x_{t1} - \overline{x}_{t1})(x_{t2} - \overline{x}_{t2})\right)^2}{\left(\sum_t (x_{t1} - \overline{x}_{t1})^2\right)\left(\sum_t (x_{t2} - \overline{x}_{t2})^2\right)} \\ &= \frac{\left(\sum_t x_{t1}x_{t2}\right)^2}{\left(\sum_t x_{t1}^2\right)\left(\sum_t x_{t2}^2\right)} \end{aligned}$$

簡単化のため $\overline{x}_{t1} = \overline{x}_{t2} = 0$ とする。

$$\begin{aligned} \mathrm{V}(\widehat{\beta}) &= \sigma^2(X'X)^{-1} \\ &= \sigma^2 \begin{pmatrix} \sum_t x_{t1}^2 & \sum_t x_{t1}x_{t2} \\ \sum_t x_{t1}x_{t2} & \sum_t x_{t2}^2 \end{pmatrix}^{-1} \end{aligned}$$

$$\begin{aligned} &\begin{pmatrix} \sum_t x_{t1}^2 & \sum_t x_{t1}x_{t2} \\ \sum_t x_{t1}x_{t2} & \sum_t x_{t2}^2 \end{pmatrix}^{-1} \\ &= \frac{1}{\left(\sum_t x_{t1}^2\right)\left(\sum_t x_{t2}^2\right) - \left(\sum_t x_{t1}x_{t2}\right)^2} \\ &\times \begin{pmatrix} \sum_t x_{t2}^2 & -\sum_t x_{t1}x_{t2} \\ -\sum_t x_{t1}x_{t2} & \sum_t x_{t1}^2 \end{pmatrix} \\ &= \frac{1}{\left(\sum_t x_{t1}^2\right)\left(\sum_t x_{t2}^2\right)(1 - r^2)} \\ &\times \begin{pmatrix} \sum_t x_{t2}^2 & -\sum_t x_{t1}x_{t2} \\ -\sum_t x_{t1}x_{t2} & \sum_t x_{t1}^2 \end{pmatrix} \end{aligned}$$

したがって，それぞれの推定値の分散は

$$\mathrm{V}(\widehat{\beta}_1) = \frac{\sigma^2}{\left(\sum_t x_{t1}^2\right)(1 - r^2)}$$

$$\mathrm{V}(\widehat{\beta}_2) = \frac{\sigma^2}{\left(\sum_t x_{t2}^2\right)(1 - r^2)}$$

となる。

4. x_{t1} と x_{t2} との間の相関が高いとき，すなわち，$r^2 \longrightarrow 1$ のとき，$\mathrm{V}(\widehat{\beta}_1) \longrightarrow \infty$，$\mathrm{V}(\widehat{\beta}_2) \longrightarrow \infty$ となる。このとき，t 値はゼロに近づく。$\Longrightarrow$ 必要の無い変数と誤って，その変数を落とす可能性が大きくなる。

5. 多重共線性の対応策としては，Ridge 回帰，主成分回帰が考えれれるが，基本的には対応策はないものと考えてよい。$\Longrightarrow$ 実証分析では最大の問題点

6. Ridge 回帰：

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'y \\ &= (Q'\Lambda Q)^{-1}X'y\end{aligned}$$

ただし，Λ は対角要素に固有値，非対角要素はゼロの行列，Q は対応する固有ベクトルから成る行列とする。

多重共線性の問題点：$(X'X)^{-1}$ が存在しないこと，すなわち，$X'X$ の固有値のいくつかがゼロになること。

回避のためには，対角行列 $\Delta > 0$ を用いて，

$$\begin{aligned}\widehat{\beta}_R &= \left(Q'(\Lambda+\Delta)Q\right)^{-1}X'y \\ &= (X'X+Q'\Delta Q)^{-1}X'y \\ &= (X'X+R)^{-1}X'y\end{aligned}$$

とする。$X'X$ の部分を，$X'X+R$ として，無理矢理に正値定符号行列にして推定する。

Ridge 推定量は不偏推定量ではない。

17 時系列分析

17.1 時系列分析の準備

1. 定常性 (stationarity)：

   時系列データを $y_1, y_2, \cdots, y_T$ とする。

   (a) 弱定常性：

   $$\begin{aligned}&\mathrm{E}(y_t) = \mu \\ &\mathrm{E}(y_t-\mu)(y_{t-\tau}-\mu) = \gamma(\tau), \\ &\qquad \tau = 0,1,2,\cdots\end{aligned}$$

   (b) 強定常性：

   y_{t_1}, y_{t_2}, $\cdots$, y_{t_r} の r 個のデータの同時分布を $f(y_{t_1}, y_{t_2}, \cdots, y_{t_r})$ とする。

   $$\begin{aligned}&f(y_{t_1}, y_{t_2}, \cdots, y_{t_r}) \\ &= f(y_{t_1+\tau}, y_{t_2+\tau}, \cdots, y_{t_r+\tau})\end{aligned}$$

2. エルゴード性：

   十分遠い観測量が無相関となること。

3. 自己共分散関数：

   $$\mathrm{E}(y_t-\mu)(y_{t-\tau}-\mu) = \gamma(\tau),$$

   $$\tau = 0,1,2,\cdots$$

   $$\gamma(\tau) = \gamma(-\tau)$$

4. 自己相関関数：

   $$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

5. 標本平均，標本自己共分散，コレログラム (標本自己相関関数)：

   $$\begin{aligned}\widehat{\mu} &= \frac{1}{T}\sum_{t=1}^{T} y_t \\ \widehat{\gamma}(\tau) &= \frac{1}{T}\sum_{t=\tau+1}^{T}(y_t-\widehat{\mu})(y_{t-\tau}-\widehat{\mu}) \\ \widehat{\rho}(\tau) &= \frac{\widehat{\gamma}(\tau)}{\widehat{\gamma}(0)}\end{aligned}$$

6. ラグ作要素 (Lag Operator)：

   $$L^{\tau}y_t = y_{t-\tau}, \qquad \tau = 1,2,\cdots$$

7. 尤度関数 (Innovation Form)：

   $y_1, y_2, \cdots, y_T$ の同時分布 (尤度関数) は次のように書き換えられる。

   $$\begin{aligned}&L(y_1,, y_2, \cdots, y_T) \\ &= L(y_T|y_{T-1},\cdots,y_1)L(y_{T-1},\cdots,y_1) \\ &= L(y_T|y_{T-1},\cdots,y_1) \\ &\times L(y_{T-1}|y_{T-2},\cdots,y_1)L(y_{T-2},\cdots,y_1) \\ &\quad\vdots \\ &= L(y_1)\prod_{t=2}^{T}L(y_t|y_{t-1},\cdots,y_1)\end{aligned}$$

従って，対数尤度関数は

$$\begin{aligned} &\log L(y_1, y_2, \cdots, y_T) \\ &= \log L(y_1) + \sum_{t=2}^{T} \log L(y_t | y_{t-1}, \cdots, y_1) \end{aligned}$$

となる。

$\log L(y_t|y_{t-1}, \cdots, y_1)$ は,

$\mathrm{E}(y_t|y_{t-1}, \cdots, y_1)$,

$\mathrm{V}(y_t|y_{t-1}, \cdots, y_1)$

を求めれば得られる。

17.2 AR モデル

AR モデル: Autoregressive Model

1. AR(p)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

次のように書き直される。

$$\phi(L) y_t = \epsilon_t$$

ただし,

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

とする。

2. 定常性：$\phi(x) = 0$ の p 個の解がすべて 1 より大きいとき

3. 偏自己相関係数：

y_t と y_{t-k} との偏自己相関係数 $\phi_{k,k}$ とは，途中の y_{t-1}, $\cdots$, y_{t-k+1} の影響を取り除いた後での y_t と y_{t-k} の関係の強さを表す指標である。

$$\phi_{1,1} = \rho(1)$$

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{2,1} \\ \phi_{2,2} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix}$$

$$\begin{pmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{3,1} \\ \phi_{3,2} \\ \phi_{3,3} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \end{pmatrix}$$

$$\vdots$$

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(1) & 1 \end{pmatrix} \times \begin{pmatrix} \phi_{k,1} \\ \phi_{k,2} \\ \vdots \\ \phi_{k,k-1} \\ \phi_{k,k} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix}$$

$\phi_{k,k}$ の計算はクラメールの公式を用いればよい。

$$\phi_{k,k} = \frac{\begin{vmatrix} 1 & \rho(1) & \cdots \rho(k-2) & \rho(1) \\ \rho(1) & 1 & \rho(k-3) & \rho(2) \\ \vdots & \vdots & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots \quad \rho(1) & \rho(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \cdots \rho(k-2) & \rho(k-1) \\ \rho(1) & 1 & \rho(k-3) & \rho(k-2) \\ \vdots & \vdots & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots \quad \rho(1) & 1 \end{vmatrix}}$$

4. AR(1) モデル $y_t = \phi_1 y_{t-1} + \epsilon_t$ を考える。

   (a) 定常性の条件は， $\phi(x) = 1 - \phi_1 x = 0$ の解 $x = 1/\phi_1$ が 1 より大きいとき，すなわち， $\phi_1 < 1$ のときである。

   (b) AR(1) モデルを書き直すと,

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \epsilon_t \\ &= \phi_1^2 y_{t-2} + \epsilon_t + \phi_1 \epsilon_{t-1} \\ &= \phi_1^3 y_{t-3} + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} \\ &\ \ \vdots \\ &= \phi_1^s y_{t-s} + \epsilon_t + \phi_1 \epsilon_{t-1} + \cdots + \phi_1^{s-1} \epsilon_{t-s+1} \end{aligned}$$

   となる。s が大きくなるにつれて， ϕ_1^s がゼロに近づくことが，定常性の条件となる。

   (c) 定常ならば， $y_t = \phi_1 y_{t-1} + \epsilon_t$ は

$$y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \cdots$$

   と書き直すことができる (AR モデルの MA 表現)。

(d) AR(1) モデルの平均

$$\begin{aligned}\mu &= \mathrm{E}(y_t)\\ &= \mathrm{E}(\epsilon_t + \phi_1\epsilon_{t-1} + \phi_1^2\epsilon_{t-2} + \ \cdots)\\ &= \mathrm{E}(\epsilon_t) + \phi_1\mathrm{E}(\epsilon_{t-1}) + \phi_1^2\mathrm{E}(\epsilon_{t-2}) + \ \cdots\\ &= 0\end{aligned}$$

(e) AR(1) モデルの自己相関関数

$$\begin{aligned}y_t &= \phi_1^\tau y_{t-\tau}\\ &\quad + \epsilon_t + \phi_1\epsilon_{t-1} + \ \cdots \ + \phi_1^{\tau-1}\epsilon_{t-\tau+1}\end{aligned}$$

となることに注意すると，

$$\begin{aligned}&\gamma(\tau)\\ &= \mathrm{E}(y_t-\mu)(y_{t-\tau}-\mu)\\ &= \mathrm{E}(y_t y_{t-\tau})\\ &= \mathrm{E}\Big((\phi_1^\tau y_{t-\tau} + \epsilon_t + \phi_1\epsilon_{t-1}\\ &\qquad + \ \cdots \ + \phi_1^{\tau-1}\epsilon_{t-\tau+1})y_{t-\tau}\Big)\\ &= \phi_1^\tau\mathrm{E}(y_{t-\tau}y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau})\\ &+ \phi_1\mathrm{E}(\epsilon_{t-1}y_{t-\tau}) + \ \cdots\\ &+ \phi_1^{\tau-1}\mathrm{E}(\epsilon_{t-\tau+1}y_{t-\tau})\\ &= \phi_1^\tau\gamma(0)\end{aligned}$$

が得られる。よって，自己相関関数は

$$\begin{aligned}\rho(\tau) &= \frac{\gamma(\tau)}{\gamma(0)}\\ &= \phi_1^\tau\end{aligned}$$

となる。

また，AR(1) モデルの両辺に $y_{t-\tau}$ を掛けて，期待値をとると，

$$\mathrm{E}(y_t y_{t-\tau}) = \phi_1\mathrm{E}(y_{t-1}y_{t-\tau}) + \mathrm{E}(\epsilon_t y_{t-\tau})$$

$$\gamma(\tau) = \begin{cases}\phi_1\gamma(\tau-1), & (\tau \neq 0 \text{ のとき})\\ \phi_1\gamma(\tau-1)+\sigma^2, & (\tau = 0 \text{ のとき})\end{cases}$$

となる。$\gamma(\tau) = \gamma(-\tau)$ であるので， $\tau = 0$ のとき，

$$\begin{aligned}\gamma(0) &= \phi_1\gamma(1) + \sigma^2\\ &= \phi_1^2\gamma(0) + \sigma^2\end{aligned}$$

となる。したがって，

$$\gamma(0) = \frac{\sigma^2}{1-\phi_1^2}$$

が得られる。

(f) AR(1) モデルの偏自己相関関数

$$\phi_{1,1} = \rho(1) = \phi_1$$

$$\phi_{2,2} = \frac{\begin{vmatrix}1 & \rho(1)\\ \rho(1) & \rho(2)\end{vmatrix}}{\begin{vmatrix}1 & \rho(1)\\ \rho(1) & 1\end{vmatrix}} = \frac{\rho(2)-\rho(1)^2}{1-\rho(1)^2} = 0$$

(g) AR(1) モデルの推定：

i. 尤度関数

$$\begin{aligned}&\log L(y_T,\cdots,y_1)\\ &= \log L(y_1) + \sum_{t=1}^{T}\log L(y_t|y_{t-1},\cdots,y_1)\\ &= -\frac{1}{2}\log(2\pi)\\ &\quad -\frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi_1^2}\right)\\ &\quad -\frac{1}{\sigma^2/(1-\phi_1^2)}y_1^2\\ &\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2)\\ &\quad -\frac{1}{\sigma^2}\sum_{t=2}^{T}(y_t-\phi_1 y_{t-1})^2\\ &= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2)\\ &\quad -\frac{1}{2}\log\left(\frac{1}{1-\phi_1^2}\right)\\ &\quad -\frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2\\ &\quad -\frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t-\phi_1 y_{t-1})^2\end{aligned}$$

なぜなら，

$$\begin{aligned}L(y_1) &= \frac{1}{\sqrt{2\pi\sigma^2/(1-\phi_1^2)}}\\ &\quad\times\exp\left(-\frac{1}{2\sigma^2/(1-\phi_1^2)}y_1^2\right)\end{aligned}$$

$$\begin{aligned}&L(y_t|y_{t-1},\cdots,y_1)\\ &\quad = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y_t-\phi_1 y_{t-1})^2\right)\end{aligned}$$

である。

$$\begin{aligned}&\frac{\partial\log L(y_T,\cdots,y_1)}{\partial\sigma^2}\\ &= -\frac{T}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4/(1-\phi_1^2)}y_1^2\end{aligned}$$

$$+\frac{1}{2\sigma^4}\sum_{t=2}^{T}(y_t-\phi_1 y_{t-1})^2=0$$

$$\frac{\partial \log L(y_T,\cdots,y_1)}{\partial \phi_1}$$
$$=-\frac{\phi_1}{1-\phi_1^2}+\frac{\phi_1}{\sigma^2}y_1^2$$
$$+\frac{1}{\sigma^2}\sum_{t=2}^{T}(y_t-\phi_1 y_{t-1})y_{t-1}=0$$

連立方程式を満たす ϕ_1, σ^2 が最尤推定値であり，次のように表される。

$$\widetilde{\sigma}^2=\frac{1}{T}\left((1-\widetilde{\phi}_1^2)y_1^2+\sum_{t=2}^{T}(y_t-\widetilde{\phi}_1 y_{t-1})^2\right)$$
$$\widetilde{\phi}_1=\frac{\sum_{t=2}^{T}y_t y_{t-1}}{\sum_{t=2}^{T}y_{t-1}^2}$$
$$+\left(\widetilde{\phi}_1 y_1^2-\frac{\widetilde{\sigma}^2\widetilde{\phi}_1}{1-\widetilde{\phi}_1^2}\right)\Big/\sum_{t=2}^{T}y_{t-1}^2$$

ii. 最小自乗法 (OLS)

$$S(\phi_1)=\sum_{t=2}^{T}(y_t-\phi_1 y_{t-1})^2$$

を最小にするような ϕ_1 を推定値とする。

$$\widehat{\phi}_1$$
$$=\frac{\sum_{t=2}^{T}y_{t-1}y_t}{\sum_{t=2}^{T}y_{t-1}^2}$$
$$=\phi_1+\frac{\sum_{t=2}^{T}y_{t-1}\epsilon_t}{\sum_{t=2}^{T}y_{t-1}^2}$$
$$=\phi_1$$
$$+\frac{\sum_{t=2}^{T}(\epsilon_{t-1}+\phi_1\epsilon_{t-3}+\phi_1^2\epsilon_{t-3}+\ \cdots)\epsilon_t}{\sum_{t=2}^{T}(\epsilon_{t-1}+\phi_1\epsilon_{t-2}+\phi_1^2\epsilon_{t-3}+\ \cdots)^2}$$
$$\rightarrow \phi_1$$
$$+\frac{\mathrm{E}\big((\epsilon_{t-1}+\phi_1\epsilon_{t-2}+\phi_1^2\epsilon_{t-3}+\ \cdots)\epsilon_t\big)}{\mathrm{E}(\epsilon_{t-1}+\phi_1\epsilon_{t-2}+\phi_1^2\epsilon_{t-3}+\ \cdots)^2}$$
$$=\phi_1$$

OLSE は一致推定量となる。途中の計算には，以下を利用する。

$$y_{t-1}=\epsilon_{t-1}+\phi_1\epsilon_{t-2}+\phi_1^2\epsilon_{t-3}+\ \cdots$$
$$\mathrm{E}\big((\epsilon_{t-1}+\phi_1\epsilon_{t-2}+\phi_1^2\epsilon_{t-3}+\ \cdots)\epsilon_t\big)=0$$
$$\frac{1}{T}\sum_{t=2}^{T}y_{t-1}^2\ \longrightarrow$$
$$\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t+\phi_1\epsilon_{t-1}+\phi_1^2\epsilon_{t-2}+\ \cdots)^2$$
$$=\mathrm{E}(\epsilon_t+\phi_1\epsilon_{t-1}+\phi_1^2\epsilon_{t-2}+\ \cdots)^2$$
$$=\frac{\sigma_\epsilon^2}{1-\phi_1^2}$$

iii. $\widehat{\phi}_1$ の漸近分布について

$y_{t-1}\epsilon_t$, $t=1,2,\cdots,T$, は互いに独立に，平均ゼロ，分散 $\sigma_\epsilon^2/(1-\phi_1^2$ の分布を持つ。よって，中心極限定理から，

$$\frac{(1/T)\sum_{t=1}^{T}y_{t-1}\epsilon_t}{\sqrt{\sigma_\epsilon^4/(1-\phi_1^2)}/\sqrt{T}}\ \longrightarrow\ N(0,1)$$

変形して，

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{t-1}\epsilon_t\ \longrightarrow\ N(0,\frac{\sigma_\epsilon^4}{1-\phi_1^2})$$

次に，

$$\frac{1}{T}\sum_{t=1}^{T}y_{t-1}^2\ \longrightarrow\ \mathrm{E}(y_{t-1})=\frac{\sigma_\epsilon^2}{1-\phi_1^2}$$

となるので，

$$\sqrt{T}(\widehat{\phi}_1-\phi_1)$$
$$=\frac{(1/\sqrt{T})\sum_{t=1}^{T}y_{t-1}\epsilon_t}{(1/T)\sum_{t=1}^{T}y_{t-1}^2}$$
$$\longrightarrow\ N(0,1-\phi_1^2)$$

を得る。

iv. いくつかの定理

A. 中心極限定理

確率変数 x_1, x_2, $\cdots$, x_T はそれぞれ独立に平均 μ，分散 σ^2 の分布に従う。$\overline{x}=(1/T)\sum_{t=1}^{T}x_t$ とする。このとき，

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{T}}\ \longrightarrow\ N(0,1)$$

B. 確率変数 x, y を考える。y はある分布に分布収束し，x はある点に確率収束する。このとき，xy は分布収束する。

例えば，

$$y\ \longrightarrow\ N(\mu,\sigma^2),\qquad x\ \longrightarrow\ c$$

のとき，

$$xy\ \longrightarrow\ N(\mu,c^2\sigma^2)$$

となる。

(h) AR(1) +drift: $y_t=\mu+\phi_1 y_{t-1}+\epsilon_t$

平均

$$\phi(L)y_t=\mu+\epsilon_t$$

ただし， $\phi(L) = 1 - \phi_1 L$ である。

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\epsilon_t$$

よって，

$$\begin{aligned} \mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t) \\ &= \phi(1)^{-1}\mu \\ &= \frac{\mu}{1-\phi_1} \end{aligned}$$

5. AR(2) モデル $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$ を考える。

(a) 定常性の条件は， $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 = 0$ の 2 つの解 (の半径) の絶対値が共に 1 より大きいときである。

(b) $$(1 - \phi_1 L - \phi_2 L^2)y_t = \epsilon_t$$

$\phi(x) = 0$ の解を $1/\alpha_1, 1/\alpha_2$ としたとき，AR(2) モデルは

$$(1-\alpha_1 L)(1-\alpha_2 L)y_t = \epsilon_t$$

として書くことができる。

$$\begin{aligned} y_t &= \frac{1}{(1-\alpha_1 L)(1-\alpha_2 L)}\epsilon_t \\ &= \left(\frac{\alpha_1/(\alpha_1-\alpha_2)}{1-\alpha_1 L} + \frac{-\alpha_2/(\alpha_1-\alpha_2)}{1-\alpha_2 L}\right)\epsilon_t \end{aligned}$$

(c) AR(2) モデルの平均

$$\begin{aligned} \mu &= \mathrm{E}(y_t) \\ &= \mathrm{E}(\phi(L)\epsilon_t) \\ &= 0 \end{aligned}$$

(d) AR(2) モデルの自己相関関数

$$\begin{aligned} &\gamma(\tau) \\ &= \mathrm{E}(y_t - \mu)(y_{t-\tau} - \mu) \\ &= \mathrm{E}(y_t y_{t-\tau}) \\ &= \mathrm{E}\Big((\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t)y_{t-\tau}\Big) \\ &= \phi_1 \mathrm{E}(y_{t-1}y_{t-\tau}) + \phi_2 \mathrm{E}(y_{t-2}y_{t-\tau}) \\ &+ \mathrm{E}(\epsilon_t y_{t-\tau}) \\ &= \begin{cases} \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2), & (\tau \neq 0 \text{ のとき}) \\ \phi_1\gamma(\tau-1) + \phi_2\gamma(\tau-2) + \sigma^2, & \\ & (\tau = 0 \text{ のとき}) \end{cases} \end{aligned}$$

初期条件は

$$\begin{aligned} \gamma(0) &= \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2 \\ \gamma(1) &= \phi_1\gamma(0) + \phi_2\gamma(1) \\ \gamma(2) &= \phi_1\gamma(1) + \phi_2\gamma(0) \end{aligned}$$

を解くことになり，

$$\gamma(0) = \left(\frac{1-\phi_2}{1+\phi_2}\right)^2 \frac{\sigma^2}{(1-\phi_2)^2 - \phi_1^2}$$

となる。

(e) $\tau = 0$ のとき，$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2$ から，

$$\gamma(0) = \frac{\sigma^2}{1 - \phi_1\rho(1) - \phi_2\rho(2)}$$

となる。

(f) 尤度関数

$$\begin{aligned} &\log L(y_T, \cdots, y_1) \\ &= \log L(y_2, y_1) \\ &+ \sum_{t=3}^{T} \log L(y_t | y_{t-1}, \cdots, y_1) \end{aligned}$$

ただし，

$$\begin{aligned} &L(y_2, y_1) \\ &= \frac{1}{\sqrt{2\pi}}|V|^{-1/2} \exp\left(-\frac{1}{2}(y_1\ y_2)V^{-1}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) \end{aligned}$$

$$\begin{aligned} &L(y_t | y_{t-1}, \cdots, y_1) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2\right) \end{aligned}$$

である。

$$V = \gamma(0)\begin{pmatrix} 1 & \phi_1/(1-\phi_2) \\ \phi_1/(1-\phi_2) & 1 \end{pmatrix}$$

(g) AR(2) +drift: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$
平均

$$\phi(L)y_t = \mu + \epsilon_t$$

ただし， $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$ である。

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\epsilon_t$$

よって，

$$\begin{aligned} \mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t) \\ &= \phi(1)^{-1}\mu \\ &= \frac{\mu}{1-\phi_1-\phi_2} \end{aligned}$$

6. AR(p) モデル $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$ について，

(a) AR(p) モデルの分散

$$\gamma(0) = \frac{\sigma^2}{1 - \phi_1\rho(1) - \ \cdots \ - \phi_p\rho(p)}$$

(b) AR(p) モデルの推定:

最小自乗法:

$$\min_{\phi_1,\cdots,\phi_p} \ (y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p})^2$$

最尤法:

$$\begin{aligned}&\log L(y_T,\cdots,y_1)\\ &= \log L(y_p,\cdots,y_2,y_1)\\ &+ \sum_{t=p+1}^{T} \log L(y_t|y_{t-1},\cdots,y_1)\end{aligned}$$

ただし，

$$\begin{aligned}&L(y_p,\cdots,y_2,y_1)\\ &= \frac{1}{\sqrt{2\pi}}|V|^{-1/2}\\ &\times \exp\left(-\frac{1}{2}(y_1\ y_2\ \cdots\ y_p)V^{-1}\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_p\end{pmatrix}\right)\end{aligned}$$

$$\begin{aligned}&L(y_t|y_{t-1},\cdots,y_1)\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{1}{2\sigma^2}(y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2}\\ &\qquad\qquad - \ \cdots \ - \phi_p y_{t-p} \quad)^2\Big)\end{aligned}$$

とする。

ユール・ウォーカー (Yule=Walker) 方程式:
$\phi_1 y_{t-1} + \phi_2 y_{t-2} + \ \cdots \ + \phi_p y_{t-p} + \epsilon_t = y_t$ の両辺に $y_{t-1}, y_{t-2}, \cdots, y_{t-p}$ をそれぞれ掛けて期待値をとり，標本分散 $\widehat{\gamma}(0)$ で割る。

$$\begin{pmatrix} 1 & \widehat{\rho}(1) & \cdots & \widehat{\rho}(p-2) & \widehat{\rho}(p-1)\\ \widehat{\rho}(1) & 1 & & \widehat{\rho}(p-3) & \widehat{\rho}(p-2)\\ \vdots & \vdots & & \vdots & \vdots\\ \widehat{\rho}(p-1) & \widehat{\rho}(p-2) & \cdots & \widehat{\rho}(1) & 1\end{pmatrix}$$
$$\times \begin{pmatrix}\phi_1\\ \phi_2\\ \vdots\\ \phi_{p-1}\\ \phi_p\end{pmatrix} = \begin{pmatrix}\widehat{\rho}(1)\\ \widehat{\rho}(2)\\ \vdots\\ \widehat{\rho}(p)\end{pmatrix}$$

ただし，

$$\widehat{\gamma}(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T}(y_t - \widehat{\mu})(y_{t-\tau} - \widehat{\mu})$$
$$\widehat{\rho}(\tau) = \frac{\widehat{\gamma}(\tau)}{\widehat{\gamma}(0)}$$

とする。

(c) AR(p) +drift: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \ \phi_p y_{t-p} + \epsilon_t$

平均

$$\phi(L)y_t = \mu + \epsilon_t$$

ただし， $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ \cdots \ - \phi_p L^p$ である。

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\epsilon_t$$

よって，

$$\begin{aligned}\mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t)\\ &= \phi(1)^{-1}\mu\\ &= \frac{\mu}{1 - \phi_1 - \phi_2 - \ \cdots \ - \phi_p}\end{aligned}$$

(d) AR(p) モデルの偏自己相関関数： $k = p + 1, p+2, \cdots$ について， $\phi_{k,k} = 0$ 。

AR(1), AR(2) モデルで確認すること。

17.3 MA モデル

MA モデル (Moveing Average Model)

1. MA(q)

$$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ \cdots \ + \theta_q\epsilon_{t-q}$$

次のように書き直される。

$$y_t = \theta(L)\epsilon_t$$

ただし，

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ \cdots \ + \theta_q L^q$$

とする。

2. 反転可能性：$\theta(x)=1+\theta_1 x+\theta_2 x^2+\cdots+\theta_q x^q=0$ の q 個の解 (の半径) がすべて 1 より大きいとき

AR(∞) モデルに書き直すことができるという条件を意味する。

3. MA(1) モデル $y_t=\epsilon_t+\theta_1\epsilon_{t-1}$

(a) MA(1) モデルの平均：

$$\begin{aligned}\mathrm{E}(y_t)&=\mathrm{E}(\epsilon_t+\theta_1\epsilon_{t-1})\\&=\mathrm{E}(\epsilon_t)+\theta_1\mathrm{E}(\epsilon_{t-1})\\&=0\end{aligned}$$

(b) MA(1) モデルの自己相関関数：

$$\begin{aligned}\gamma(0)&=\mathrm{E}(y_t^2)\\&=\mathrm{E}(\epsilon_t+\theta_1\epsilon_{t-1})^2\\&=\mathrm{E}(\epsilon_t^2+2\theta_1\epsilon_t\epsilon_{t-1}+\theta_1^2\epsilon_{t-1}^2)\\&=\mathrm{E}(\epsilon_t^2)+2\theta_1\mathrm{E}(\epsilon_t\epsilon_{t-1})+\theta_1^2\mathrm{E}(\epsilon_{t-1}^2)\\&=(1+\theta_1^2)\sigma^2\end{aligned}$$

$$\begin{aligned}\gamma(1)&=\mathrm{E}(y_ty_{t-1})\\&=\mathrm{E}(\epsilon_t+\theta_1\epsilon_{t-1})(\epsilon_{t-1}+\theta_1\epsilon_{t-2})\\&=\theta_1\sigma^2\end{aligned}$$

$$\begin{aligned}\gamma(2)&=\mathrm{E}(y_ty_{t-2})\\&=\mathrm{E}(\epsilon_t+\theta_1\epsilon_{t-1})(\epsilon_{t-2}+\theta_1\epsilon_{t-3})\\&=0\end{aligned}$$

まとめて，

$$\rho(\tau)=\begin{cases}\dfrac{\theta_1}{1+\theta_1^2},&(\tau=1\ \text{のとき})\\0,&(\tau=2,3,\cdots\ \text{のとき})\end{cases}$$

となる。

MA(1) モデルの場合，

$$-\frac{1}{2}\le\rho(1)\le\frac{1}{2}$$

となる。

(c) 反転可能性の条件：

$$\begin{aligned}\epsilon_t&=-\theta_1\epsilon_{t-1}+y_t\\&=(-\theta_1)^2\epsilon_{t-2}+y_t+(-\theta_1)y_{t-1}\\&=(-\theta_1)^3\epsilon_{t-3}+y_t+(-\theta_1)y_{t-1}+(-\theta_1)^2y_{t-2}\\&\ \ \vdots\\&=(-\theta_1)^s\epsilon_{t-s}+y_t+(-\theta_1)y_{t-1}+(-\theta_1)^2y_{t-2}\\&\quad+\cdots+(-\theta_1)^{t-s+1}y_{t-s+1}\end{aligned}$$

$(-\theta_1)^s\epsilon_{t-s}\longrightarrow 0$ のとき，MA(1) モデルは次のように AR(∞) で書き直される。

$$\begin{aligned}y_t=&-(-\theta_1)y_{t-1}-(-\theta_1)^2y_{t-2}-\cdots\\&-(-\theta_1)^{t-s+1}y_{t-s+1}-\cdots+\epsilon_t\end{aligned}$$

(d) 尤度関数：

自己共分散関数は $\gamma(0)=(1+\theta_1^2)\sigma^2$，$\gamma(1)=\theta_1\sigma^2$，その他の τ について，$\gamma(\tau)=0$ となるので，$y_1,y_2,\cdots,y_T$ の同時分布は

$$\begin{aligned}&f(y_1,y_2,\cdots,y_T)\\&=\frac{1}{(2\pi)^{T/2}}|V|^{-1/2}\exp\left(-\frac{1}{2}Y'V^{-1}Y\right)\end{aligned}$$

ただし，

$$Y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_T\end{pmatrix},$$

$$V=\sigma^2\begin{pmatrix}1+\theta_1^2&\theta_1&0&\cdots&0\\\theta_1&1+\theta_1^2&\theta_1&\ddots&\vdots\\0&\theta_1&\ddots&\ddots&0\\\vdots&\ddots&\ddots&1+\theta_1^2&\theta_1\\0&\cdots&0&\theta_1&1+\theta_1^2\end{pmatrix}$$

となる。

4. MA(1) +drift: $y_t=\mu+\epsilon_t+\theta_1\epsilon_{t-1}$

平均

$$y_t=\mu+\theta(L)\epsilon_t$$

ただし，$\theta(L)=1+\theta_1L$ である。よって，

$$\begin{aligned}\mathrm{E}(y_t)&=\mu+\theta(L)\mathrm{E}(\epsilon_t)\\&=\mu\end{aligned}$$

5. MA(2) モデル $y_t=\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}$

(a) MA(2) モデルの自己共分散関数：

$$\gamma(\tau)=\begin{cases}(1+\theta_1^2+\theta_2^2)\sigma^2 & (\tau=0)\\ (\theta_1+\theta_1\theta_2)\sigma^2 & (\tau=1)\\ \theta_2\sigma^2 & (\tau=2)\\ 0 & (\text{その他})\end{cases}$$

(b) $\theta(x)=0$ の 2 つの解を $-1/\beta_1$, $-1/\beta_2$ とする。定常であるためには，β_1, β_2 共に絶対値で 1 より大きいことが必要である。このとき，MA(2) モデルは次のように表される。

$$\begin{aligned}y_t &= \epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}\\ &=(1+\theta_1L+\theta_2L^2)\epsilon_t\\ &=(1+\beta_1L)(1+\beta_2L)\epsilon_t\end{aligned}$$

MA(2) モデルの AR(∞) 表現は

$$\begin{aligned}\epsilon_t &= \frac{1}{(1+\beta_1L)(1+\beta_2L)}y_t\\ &=\left(\frac{\beta_1/(\beta_1-\beta_2)}{1+\beta_1L}-\frac{\beta_2/(\beta_1-\beta_2)}{1+\beta_2L}\right)y_t\end{aligned}$$

(c) 尤度関数：

$$V=\sigma^2\begin{pmatrix}1+\theta_1^2+\theta_2^2 & \theta_1+\theta_1\theta_2 & \theta_2 & & 0\\ \theta_1+\theta_1\theta_2 & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_1\theta_2 & \ddots & \\ \theta_2 & \theta_1+\theta_1\theta_2 & \ddots & \ddots & \theta_2\\ & \ddots & \ddots & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_1\theta_2\\ 0 & & \theta_2 & \theta_1+\theta_1\theta_2 & 1+\theta_1^2+\theta_2^2\end{pmatrix}$$

(d) MA(2) +drift: $y_t=\mu+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}$
平均

$$y_t=\mu+\theta(L)\epsilon_t$$

ただし，$\theta(L)=1+\theta_1L+\theta_2L^2$ である。よって，

$$\begin{aligned}\mathrm{E}(y_t)&=\mu+\theta(L)\mathrm{E}(\epsilon_t)\\ &=\mu\end{aligned}$$

6. MA(q) モデル $y_t=\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ \cdots\ +\theta_q\epsilon_{t-q}$

(a) MA(q) モデルの平均：

$$\begin{aligned}&\mathrm{E}(y_t)\\ &=\mathrm{E}(\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ \cdots\ +\theta_q\epsilon_{t-q})\\ &=0\end{aligned}$$

(b) MA(q) モデルの自己共分散：

$$\gamma(\tau)=\begin{cases}\sigma^2(\theta_0\theta_\tau+\theta_1\theta_{\tau+1}+\ \cdots\ +\theta_{q-\tau}\theta_q),\\ \qquad \tau=1,2,\cdots,q,\\ 0,\ \tau=q+1,q+2,\cdots,\end{cases}$$

すなわち，

$$\gamma(\tau)=\begin{cases}\sigma^2\sum_{i=0}^{q-\tau}\theta_i\theta_{\tau+i}, & \tau=1,2,\cdots,q\\ 0, & \tau=q+1,q+2,\cdots\end{cases}$$

ただし，$\theta_0=1$ とする。

(c) MA(q) モデルは定常。

(d) MA(q) +drift: $y_t=\mu+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ \cdots\ +\theta_q\epsilon_{t-q}$
平均

$$y_t=\mu+\theta(L)\epsilon_t$$

ただし，$\theta(L)=1+\theta_1L+\theta_2L^2+\ \cdots\ +\theta_qL^q$ である。よって，

$$\begin{aligned}\mathrm{E}(y_t)&=\mu+\theta(L)\mathrm{E}(\epsilon_t)\\ &=\mu\end{aligned}$$

を得る。

17.4 ARMA モデル

ARMA モデル: Autoregressive Moving Average Model

1. ARMA(p,q)

$$\begin{aligned}y_t&=\phi_1y_{t-1}+\phi_2y_{t-2}+\ \cdots\ +\phi_py_{t-p}\\ &\quad+\epsilon_t+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+\ \cdots\ +\theta_q\epsilon_{t-q}\end{aligned}$$

次のように書き直される。

$$\phi(L)y_t=\theta(L)\epsilon_t$$

2. 尤度関数：
Y の分散共分散行列 V を求める。

3. ARMA(1,1) モデル: $y_t=\phi_1y_{t-1}+\epsilon_t+\theta_1\epsilon_{t-1}$
自己相関係数を求める。

y_t の期待値は，両辺に期待値をとって，

$$E(y_t) = \phi_1 E(y_{t-1}) + E(\epsilon_t) + \theta_1 E(\epsilon_{t-1})$$

となり，右辺第 2, 3 項はゼロとなる。よって，

$$E(y_t) = 0$$

を得る。

自己共分散を求める。両辺に $y_{t-\tau}$ を掛けて，期待値をとる。

$$\begin{aligned} E(y_t y_{t-\tau}) &= \phi_1 E(y_{t-1} y_{t-\tau}) \\ &\quad + E(\epsilon_t y_{t-\tau}) + \theta_1 E(\epsilon_{t-1} y_{t-\tau}) \end{aligned}$$

各項は

$$\begin{aligned} &E(y_t y_{t-\tau}) = \gamma(\tau) \\ &E(y_{t-1} y_{t-\tau}) = \gamma(\tau-1) \\ &E(\epsilon_t y_{t-\tau}) = \begin{cases} \sigma^2, \tau = 0 \\ 0, \quad \tau = 1, 2, \cdots \end{cases} \\ &E(\epsilon_{t-1} y_{t-\tau}) = \begin{cases} (\phi_1 + \theta_1)\sigma^2, \tau = 0 \\ \sigma^2, & \tau = 1 \\ 0, & \tau = 2, 3, \cdots \end{cases} \end{aligned}$$

となる。よって，

$$\begin{aligned} &\gamma(0) = \phi_1 \gamma(1) + (1 + \phi_1\theta_1 + \theta_1^2)\sigma^2 \\ &\gamma(1) = \phi_1 \gamma(0) + \theta_1 \sigma^2 \\ &\gamma(\tau) = \phi_1 \gamma(\tau - 1), \qquad \tau = 2, 3, \cdots \end{aligned}$$

となる。上の 2 つの式から，$\gamma(0), \gamma(1)$ を計算すると，

$$\begin{aligned} &\begin{pmatrix} 1 & -\phi_1 \\ -\phi_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} 1 + \phi_1\theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} &\begin{pmatrix} \gamma(0) \\ \gamma(1) \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} 1 & -\phi_1 \\ -\phi_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 + \phi_1\theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix} \\ &= \frac{\sigma^2}{1 - \phi_1^2} \begin{pmatrix} 1 & \phi_1 \\ \phi_1 & 1 \end{pmatrix} \begin{pmatrix} 1 + \phi_1\theta_1 + \theta_1^2 \\ \theta_1 \end{pmatrix} \\ &= \frac{\sigma^2}{1 - \phi_1^2} \begin{pmatrix} 1 + 2\phi_1\theta_1 + \theta_1^2 \\ (1 + \phi_1\theta_1)(\phi_1 + \theta_1) \end{pmatrix} \end{aligned}$$

したがって，自己相関係数の初期条件は

$$\rho(1) = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 + 2\phi_1\theta_1 + \theta_1^2}$$

であり，また，

$$\rho(\tau) = \phi_1 \rho(\tau - 1)$$

が得られる。

17.5 ARIMA モデル

ARIMA モデル: Autoregressive Integrated Moving Average Model

1. ARIMA(p, d, q)

$$\phi(L)\Delta^d y_t = \theta(L)\epsilon_t$$

ただし，

$$\begin{aligned} \Delta^d y_t &= \Delta^{d-1}(1 - L) y_t \\ &= \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1} \\ &= (1 - L)^d y_t, \qquad d = 1, 2, \cdots \end{aligned}$$

$$\Delta^0 y_t = y_t$$

2. ARMA(p, q) +drift: $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$

平均

$$\phi(L) y_t = \mu + \theta(L)\epsilon_t$$

ただし， $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ である。

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\theta(L)\epsilon_t$$

よって，

$$\begin{aligned} E(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\theta(L) E(\epsilon_t) \\ &= \phi(1)^{-1}\mu \\ &= \frac{\mu}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \end{aligned}$$

17.6 SARIMA モデル

SARIMA モデル: Seasonal ARIMA Model

1. SARIMA(p, d, q)

$$\phi(L)\Delta^d\Delta_s y_t = \theta(L)\epsilon_t$$

ただし，

$$\begin{aligned}\Delta_s y_t &= (1-L^s)y_t \\ &= y_t - y_{t-s}\end{aligned}$$

四半期データのとき $s=4$，月次データのとき $s=12$ となる。

17.7 最適予測

1. AR(p) モデル：$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t$

   (a) 次を定義する。

   $$\mathrm{E}(y_{t+k}|I_t) = y_{t+k|t}$$

   $y_{t+k} = \phi_1 y_{t+k-1} + \cdots + \phi_p y_{t+k-p} + \epsilon_{t+k}$ の両辺に期待値をとって，

   $$y_{t+k|t} = \phi_1 y_{t+k-1|t} + \cdots + \phi_p y_{t+k-p|t}$$

   ただし， $s \le t$ のとき， $y_{s|t} = y_s$ となる。

   (b) 最適予測量は上の定差方程式となる。

2. MA(q) モデル：$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$

   (a) 推定された攪乱項を $\widehat{\epsilon}_T, \widehat{\epsilon}_{T-1}, \cdots, \widehat{\epsilon}_1$ とする。

   (b) $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$ により，$y_{t+k} = \epsilon_{t+k} + \theta_1\epsilon_{t+k-1} + \cdots + \theta_q\epsilon_{t+k-q}$ となる。

   (c) したがって，

   $$\begin{aligned}y_{t+k|t} &= \epsilon_{t+k|t} + \theta_1\epsilon_{t+k-1|t} \\ &\quad + \cdots + \theta_q\epsilon_{t+k-q|t}\end{aligned}$$

   ただし，$s > t$ のとき， $\epsilon_{s|t} = 0$ となる。また，$s \le t$ のとき， $\epsilon_{s|t} = \widehat{\epsilon}_s$ である。

3. ARMA(p, q) モデル：$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$

   (a) $y_{t+k} = \phi_1 y_{t+k-1} + \cdots + \phi_p y_{t+k-p} + \epsilon_{t+k} + \theta_1\epsilon_{t+k-1} + \cdots + \theta_q\epsilon_{t+k-q}$

   (b) 最適予測量は

   $$\begin{aligned}y_{t+k|t} &= \phi_1 y_{t+k-1|t} + \cdots + \phi_p y_{t+k-p|t} \\ &\quad + \epsilon_{t+k|t} + \theta_1\epsilon_{t+k-1|t} + \cdots + \theta_q\epsilon_{t+k-q|t}\end{aligned}$$

   となる。ただし， $s \le t$ のとき， $y_{s|t} = y_s$，$\epsilon_{s|t} = \widehat{\epsilon}_s$ であり，$s > t$ のとき， $\epsilon_{s|t} = 0$ となる。

17.8 識別 (同定, Identification) ・推定問題

1. d, s を与えて，AIC (または，SBIC) 基準で p, q を求める方法：

   (a) AIC (Akaike's Information Criterion)

   $$\mathrm{AIC} = \log s^2 + \frac{2(p+q)}{T}$$

   (b) SBIC (Shwarz's Bayesian Information Criterion)

   $$\mathrm{SBIC} = \log s^2 + \frac{(p+q)\log T}{T}$$

2. 標本自己相関関数 $\widehat{\rho}(\tau)$，$\tau = 1, 2, \cdots$，のグラフから p, d, q, s を決める方法：

	AR(p) モデル	MA(q) モデル
自己相関関数	減衰	$\rho(k) = 0$, $k = q+1, q+2, \cdots$
偏自己相関関数	$\phi(k, k) = 0$, $k = p+1, p+2, \cdots$	減衰

   (a) 季節性を除くために， $\Delta_s y_t$ を計算する。$\Delta_s y_t$ の自己相関関数を計算する。自己相関関数が s 期ごとに周期を持てば，さらに， $(1-L^s)$ をとる。

   (b) 階差の次数を決める。その都度，自己相関関数を計算する。自己相関関数が， τ が大きくなるにつれて，減衰すれば次のステップへ。

   (c) AR 項の次数を決める。その都度，自己相関関数を計算する。ある τ 以降，自己相関関数がゼロに近くなっていれば，次のステップへ。

   (d) MA 項の次数を決める。その都度，自己相関関数を計算する。自己相関関数がゼロの回りでランダムに散らばっていればプロセスを終了する。

ARIMA モデルの推定例

家計消費支出 (原系列，1985 年価格)

```
IDENT RC
Date: 4-30-1995 / Time: 17:03
SMPL range: 1955.1 - 1994.1
Number of observations: 157
==========================================================================
       Autocorrelations        Partial Autocorrelations        ac     pac
==========================================================================
|           . |*************|           . |*************|  1  0.970  0.970
|           . |************ |           . |*.           |  2  0.947  0.096
|           . |************ |           . |**           |  3  0.934  0.168
|           . |************ |           . |***          |  4  0.932  0.244
|           . |************ |      ******| .            |  5  0.902 -0.427
|           . |***********  |           . |*.           |  6  0.879  0.094
|           . |***********  |           . |*.           |  7  0.864  0.085
|           . |***********  |           . |*.           |  8  0.861  0.073
|           . |***********  |        ****| .            |  9  0.830 -0.280
|           . |**********   |           . |*.           | 10  0.805  0.084
|           . |**********   |           . |*.           | 11  0.790  0.062
|           . |**********   |           . | .           | 12  0.786  0.022
|           . |**********   |          **| .            | 13  0.755 -0.181
|           . |*********    |           . |*.           | 14  0.730  0.072
|           . |*********    |           . | .           | 15  0.714  0.026
|           . |*********    |           . | .           | 16  0.708 -0.018
|           . |*********    |          **| .            | 17  0.677 -0.127
|           . |********     |           . |*.           | 18  0.652  0.068
|           . |********     |           . | .           | 19  0.636  0.038
|           . |********     |           . | .           | 20  0.631 -0.019
==========================================================================
 Box-Pierce Q-Stat 2046.36    Prob   0.0000      SE of Correlations  0.080
 Ljung-Box  Q-Stat 2201.64    Prob   0.0000
==========================================================================
```

```
DC=RC-RC (-1)
IDENT DC
Date: 4-30-1995 / Time: 17:03
SMPL range: 1955.2 - 1994.1
```

```
Number of observations: 156
==========================================================================
       Autocorrelations        Partial Autocorrelations        ac      pac
==========================================================================
|        *****| .            |        *****| .            |  1 -0.367 -0.367
|          ***| .            |       ******| .            |  2 -0.250 -0.444
|        *****| .            | ************| .            |  3 -0.351 -0.931
|           . |************  |           . |*******       |  4  0.957  0.543
|        *****| .            |           . |*.            |  5 -0.358  0.079
|          ***| .            |           . |*.            |  6 -0.239  0.097
|         ****| .            |           **| .            |  7 -0.346 -0.118
|           . |************  |           . |**            |  8  0.934  0.134
|        *****| .            |           . | .            |  9 -0.352  0.027
|          ***| .            |           . | .            | 10 -0.229  0.035
|         ****| .            |           . | .            | 11 -0.336  0.024
|           . |************  |           .*| .            | 12  0.904 -0.053
|         ****| .            |           . | .            | 13 -0.344 -0.019
|          ***| .            |           . | .            | 14 -0.219 -0.025
|         ****| .            |           . |**            | 15 -0.320  0.120
|           . |***********   |           . |*.            | 16  0.874  0.060
|         ****| .            |           . | .            | 17 -0.342 -0.011
|          ***| .            |           . | .            | 18 -0.209 -0.027
|         ****| .            |           .*| .            | 19 -0.309 -0.066
|           . |***********   |           **| .            | 20  0.840 -0.152
==========================================================================
 Box-Pierce Q-Stat  860.27    Prob   0.0000      SE of Correlations  0.080
 Ljung-Box  Q-Stat  938.60    Prob   0.0000
==========================================================================


SDC=DC-DC (-4)
IDENT SDC
Date: 4-30-1995 / Time: 17:04
SMPL range: 1956.2 - 1994.1
Number of observations: 152
==========================================================================
       Autocorrelations        Partial Autocorrelations        ac      pac
==========================================================================
|          ***| .            |          ***| .            |  1 -0.253 -0.253
|           . | .            |           .*| .            |  2 -0.024 -0.094
|           . |****          |           . |***           |  3  0.273  0.261
|      *******| .            |       ******| .            |  4 -0.504 -0.428
```

```
|          . |*.          |          **| .          |  5  0.051 -0.164
|          . |*.          |          . | .          |  6  0.069 -0.023
|          **| .          |          . |*.          |  7 -0.145  0.088
|          . |*.          |          **| .          |  8  0.092 -0.168
|          .*| .          |          **| .          |  9 -0.043 -0.177
|          .*| .          |          .*| .          | 10 -0.077 -0.100
|          . | .          |          .*| .          | 11 -0.017 -0.087
|          . | .          |          . | .          | 12  0.027 -0.033
|          . |*.          |          . |*.          | 13  0.106  0.077
|          . | .          |          .*| .          | 14 -0.023 -0.064
|          . |*.          |          . | .          | 15  0.078 -0.034
|          . | .          |          . | .          | 16  0.007 -0.009
|          .*| .          |          . |*.          | 17 -0.078  0.051
|          . | .          |          .*| .          | 18 -0.003 -0.096
|          .*| .          |          **| .          | 19 -0.077 -0.136
|          . | .          |          .*| .          | 20 -0.014 -0.046
==========================================================================
 Box-Pierce Q-Stat   71.21    Prob   0.0000      SE of Correlations  0.081
 Ljung-Box  Q-Stat   74.39    Prob   0.0000
==========================================================================
```

```
 LS // Dependent Variable is SDC
 Date: 4-30-1995 / Time: 17:04
 SMPL range: 1956.2 - 1994.1
 Number of observations: 152
 Convergence achieved after 7 iterations
 ========================================================================
       VARIABLE      COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
 ========================================================================
        MA(4)        -0.5167656     0.0701419    -7.3674287     0.0000
 ========================================================================
 R-squared                0.268649    Mean of dependent var     7.066744
 Adjusted R-squared       0.268649    S.D. of dependent var     571.0115
 S.E. of regression       488.3237    Sum of squared resid      36007470
 Log likelihood          -1156.206    Durbin-Watson stat        2.464279
 ========================================================================
```

```
IDENT RESID
Date: 4-30-1995 / Time: 17:04
SMPL range: 1956.2 - 1994.1
```

```
Number of observations: 152
==========================================================================
      Autocorrelations       Partial Autocorrelations        ac     pac
==========================================================================
|          ***| .          |           ***| .          |  1 -0.236 -0.236
|            . | .         |            .*| .          |  2 -0.033 -0.094
|            . |**         |            . |**          |  3  0.150  0.127
|           .*| .          |            . | .          |  4 -0.099 -0.037
|           .*| .          |            .*| .          |  5 -0.063 -0.090
|            . | .         |            .*| .          |  6  0.014 -0.051
|           **| .          |            **| .          |  7 -0.117 -0.123
|            . |*.         |            . |*.          |  8  0.103  0.068
|           .*| .          |            . | .          |  9 -0.047 -0.024
|           .*| .          |            .*| .          | 10 -0.105 -0.109
|           .*| .          |            **| .          | 11 -0.052 -0.165
|            . |*.         |            . |*.          | 12  0.107  0.058
|            . |*.         |            . |**          | 13  0.047  0.133
|           .*| .          |            .*| .          | 14 -0.083 -0.047
|            . | .         |            .*| .          | 15  0.028 -0.063
|            . |*.         |            . |*.          | 16  0.095  0.040
|           .*| .          |            . | .          | 17 -0.103 -0.037
|            . | .         |            .*| .          | 18 -0.031 -0.056
|           .*| .          |            **| .          | 19 -0.091 -0.151
|            . |*.         |            . | .          | 20  0.092  0.038
==========================================================================
 Box-Pierce Q-Stat   29.16    Prob   0.0847      SE of Correlations  0.081
 Ljung-Box  Q-Stat   31.17    Prob   0.0530
==========================================================================
```

```
 LS // Dependent Variable is SDC
 Date: 4-30-1995 / Time: 17:04
 SMPL range: 1956.2 - 1994.1
 Number of observations: 152
 Convergence achieved after 7 iterations
=========================================================================
      VARIABLE      COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
=========================================================================
       MA(1)        -0.3092160     0.0608381    -5.0826070     0.0000
       MA(4)        -0.5727303     0.0605915    -9.4523260     0.0000
=========================================================================
 R-squared                0.339172    Mean of dependent var     7.066744
```

```
 Adjusted R-squared        0.334767    S.D. of dependent var     571.0115
 S.E. of regression        465.7274    Sum of squared resid      32535304
 Log likelihood           -1148.499    F-statistic               76.98805
 Durbin-Watson stat        1.937266    Prob (F-statistic)          0.000000
 ========================================================================


IDENT RESID
Date: 4-30-1995 / Time: 17:05
SMPL range: 1956.2 - 1994.1
Number of observations: 152
==========================================================================
       Autocorrelations        Partial Autocorrelations        ac     pac
==========================================================================
|           . | .          |           . | .          |  1  0.022  0.022
|           . | .          |           . | .          |  2  0.006  0.006
|           . |**          |           . |**          |  3  0.133  0.133
|           .*| .          |           .*| .          |  4 -0.061 -0.068
|           . | .          |           . | .          |  5  0.028  0.031
|           . | .          |           . | .          |  6 -0.001 -0.021
|           .*| .          |           .*| .          |  7 -0.081 -0.065
|           . |*.          |           . |*.          |  8  0.088  0.083
|           . | .          |           . | .          |  9 -0.021 -0.021
|           **| .          |           .*| .          | 10 -0.121 -0.106
|           .*| .          |           .*| .          | 11 -0.046 -0.072
|           . |*.          |           . |**          | 12  0.092  0.122
|           . |*.          |           . |*.          | 13  0.077  0.098
|           .*| .          |           .*| .          | 14 -0.069 -0.084
|           . | .          |           . | .          | 15  0.017 -0.000
|           . |*.          |           . |*.          | 16  0.055  0.041
|           .*| .          |           .*| .          | 17 -0.087 -0.083
|           .*| .          |           .*| .          | 18 -0.073 -0.081
|           .*| .          |           .*| .          | 19 -0.090 -0.072
|           . |*.          |           . |*.          | 20  0.050  0.072
==========================================================================
 Box-Pierce Q-Stat   15.23    Prob   0.7628      SE of Correlations  0.081
 Ljung-Box  Q-Stat   16.64    Prob   0.6764
==========================================================================
```

17.9 周波数領域

周波数領域: Frequency Domain

1. スペクトラム (パワー・スペクトラム) の定義：

$$f(\lambda) = (2\pi)^{-1}\sum_{\tau=-\infty}^{\infty}\gamma(\tau)\cos\lambda\tau$$
$$= (2\pi)^{-1}\sum_{\tau=-\infty}^{\infty}\gamma(\tau)\exp(-i\lambda\tau)$$

2. y_t がホワイト・ノイズであれば， $f(\lambda) = (2\pi)^{-1}\sigma^2$ 。

3. スペクトラムと自己相関関数との関係

$$\gamma(\tau) = \int_{-\pi}^{\pi} f(\lambda)\cos\lambda\tau \mathrm{d}\lambda$$

従って，スペクトラムは自己相関関数のすべての情報を持っている。

4. $\sum w_j^2 < \infty$ とする。

$$y_t = \sum_{j=-r}^{s} w_j x_{t-j}$$

$f_x(\lambda)$ を x_t のスペクトラムとする。 $W(\lambda)$ を次のように定義する。

$$W(\lambda) = \sum_{j=-r}^{s} w_j e^{-i\lambda j}$$

このとき， y_t のスペクトラムは以下のようになる。

$$f_y(\lambda) = |W(\lambda)|^2 f_x(\lambda)$$

$|W(\lambda)|^2$ は伝達関数 (transfer function) と呼ばれ，

$$|W(\lambda)|^2 = W(\lambda)\overline{W(\lambda)}$$
$$= \sum_{j=-r}^{s} w_j e^{-i\lambda j}\sum_{j=-r}^{s} w_j e^{i\lambda j}$$

$\overline{W(\lambda)}$ は $W(\lambda)$ の共役複素数とする。

5. MA(q) モデルの場合：

$$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$$
$$= (1 + \theta_1 L + \cdots + \theta_q L^q)\epsilon_t$$
$$= \theta(L)\epsilon_t$$

$$f_y(\lambda) = \theta(e^{-i\lambda})\theta(e^{i\lambda})f_\epsilon(\lambda)$$
$$= \theta(e^{-i\lambda})\theta(e^{i\lambda})\frac{\sigma^2}{2\pi}$$

6. AR(p) モデルの場合:

$$\phi(L)y_t = \epsilon_t$$

$$y_t = \phi(L)^{-1}\epsilon_t$$

$$f_y(\lambda) = \frac{1}{\phi(e^{-i\lambda})\phi(e^{i\lambda})}f_\epsilon(\lambda)$$
$$= \frac{1}{\phi(e^{-i\lambda})\phi(e^{i\lambda})}\frac{\sigma^2}{2\pi}$$

7. ARMA(p,q) モデルの場合:

$$\phi(L)y_t = \theta(L)\epsilon_t$$

$$y_t = \phi(L)^{-1}\theta(L)\epsilon_t$$

$$f_y(\lambda) = \frac{\theta(e^{-i\lambda})\theta(e^{i\lambda})}{\phi(e^{-i\lambda})\phi(e^{i\lambda})}f_\epsilon(\lambda)$$
$$= \frac{\theta(e^{-i\lambda})\theta(e^{i\lambda})}{\phi(e^{-i\lambda})\phi(e^{i\lambda})}\frac{\sigma^2}{2\pi}$$

17.10 ARCH モデル

Autoregressive Conditional Heteroskedasticity

1. ARCH (p) モデル

$$\epsilon_t|\epsilon_{t-1}, \epsilon_{t-2}, \cdots, \epsilon_1 \sim N(0, h_t)$$

ただし，

$$h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_p\epsilon_{t-p}^2$$

とする。

ϵ_t の条件なしの分散 (Unconditional Variance) :

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p}$$

2. GARCH (p,q)

$$\epsilon_t|\epsilon_{t-1},\epsilon_{t-2},\cdots,\epsilon_1 \sim N(0,h_t)$$

ただし，

$$h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_p\epsilon_{t-p}^2 + \beta_1 h_{t-1} + \cdots + \beta_q h_{t-q}$$

とする。

3. 回帰モデルへの応用 (ARCH (1) のケース) ：

$$y_t = x_t\beta + \epsilon_t,$$

$$\epsilon_t|\epsilon_{t-1},\epsilon_{t-2},\cdots,\epsilon_1 \sim N(0,\alpha_0+\alpha_1\epsilon_{t-1}^2)$$

$\epsilon_1,\cdots,\epsilon_T$ の同時分布は

$$\begin{aligned}
&f(\epsilon_1,\cdots,\epsilon_T)\\
&= f(\epsilon_1)\prod_{t=2}^{T} f(\epsilon_t|\epsilon_{t-1},\cdots,\epsilon_1)\\
&= (2\pi)^{-1/2}\left(\frac{\alpha_0}{1-\alpha_1}\right)^{-1/2}\\
&\times \exp\left(-\frac{1}{2\alpha_0/(1-\alpha_1)}\epsilon_1^2\right)\\
&\times (2\pi)^{-\frac{T-1}{2}}\prod_{t=2}^{T}(\alpha_0+\alpha_1\epsilon_{t-1}^2)^{-1/2}\\
&\times \exp\left(-\frac{1}{2}\sum_{t=2}^{T}\frac{\epsilon_t^2}{\alpha_0+\alpha_1\epsilon_{t-1}^2}\right)
\end{aligned}$$

となる。したがって，対数尤度関数は

$$\begin{aligned}
&\log L(\beta,\alpha_0,\alpha_1;y_1,\cdots,y_T)\\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\alpha_0}{1-\alpha_1}\right)\\
&\quad -\frac{1}{2\alpha_0/(1-\alpha_1)}(y_1-x_1\beta)^2\\
&\quad -\frac{T-1}{2}\log(2\pi)\\
&\quad -\frac{1}{2}\sum_{t=2}^{T}\log\left(\alpha_0+\alpha_1(y_{t-1}-x_{t-1}\beta)^2\right)\\
&\quad -\frac{1}{2}\sum_{t=2}^{T}\frac{(y_t-x_t\beta)^2}{\alpha_0+\alpha_1(y_{t-1}-x_{t-1}\beta)^2}
\end{aligned}$$

によって表される。対数尤度関数を最大にするような α_0, α_1, β を求める。

$\alpha_0>0,\ \alpha_1>0$ でなければならないので，明示的にこの条件を含めて推定することも可能。この場合，モデルは，$\mathrm{E}(\epsilon_t^2|\epsilon_{t-1},\epsilon_{t-2},\cdots,\epsilon_1)=\alpha_0^2+\alpha_1^2\epsilon_{t-1}^2$ と修正される。

ARCH (1) 効果の検定：

(a) $y_t = x_t\beta + u_t$ を最小自乗法 (OLS) で $\widehat{\beta}$ と残差 $\widehat{u}_t = y_t - x_t\widehat{\beta}$ を求める。

(b) $\widehat{u}_t^2 = \alpha_0 + \alpha_1\widehat{u}_{t-1}^2$ を推定する。α_1 が有意であれば ARCH (1) 効果は存在する。これは，LM 検定に相当する。

18　単位根，共和分

18.1　単位根 (Unit Root)

参考文献：　J.D. Hamilton, (1994) *Time Series Analysis*, Princeton University Press

1. なぜ単位根問題が重要か？

(a) 経済変数は時間と共に増加傾向にある。最小自乗法の前提の一つとして，データが定常であるという仮定がある。これは，$\dfrac{1}{T}X'X$ がある行列に収束することを意味する。この仮定が成り立たなければ，これまでの議論が漸近的にも成り立たなくなる。

(b) 非定常時系列の中で例外なのが，単位根の場合である。単位根のとき，自己相関係数の最小自乗法の推定値は一致性を持つ。AR(1) モデルの自己相関係数の最小自乗法の推定値は $\sqrt{T}$ のスピードで真の値に近づいていくが ($\sqrt{T}$-consistent), 単位根の場合は T のスピードで真の値に近づく (T-consistent)。

(c) 経済変数は時間と共に増加傾向にあが，経済変数がトレンド定常 ($y_t = a_0 + a_1 t + \epsilon_t$) か階差定常 ($y_t = b_0 + y_{t-1} + \epsilon_t$) かを調べることは重要である。$k$ 期先の予測の場合，次のような違いがある。

$$(\text{トレンド定常})\quad y_{t+1|t} = a_0 + a_1(t+1)$$

(階差定常) $\quad y_{t+k|t} = b_0 k + y_t$

2. $|\phi_1| < 1$ の場合:

$y_t = \phi_1 y_{t-1} + \epsilon_t,\ \epsilon_t \sim \text{i.i.d. } N(0,\sigma^2),\ y_0 = 0,\ t = 1,\cdots,T$

このとき，ϕ_1 の OLSE は

$$\widehat{\phi}_1 = \frac{\displaystyle\sum_{t=1}^{T} y_{t-1} y_t}{\displaystyle\sum_{t=1}^{T} y_{t-1}^2}$$

となる。$|\phi_1| < 1$ としたとき，

$$\widehat{\phi}_1 = \phi_1 + \frac{\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \phi_1,$$

$$\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t \ \longrightarrow\ 0\right)$$

漸近分散は，

$$\left(\sqrt{T}(\widehat{\phi}_1 - \phi_1)\right)^2 = \frac{\left(\displaystyle\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t\right)^2}{\left(\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^2}$$

分子・分母はそれぞれ，

$$\begin{aligned}&\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t\right)^2 \\ &= \frac{1}{T}\sum_{t=2}^{T}\sum_{s=2}^{T} y_{t-1}y_{s-1}\epsilon_t\epsilon_s \\ &\longrightarrow \sigma^2\gamma(0)\end{aligned}$$

$$\begin{aligned}&\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^2 \\ &\longrightarrow (\gamma(0))^2\end{aligned}$$

となる。よって，

$$\begin{aligned}&\sqrt{T}(\widehat{\phi}_1 - \phi_1) \\ &\longrightarrow N\left(0, \frac{\sigma^2}{\gamma(0)}\right) \\ &= N\left(0, 1-\phi_1^2\right)\end{aligned}$$

となる。$\gamma(0) = \dfrac{\sigma^2}{1-\phi_1^2}$ であることに注意せよ。

3. $\phi_1 = 1$ の場合，予想として，

$$\sqrt{T}(\widehat{\phi}_1 - 1)\ \longrightarrow\ 0$$

となる。すなわち， $\phi_1 = 1$ に確率収束するような分布 (degenetate distribution) になるのか？

4. $\phi_1 = 1$ の場合:

$y_t = y_{t-1} + \epsilon_t$, $y_0 = 0$ は次のように書き直される。

$$y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \ \cdots\ + \epsilon_1$$

よって，

$$y_t \sim N(0, \sigma^2 t)$$

が得られる。

(a) $\sum y_{t-1}\epsilon_t$ を考える。$y_t^2 = (y_{t-1}+\epsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\epsilon_t + \epsilon_t^2$ になるので，

$$y_{t-1}\epsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \epsilon_t^2)$$

と書き換えられる。$y_0 = 0$ を考慮にいれると，

$$\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}y_T^2 - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t^2$$

が得られる。$\sigma^2 T$ で両辺を割って，

$$\begin{aligned}&\frac{1}{\sigma^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t \\ &= \frac{1}{2}\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\end{aligned}$$

となる。$y_t \sim N(0, \sigma^2 t)$ から，

$$\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2 \ \sim\ \chi^2(1)$$

また，

$$\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \ \longrightarrow\ \sigma^2$$

よって，

$$\begin{aligned}&\frac{1}{\sigma^2 T}\sum_{t=1}^{T}y_{t-1}\epsilon_t\\ &=\frac{1}{2}\left(\frac{y_T}{\sigma\sqrt{T}}\right)^2-\frac{1}{2\sigma^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2\\ &\longrightarrow \frac{1}{2}(X-1)\end{aligned}$$

となる。ただし，$X\sim\chi^2(1)$。

(b) $\sum y_{t-1}^2$ を考える。

$$\begin{aligned}\mathrm{E}\left(\sum_{t=1}^{T}y_{t-1}^2\right)&=\sum_{t=1}^{T}\mathrm{E}(y_{t-1}^2)\\ &=\sum_{t=1}^{T}\sigma^2(t-1)\\ &=\sigma^2\frac{T(T-1)}{2}\end{aligned}$$

よって，

$$\frac{1}{T^2}\sum_{t=1}^{T}y_{t-1}^2 \ \longrightarrow\ \text{ある分布}$$

5. $\sqrt{T}(\widehat{\phi}_1-1)$ ではなく，$T(\widehat{\phi}_1-1)$ がある極限分布を持つ。

6. ランダム・ウォークに関する基礎知識：

(a) モデル： $y_t=y_{t-1}+\epsilon_t$, $y_0=0$, $\epsilon_t\sim N(0,1)$ とする。このとき，

$$\begin{aligned}&y_t=\epsilon_t+\epsilon_{t-1}+\ \cdots\ +\epsilon_1\\ &y_t\sim N(0,t)\end{aligned}$$

y_s と y_t $(s>t)$ との差は

$$y_s-y_t=\epsilon_s+\epsilon_{s-1}+\ \cdots\ +\epsilon_{t+2}+\epsilon_{t+1}$$

となり，その分布は

$$y_s-y_t\sim N(0,s-t)$$

となる。

(b) 次のように書き換えられる。

$$y_t=y_{t-1}+e_{1,t}+e_{2,t}+\ \cdots\ +e_{N,t}$$

ただし，上のランダム・ウォーク過程との関連では，$\epsilon_t=e_{1,t}+e_{2,t}+\cdots+e_{N,t}$ である。すなわち，t 期から $t+1$ 期の間を N 期に分割する。$N\to\infty$ の連続時間のときを Standard Brownian Motion と呼び，そのときの t 期の値を $W(t)$ とする。すなわち，t 期から $t+\Delta$ 期への変化は $N(0,\Delta)$ の分布となる。$\Delta\to 0$ のときの値を $W(t)$ とする。

定義:

Standard Brownian motion $W(\cdot)$ は連続時間の確率過程であり，$W(t)$, $t\in[0,1]$ は次の条件を満たす。

i. $W(0)=0$

ii. $0\le r_1<r_2<\ \cdots\ <r_k\le 1$ となる任意の時点について，$W(r_2)-W(r_1)$, $W(r_3)-W(r_2)$, $\cdots$, $W(r_k)-W(r_{k-1})$ はそれぞれ独立に $W(s)-W(t)\sim N(0,s-t)$ の正規分布をする。

iii. $W(r)$ は r で (確率 1 で) 連続である。

例えば，

$$Z(t)=\sigma W(t)\sim N(0,\sigma^2 t)$$

これは，分散 σ^2 の Brownian motion

$$Z(t)=(W(t))^2\sim t\times\chi^2(1)$$

(c) $X_T(r)$ を定義する。ただし，$r\in[0,1]$ とする。

$$X_T(r)=\begin{cases}0, & 0\le r<\dfrac{1}{T}\\ \dfrac{u_1}{T}, & \dfrac{1}{T}\le r<\dfrac{2}{T}\\ \dfrac{u_1+u_2}{T}, & \dfrac{2}{T}\le r<\dfrac{3}{T}\\ \quad\vdots & \\ \dfrac{u_1+u_2+\ \cdots\ +u_T}{T}, & r=1\end{cases}$$

$$\begin{aligned}&X_T(r)\equiv\frac{1}{T}\sum_{t=1}^{[Tr]}u_t\\ &\sqrt{T}X_T(r)\ \longrightarrow\ N(0,r\sigma^2)\end{aligned}$$

ただし，$[Tr]$ は $T \times r$ を越えない最大の整数とする。

$$\frac{\sqrt{T}\left(X_T(r_2) - X_T(r_1)\right)}{\sigma} \longrightarrow N(0, r_2 - r_1)$$

$$\frac{\sqrt{T}X_T(\cdot)}{\sigma} \longrightarrow W(\cdot)$$

例えば,

$$X_T = \frac{1}{T}\sum_{t=1}^{T} u_t$$

のとき,

$$\frac{\sqrt{T}X_T(1)}{\sigma} = \frac{1}{\sigma\sqrt{T}}\sum_{t=1}^{T} u_t \longrightarrow W(1) \sim N(0,1)$$

(d) $y_t = y_{t-1} + \epsilon_t$ のとき,

$$X_T(r) = \begin{cases} 0, & 0 \le r < \dfrac{1}{T} \\ \dfrac{y_1}{T}, & \dfrac{1}{T} \le r < \dfrac{2}{T} \\ \dfrac{y_2}{T}, & \dfrac{2}{T} \le r < \dfrac{3}{T} \\ \quad\vdots & \\ \dfrac{y_{T-1}}{T}, & \dfrac{T-1}{T} \le r < 1 \\ \dfrac{y_T}{T}, & r = 1 \end{cases}$$

$$S_T(r) = \begin{cases} 0, & 0 \le r < \dfrac{1}{T} \\ \dfrac{y_1^2}{T}, & \dfrac{1}{T} \le r < \dfrac{2}{T} \\ \dfrac{y_2^2}{T}, & \dfrac{2}{T} \le r < \dfrac{3}{T} \\ \quad\vdots & \\ \dfrac{y_{T-1}^2}{T}, & \dfrac{T-1}{T} \le r < 1 \\ \dfrac{y_T^2}{T}, & r = 1 \end{cases}$$

それぞれの区間について，長方形の面積の和を求めると,

$$\begin{aligned} &\int_0^1 X_T(r)\mathrm{d}r \\ &= \frac{y_1}{T}\left(\frac{2}{T} - \frac{1}{T}\right) + \frac{y_2}{T}\left(\frac{3}{T} - \frac{2}{T}\right) \\ &+ \ \cdots \ + \frac{y_{T-1}}{T}\left(1 - \frac{T-1}{T}\right) \\ &= \frac{y_1}{T^2} + \frac{y_2}{T^2} + \ \cdots \ + \frac{y_{T-1}}{T^2} \\ &= \frac{1}{T^2}\sum_{t=1}^{T} y_t \end{aligned}$$

$$\begin{aligned} \int_0^1 S_T(r)\mathrm{d}r &= \frac{y_1^2}{T^2} + \frac{y_2^2}{T^2} + \ \cdots \ + \frac{y_{T-1}^2}{T^2} \\ &= \frac{1}{T^2}\sum_{t=1}^{T} y_t^2 \end{aligned}$$

一方,

$$\int_0^1 \sqrt{T}X_T(r)\mathrm{d}r \longrightarrow \sigma\int_0^1 W(r)\mathrm{d}r$$

から,

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} y_t \longrightarrow \sigma\int_0^1 W(r)\mathrm{d}r$$

となる。

$\sqrt{T}X_T(\cdot) \longrightarrow \sigma W(\cdot)$,

$S_T(r) \equiv \left(\sqrt{T}X_T(r)\right)^2$

から,

$$S_T(\cdot) \longrightarrow \sigma^2\left(W(\cdot)\right)^2$$

となる。(Continuous Mapping Theorem)

したがって,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_t^2 = \int_0^1 S_T(r)\mathrm{d}r \longrightarrow \sigma^2\int_0^1 \left(W(r)\right)^2 \mathrm{d}r$$

となる。

(e) $T^{-3/2}\sum_{t=1}^{T} y_{t-1}$ を次のように分解する。

$$\begin{aligned} &T^{-3/2}\sum_{t=1}^{T} y_{t-1} \\ &= T^{-3/2}\big(u_1 + (u_1+u_2) + (u_1+u_2+u_3) \\ &+ \ \cdots \ + (u_1+u_2+ \ \cdots \ + u_{T-1})\big) \\ &= T^{-3/2}\big((T-1)u_1 + (T-2)u_2 + (T-3)u_3 \\ &+ \ \cdots \ + 2u_{T-2} + u_{T-1}\big) \\ &= T^{-3/2}\sum_{t=1}^{T}(T-t)u_t \\ &= T^{-1/2}\sum_{t=1}^{T} u_t - T^{-3/2}\sum_{t=1}^{T} tu_t \end{aligned}$$

次の事実を利用すると，

$$\begin{pmatrix} T^{-1/2}\sum_{t=1}^{T} u_t \\ T^{-3/2}\sum_{t=1}^{T} t\ u_t \end{pmatrix} \longrightarrow$$

$$N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}\right)$$

$T^{-3/2}\sum_{t=1}^{T} y_{t-1}$ の分散は $\dfrac{\sigma^2}{3}$ となる。よって，

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} = \sigma \int_0^1 W(r)\mathrm{d}r \longrightarrow N(0, \frac{\sigma^2}{3})$$

(f) $W(1) \sim N(0,1)$ であるので，$(W(1))^2 \sim \chi^2(1)$ となる。

(g) **いくつかの公式:**

モデル: $y_t = y_{t-1} + \epsilon_t$

i. $T^{-1/2}\sum_{t=1}^{T}\epsilon_t \longrightarrow \sigma\ W(1)$ $= N(0, \sigma^2)$

ii. $T^{-1}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow \frac{1}{2}\sigma^2\left((W(1))^2 - 1\right)$ $= \frac{1}{2}\sigma^2\left(\chi^2(1) - 1\right)$

iii. $T^{-3/2}\sum_{t=1}^{T} t\ \epsilon_t \longrightarrow \sigma\ W(1) - \sigma\int_0^1 W(r)\mathrm{d}r$, $(= N(0, \sigma^2/3))$

iv. $T^{-3/2}\sum_{t=1}^{T} y_{t-1} \longrightarrow \sigma\int_0^1 W(r)\mathrm{d}r$ $= N(0, \sigma^2)$

v. $T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \sigma^2\int_0^1 (W(r))^2\,\mathrm{d}r$

vi. $T^{-5/2}\sum_{t=1}^{T} t\ y_{t-1} \longrightarrow \sigma\int_0^1 r\ W(r)\mathrm{d}r$

vii. $T^{-3}\sum_{t=1}^{T} t\ y_{t-1}^2 \longrightarrow \sigma^2\int_0^1 r\ (W(r))^2\,\mathrm{d}r$

7. 正確な分布の導出

(a) 「真のモデル： $y_t = y_{t-1} + \epsilon_t$」，「推定するモデル： $y_t = \phi_1 y_{t-1} + \epsilon_t$」 の場合：
ϕ_1 の推定値を $\widehat{\phi}_1$ とする。

$$\widehat{\phi}_1 = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2}$$

$$T(\widehat{\phi}_1 - 1) = \frac{T^{-1}\sum_{t=1}^{T} y_{t-1}u_t}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}$$

$$T(\widehat{\phi}_1 - 1) \longrightarrow \frac{\frac{1}{2}\left((W(1))^2 - 1\right)}{\int_0^1 (W(r))^2\,\mathrm{d}r}$$

$(W(1))^2 \sim \chi^2(1)$ であることに注意せよ。$\widehat{\phi}_1$ は super-consistent であるという。通常の t 検定の統計量は

$$t_T = \frac{\widehat{\phi}_1 - 1}{s_\phi},$$

$$s_\phi = \left(s_T^2 \Big/ \sum_{t=1}^{T} y_{t-1}^2\right)^{1/2},$$

$$s_T^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \widehat{\phi}_1 y_{t-1})^2$$

と表される。t 統計量 t_T を次のように変形する。

$$t_T = \frac{\widehat{\phi}_1 - 1}{s_\phi} = \frac{T(\widehat{\phi}_1 - 1)}{Ts_\phi}$$

分母は

$$Ts_\phi = \left(s_T^2 \Big/ \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2\right)^{1/2} \longrightarrow \left(\sigma^2 \Big/ \left(\sigma^2\int_0^1 (W(r))^2\,\mathrm{d}r\right)\right)^{1/2} = \left(\int_0^1 (W(r))^2\,\mathrm{d}r\right)^{-1/2}$$

となる。$s^2 \longrightarrow \sigma^2$ が利用されている。

よって，

$$t_T=\frac{\widehat{\phi}_1-1}{s_\phi}\longrightarrow$$

$$\frac{\frac{1}{2}\left((W(1))^2-1\right)}{\int_0^1 (W(r))^2\,\mathrm{d}r}\Big/\left(\int_0^1 (W(r))^2\,\mathrm{d}r\right)^{-1/2}$$

$$=\frac{\frac{1}{2}\left((W(1))^2-1\right)}{\left(\int_0^1 (W(r))^2\,\mathrm{d}r\right)^{1/2}}$$

となり，通常の t_T 統計量は t 分布に従わない。

(b) 「真のモデル： $y_t=y_{t-1}+\epsilon_t$」,「推定するモデル： $y_t=\alpha_0+\phi_1 y_{t-1}+\epsilon_t$」 の場合：

$$\begin{aligned}
&\begin{pmatrix}\widehat{\alpha}_0\\ \widehat{\phi}_1\end{pmatrix}\\
&=\begin{pmatrix}T & \sum y_{t-1}\\ \sum y_{t-1} & \sum y_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum y_t\\ \sum y_{t-1}y_t\end{pmatrix}\\
&=\begin{pmatrix}\alpha_0\\ \phi_1\end{pmatrix}\\
&\quad+\begin{pmatrix}T & \sum y_{t-1}\\ \sum y_{t-1} & \sum y_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum u_t\\ \sum y_{t-1}u_t\end{pmatrix}
\end{aligned}$$

真のモデルは $\alpha_0=0$, $\phi_1=1$ であるので,

$$\begin{aligned}
&\begin{pmatrix}\widehat{\alpha}_0\\ \widehat{\phi}_1-1\end{pmatrix}\\
&=\begin{pmatrix}T & \sum y_{t-1}\\ \sum y_{t-1} & \sum y_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\sum u_t\\ \sum y_{t-1}u_t\end{pmatrix}\\
&=\begin{pmatrix}O_p(T) & O_p(T^{3/2})\\ O_p(T^{3/2}) & O_p(T^2)\end{pmatrix}^{-1}\begin{pmatrix}O_p(T^{1/2})\\ O_p(T)\end{pmatrix}
\end{aligned}$$

となる。

注) 確率変数 x, 定数 k にいついて，$x=O_p(k)$ とは x/k がある分布に収束することを意味する。

行列の各要素を $O_p(1)$ にするためには,

$$\Gamma=\begin{pmatrix}T^{1/2} & 0\\ 0 & T\end{pmatrix}$$

を用いて,

$$\begin{aligned}
&\Gamma\begin{pmatrix}\widehat{\alpha}_0\\ \widehat{\phi}_1-1\end{pmatrix}\\
&=\begin{pmatrix}T^{1/2}\widehat{\alpha}_0\\ T\widehat{\phi}_1-1\end{pmatrix}\\
&=\Gamma\begin{pmatrix}O_p(T) & O_p(T^{3/2})\\ O_p(T^{3/2}) & O_p(T^2)\end{pmatrix}^{-1}\\
&\quad\times\Gamma\Gamma^{-1}\begin{pmatrix}O_p(T^{1/2})\\ O_p(T)\end{pmatrix}\\
&=\left(\Gamma^{-1}\begin{pmatrix}O_p(T) & O_p(T^{3/2})\\ O_p(T^{3/2}) & O_p(T^2)\end{pmatrix}\Gamma^{-1}\right)^{-1}\\
&\quad\times\Gamma^{-1}\begin{pmatrix}O_p(T^{1/2})\\ O_p(T)\end{pmatrix}\\
&=\left(\Gamma^{-1}\begin{pmatrix}T & \sum y_{t-1}\\ \sum y_{t-1} & \sum y_{t-1}^2\end{pmatrix}\Gamma^{-1}\right)^{-1}\\
&\quad\times\Gamma^{-1}\begin{pmatrix}\sum u_t\\ \sum y_{t-1}u_t\end{pmatrix}\\
&=\begin{pmatrix}1 & T^{-3/2}\sum y_{t-1}\\ T^{-3/2}\sum y_{t-1} & T^{-2}\sum y_{t-1}^2\end{pmatrix}^{-1}\\
&\quad\times\begin{pmatrix}T^{-1/2}\sum u_t\\ T^{-1}\sum y_{t-1}u_t\end{pmatrix}
\end{aligned}$$

と変形する。それぞれは，次に様に収束する。

$$\begin{aligned}
&\begin{pmatrix}1 & T^{-3/2}\sum y_{t-1}\\ T^{-3/2}\sum y_{t-1} & T^{-2}\sum y_{t-1}^2\end{pmatrix}\longrightarrow\\
&\begin{pmatrix}1 & 0\\ 0 & \sigma\end{pmatrix}\\
&\times\begin{pmatrix}1 & \int_0^1 W(r)\mathrm{d}r\\ \int_0^1 W(r)\mathrm{d}r & \int_0^1 (W(r))^2\,\mathrm{d}r\end{pmatrix}\\
&\times\begin{pmatrix}1 & 0\\ 0 & \sigma\end{pmatrix}
\end{aligned}$$

なぜなら,

$$\begin{aligned}
&\begin{pmatrix}1 & \sigma\int_0^1 W(r)\mathrm{d}r\\ \sigma\int_0^1 W(r)\mathrm{d}r & \sigma^2\int_0^1 (W(r))^2\,\mathrm{d}r\end{pmatrix}\\
&=\begin{pmatrix}1 & 0\\ 0 & \sigma\end{pmatrix}\\
&\quad\times\begin{pmatrix}1 & \int_0^1 W(r)\mathrm{d}r\\ \int_0^1 W(r)\mathrm{d}r & \int_0^1 (W(r))^2\,\mathrm{d}r\end{pmatrix}\\
&\quad\times\begin{pmatrix}1 & 0\\ 0 & \sigma\end{pmatrix}
\end{aligned}$$

となることに注意せよ。

$$\begin{pmatrix}T^{-1/2}\sum u_t\\ T^{-1}\sum y_{t-1}u_t\end{pmatrix}\longrightarrow$$

$$\sigma\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\begin{pmatrix}W(1)\\ \frac{1}{2}\left((W(1))^2-1\right)\end{pmatrix}$$

なぜなら，

$$\begin{pmatrix}\sigma\, W(1)\\ \frac{1}{2}\sigma^2\left((W(1))^2-1\right)\end{pmatrix}$$

$$=\sigma\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\begin{pmatrix}W(1)\\ \frac{1}{2}\left((W(1))^2-1\right)\end{pmatrix}$$

が成り立つ。

よって，

$$\begin{pmatrix}T^{1/2}\widehat{\alpha}_0\\ T\widehat{\phi}_1-1\end{pmatrix}\longrightarrow$$

$$\left(\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\begin{pmatrix}1&\int_0^1 W(r)\mathrm{d}r\\ \int_0^1 W(r)\mathrm{d}r&\int_0^1 (W(r))^2\,\mathrm{d}r\end{pmatrix}\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\right)^{-1}$$

$$\times\sigma\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\begin{pmatrix}W(1)\\ \frac{1}{2}\left((W(1))^2-1\right)\end{pmatrix}$$

$T(\widehat{\phi}_1-1)$ は漸近的に次の分布に収束する。

$$T(\widehat{\phi}_1-1)\longrightarrow$$

$$\frac{\frac{1}{2}\left((W(1))^2-1\right)-W(1)\int_0^1 W(r)\mathrm{d}r}{\int_0^1 (W(r))^2\,\mathrm{d}r-\left(\int_0^1 W(r)\mathrm{d}r\right)^2}$$

通常の t 統計量との関連

$$t_T=\frac{\widehat{\phi}_1-1}{\left(s_\phi^2\right)^{1/2}}=\frac{T(\widehat{\phi}_1-1)}{\left(T^2s_\phi^2\right)^{1/2}}$$

分母について，

$$s_\phi^2=s_T^2\,(0\quad 1)\begin{pmatrix}T&\sum y_{t-1}\\ \sum y_{t-1}&\sum y_{t-1}^2\end{pmatrix}^{-1}\begin{pmatrix}0\\1\end{pmatrix}$$

$$s_T^2=\frac{1}{T-2}\sum_{t=1}^T(y_t-\widehat{\alpha}_0-\widehat{\phi}_1y_{t-1})^2$$

$T^2s_\phi^2$ は次のように確率収束する。

$$T^2s_\phi^2\longrightarrow$$

$$\sigma^2\,(0\quad 1)\left(\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\right.$$

$$\times\begin{pmatrix}1&\int_0^1 W(r)\mathrm{d}r\\ \int_0^1 W(r)\mathrm{d}r&\int_0^1 (W(r))^2\,\mathrm{d}r\end{pmatrix}$$

$$\left.\times\begin{pmatrix}1&0\\0&\sigma\end{pmatrix}\right)^{-1}\begin{pmatrix}0\\1\end{pmatrix}$$

$$=\frac{1}{\int_0^1 (W(r))^2\,\mathrm{d}r-\left(\int_0^1 W(r)\mathrm{d}r\right)^2}$$

よって，t 統計量は

$$t_T\longrightarrow$$

$$\frac{\frac{1}{2}\left((W(1))^2-1\right)-W(1)\int_0^1 W(r)\mathrm{d}r}{\left(\int_0^1 (W(r))^2\,\mathrm{d}r-\left(\int_0^1 W(r)\mathrm{d}r\right)^2\right)^{1/2}}$$

(c) 「真のモデル： $y_t=\alpha_0+y_{t-1}+\epsilon_t$」,「推定するモデル： $y_t=\alpha_0+\phi_1y_{t-1}+\epsilon_t$」 の場合：
モデルを次のように書き直す。

$$\begin{aligned}y_t&=y_0+\alpha_0t+(\epsilon_1+\epsilon_2+\ \cdots\ +\epsilon_t)\\&=y_0+\alpha_0t+u_t\end{aligned}$$

ただし，$u_t=\epsilon_1+\epsilon_2+\ \cdots\ +\epsilon_t$ とする。

○ $\sum_{t=1}^T y_{t-1}$ について

$$\begin{aligned}\sum_{t=1}^T y_{t-1}&=\sum_{t=1}^T y_0+\sum_{t=1}^T\alpha_0(t-1)+\sum_{t=1}^T u_{t-1}\\&=O_p(T)+O_p(T^2)+O_p(T^{3/2})\end{aligned}$$

$$T^{-2}\sum_{t=1}^T y_{t-1}\ \longrightarrow\ \frac{\alpha_0}{2}$$

○ $\sum_{t=1}^T y_{t-1}^2$ について

$$\begin{aligned}\sum_{t=1}^T y_{t-1}^2&=\sum_{t=1}^T(y_0+\alpha_0(t-1)+u_{t-1})^2\\&=\sum_{t=1}^T y_0^2+\sum_{t=1}^T\alpha_0^2(t-1)^2+\sum_{t=1}^T u_{t-1}^2\\&\quad+\sum_{t=1}^T 2y_0\alpha_0(t-1)+\sum_{t=1}^T 2y_0u_{t-1}\\&\quad+\sum_{t=1}^T 2\alpha_0(t-1)u_{t-1}\\&=O_p(T)+O_p(T^3)+O_p(T^2)\\&\quad+O_p(T^2)+O_p(T^{3/2})\end{aligned}$$

$$+ O_p(T^{5/2})$$

$$T^{-3}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \frac{\alpha_0^2}{3}$$

○ $\displaystyle\sum_{t=1}^{T} y_{t-1}\epsilon_t$ について

$$\begin{aligned}
&\sum_{t=1}^{T} y_{t-1}\epsilon_t \\
&= \sum_{t=1}^{T} \left(y_0 + \alpha_0(t-1) + u_{t-1}\right)\epsilon_t \\
&= \sum_{t=1}^{T} y_0\epsilon_t + \sum_{t=1}^{T}\alpha_0(t-1)\epsilon_t + \sum_{t=1}^{T} u_{t-1}\epsilon_t \\
&= O_p(T^{1/2}) + O_p(T^{3/2}) + O_p(T)
\end{aligned}$$

したがって，OLS 推定値は

$$\begin{aligned}
&\begin{pmatrix} \widehat{\alpha}_0 - \alpha_0 \\ \widehat{\phi}_1 - 1 \end{pmatrix} \\
&= \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum u_t \\ \sum y_{t-1}u_t \end{pmatrix} \\
&= \begin{pmatrix} O_p(T) & O_p(T^2) \\ O_p(T^2) & O_p(T^3) \end{pmatrix}^{-1} \begin{pmatrix} O_p(T^{1/2}) \\ O_p(T^{3/2}) \end{pmatrix}
\end{aligned}$$

となる。

$$\Gamma = \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix}$$

を用いて，同様の計算を行うと，

$$\begin{pmatrix} T^{1/2}(\widehat{\alpha}_0 - \alpha_0) \\ T^{3/2}(\widehat{\phi}_1 - 1) \end{pmatrix} \longrightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \dfrac{\alpha_0}{2} \\ \dfrac{\alpha_0}{2} & \dfrac{\alpha_0^2}{3} \end{pmatrix}\right)$$

を得る。

(d) その他のケース：

i. 「真のモデル： $y_t = y_{t-1} + \epsilon_t$」,「推定するモデル： $y_t = \alpha_0 + \alpha_1 t + \phi_1 y_{t-1} + \epsilon_t$」 の場合：

ii. 「真のモデル： $y_t = \alpha_0 + y_{t-1} + \epsilon_t$」,「推定するモデル： $y_t = \alpha_0 + \alpha_1 t + \phi_1 y_{t-1} + \epsilon_t$」の場合：

iii. 「真のモデル：$y_t = \alpha_0 + \alpha_1 t + y_{t-1} + \epsilon_t$」,「推定するモデル：$y_t = \alpha_0 + \alpha_1 t + \phi_1 y_{t-1} + \epsilon_t$」の場合：

18.2 共和分 (Cointegration)

1. y_t (スカラー) について，$\Delta y_t = y_t - y_{t-1}$ がホワイト・ノイズのとき，$\Delta y_t \sim I(1)$ と書く。

2. **Cointegration の定義：**

   $g \times 1$ のベクトル y_t のそれぞれの系列 (要素) が個々に $I(1)$ であり (言い換えると，各系列が単位根を持つ), 各系列の線形結合 (すなわち，$g \times 1$ の nonzero ベクトル a について，$a'y_t$) が $I(0)$ であるとき (すなわち，定常)，y_t は cointegration が生じているという。

3. **例：**

   $y_t = (y_{1,t},\ y_{2,t})'$ が次の Vector Autoregressive Process に従っているとする。

$$\begin{aligned} y_{1,t} &= \phi_1 y_{2,t} + \epsilon_{1,t} \\ y_{2,t} &= y_{2,t-1} + \epsilon_{2,t} \end{aligned}$$

   このとき，

$$\begin{aligned} &\Delta y_{1,t} = \phi_1 u_{2,t} + u_{1,t} - u_{1,t-1}, \quad \text{(MA(1) 過程)} \\ &\Delta y_{2,t} = u_{2,t} \end{aligned}$$

   となり，$y_{1,t}$，$y_{2,t}$ は共に $I(1)$ である。一方，$y_{1,t} - \phi_1 y_{2,t}$ は $I(0)$ である。

   よって，$y_t = (y_{1,t},\ y_{2,t})'$ は cointegration の関係があるという。そのときの cointegrating vector は $a = (1,\ -\phi_1)$ である。Cointegrating vector は一意でない。よって，通常，a の一番目の要素は 1 と置かれる。

4. y_t, x_t 定常である必要はない。回帰モデル $y_t = x_t\beta + u_t$ について，$u_t \sim I(0)$ となるような β が存在しなければ，OLS は信頼性はない (spurious regression, 見せかけの回帰)。

5. $y_t \sim I(1)$，y_t は $g \times 1$ ベクトル，$y_t = (y_{1,t}, y_{2,t}')'$, $y_{2,t}$ は $k \times 1$ ベクトル，$k = g - 1$ とする。次の回帰モデルを考える。

$$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t, \qquad t = 1, \cdots, T$$

   OLS 推定値は

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y_{2,t}' \\ \sum y_{2,t} & \sum y_{2,t}y_{2,t}' \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t}y_{2,t} \end{pmatrix}$$

となる。

次に，帰無仮説 H_0 : $R\gamma = r$ の検定を考える。ただし，R は $m \times k$ の行列，r は $m \times 1$ のベクトルとする。F 検定の統計量 (F_T とする) は

$$F_T = \frac{1}{m}(R\widehat{\gamma} - r)' \left(s_T^2 \begin{pmatrix} 0 & R \end{pmatrix} \begin{pmatrix} T & \sum y_{2,t}' \\ \sum y_{2,t} & \sum y_{2,t} y_{2,t}' \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ R' \end{pmatrix} \right)^{-1} (R\widehat{\gamma} - r)$$

$$s_T^2 = \frac{1}{T-g} \sum_{t=1}^{T} (y_{1,t} - \widehat{\alpha} - \widehat{\gamma}' y_{2,t})^2$$

となる。

$y_{1,t} - \gamma y_{2,t}$ が定常となるような γ が存在しないとき，OLS 推定値 $\widehat{\gamma}$ は有意にゼロとは異なり，標本が十分大きいとき，F 検定は帰無仮説を棄却する。

6. **Cointegrating Vector の推定：**

$z_t = a' y_t$ と置く。z_t はスカラー，a, y_t は $g \times 1$ ベクトルとする。

$z_t \sim I(0)$ のとき (定常),

$$T^{-1} \sum_{t=1}^{T} z_t^2 = T^{-1} \sum_{t=1}^{T} (a' y_t)^2 \ \longrightarrow\ \mathrm{E}(z_t^2)$$

$z_t \sim I(1)$ のとき (非定常，a は cointegrating vector でないとき),

$$T^{-2} \sum_{t=1}^{T} (a' y_t)^2 \ \longrightarrow\ \lambda^2 \int_0^1 (W(r))^2 \, \mathrm{d}r$$

$W(r)$ は standard Brownian motion，λ^2 は $(1-L)z_t$ の分散とする。

もし，a が cointegrating vector でなければ，

$T^{-1} \sum z_t$ は発散する。

定理:

$y_{1,t}$ はスカラー，$y_{2,t}$ は $k \times 1$ ベクトル，$g = k+1$，$(y_{1,t}, y_{2,t}')'$ は $g \times 1$ ベクトル。次のモデルを考える。

$$y_{1,t} = \alpha + \gamma' y_{2,t} + z_t^*$$

$$\Delta y_{2,t} = u_{2,t}$$

$$\begin{pmatrix} z_t^* \\ u_{2,t} \end{pmatrix} = \Psi^*(L) \epsilon_t$$

ϵ_t は $g \times 1$ i.i.d. ベクトルで $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$。

OLS 推定値は

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y_{2,t}' \\ \sum y_{2,t} & \sum y_{2,t} y_{2,t}' \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}$$

次の λ_1^* $(g \times 1)$, Λ_2^* $(k \times g)$ を定義する。

$$\Psi^*(1)\ P = \begin{pmatrix} \lambda_1^{*\prime} \\ \Lambda_2^* \end{pmatrix}$$

このとき，

$$\begin{pmatrix} T^{1/2}(\widehat{\alpha} - \alpha) \\ T(\widehat{\gamma} - \gamma) \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \left(\Lambda_2^* \int W(r) \mathrm{d}r \right)' \\ \Lambda_2^* \int W(r) \mathrm{d}r & \Lambda_2^* \left(\int (W(r)) (W(r))' \, \mathrm{d}r \right) \Lambda_2^{*\prime} \end{pmatrix}^{-1} \times \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

となる。ただし，

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \lambda_1^{*\prime} W(1) \\ \Lambda_2^* \left(\int W(r) \, (\mathrm{d}W(r))' \right) \lambda_1^* + \sum_{\tau=0}^{\infty} \mathrm{E}(u_{2,t} z_{t+\tau}^*) \end{pmatrix}$$

である。$W(r)$ は g 次元の standard Brownian motion である。

1) Cointegrating vector の OLS 推定値は，たとえ攪乱項 u_t に系列相関が存在しても，一致性を持つ。

2) OLS 推定値は一致性を持つので，$T^{-1} \sum \widehat{u}_t^2 \ \longrightarrow\ \sigma^2$ となる。

3) $T^{-1} \sum (y_{1,t} - \overline{y}_1)^2$ は無限大に大きくなるので，決定係数 R^2 は 1 に近づく。

19 GMM (Generalized Mothod of Moments)

1. 積率法 (Method of Moments):

回帰モデル: $y_t = x_t \beta + \epsilon_t$

仮定より，$\mathrm{E}(x_t'\epsilon_t)=0$。

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}x_t'\epsilon_t &= \frac{1}{T}\sum_{t=1}^{T}x_t'(y_t-x_t\beta)\\ &=0\end{aligned}$$

したがって，

$$\beta_{MM}=\left(\frac{1}{T}\sum_{t=1}^{T}x_t'x_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}x_t'y_t\right)$$

となり，最小自乗法 (OLS) に一致する。

2. 一般化積率法 (Generalized Mothod of Moments, GMM):

$$\mathrm{E}\left(h(\theta;w_t)\right)=0$$

θ は推定されるべき未知パラメータ (k×1 ベクトル), w_t は観測値ベクトルで $w_t=(y_t,x_t)$，$h(\theta;w_t)$ は $r\times 1$ ベクトル $(r\geq k)$ とする。

$g(\theta;W_T)$ を次のように定義する。

$$g(\theta;W_T)=\frac{1}{T}\sum_{t=1}^{T}h(\theta;w_t)$$

ただし，$W_T=\{w_T,w_{T-1},\cdots,w_1\}$ とする。

$$\Big(g(\theta;W_T)\Big)'S^{-1}\Big(g(\theta;W_T)\Big)$$

を最小にする θ の推定値 $\widehat{\theta}_T$ を GMM 推定量とする。ただし，S は次のように定義される。

$$S=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=-\infty}^{\infty}\mathrm{E}\left(\left(h(\theta;w_t)\right)\left(h(\theta;w_{t-\tau})\right)'\right)$$

実際には，S はその推定値によって置き換えられ $\widehat{S}_T$ が用いられる。

$$\widehat{S}_T=\frac{1}{T}\sum_{t=1}^{T}\Big(h(\widehat{\theta}_T;w_t)\Big)\Big(h(\widehat{\theta}_T;w_t)\Big)' \longrightarrow S$$

このとき，

$$\sqrt{T}(\widehat{\theta}_T-\theta) \longrightarrow N\left(0,(DS^{-1}D')^{-1}\right)$$

ただし，

$$D=\frac{\partial g(\theta;W_T)}{\partial\theta'}$$

D は $r\times k$ 行列。$\widehat{D}_T$ を D の推定値とすると，$\widehat{\theta}_T$ の分散の推定値は

$$\widehat{D}_T=\frac{\partial g(\widehat{\theta}_T;W_T)}{\partial\theta'}$$

となる。

漸近的正規性の証明:

仮定 1： $\widehat{\theta}_T \longrightarrow \theta$

仮定 2： $\sqrt{T}g(\theta;W_T) \longrightarrow N(0,S)$

このとき，

$$\begin{aligned}g(\theta;W_T) &\approx g(\widehat{\theta}_T;W_T)+\frac{\partial g(\widehat{\theta}_T;W_T)}{\partial\theta'}(\theta-\widehat{\theta}_T)\\ &= g(\widehat{\theta}_T;W_T)+\widehat{D}_T(\theta-\widehat{\theta}_T)\end{aligned}$$

となる。一方，GMM の一階の条件は，

$$\Big(\frac{\partial g(\theta;W_T)}{\partial\theta'}\Big)'S^{-1}\Big(g(\theta;W_T)\Big)=0$$

となる。

さらに，近似式を代入して

$$\begin{aligned}&D'S^{-1}\Big(g(\theta;W_T)\Big)\\ &=D'S^{-1}\Big(g(\theta;W_T)+\widehat{D}_T(\theta-\widehat{\theta}_T)\Big)\\ &=D'S^{-1}g(\widehat{\theta}_T;W_T)\\ &\quad+D'S^{-1}\widehat{D}_T(\theta-\widehat{\theta}_T)\end{aligned}$$

を得る。

したがって，

$$\begin{aligned}&\sqrt{T}(\widehat{\theta}_T-\theta)\\ &\approx(D'S^{-1}\widehat{D}_T)^{-1}D'S^{-1}\\ &\quad\times\sqrt{T}\Big(g(\widehat{\theta}_T;W_T)-g(\theta;W_T)\Big)\end{aligned}$$

よって，GMM 推定値 $\widehat{\theta}_T$ は次の漸近分布を持つ。

$$\sqrt{T}(\widehat{\theta}_T-\theta) \longrightarrow N\left(0,(D'S^{-1}D)^{-1}\right)$$

ここでは，$\widehat{D}_T \longrightarrow D$ が利用されている。

さらに，仮定 2 より，

$$\left(\sqrt{T}g(\widehat{\theta}_T;W_T)\right)'\widehat{S}_T^{-1}\left(\sqrt{T}g(\widehat{\theta}_T;W_T)\right)$$
$$\longrightarrow\ \chi^2(r)$$

が成り立つことにも注意せよ。

例：

(a) OLS:
回帰モデル：　$y_t = x_t\beta + \epsilon_t$, $\mathrm{E}(x_t\epsilon_t) = 0$

$$h(\theta; w_t) = x_t(y_t - x_t\beta)$$

(b) IV:
回帰モデル：　$y_t = x_t\beta + \epsilon_t$, $\mathrm{E}(x_t\epsilon_t) \neq 0$, $\mathrm{E}(z_t\epsilon_t) = 0$

$$h(\theta; w_t) = z_t(y_t - x_t\beta)$$

(c) NLS:
回帰モデル：　$f(y_t, x_t, \beta) = \epsilon_t$, $\mathrm{E}(z_t\epsilon_t) = 0$

$$h(\theta; w_t) = z_t f(y_t, x_t, \beta)$$

20　その他のトピック

1. パネル・データ (Panel Data)
2. Discrete Dependent Variable, Limited Dependent Variable
3. SUR (Seemingly Unrelated Regression model)
4. 同時方程式 (Simultaneous Equation)
5. ベイズ推定 (Bayesian Estimation)
6. ノンパラメトリック推定 (Nonparametric Regression)