ミクロ計量経済

Thu., 8:50-10:20

Room # 4 (文法経講義棟)

- The prerequisite of this class is **Basic Statistics (統計基礎)** and **Econometrics (エコノメトリックス)** (undergraduate level, next semester,『計量経済学』山本 拓 著, 新世社).
- The class of **Introductory Econometrics** (計量経済学基礎) should be registered.

http://www2.econ.osaka-u.ac.jp/˜tanizaki/class/2017/micro_econome

予定

1. 最小二乗法（復習）

2. 最尤法（復習）

3. 質的データ

4. パネルデータ

5. .....

1　最小二乗法について

経済理論に基づいた線型モデルの係数の値をデータから求める時に用いられる手法 $\Longrightarrow$ 最小二乗法

1.1　最小二乗法と回帰直線

$(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ のように n 組のデータがあり，X_i と Y_i との間に以下の線型関係を想定する。

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

X_i は説明変数，Y_i は被説明変数，α, β はパラメータとそれぞれ呼ばれる。
上の式は回帰モデル (または，回帰式) と呼ばれる。目的は，切片 α と傾き β をデータ $\{(X_i, Y_i), i = 1, 2, \cdots, n\}$ から推定すること，

データについて：

1. タイム・シリーズ (時系列)・データ： i が時間を表す (第 i 期)。
2. クロス・セクション (横断面)・データ： i が個人や企業を表す (第 i 番目の家計，第 i 番目の企業)。

1.2　切片 α と傾き β の推定

次のような関数 $S(\alpha,\beta)$ を定義する。

$$S(\alpha,\beta) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

このとき,

$$\min_{\alpha,\beta} S(\alpha,\beta)$$

となるような α,β を求める (最小自乗法)。このときの解を $\widehat{\alpha},\widehat{\beta}$ とする。

最小化のためには,

$$\frac{\partial S(\alpha,\beta)}{\partial \alpha} = 0$$

$$\frac{\partial S(\alpha,\beta)}{\partial \beta} = 0$$

を満たす α,β が $\widehat{\alpha},\widehat{\beta}$ となる。　すなわち，$\widehat{\alpha},\widehat{\beta}$ は,

$$\sum_{i=1}^{n}(Y_i - \widehat{\alpha} - \widehat{\beta}X_i) = 0, \tag{1}$$

$$\sum_{i=1}^{n} X_i(Y_i - \widehat{\alpha} - \widehat{\beta}X_i) = 0, \tag{2}$$

を満たす。　さらに,

$$\sum_{i=1}^{n} Y_i = n\widehat{\alpha} + \widehat{\beta}\sum_{i=1}^{n} X_i, \tag{3}$$

$$\sum_{i=1}^{n} X_iY_i = \widehat{\alpha}\sum_{i=1}^{n} X_i + \widehat{\beta}\sum_{i=1}^{n} X_i^2,$$

行列表示によって，

$$\begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_iY_i \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix},$$

逆行列の公式：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$\widehat{\alpha}, \widehat{\beta}$ について，まとめて，

$$\begin{aligned} \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} &= \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_iY_i \end{pmatrix} \\ &= \frac{1}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \begin{pmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_iY_i \end{pmatrix} \end{aligned}$$

さらに，$\widehat{\beta}$ について解くと，

$$\widehat{\beta} = \frac{n\sum_{i=1}^n X_iY_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}$$

$$= \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

連立方程式の (3) 式から，

$$\widehat{\alpha} = \overline{Y} - \widehat{\beta X}$$

となる。ただし，

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

とする。

数値例：　以下の数値例を使って，回帰式 $Y_i = \alpha + \beta X_i$ の α，β の推定値 $\widehat{\alpha}$，$\widehat{\beta}$ を求める。

i	Y_i	X_i
1	6	10
2	9	12
3	10	14
4	10	16

$\widehat{\alpha}$，$\widehat{\beta}$ を求めるための公式は

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$
$$\widehat{\alpha} = \overline{Y} - \widehat{\beta\overline{X}}$$

なので，必要なものは $\overline{X}$，$\overline{Y}$，$\displaystyle\sum_{i=1}^{n} X_i^2$，$\displaystyle\sum_{i=1}^{n} X_i Y_i$ である。

i	Y_i	X_i	X_iY_i	X_i^2
1	6	10	60	100
2	9	12	108	144
3	10	14	140	196
4	10	16	160	256
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$
	35	52	468	696
平均	$\overline{Y}$	$\overline{X}$		
	8.75	13		

よって,

$$\widehat{\beta} = \frac{468 - 4 \times 13 \times 8.75}{696 - 4 \times 13^2} = \frac{13}{20} = 0.65$$
$$\widehat{\alpha} = 8.75 - 0.65 \times 13 = 0.3$$

となる。

注意事項：

1. α, β は真の値で未知
2. $\widehat{\alpha}, \widehat{\beta}$ は α, β の推定値でデータから計算される

回帰直線は

$$\widehat{Y_i} = \widehat{\alpha} + \widehat{\beta} X_i,$$

として与えられる。

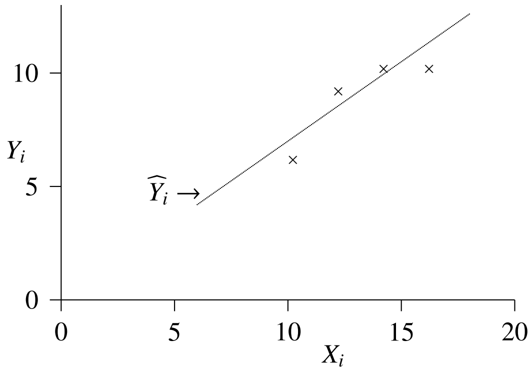
上の数値例では,

$$\widehat{Y_i} = 0.3 + 0.65 X_i$$

となる。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$
1	6	10	60	100	6.8
2	9	12	108	144	8.1
3	10	14	140	196	9.4
4	10	16	160	256	10.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$
	35	52	468	696	35.0
平均	$\overline{Y}$	$\overline{X}$			
	8.75	13			

図 2： Y_i， X_i， $\widehat{Y_i}$



$\widehat{Y_i}$ を実績値 Y_i の予測値または理論値と呼ぶ。

$$\widehat{u_i} = Y_i - \widehat{Y_i},$$

$\widehat{u}_i$ を残差と呼ぶ。

$$Y_i = \widehat{Y}_i + \widehat{u}_i = \widehat{\alpha} + \widehat{\beta}X_i + \widehat{u}_i,$$

さらに，$\overline{Y}$ を両辺から引いて，

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + \widehat{u}_i,$$

1.3　残差 $\widehat{u}_i$ の性質について

$\widehat{u}_i = Y_i - \widehat{\alpha} - \widehat{\beta}X_i$ に注意して，(1) 式から，

$$\sum_{i=1}^{n} \widehat{u}_i = 0,$$

を得る。　　　(2) 式から，

$$\sum_{i=1}^{n} X_i\widehat{u}_i = 0,$$

を得る。　$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} X_i$ から，

$$\sum_{i=1}^{n} \widehat{Y}_i \widehat{u}_i = 0,$$

を得る。なぜなら，

$$\begin{aligned}\sum_{i=1}^{n} \widehat{Y}_i \widehat{u}_i &= \sum_{i=1}^{n} (\widehat{\alpha} + \widehat{\beta} X_i) \widehat{u}_i \\ &= \widehat{\alpha} \sum_{i=1}^{n} \widehat{u}_i + \widehat{\beta} \sum_{i=1}^{n} X_i \widehat{u}_i \\ &= 0\end{aligned}$$

である。

i	Y_i	X_i	$\widehat{Y}_i$	$\widehat{u}_i$	$X_i\widehat{u}_i$	$\widehat{Y}_i\widehat{u}_i$
1	6	10	6.8	−0.8	−8.0	−5.44
2	9	12	8.1	0.9	10.8	7.29
3	10	14	9.4	0.6	8.4	5.64
4	10	16	10.7	−0.7	−11.2	−7.49
合計	$\sum Y_i$	$\sum X_i$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$	$\sum X_i\widehat{u}_i$	$\sum \widehat{Y}_i\widehat{u}_i$
	35	52	35.0	0.0	0.0	0.00

1.4 決定係数 R^2 について

次の式

$$(Y_i - \overline{Y}) = (\widehat{Y}_i - \overline{Y}) + \widehat{u}_i,$$

の両辺を二乗して，総和すると，

$$\begin{aligned}\sum_{i=1}^{n}(Y_i-\overline{Y})^2 &= \sum_{i=1}^{n}\Big((\widehat{Y_i}-\overline{Y})+\widehat{u_i}\Big)^2 \\ &= \sum_{i=1}^{n}(\widehat{Y_i}-\overline{Y})^2 + 2\sum_{i=1}^{n}(\widehat{Y_i}-\overline{Y})\widehat{u_i} + \sum_{i=1}^{n}\widehat{u_i^2} \\ &= \sum_{i=1}^{n}(\widehat{Y_i}-\overline{Y})^2 + \sum_{i=1}^{n}\widehat{u_i^2}\end{aligned}$$

となる。まとめると，

$$\sum_{i=1}^{n}(Y_i-\overline{Y})^2 = \sum_{i=1}^{n}(\widehat{Y_i}-\overline{Y})^2 + \sum_{i=1}^{n}\widehat{u_i^2}$$

を得る。さらに，

$$1 = \frac{\sum_{i=1}^{n}(\widehat{Y_i}-\overline{Y})^2}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2} + \frac{\sum_{i=1}^{n}\widehat{u_i^2}}{\sum_{i=1}^{n}(Y_i-\overline{Y})^2}$$

それぞれの項は，

1. $\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \Longrightarrow y$ の全変動

2. $\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 \Longrightarrow \widehat{Y}_i$ (回帰直線) で説明される部分

3. $\sum_{i=1}^{n}\widehat{u}_i^2 \Longrightarrow \widehat{Y}_i$ (回帰直線) で説明されない部分

となる。

回帰式の当てはまりの良さを示す指標として，決定係数 R^2 を以下の通りに定義する。

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

または,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2},$$

として書き換えられる。

または，$Y_i = \widehat{Y}_i + \widehat{u}_i$ と

$$
\begin{aligned}
\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y} - \widehat{u}_i) \\
&= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y}) - \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})\widehat{u}_i \\
&= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y})
\end{aligned}
$$

を用いて，

$$
\begin{aligned}
R^2 &= \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \\
&= \frac{\left(\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2\right)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2} \\
&= \left(\frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}}\right)^2
\end{aligned}
$$

と書き換えられる。　すなわち，R^2 は Y_i と $\widehat{Y}_i$ の相関係数の二乗と解釈される。

$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n}\widehat{u}_i^2$ から，明らかに，

$$0 \leq R^2 \leq 1,$$

となる。R^2 が 1 に近づけば回帰式の当てはまりは良いと言える。しかし，t 分布のような数表は存在しない。したがって,「どの値よりも大きくなるべき」というような基準はない。

慣習的には，メドとして 0.9 以上を判断基準にする。

数値例：　決定係数の計算には以下の公式を用いる。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n}\widehat{u}_i^2}{\sum_{i=1}^{n}Y_i^2 - n\overline{Y}^2}$$

計算に必要なものは，$\widehat{u}_i = Y_i - (\widehat{\alpha} + \widehat{\beta} X_i)$，$\overline{Y}$，$\sum_{i=1}^{n} Y_i^2$ である。

i	Y_i	X_i	$\widehat{Y}_i$	$\widehat{u}_i$	$\widehat{u}_i$	Y_i^2
1	6	10	6.8	−0.8	0.64	36
2	9	12	8.1	0.9	0.81	81
3	10	14	9.4	0.6	0.36	100
4	10	16	10.7	−0.7	0.49	100
合計	$\sum Y_i$	$\sum X_i$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$	$\sum \widehat{u}_i^2$	$\sum Y_i^2$
	35	52	35.0	0.0	2.30	317

$\sum \widehat{u}_i^2 = 2.30$，$\overline{X} = 13$，$\overline{Y} = 8.75$，$\sum_{i=1}^{n} Y_i^2 = 317$ なので，

$$R^2 = 1 - \frac{2.30}{317 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

1.5　まとめ

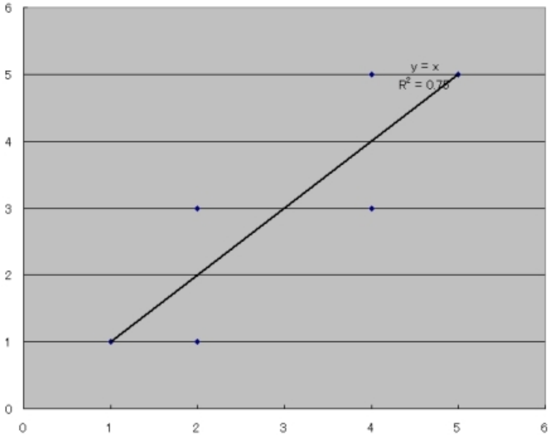
$\widehat{\alpha}$，$\widehat{\beta}$ を求めるための公式は

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{XY}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}$$
$$\widehat{\alpha} = \overline{Y} - \widehat{\beta X}$$

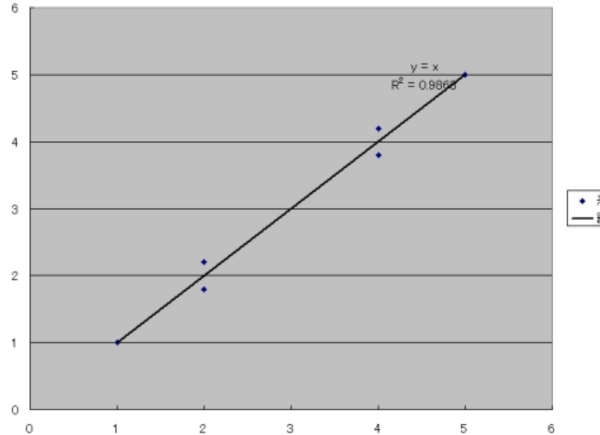
なので，必要なものは $\overline{X}$，$\overline{Y}$，$\displaystyle\sum_{i=1}^{n} X_i^2$，$\displaystyle\sum_{i=1}^{n} X_i Y_i$ である。
決定係数の計算には以下の公式を用いる。

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2}$$

計算に必要なものは，$\sum \widehat{u}_i^2$，$\overline{Y}$，$\displaystyle\sum_{i=1}^{n} Y_i^2$ である。



y = x
R2 = 0.75
0 1 2 3 4 5 6
0 1 2 3 4 5 6



y = x
R2 = 0.9868
0 1 2 3 4 5 6
0 1 2 3 4 5 6

2 Regression Analysis (回帰分析)

2.1 Setup of the Model

When (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available, suppose that there is a linear relationship between y and x, i.e.,

$$y_i = \beta_1 + \beta_2 x_i + u_i, \tag{4}$$

for $i = 1, 2, \cdots, n$. x_i and y_i denote the ith observations.

$\longrightarrow$ **Single (or simple) regression model (単回帰モデル)**

y_i is called the **dependent variable (従属変数)** or the **explained variable (被説明変数)**, while x_i is known as the **independent variable (独立変数)** or the **explanatory (or explaining) variable (説明変数)**.

β_1 = **Intercept (切片)**, β_2 = **Slope (傾き)**

β_1 and β_2 are unknown **parameters (パラメータ，母数)** to be estimated.

β_1 and β_2 are called the **regression coefficients (回帰係数)**.

u_i is the unobserved **error term (誤差項)** assumed to be a random variable with mean zero and variance σ^2.
σ^2 is also a parameter to be estimated.

x_i is assumed to be **nonstochastic (非確率的)**, but y_i is **stochastic (確率的)** because y_i depends on the error u_i.

The error terms $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed, which is called ***iid***.

It is assumed that u_i has a distribution with mean zero, i.e., $\mathrm{E}(u_i) = 0$ is assumed.

Taking the expectation on both sides of (4), the expectation of y_i is represented as:

$$\begin{aligned} \mathrm{E}(y_i) &= \mathrm{E}(\beta_1 + \beta_2 x_i + u_i) = \beta_1 + \beta_2 x_i + \mathrm{E}(u_i) \\ &= \beta_1 + \beta_2 x_i, \end{aligned} \tag{5}$$

for $i = 1, 2, \cdots, n$.

Using $\mathrm{E}(y_i)$ we can rewrite (4) as $y_i = \mathrm{E}(y_i) + u_i$.

(5) represents the true regression line.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be estimates of β_1 and β_2.

Replacing β_1 and β_2 by $\hat{\beta}_1$ and $\hat{\beta}_2$, (4) turns out to be:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + e_i, \tag{6}$$

for $i = 1, 2, \cdots, n$, where e_i is called the **residual (残差)**.

The residual e_i is taken as the experimental value (or realization) of u_i.

We define $\hat{y}_i$ as follows:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i, \tag{7}$$

for $i = 1, 2, \cdots, n$, which is interpreted as the **predicted value (予測値)** of y_i.

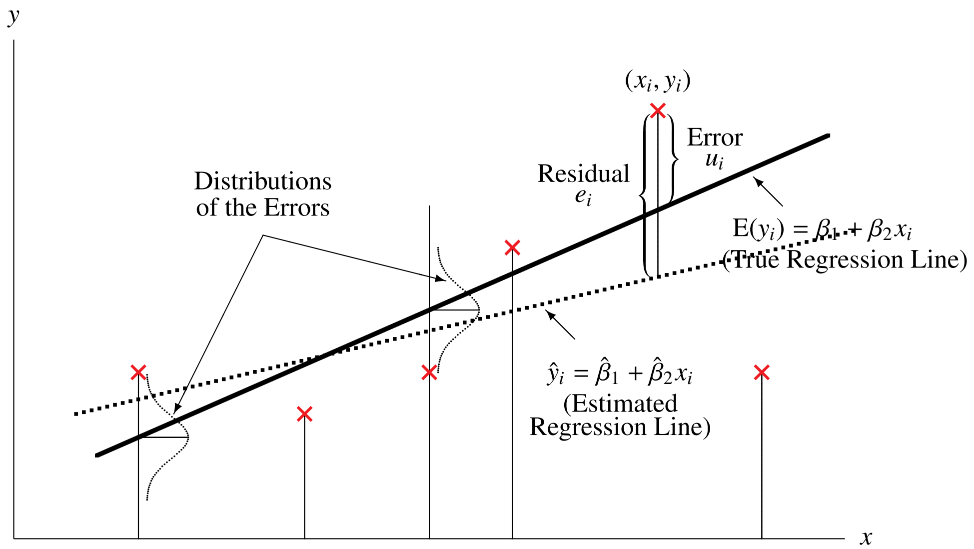
(7) indicates the estimated regression line, which is different from (5).

Moreover, using $\hat{y}_i$ we can rewrite (6) as $y_i = \hat{y}_i + e_i$.

(5) and (7) are displayed in Figure 1.

Consider the case of $n = 6$ for simplicity. $\times$ indicates the observed data series.

Figure 1. True and Estimated Regression Lines (回帰直線)



The true regression line (5) is represented by the solid line, while the estimated regression line (7) is drawn with the dotted line.

Based on the observed data, β_1 and β_2 are estimated as: $\hat{\beta}_1$ and $\hat{\beta}_2$.

In the next section, we consider how to obtain the estimates of β_1 and β_2, i.e., $\hat{\beta}_1$ and $\hat{\beta}_2$.

2.2 Ordinary Least Squares Estimation

Suppose that (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are available.
For the regression model (4), we consider estimating β_1 and β_2.
Replacing β_1 and β_2 by their estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, remember that the residual e_i is given by:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i.$$

The sum of squared residuals is defined as follows:

$$S(\hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2.$$

It might be plausible to choose the $\hat{\beta}_1$ and $\hat{\beta}_2$ which minimize the sum of squared residuals, i.e., $S(\hat{\beta}_1, \hat{\beta}_2)$.

This method is called the **ordinary least squares estimation (最小二乗法，OLS)**.

To minimize $S(\hat{\beta}_1, \hat{\beta}_2)$ with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$, we set the partial derivatives equal to zero:

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0,$$

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0.$$

The second order condition for minimization is:

$$\begin{pmatrix} \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_1^2} & \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_1 \partial \hat{\beta}_2} \\ \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_2 \partial \hat{\beta}_1} & \frac{\partial^2 S(\hat{\beta}_1,\hat{\beta}_2)}{\partial \hat{\beta}_2^2} \end{pmatrix} = \begin{pmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{pmatrix}$$

should be a positive definite matrix.

The diagonal elements $2n$ and $2\sum_{i=1}^n x_i^2$ are positive.

The determinant:

$$\begin{vmatrix} 2n & 2\sum_{i=1}^n x_i \\ 2\sum_{i=1}^n x_i & 2\sum_{i=1}^n x_i^2 \end{vmatrix} = 4n\sum_{i=1}^n x_i^2 - 4(\sum_{i=1}^n x_i)^2 = 4n\sum_{i=1}^n (x_i - \overline{x})^2$$

is positive. $\Longrightarrow$ The second-order condition is satisfied.

The first two equations yield the following two equations:

$$\overline{y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{x}, \tag{8}$$

$$\sum_{i=1}^n x_i y_i = n\overline{x}\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n x_i^2, \tag{9}$$

where $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$.

Multiplying (8) by $n\overline{x}$ and subtracting (9), we can derive $\hat{\beta}_2$ as follows:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{10}$$

From (8), $\hat{\beta}_1$ is directly obtained as follows:

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2\overline{x}. \tag{11}$$

When the observed values are taken for y_i and x_i for $i = 1, 2, \cdots, n$, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimate**s (or simply the **least squares estimate**s, 最小二乗推定値) of β_1 and β_2.

When y_i for $i = 1, 2, \cdots, n$ are regarded as the random sample, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimator**s (or the **least squares estimator**s, 最小二乗推定量) of β_1 and β_2.

2.3 Properties of Least Squares Estimator

Equation (10) is rewritten as:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (x_i - \overline{x}) y_i}{\sum_{i=1}^n (x_i - \overline{x})^2} - \frac{\overline{y} \sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$
$$= \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} y_i = \sum_{i=1}^n \omega_i y_i. \tag{12}$$

In the third equality, $\sum_{i=1}^n (x_i - \overline{x}) = 0$ is utilized because of $\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

In the fourth equality, ω_i is defined as: $\omega_i = \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}$.

ω_i is nonstochastic because x_i is assumed to be nonstochastic.

ω_i has the following properties:

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^n \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sum_{i=1}^n (x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2} = 0, \tag{13}$$

$$\sum_{i=1}^{n} \omega_i x_i = \sum_{i=1}^{n} \omega_i (x_i - \overline{x}) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = 1, \tag{14}$$

$$\sum_{i=1}^{n} \omega_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \right)^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2} = \frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{15}$$

The first equality of (14) comes from (13).

From now on, we focus only on $\hat{\beta}_2$, because usually β_2 is more important than β_1 in the regression model (4).

In order to obtain the properties of the least squares estimator $\hat{\beta}_2$, we rewrite (12) as:

$$\begin{aligned} \hat{\beta}_2 &= \sum_{i=1}^{n} \omega_i y_i = \sum_{i=1}^{n} \omega_i (\beta_1 + \beta_2 x_i + u_i) \\ &= \beta_1 \sum_{i=1}^{n} \omega_i + \beta_2 \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n} \omega_i u_i = \beta_2 + \sum_{i=1}^{n} \omega_i u_i. \end{aligned} \tag{16}$$

In the fourth equality of (16), (13) and (14) are utilized.

[Review] Random Variables:

Let X_1, X_2, $\cdots$, X_n be n random variavles, which are mutually independently and identically distributed.

mutually independent $\Longrightarrow$ $f(x_i, x_j) = f_i(x_i) f_j(x_j)$ for $i \neq j$.

$f(x_i, x_j)$ denotes a joint distribution of X_i and X_j.

$f_i(x)$ indicates a marginal distribution of X_i.

identical $\Longrightarrow$ $f_i(x) = f_j(x)$ for $i \neq j$.

[End of Review]

[Review] Mean and Variance:

Let X and Y be random variables (continuous type), which are independently distributed.

Definition and Formulas:

- $\mathrm{E}(g(X)) = \displaystyle\int g(x)f(x)\mathrm{d}x$ for a function $g(\cdot)$ and a density function $f(\cdot)$.
- $\mathrm{V}(X) = \mathrm{E}((X-\mu)^2) = \displaystyle\int (x-\mu)^2 f(x)\mathrm{d}x$ for $\mu = \mathrm{E}(X)$.
- $\mathrm{E}(aX+b) = a\mathrm{E}(X)+b$ and $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$.
- $\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$ and $\mathrm{V}(X \pm Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

[End of Review]

Mean and Variance of $\hat{\beta}_2$: $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2, but they are not necessarily normal.

Remember that we do not need normality assumption to obtain mean and variance but the normality assumption is required to test a hypothesis.

From (16), the expectation of $\hat{\beta}_2$ is derived as follows:

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \mathrm{E}(\sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) = \beta_2. \qquad (17)$$

It is shown from (17) that the ordinary least squares estimator $\hat{\beta}_2$ is an **unbiased estimator (不偏推定量)** of β_2.

From (16), the variance of $\hat{\beta}_2$ is computed as:

$$\begin{aligned} \mathrm{V}(\hat{\beta}_2) &= \mathrm{V}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sum_{i=1}^{n} \mathrm{V}(\omega_i u_i) = \sum_{i=1}^{n} \omega_i^2 \mathrm{V}(u_i) \\ &= \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}. \end{aligned} \tag{18}$$

The third equality holds because $u_1, u_2, \cdots, u_n$ are mutually independent.

The last equality comes from (15).

Thus, $\mathrm{E}(\hat{\beta}_2)$ and $\mathrm{V}(\hat{\beta}_2)$ are given by (17) and (18).

Gauss-Markov Theorem (ガウス・マルコフ定理): $\hat{\beta}_2$ has minimum variance within a class of the linear unbiased estimators.

$\longrightarrow$ **best linear unbiased estimator (BLUE, 最良線型不偏推定量)**

(Proof is omitted.)

Distribution of $\hat{\beta}_2$: We discuss the small sample properties of $\hat{\beta}_2$.

In order to obtain the distribution of $\hat{\beta}_2$ in small sample, the distribution of the error term has to be assumed.

Therefore, the extra assumption is that $u_i \sim N(0, \sigma^2)$.

Writing (16), again, $\hat{\beta}_2$ is represented as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i.$$

First, we obtain the distribution of the second term in the above equation.

It is well known that sum of normal random variables results in a normal distribution.

Therefore, $\sum_{i=1}^{n} \omega_i u_i$ is distributed as:

$$\sum_{i=1}^{n} \omega_i u_i \sim N(0,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2).$$

Therefore, $\hat{\beta}_2$ is distributed as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i \sim N(\beta_2,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2),$$

or equivalently,

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{\sum_{i=1}^{n} \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim N(0, 1),$$

for any n.

Moreover, replacing σ^2 by its estimator $s^2 = \dfrac{1}{n-2} \displaystyle\sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$, it is known that we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim t(n-2),$$

where $t(n-2)$ denotes t distribution with $n-2$ degrees of freedom.

Thus, under normality assumption on the error term u_i, the $t(n-2)$ distribution is used for the confidence interval and the testing hypothesis in small sample.

Or, taking the square on both sides,

$$\left(\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right)^2 \sim F(1, n-2).$$

[Review] Confidence Interval (信頼区間，区間推定)):

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we can obtain: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

That is,

$$P\Big(-t_{\alpha/2}(n-1) < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\Big) = 1 - \alpha$$

i.e.,

$$P\Big(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\Big) = 1 - \alpha.$$

Note that $t_{\alpha/2}(n-1)$ is obtained from the t distribution table, given α and $n-1$.

Then, replacing $\overline{X}$ by $\overline{x}$, we obtain the $100(1-\alpha)\%$ confidence interval of μ as follows:

$$(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}).$$

[End of Review]

In the case of OLS,

$$P\Big(-t_{\alpha/2}(n-2) < \frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} < t_{\alpha/2}(n-2)\Big) = 1 - \alpha,$$

where $t_{\alpha/2}(n-2)$ denotes $100 \times \alpha/2\%$ point from the $t(n-2)$ distribution.

Rewriting,

$$P\Big(\hat{\beta}_2 - t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} < \beta_2 < \hat{\beta}_2 + t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\Big) = 1 - \alpha.$$

Replacing $\hat{\beta}_2$ and s^2 by observed data, the $100(1-\alpha)\%$ confidence interval of β_2 is given by:

$$\Big(\hat{\beta}_2 - t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}},\ \hat{\beta}_2 + t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\Big).$$

[Review] Testing the Hypothesis (仮説検定):

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we obtain: $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, which is known as the unbiased estimator of σ^2.

- The null hypothesis $H_0: \mu = \mu_0$, where μ_0 is a fixed number.
- The alternative hypothesis $H_1: \mu \neq \mu_0$

Under the null hypothesis, we have the disribution: $\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$.

Replacing $\overline{X}$ and S^2 by $\overline{x}$ and s^2, compare $\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$ and $t(n-1)$.

H_0 is rejected when $\left|\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| > t_{\alpha/2}(n-1)$.

$t_{\alpha/2}(n-1)$ is obtained from the significance level α and the degrees of freedom $n-1$.

[End of Review]

In the case of OLS, the hypotheses are as follows:

- The null hypothesis $H_0 : \beta_2 = \beta_2^*$
- The alternative hypothesis $H_1 : \beta_2 \neq \beta_2^*$

Under H_0,

$$\frac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} \sim t(n-2).$$

Replacing $\hat{\beta}_2$ and s^2 by the observed data, compare $\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ and $t(n-2)$.

H_0 is rejected at significance level α when $\left|\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right| > t_{\alpha/2}(n-1)$.

(*) $\hat{\beta}_2$ = Coefficient, $\dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ = Standard Error,

s = Standard Error of Regression

3　多重回帰

n 組のデータ $(Y_i, X_{1i}, X_{2i}, \cdots, X_{ki})$, $i = 1, 2, \cdots, n$ を用いて，k 変数の多重回帰モデルを考える。

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

ただし，X_{ji} は j 番目の説明変数の第 i 番目の観測値を表す。u_i は誤差項 (または，攪乱項) で，同じ仮定を用いる (すなわち，$u_1, u_2, \cdots, u_n$ は互いに独立に，平均ゼロ，分散 σ^2 の正規分布に従う)。

$\beta_1, \beta_2, \cdots, \beta_k$ は推定されるべきパラメータである。

すべての i について，$X_{1i} = 1$ とすれば，β_1 は定数項として表される。

次のような関数 $S(\beta_1, \beta_2, \cdots, \beta_k)$ を定義する。

$$S(\beta_1, \beta_2, \cdots, \beta_k) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2$$

このとき，

$$\min_{\beta_1,\beta_2,\cdots,\beta_k} S(\beta_1, \beta_2, \cdots, \beta_k)$$

となるような $\beta_1, \beta_2, \cdots, \beta_k$ を求める。$\Longrightarrow$ 最小自乗法

このときの解を $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

最小化のためには，

$$\frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_1} = 0, \quad \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_2} = 0, \quad \cdots, \quad \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_k} = 0$$

を満たす $\beta_1, \beta_2, \cdots, \beta_k$ が $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ となる。

すなわち，$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ は，

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}) X_{1i} = 0,$$

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}) X_{2i} = 0,$$

$$\vdots$$

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})X_{ki} = 0,$$

を満たす。

さらに,

$$\sum_{i=1}^{n} X_{1i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} X_{1i}X_{2i} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{1i}X_{ki},$$

$$\sum_{i=1}^{n} X_{2i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{2i}X_{ki},$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ki}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{ki} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}X_{ki} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{ki}^2,$$

行列表示によって，

$$\begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix},$$

が得られ，$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ についてまとめると，

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix},$$

を解くことになる。$\Longrightarrow$ コンピュータによって計算

3.1 推定量の性質

$\beta_1, \beta_2, \cdots, \beta_k$ の最小二乗推定量は $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

誤差項 (または，攪乱項) u_i の分散 σ^2 の推定量 s^2 は,

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})^2$$

として表される。

このとき,

$$\mathrm{E}(\widehat{\beta}_j) = \beta_j, \qquad \mathrm{E}(s^2) = \sigma^2,$$

を証明することが出来る。(証明略)

分布について： $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ の分散は以下のように表される。

$$
\mathrm{V}\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} = \begin{pmatrix} \mathrm{V}(\widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_k) \\ \mathrm{Cov}(\widehat{\beta}_2, \widehat{\beta}_1) & \mathrm{V}(\widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_2, \widehat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_k, \widehat{\beta}_2) & \cdots & \mathrm{V}(\widehat{\beta}_k) \end{pmatrix}
$$

$$
= \sigma^2 \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1}
$$

$\widehat{\beta}_j$ の分散 (すなわち，上の逆行列の j 番目の対角要素) を，

$$
\mathrm{V}(\widehat{\beta}_j) = \sigma^2_{\widehat{\beta}_j},
$$

として，その推定量を $s^2_{\widehat{\beta}_j}$ とする。

このとき，

$$\widehat{\beta}_j \sim N(\beta_j, \sigma^2_{\widehat{\beta}_j}),$$

となり，標準化すると，

$$\frac{\widehat{\beta}_j - \beta_j}{\sigma_{\widehat{\beta}_j}} \sim N(0, 1),$$

が得られる。さらに，

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k),$$

となり (証明略)，しかも，$\widehat{\beta}_j$ と s^2 の独立性から (証明略)，

$$\frac{\widehat{\beta}_j - \beta_j}{s_{\widehat{\beta}_j}} \sim t(n-k)$$

となる。

よって，通常の区間推定や仮説検定を行うことが出来る。

決定係数について：　また，決定係数 R^2 についても同様に表される。

$$R^2 = \frac{\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^n \widehat{u}_i^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2}$$

ただし，$\widehat{Y}_i = \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki}$，$Y_i = \widehat{Y}_i + \widehat{u}_i$ である。

R^2 は，説明変数を増やすことによって，必ず大きくなる。なぜなら，説明変数が増えることによって，$\sum_{i=1}^n \widehat{u}_i^2$ が必ず減少するからである。

R^2 を基準にすると，被説明変数にとって意味のない変数でも，説明変数が多いほど，よりよいモデルということになる。この点を改善するために，自由度修正済み決定係数 $\overline{R}^2$ を用いる。

$$\overline{R}^2 = 1 - \frac{\sum_{i=1}^n \widehat{u}_i^2/(n-k)}{\sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)},$$

$\sum_{i=1}^n \widehat{u}_i^2/(n-k)$ は u_i の分散 σ^2 の不偏推定量であり，$\sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)$ は Y_i の分散の不偏推定量である。

R^2 と $\overline{R}^2$ との関係は，

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k},$$

となる。さらに，

$$\frac{1-\overline{R}^2}{1-R^2} = \frac{n-1}{n-k} \geq 1,$$

という関係から，$\overline{R}^2 \leq R^2$ という結果を得る。($k = 1$ のときのみに，等号が成り立つ。)

数値例：　今までと同じ数値例で，$\overline{R}^2$ を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

まず R^2 は,

$$R^2 = 1 - \frac{\sum \widehat{u}_i^2}{\sum Y_i^2 - n\overline{Y}^2} = 1 - \frac{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2}{35 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

となり，$\overline{R}^2$ は，

$$\overline{R}^2 = 1 - \frac{\sum \widehat{u}_i^2/(n-k)}{(\sum Y_i^2 - n\overline{Y}^2)/(n-1)} = 1 - \frac{2.30/(4-2)}{10.75/(4-1)} = 0.679$$

となる。

注意： R^2 や $\overline{R}^2$ を比較する場合，被説明変数が同じことが必要である。被説明変数が異なる場合 (例えば，被説明変数を上昇率とするかそのままの値を用いるかによって，被説明変数が異なる)，誤差項 u_i の標準誤差で比較すべきである (標準誤差の小さいモデルを採用する)。 $\Longrightarrow$ 関数型の選択

4　系列相関：*DW* について

4.1　*DW* について

最小自乗法の仮定の一つに,「攪乱項 $u_1, u_2, \cdots, u_n$ 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_i と u_{i-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

$u_1, u_2, \cdots, u_n$ 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, $u_1, u_2, \cdots, u_n$ 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，$u_1, u_2, \cdots, u_n$ 負の系列相関があると言う。

特徴： $u_1, u_2, \cdots, u_i$ から u_{i+1} の符号が予想できる。$\Longrightarrow$ 「$u_1, u_2, \cdots, u_n$ 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

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

$$u_i = \rho u_{i-1} + \epsilon_i,$$

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

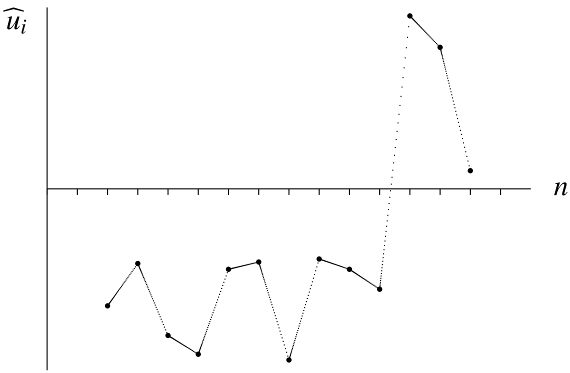


図 4： 正の系列相関

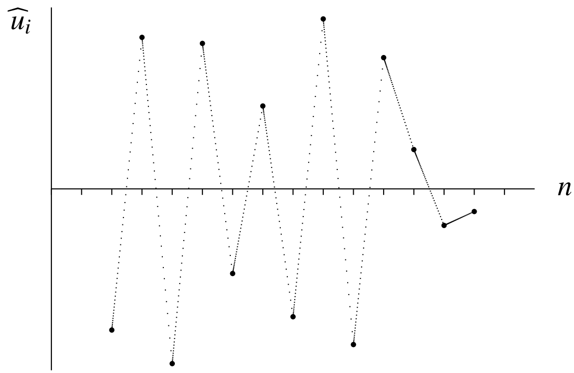


図 5： 負の系列相関

ダービン・ワトソン比の定義は次の通りである。

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u_{i-1}})^2}{\sum_{i=1}^{n}\widehat{u_i^2}}$$

DW は近似的に，次のように表される。

$$\begin{aligned} DW &= \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u_{i-1}})^2}{\sum_{i=1}^{n}\widehat{u_i^2}} = \frac{\sum_{i=2}^{n}\widehat{u_i^2} - 2\sum_{i=2}^{n}\widehat{u_i}\widehat{u_{i-1}} + \sum_{i=2}^{n}\widehat{u_{i-1}^2}}{\sum_{i=1}^{n}\widehat{u_i^2}} \\ &= \frac{2\sum_{i=1}^{n}\widehat{u_i^2} - (\widehat{u_1^2} + \widehat{u_n^2})}{\sum_{i=1}^{n}\widehat{u_i^2}} - 2\frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u_{i-1}}}{\sum_{i=1}^{n}\widehat{u_i^2}} \approx 2(1-\widehat{\rho}), \end{aligned}$$

以下の 2 つの近似が用いられる。

$$\frac{\widehat{u_1^2} + \widehat{u_n^2}}{\sum_{i=1}^{n}\widehat{u_i^2}} \approx 0,$$

$$\frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u_{i-1}}}{\sum_{i=1}^{n}\widehat{u_i^2}} = \frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u_{i-1}}}{\sum_{i=2}^{n}\widehat{u_{i-1}^2} + \widehat{u_n^2}} \approx \frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u_{i-1}}}{\sum_{i=2}^{n}\widehat{u_{i-1}^2}} = \widehat{\rho},$$

すなわち，$\widehat{\rho}$ は $\widehat{u_i}$ と $\widehat{u_{i-1}}$ の回帰係数である。$u_i = \rho u_{i-1} + \epsilon_i$ において，u_i, u_{i-1} の代わりに $\widehat{u_i}, \widehat{u_{i-1}}$ 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 n とパラメータ数 k に依存する。表 1 を参照せよ。

k' は定数項を除くパラメータ数を表すものとする。

See

`http://www.stanford.edu/~clint/bench/dwcrit.htm`

for the DW table.

Table 1: ダービン・ワトソン統計量の 5 % 点の上限と下限

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$		$k'=6$		$k'=7$		$k'=8$		$k'=9$		$k'=10$		$k'=11$		$k'=12$		$k'=13$	
n	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du
6	0.610	1.400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.700	1.356	0.467	1.896	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.763	1.332	0.559	1.777	0.367	2.287	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	—	—	—	—	—	—	—	—	—	—	—	—	—	—
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	—	—	—	—	—	—	—	—	—	—	—	—
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	—	—	—	—	—	—	—	—	—	—
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	—	—	—	—	—	—	—	—
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438	—	—	—	—	—	—
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304	0.098	3.503	—	—	—	—
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184	0.138	3.378	0.087	3.557	—	—
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073	0.177	3.265	0.123	3.441	0.078	3.603
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974	0.220	3.159	0.160	3.335	0.111	3.496
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885	0.263	3.063	0.200	3.234	0.145	3.395
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806	0.307	2.976	0.240	3.141	0.182	3.300
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735	0.349	2.897	0.281	3.057	0.220	3.211
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670	0.391	2.826	0.322	2.979	0.259	3.129
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613	0.431	2.761	0.362	2.908	0.297	3.053
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560	0.470	2.702	0.400	2.844	0.335	2.983
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513	0.508	2.649	0.438	2.784	0.373	2.919
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470	0.544	2.600	0.475	2.730	0.409	2.860
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431	0.578	2.555	0.510	2.680	0.445	2.805
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.681	2.396	0.612	2.515	0.544	2.634	0.479	2.754
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363	0.643	2.477	0.577	2.592	0.513	2.708
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333	0.674	2.443	0.608	2.553	0.545	2.665
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306	0.703	2.411	0.638	2.518	0.576	2.625
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.796	2.281	0.731	2.382	0.667	2.484	0.606	2.588
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.079	1.891	1.015	1.978	0.950	2.069	0.885	2.162	0.821	2.257	0.758	2.355	0.695	2.454	0.634	2.553
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236	0.783	2.330	0.722	2.425	0.662	2.521
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.876	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216	0.808	2.306	0.748	2.398	0.689	2.492
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.197	0.831	2.285	0.772	2.374	0.714	2.464
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180	0.854	2.265	0.796	2.351	0.739	2.438
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164	0.875	2.246	0.819	2.329	0.763	2.413
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.150	0.896	2.228	0.840	2.309	0.785	2.391
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088	0.988	2.156	0.938	2.225	0.887	2.296
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044	1.064	2.103	1.019	2.163	0.973	2.225
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010	1.129	2.062	1.087	2.116	1.045	2.170
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984	1.184	2.031	1.145	2.079	1.106	2.127
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964	1.231	2.006	1.195	2.049	1.160	2.093
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.838	1.369	1.874	1.337	1.910	1.305	1.948	1.272	1.987	1.239	2.026	1.206	2.066
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935	1.308	1.970	1.277	2.006	1.247	2.043
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925	1.340	1.957	1.312	1.990	1.283	2.024
85	1.623	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916	1.369	1.946	1.342	1.977	1.315	2.008
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909	1.395	1.937	1.369	1.966	1.344	1.995
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903	1.418	1.930	1.394	1.956	1.370	1.984
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898	1.439	1.923	1.416	1.948	1.393	1.974
150	1.720	1.747	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.846	1.608	1.862	1.593	1.877	1.579	1.892	1.564	1.908	1.549	1.924
200	1.758	1.779	1.748	1.789	1.738	1.799	1.728	1.809	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874	1.654	1.885	1.643	1.897	1.632	1.908

n は標本数，k' は定数項を除く説明変数の数とする。

（出所）http://www.stanford.edu/~clint/bench/dwcrit.htm

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \approx 2(1-\widehat{\rho}) \longrightarrow 2(1-\rho)$$

$-1 < \rho < 1$ なので (証明略)，近似的に $0 \le DW \le 4$ となる。

- $0 \le DW \le dl \qquad \longrightarrow$ u_i に正の系列相関
- $dl \le DW \le du \qquad \longrightarrow$ u_i に正の系列相関と判定できない
- $du \le DW \le 4-du \qquad \longrightarrow$ u_i に系列相関なし
- $4-du \le DW \le 4-dl \qquad \longrightarrow$ u_i に負の系列相関と判定できない
- $4-dl \le DW \le 4 \qquad \longrightarrow$ u_i に負の系列相関

数値例：　今までと同じ数値例で，DW を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2}$$

$$= \frac{(-0.8-0.9)^2 + (0.9-0.6)^2 + (0.6-(-0.7))^2}{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2} = \frac{4.67}{2.30} = 2.03$$

推定結果の表記方法：　回帰モデル：

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

の推定の結果，$\widehat{\alpha} = 0.3, \widehat{\beta} = 0.65$, $s_{\widehat{\alpha}} = \sqrt{10.0005} = 3.163$, $s_{\widehat{\beta}} = \sqrt{0.0575} = 0.240$, $\dfrac{\widehat{\alpha}}{s_{\widehat{\alpha}}} = 0.095$, $\dfrac{\widehat{\beta}}{s_{\widehat{\beta}}} = 2.708$, $s^2 = 1.15$ (すなわち，$s = 1.07$), $R^2 = 0.786$, $\overline{R}^2 = 0.679$, $DW = 2.03$ を得た。

これらをまとめて，

$$Y_i = \underset{(0.095)}{0.3} + \underset{(2.708)}{0.65} X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は t 値を表すものとする。

または,

$$Y_i = \underset{(3.163)}{0.3} + \underset{(0.240)}{0.65} X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は標準誤差を表すものとする。

のように書く。$s = \sqrt{1.15} = 1.07$ に注意。

4.2 系列相関のもとで回帰式の推定

回帰式が

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

$$u_i = \rho u_{i-1} + \epsilon_i,$$

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

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \alpha(1-\rho) + \beta(X_i - \rho X_{i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_i^* = (X_i - \rho X_{i-1})$

を新たな変数として，

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

に最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 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

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

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

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \beta_1(X_{1i} - \rho X_{1,i-1}) + \beta_2(X_{1i} - \rho X_{2,i-1}) + \cdots + \beta_k(X_{1i} - \rho X_{k,i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho X_{2,i-1}),\ \cdots,\ X_{ki}^* = (X_{ki} - \rho X_{k,i-1})$
を新たな変数として，

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i$$

最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 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 (その 1): 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，

$Y_i^* = (Y_i - \widehat{\rho} Y_{i-1})$, $X_{1i}^* = (X_{1i} - \widehat{\rho} X_{1,i-1})$, $X_{2i}^* = (X_{2i} - \widehat{\rho} X_{2,i-1})$, $\cdots$, $X_{ki}^* = (X_{ki} - \widehat{\rho} X_{k,i-1})$ を新たな変数として,

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i,$$

に最小二乗法を適用する。

ρ の求め方について (その 2): 収束計算によって求める。 $\longrightarrow$ コクラン・オーカット法

1. $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, 2, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\beta}_1, \cdots, \widehat{\beta}_k, \widehat{u}_i$ を得る。

2. $\widehat{u}_i = \rho \widehat{u}_{i-1} + \epsilon_i, \quad i = 2, 3, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\rho}$ を得る。

3. $\rho^{(m-1)} = \widehat{\rho}$ とおく。

4. $Y_i^* = (Y_i - \rho^{(m-1)} Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho^{(m-1)} X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho^{(m-1)} X_{2,i-1}),\ \cdots,$ $X_{ki}^* = (X_{ki} - \rho^{(m-1)} X_{k,i-1})$ を計算する。

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i, \quad i = 2, 3, \cdots, n$$

を最小二乗法で推定する。$\longrightarrow \widehat{\beta}_1, \cdots, \widehat{\beta}_k$ を得る。

5. $\widehat{u}_i = Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}, \quad i = 1, 2, \cdots, n$

を計算する。

6. ステップ 2 に戻り，$m = 1, 2, \cdots$ について繰り返す。

収束先を $\beta_1, \beta_2, \cdots, \beta_k, \rho$ の推定値とする。

5　不均一分散 (不等分散)

回帰式が

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

の場合を考える。X_i が外生変数，Y_i は内生変数，u_i 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 $u_1, u_2, \cdots, u_n$ 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_i) に依存する場合，すなわち，u_i の平均はゼロ，分散は $\sigma_*^2 z_i^2$ の場合は，最小二乗法の仮定に反する。そのため，単純には，$Y_i = \alpha + \beta X_i + u_i$ 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。

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

このとき，新たな攪乱項 u_i^* は平均ゼロ，分散 σ_*^2 の分布となる (すなわち，「同

一の」分布)。

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

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

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

u_i の仮定 $\mathrm{V}(u_i) = \sigma_*^2 z_i^2$ が最後に使われている。
よって，$\dfrac{Y_i}{z_i}, \dfrac{1}{z_i}, \dfrac{X_i}{z_i}$ 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

$$\widehat{u_i^2} = \gamma z_i + \epsilon_i$$

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_i は回帰式に含まれる変数でもよい。例えば，u_i の平均はゼロ，分散は $\sigma_*^2 X_i^2$ の

場合，各変数を X_i で割って，

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

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。

6　最尤法について

標本 $X_1, X_2, \cdots, X_n$ の密度関数：

$$f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

θ は未知母数 $\Longrightarrow \widehat{\theta}_n(x_1, x_2, \cdots, x_n)$ によって推定

$$l(\theta) = l(\theta; x) = l(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$$

のように，θ の関数として考える。

$l(\theta)$： 尤度関数

尤度関数を最大にする θ を $\widehat{\theta}_n$ とする。

$\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, X_2, \cdots, X_n) \Longrightarrow$ 最尤推定量

$\widehat{\theta}_n(x_1, x_2, \cdots, x_n) \Longrightarrow$ 最尤推定値

すなわち，

$$\frac{\partial l(\theta)}{\partial \theta} = 0$$

を解くことによって，最尤推定量 $\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, X_2, \cdots, X_n)$ が得られる。

最尤推定量の性質：

小標本について (n が小さいとき)：

- 一般に，最尤推定量は不偏性を持っていないが，適当な変換によって，不偏推定量を作ることが出来る場合が多い。

- 有効推定量が存在すれば (すなわち，クラメール・ラオの不等式の等号を満たすような推定量が存在するならば)，最尤推定量は有効推定量に一致する。

- 十分統計量が存在すれば，最尤推定量は十分統計量の関数となる。

大標本について (n が大きいとき)：

$n \longrightarrow \infty$ のとき,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \longrightarrow N\Big(0, \sigma^2\Big)$$

となる。$\Longrightarrow$ 一致性，漸近的正規性，漸近的有効性

ただし,

$$\sigma^2 = \sigma^2(\theta) = \frac{1}{\mathrm{E}\left[\left(\dfrac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right]}$$

すなわち，$n \longrightarrow \infty$ のとき,

$$\frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sigma(\theta)} = \frac{\widehat{\theta}_n - \theta}{\sigma(\theta)/\sqrt{n}} \longrightarrow N(0, 1)$$

となる。

したがって，厳密ではないが，n が大きいとき,

$$\widehat{\theta}_n \sim N\left(\theta, \frac{\sigma^2(\theta)}{n}\right)$$

と近似できる。

すなわち，$n \longrightarrow \infty$ のとき，$\widehat{\theta}_n$ の分散はクラメール・ラオの不等式の下限 $\dfrac{\sigma^2(\theta)}{n}$ に近づくことを意味する。

$\Longrightarrow$ 漸近的に有効推定量

さらに，分母の θ を最尤推定量 $\widehat{\theta}_n$ で置き換えて，$n \longrightarrow \infty$ のとき，

$$\frac{\widehat{\theta}_n - \theta}{\sigma(\widehat{\theta}_n)/\sqrt{n}} \longrightarrow N(0, 1)$$

となる。

実際には，n が大きいとき，

$$\widehat{\theta}_n \sim N\left(\theta, \frac{\sigma^2(\widehat{\theta}_n)}{n}\right)$$

と近似して用いる。

例：

$X_1, X_2, \cdots, X_n$ は互いに独立で，すべて同一の正規分布 (すなわち，平均 μ, 分散 σ^2 ですべて同一の分布) に従うものとする。μ, σ^2 の最尤推定量を求める。

$$f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$
$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) = l(\mu, \sigma^2)$$

対数をとる。(最大化しやすくなる場合が多い)

$$\log l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

対数尤度関数 $\log l(\mu, \sigma^2)$ を μ と σ^2 について微分して，ゼロと置く。

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

この 2 つの連立方程式を解く。

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

μ, σ^2 の最尤推定量は,

$$\overline{X}, \quad S^{**2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

となる。

$\mathrm{E}(\overline{X}) = \mu$ なので，μ の最尤推定量 $\overline{X}$ は不偏推定量である。

$\mathrm{E}(S^{**2}) = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$ なので，σ^2 の最尤推定量 S^{**2} は不偏推定量でない。

例：

$X_1, X_2, \cdots, X_n$ は互いに独立で，すべて同一のベルヌイ分布ですべて同一の分布) に従うものとする。すなわち，X の確率関数は $P(X = x) = f(x; p) = p^x(1 - p)^{1-p}$, $x = 0, 1$, となる。p の最尤推定量を求める。

$$f(x_1, x_2, \cdots, x_n; p) = \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i} = l(p)$$

対数をとる。

$$\log l(p) = (\sum_i x_i)\log(p) + (n - \sum_i x_i)\log(1-p)$$

対数尤度関数 $\log l(p)$ を p について微分して，ゼロと置く。

$$\frac{\partial \log l(p)}{\partial p} = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} = \frac{\sum_i x_i - np}{p(1-p)} = 0$$

この方程式を解く。

$$p = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

p の最尤推定量は,

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

となる。

$\mathrm{E}(\overline{X}) = p$ なので，p の最尤推定量 $\overline{X}$ は不偏推定量である。

X がベルヌイ分布 $f(x; p)$ のとき，$\mathrm{E}(X) = p$ に注意。

例：

$X_1, X_2, \cdots, X_n$ は互いに独立で，すべて同一のポアソン分布 (すなわち，平均 λ ですべて同一の分布) に従うものとする。λ の最尤推定量を求める。

ポアソン分布の確率関数は,

$$P(X = x) = f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \cdots$$

なので，尤度関数は,

$$l(\lambda) = \prod_{i=1}^{n} f(x_i; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

対数尤度関数は,

$$\log l(\lambda) = \log(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \log(\prod_{i=1}^{n} x_i!)$$

となる。

$$\frac{\partial \log l(\lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0$$

これを解いて，λ の最尤推定量 $\widehat{\lambda}$ は,

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

となる。

$\widehat{\lambda}$ は，λ の不偏推定量，有効推定量，十分推定量，一致推定量である。

証明：

X がパラメータ λ のポアソン分布に従うとき,

$$\mathrm{E}(X) = \mathrm{V}(X) = \lambda$$

となる。

不偏性：

$$\mathrm{E}(\widehat{\lambda}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\lambda = \lambda$$

有効性：

$$\mathrm{V}(\widehat{\lambda}) = \mathrm{V}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\lambda = \frac{\lambda}{n}$$

$$\begin{aligned}\frac{1}{n\mathrm{E}\left[\left(\frac{\partial \log f(X;\lambda)}{\partial\lambda}\right)^2\right]} &= \frac{1}{n\mathrm{E}\left[\left(\frac{\partial(X\log\lambda - \lambda - \log X!)}{\partial\lambda}\right)^2\right]} = \frac{1}{n\mathrm{E}\left[\left(\frac{X}{\lambda} - 1\right)^2\right]} \\ &= \frac{\lambda^2}{n\mathrm{E}[(X-\lambda)^2]} = \frac{\lambda^2}{n\mathrm{V}(X)} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n}\end{aligned}$$

したがって,

$$\mathrm{V}(\widehat{\lambda}) = \frac{1}{n\mathrm{E}\left[\left(\frac{\partial \log f(X;\lambda)}{\partial\lambda}\right)^2\right]}$$

となり，$\mathrm{V}(\widehat{\lambda})$ は，クラメール・ラオの下限に一致する。よって，$\widehat{\lambda}$ は有効推定量である。

十分性：

$$\prod_{i=1}^{n} f(x_i;\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} = \frac{\lambda^{n\overline{x}} e^{-n\lambda}}{(n\overline{x})!} \frac{(n\overline{x})!}{\prod_{i=1}^{n} x_i!} = g(\overline{x};\lambda)\, h(x_1, x_2, \cdots, x_n)$$

と分解できる。

一致性：

$$\mathrm{E}(\overline{X}) = \lambda, \qquad \mathrm{V}(\overline{X}) = \frac{\lambda}{n}$$

なので，チェビシェフの不等式に当てはめる。

$$P(|\overline{X} - \lambda| > \epsilon) < \frac{\lambda}{n\epsilon^2} \longrightarrow \infty$$

したがって，一致性も成り立つ。

6.1 最尤法の例：AR(1) モデル

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

1. Mean of y_t given $y_{t-1}, y_{t-2}, \cdots$

$$\mathrm{E}(y_t | y_{t-1}, y_{t-2}, \cdots) = \phi y_{t-1}$$

2. Variance of y_t given $y_{t-1}, y_{t-2}, \cdots$

$$\mathrm{V}(y_t | y_{t-1}, y_{t-2}, \cdots) = \sigma^2$$

3. Thus, $y_t | y_{t-1}, y_{t-2}, \cdots \sim N(0, \sigma^2)$. $\Longrightarrow$ Conditional distribution of y_t given $y_{t-1}, y_{t-2}, \cdots$

4. The stationarity condition is: the solution of $\phi(x) = 1 - \phi x = 0$, i.e., $x = 1/\phi$, is greater than one in absolute value, or equivalently, $|\phi| < 1$.

5. Rewriting the AR(1) model,

$$y_t = \phi y_{t-1} + \epsilon_t$$

$$
\begin{aligned}
&= \phi^2 y_{t-2} + \epsilon_t + \phi\epsilon_{t-1} \\
&= \phi^3 y_{t-3} + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} \\
&\quad \vdots \\
&= \phi^s y_{t-s} + \epsilon_t + \phi\epsilon_{t-1} + \cdots + \phi^{s-1}\epsilon_{t-s+1}.
\end{aligned}
$$

As s is large, ϕ^s approaches zero. $\Longrightarrow$ Stationarity condition

6. For stationarity, $y_t = \phi y_{t-1} + \epsilon_t$ is rewritten as:

$$
y_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots
$$

7. Mean of y_t

$$
\begin{aligned}
\mathrm{E}(y_t) &= \mathrm{E}(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots) \\
&= \mathrm{E}(\epsilon_t) + \phi\mathrm{E}(\epsilon_{t-1}) + \phi^2\mathrm{E}(\epsilon_{t-2}) + \cdots = 0
\end{aligned}
$$

8. Variance of y_t

$$\begin{aligned} V(y_t) &= V(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots) \\ &= V(\epsilon_t) + V(\phi\epsilon_{t-1}) + V(\phi^2\epsilon_{t-2}) + \cdots \\ &= \sigma^2(1 + \phi^2 + \phi^4 + \cdots) = \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

9. Thus, $y_t \sim N\Big(0, \dfrac{\sigma^2}{1-\rho^2}\Big)$. $\Longrightarrow$ Unconditional distribution of y_t

10. Estimation of AR(1) model:

    (a) Log-likelihood function

$$\begin{aligned} \log f(y_T, \cdots, y_1) &= \log f(y_1) + \sum_{t=1}^{T} \log f(y_t|y_{t-1}, \cdots, y_1) \\ &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{1}{\sigma^2/(1-\phi^2)}y_1^2 \end{aligned}$$

$$-\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2$$

$$= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\frac{1}{1-\phi^2}\right)$$
$$-\frac{1}{2\sigma^2/(1-\phi^2)}y_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2$$

Note as follows:

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

$$\frac{\partial \log f(y_T,\cdots,y_1)}{\partial\sigma^2} = -\frac{T}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4/(1-\phi^2)}y_1^2 + \frac{1}{2\sigma^4}\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2 = 0$$

$$\frac{\partial \log f(y_T, \cdots, y_1)}{\partial \phi} = -\frac{\phi}{1-\phi^2} + \frac{\phi}{\sigma^2} y_1^2 + \frac{1}{\sigma^2} \sum_{t=2}^{T} (y_t - \phi y_{t-1}) y_{t-1} = 0$$

The MLE of ϕ and σ^2 satisfies the above two equation.

6.2 最尤法の例：系列相関のもとで回帰式の推定：その 2

$$y_t = X_t\beta + u_t, \qquad u_t = \rho u_{t-1} + \epsilon, \qquad \epsilon_t \sim N(0, \sigma^2)$$

Log of distribution function of u_t

$$\begin{aligned}
\log f(u_T, \cdots, u_1) &= \log f(u_1) + \sum_{t=1}^{T} \log f(u_t | u_{t-1}, \cdots, y_1) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{1}{\sigma^2/(1-\rho^2)}u_1^2 \\
&\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{t=2}^{T}(u_t - \rho u_{t-1})^2 \\
&= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) \\
&\quad -\frac{1}{2\sigma^2/(1-\rho^2)}u_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(u_t - \rho u_{t-1})^2
\end{aligned}$$

Log of distribution function of y_t

$$
\begin{aligned}
&\log f(y_T, \cdots, y_1) \\
&= \log f(y_1) + \sum_{t=1}^{T} \log f(y_t | y_{t-1}, \cdots, y_1) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{1}{\sigma^2/(1-\rho^2)}(y_1 - X_1\beta)^2 \\
&\quad -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \frac{1}{\sigma^2}\sum_{t=2}^{T}\Big((y_t - X_t\beta) - \rho(y_{t-1} - X_{t-1}\beta)\Big)^2
\end{aligned}
$$

$$
= -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t^* - X_t^*\beta)^2,
$$

where

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

$\log f(y_T, \cdots, y_1)$ is maximized with respect to β, ρ and σ^2.

推定例： OLS, AR(1), AR(1)+X

StataSE をクリック

● データの編集

「Data」「Data Editor」を選択

Excel からデータのコピー
123,456 という形式でなく，123456 のようにコンマのない形式に設定すること。
方法： 「書式」「セル」のところで「表示形式」のタブの「標準」を選択
データ名は var1, var2, var3, ... となるので，出来れば変更

● command の欄にコマンドを入力

例えば，Y=$\alpha+\beta$ X+γ Z で，α，β，γ を推定するとき，
「reg Y X Z」リターン
とタイプする。 結果は results の欄に出力

Y, X, Z が時系列データのとき，

　「gen t=_n」リターン
　「tsset t」リターン
として，時系列データを扱っているということを宣言する。　t は他の名前でも構わない。
そして，
　「reg Y X Z」リターン
とする。
　「dwstat」リターン
とすると，ダービンワトソン比が出力される。

グラフについて：
　「scatter Y X」リターン
とすると，横軸 X，縦軸 Y のグラフ。
　「line Y X time」リターン
とすると，横軸 time，縦軸 X と Y のグラフ。

● 参考書

筒井淳也、秋吉美都、水落正明、　福田亘孝著
『Stata で計量経済学入門』(2007 年 3 月) ミネルヴァ書房 \2,940

● データ：　山本拓 (1995)『計量経済学』の数値例

t	x	y
1	10	6
2	12	9
3	14	10
4	16	10

● 出力結果

```
. gen t=_n

. tsset t

. reg y x

      Source |       SS       df       MS              Number of obs =       4
-------------+------------------------------           F(  1,     2) =    7.35
       Model |      8.45       1        8.45           Prob > F      =  0.1134
    Residual |       2.3       2        1.15           R-squared     =  0.7860
-------------+------------------------------           Adj R-squared =  0.6791
       Total |     10.75       3  3.58333333           Root MSE      =  1.0724

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |        .65   .2397916     2.71   0.113    -.3817399     1.68174
       _cons |         .3   3.163068     0.09   0.933    -13.30958    13.90958
------------------------------------------------------------------------------

. arima y, ar(1) nocons

(setting optimization to BHHH)
Iteration 0:   log likelihood = -10.213007
```

```
Iteration 1:   log likelihood = -9.8219683
Iteration 2:   log likelihood = -9.7761938
Iteration 3:   log likelihood = -9.6562972
Iteration 4:   log likelihood = -9.5973095
(switching optimization to BFGS)
Iteration 5:   log likelihood = -9.5850964
Iteration 6:   log likelihood = -9.5799049
Iteration 7:   log likelihood = -9.5770119
Iteration 8:   log likelihood = -9.5770099
Iteration 9:   log likelihood = -9.5770099

ARIMA regression

Sample:  1 - 4                                  Number of obs      =          4
                                                Wald chi2(1)       =     101.94
Log likelihood =  -9.57701                      Prob > chi2        =     0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .9759129    .096657    10.10   0.000     .7864686    1.165357
-------------+----------------------------------------------------------------
      /sigma |   1.812458   .8837346     2.05   0.020     .0803696    3.544545
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

```
. arima y x,ar(1)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -4.3799561
Iteration 1:   log likelihood = -4.3799068  (backed up)
Iteration 2:   log likelihood =  -4.379678  (backed up)
Iteration 3:   log likelihood = -4.3796767  (backed up)
Iteration 4:   log likelihood = -4.3796761  (backed up)
(switching optimization to BFGS)
Iteration 5:   log likelihood = -4.3796757  (backed up)
Iteration 6:   log likelihood = -4.3235592
Iteration 7:   log likelihood = -4.2798453
Iteration 8:   log likelihood = -4.2471467
Iteration 9:   log likelihood =  -4.239353
Iteration 10:  log likelihood = -4.2384456
Iteration 11:  log likelihood =  -4.238435
Iteration 12:  log likelihood =  -4.238435

ARIMA regression

Sample:  1 - 4                                  Number of obs      =         4
                                                Wald chi2(2)       =   1001.98
Log likelihood = -4.238435                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
y            |
           x |    .635658   .0583723    10.89   0.000     .5212505    .7500656
       _cons |   .6512199          .        .       .            .           .
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.5631492   2.177484    -0.26   0.796    -4.830939    3.704641
-------------+----------------------------------------------------------------
      /sigma |   .6656358   .7509811     0.89   0.188            0    2.137532
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

7 Qualitative Dependent Variable (質的従属変数)

1. **Discrete Choice Model (離散選択モデル)**

2. **Limited Dependent Variable Model (制限従属変数モデル)**

3. **Count Data Model (計数データモデル)**

Usually, the regression model is given by:

$$y_i = X_i\beta + u_i, \qquad u_i \sim N(0, \sigma^2), \qquad i = 1, 2, \cdots, n,$$

where y_i is a continuous type of random variable within the interval from $-\infty$ to ∞.

When y_i is discrete or truncated, what happens?

7.1 Discrete Choice Model (離散選択モデル)

7.1.1 Binary Choice Model (二値選択モデル)

Example 1: Consider the regression model:

$$y_i^* = X_i\beta + u_i, \qquad u_i \sim (0, \sigma^2), \qquad i = 1, 2, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as 0 or 1, i.e.,

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0, \\ 0, & \text{if } y_i^* \le 0. \end{cases}$$

Consider the probability that y_i takes 1, i.e.,

$$\begin{aligned} P(y_i = 1) &= P(y_i^* > 0) = P(u_i > -X_i\beta) = P(u_i^* > -X_i\beta^*) = 1 - P(u_i^* \le -X_i\beta^*) \\ &= 1 - F(-X_i\beta^*) = F(X_i\beta^*), \quad \text{(if the dist. of } u_i^* \text{ is symmetric.)}, \end{aligned}$$

where $u_i^* = \dfrac{u_i}{\sigma}$, and $\beta^* = \dfrac{\beta}{\sigma}$ are defined.

(*) β^* can be estimated, but β and σ^2 cannot be estimated separately (i.e., β and σ^2 are not identified).

The distribution function of u_i^* is given by $F(x) = \int_{-\infty}^{x} f(z)\mathrm{d}z$.

If u_i^* is standard normal, i.e., $u_i^* \sim N(0, 1)$, we call **probit model**.

$$F(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-\frac{1}{2}z^2)\mathrm{d}z, \qquad f(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2).$$

If u_i^* is logistic, we call **logit model**.

$$F(x) = \frac{1}{1 + \exp(-x)}, \qquad f(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2}.$$

We can consider the other distribution function for u_i^*.

Likelihood Function: y_i is the following Bernoulli distribution:

$$f(y_i) = (P(y_i = 1))^{y_i}(P(y_i = 0))^{1-y_i} = (F(X_i\beta^*))^{y_i}(1 - F(X_i\beta^*))^{1-y_i}, \qquad y_i = 0, 1.$$

[Review — Bernoulli Distribution (ベルヌイ分布)]

Suppose that X is a Bernoulli random variable. the distribution of X, denoted by $f(x)$, is:

$$f(x) = p^x(1 - p)^{1-x}, \qquad x = 0, 1.$$

The mean and variance are:

$$\mu = \mathrm{E}(X) = \sum_{x=0}^{1} x f(x) = 0 \times (1 - p) + 1 \times p = p,$$

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}((X - \mu)^2) = \sum_{x=0}^{1} (x - \mu)^2 f(x) = (0 - p)^2(1 - p) + (1 - p)^2 p = p(1 - p).$$

[End of Review]

The likelihood function is given by:

$$L(\beta^*) = f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} (F(X_i\beta^*))^{y_i}(1 - F(X_i\beta^*))^{1-y_i},$$

The log-likelihood function is:

$$\log L(\beta^*) = \sum_{i=1}^{n} \Big(y_i \log F(X_i\beta^*) + (1 - y_i) \log(1 - F(X_i\beta^*)) \Big),$$

Solving the maximization problem of $\log L(\beta^*)$ with respect to β^*, the first order condition is:

$$\frac{\partial \log L(\beta^*)}{\partial \beta^*} = \sum_{i=1}^{n} \Big(\frac{y_i X_i' f(X_i\beta^*)}{F(X_i\beta^*)} - \frac{(1 - y_i) X_i' f(X_i\beta^*)}{1 - F(X_i\beta^*)} \Big)$$

$$= \sum_{i=1}^{n} \frac{X_i' f(X_i\beta^*)(y_i - F(X_i\beta^*))}{F(X_i\beta^*)(1 - F(X_i\beta^*))} = \sum_{i=1}^{n} \frac{X_i' f_i (y_i - F_i)}{F_i(1 - F_i)} = 0,$$

where $f_i \equiv f(X_i\beta^*)$ and $F_i \equiv F(X_i\beta^*)$. Remember that $f(x) \equiv \dfrac{\mathrm{d}F(x)}{\mathrm{d}x}$.

The second order condition is:

$$
\begin{aligned}
\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}} &= \sum_{i=1}^n \frac{X_i' \dfrac{\partial f_i}{\partial\beta^*}(y_i - F_i)}{F_i(1-F_i)} + \sum_{i=1}^n \frac{X_i' f_i \dfrac{\partial(f_i - F_i)}{\partial\beta^*}}{F_i(1-F_i)} \\
&\quad + \sum_{i=1}^n X_i' f_i (y_i - F_i) \frac{\partial(F_i(1-F_i))^{-1}}{\partial\beta^*} \\
&= \sum_{i=1}^n \frac{X_i' X_i f_i' (y_i - F_i)}{F_i(1-F_i)} - \sum_{i=1}^n \frac{X_i' X_i f_i^2}{F_i(1-F_i)} + \sum_{i=1}^n X_i' f_i (y_i - F_i) \frac{X_i f_i (1 - 2F_i)}{(F_i(1-F_i))^2}
\end{aligned}
$$

is a negative definite matrix.

For maximization, the method of scoring is given by:

$$
\begin{aligned}
\beta^{*(j+1)} &= \beta^{*(j)} + \left(-\mathrm{E}\Big(\frac{\partial^2 \log L(\beta^{*(j)})}{\partial\beta^*\partial\beta^{*\prime}} \Big) \right)^{-1} \frac{\partial \log L(\beta^{*(j)})}{\partial\beta^*} \\
&= \beta^{*(j)} + \left(\sum_{i=1}^n \frac{X_i' X_i (f_i^{(j)})^2}{F_i^{(j)}(1-F_i^{(j)})} \right)^{-1} \sum_{i=1}^n \frac{X_i' f_i^{(j)} (y_i - F_i^{(j)})}{F_i^{(j)}(1-F_i^{(j)})},
\end{aligned}
$$

where $F_i^{(j)} = F(X_i\beta^{*(j)})$ and $f_i^{(j)} = f(X_i\beta^{*(j)})$. Note that

$$I(\beta^*) = \mathrm{E}\Big(\frac{\partial^2 \log L(\beta^{*(j)})}{\partial\beta^*\partial\beta^{*\prime}}\Big) = -\sum_{i=1}^{n} \frac{X_i'X_i f_i^2}{F_i(1-F_i)}.$$

because of $\mathrm{E}(y_i) = F_i$.

It is known that

$$\sqrt{n}(\hat{\beta}^* - \beta^*) \longrightarrow N\left(0,\ \lim_{n\to\infty}\Big(-\frac{1}{n}\mathrm{E}\Big(\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\Big)\Big)^{-1}\right),$$

where $\hat{\beta}^* \equiv \lim_{j\to\infty}\beta^{*(j)}$ denotes MLE of β^*.

Practically, we use the following normal distribution:

$$\hat{\beta}^* \sim N\Big(\beta^*,\ I(\hat{\beta}^*)^{-1}\Big),$$

where $I(\beta^*) = -\mathrm{E}\left(\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\right) = \sum_{i=1}^{n} \frac{X_i'X_i \hat{f}_i^2}{\hat{F}_i(1-\hat{F}_i)}$, $\hat{f}_i = f(X_i\hat{\beta}^*)$ and $\hat{F}_i = F(X_i\hat{\beta}^*)$.

Thus, the significance test for β^* and the confidence interval for β^* can be constructed.

Another Interpretation: This maximization problem is equivalent to the nonlinear least squares estimation problem from the following regression model:

$$y_i = F(X_i\beta^*) + u_i,$$

where $u_i = y_i - F_i$ takes $u_i = 1 - F_i$ with probability $P(y_i = 1) = F(X_i\beta^*) = F_i$ and $u_i = -F_i$ with probability $P(y_i = 0) = 1 - F(X_i\beta^*) = 1 - F_i$.

Therefore, the mean and variance of u_i are:

$$\mathrm{E}(u_i) = (1 - F_i)F_i + (-F_i)(1 - F_i) = 0,$$

$$\sigma_i^2 = \mathrm{V}(u_i) = \mathrm{E}(u_i^2) - (\mathrm{E}(u_i))^2 = (1 - F_i)^2F_i + (-F_i)^2(1 - F_i) = F_i(1 - F_i).$$

The weighted least squares method solves the following minimization problem:

$$\min_{\beta^*} \sum_{i=1}^{n} \frac{(y_i - F(X_i\beta^*))^2}{\sigma_i^2}.$$

The first order condition is:

$$\sum_{i=1}^{n} \frac{X_i' f(X_i\beta^*)(y_i - F(X_i\beta^*))}{\sigma_i^2} = \sum_{i=1}^{n} \frac{X_i' f_i(y_i - F_i)}{F_i(1 - F_i)} = 0,$$

which is equivalent to the first order condition of MLE.

Thus, the binary choice model is interpreted as the nonlinear least squares.

Prediction: $\mathrm{E}(y_i) = 0 \times (1 - F_i) + 1 \times F_i = F_i \equiv F(X_i\beta^*)$.

Empirical Application of Example 1: Excess Demand Function.

Demand is observed as both amount and quantity, while supply is not.

Therefore, excess demand is not observed,

Data are taken from household expenditure survey as follows:

y	実収入【円】
s	清酒【円】
b	ビール【円】
sml	清酒【1ml】
bl	ビール【1l】
CPI	消費者物価指数・総合

year	y	s	b	sml	bl	CPI
2000.01	458911	716	1350	828	2.67	102.8
2000.02	486601	643	1527	728	3.01	102.5
2000.03	494395	661	1873	775	3.69	102.7
2000.04	505409	614	1967	749	3.93	102.9
2000.05	460116	567	2311	679	4.64	103.0

2000.06	772611	518	2225	596	4.40	102.8
2000.07	640258	459	3419	511	6.57	102.5
2000.08	506757	455	2976	530	5.91	102.8
2000.09	446405	477	2160	580	4.27	102.7
2000.10	488921	626	1805	750	3.59	102.7
2000.11	457054	680	1674	831	3.36	102.4
2000.12	1035616	1623	2546	1688	5.04	102.5
2001.01	453748	689	1363	806	2.71	102.5
2001.02	475556	554	1299	688	2.59	102.1
2001.03	481198	567	1467	708	2.99	101.9
2001.04	498080	532	1641	637	3.33	102.1
2001.05	447510	486	1825	608	3.63	102.2
2001.06	766471	446	2003	535	3.99	101.9
2001.07	614715	493	2656	568	5.25	101.6
2001.08	496482	436	2326	492	4.60	102.0
2001.09	447397	479	1546	617	3.06	101.8
2001.10	489834	568	1426	733	2.87	101.8
2001.11	461094	646	1222	818	2.42	101.3
2001.12	1000728	1609	2274	1710	4.54	101.2
2002.01	462389	637	1040	716	2.01	101.0
2002.02	477622	570	1040	778	2.15	100.5
2002.03	496351	552	1418	748	2.81	100.7
2002.04	485770	502	1427	689	2.93	101.0
2002.05	444612	497	1623	602	3.40	101.3
2002.06	745480	442	1900	537	3.74	101.2
2002.07	583862	499	2437	554	4.72	100.8
2002.08	488257	472	2358	508	4.72	101.1
2002.09	440319	437	1522	536	2.98	101.1
2002.10	475494	561	1378	757	2.85	100.9

2002.11	439186	730	1347	888	2.59	100.9
2002.12	939747	1589	2177	1936	4.21	100.9
2003.01	435989	549	1025	632	1.98	100.6
2003.02	455309	519	1089	670	2.19	100.3
2003.03	456873	531	1343	686	2.55	100.6
2003.04	475037	514	1369	576	2.69	100.9
2003.05	429669	518	1396	724	2.73	101.1
2003.06	730617	484	1609	597	3.17	100.8
2003.07	574574	492	2013	636	3.93	100.6
2003.08	474973	503	2146	641	4.33	100.8
2003.09	429301	395	1331	463	2.65	100.9
2003.10	467408	498	1312	560	2.64	100.9
2003.11	435079	560	1230	760	2.42	100.4
2003.12	932887	1484	2012	1621	3.97	100.5
2004.01	445133	530	1062	595	2.10	100.3
2004.02	474143	591	1086	705	2.20	100.3
2004.03	456288	455	1239	621	2.43	100.5
2004.04	488217	441	1273	539	2.47	100.5
2004.05	446758	438	1530	524	3.06	100.6
2004.06	723370	391	1729	447	3.37	100.8
2004.07	599045	414	2166	432	4.18	100.5
2004.08	476264	403	2032	474	4.05	100.6
2004.09	440187	387	1414	450	2.82	100.9
2004.10	467895	454	1269	551	2.54	101.4
2004.11	442885	482	1266	619	2.54	101.2
2004.12	920100	1262	1912	1272	3.83	100.7
2005.01	448635	542	999	678	1.94	100.5
2005.02	469673	497	917	630	1.84	100.2
2005.03	451360	485	1060	714	2.05	100.5

2005.04	495036	406	1226	505	2.47	100.6
2005.05	440388	386	1437	443	2.84	100.7
2005.06	720667	375	1472	430	2.96	100.3
2005.07	576129	451	2214	555	4.36	100.2
2005.08	463034	370	2001	440	4.04	100.3
2005.09	427753	323	1321	390	2.56	100.6
2005.10	463838	500	1246	624	2.39	100.6
2005.11	433036	519	1064	678	2.12	100.2
2005.12	905473	1173	2090	1152	4.06	100.3
2006.01	437787	466	921	501	1.82	100.4
2006.02	461368	433	884	580	1.71	100.1
2006.03	429948	416	1060	517	2.06	100.3
2006.04	472583	444	1269	536	2.43	100.5
2006.05	426680	426	1367	544	2.55	100.8
2006.06	684632	431	1360	529	2.60	100.8
2006.07	613269	358	1803	395	3.47	100.5
2006.08	475866	400	1843	448	3.50	101.2
2006.09	429017	341	1139	444	2.20	101.2
2006.10	467163	479	1183	696	2.35	101.0
2006.11	442147	533	1053	660	2.01	100.5
2006.12	968162	1144	1882	1200	3.76	100.6
2007.01	441039	505	941	695	1.82	100.4
2007.02	471681	428	899	580	1.69	99.9
2007.03	445076	434	1071	528	2.07	100.2
2007.04	472446	413	1291	506	2.60	100.5
2007.05	431013	346	1302	450	2.42	100.8
2007.06	735579	374	1532	490	2.93	100.6
2007.07	592452	414	1845	530	3.63	100.5
2007.08	467786	368	2121	511	4.10	101.0

2007.09	431793	329	1446	425	2.80	101.0
2007.10	469981	445	1108	542	2.15	101.3
2007.11	435640	541	1116	594	2.20	101.1
2007.12	950654	1085	1892	1209	3.56	101.3
2008.01	438998	509	1000	707	1.99	101.1
2008.02	476282	445	1008	558	1.98	100.9
2008.03	453482	400	1199	573	2.35	101.4
2008.04	469774	376	1234	492	2.44	101.3
2008.05	435076	329	1404	406	2.72	102.1
2008.06	737166	356	1410	395	2.72	102.6
2008.07	587732	298	1832	338	3.48	102.8
2008.08	488216	334	1767	413	3.36	103.1
2008.09	433502	293	1086	423	2.03	103.1
2008.10	481746	346	1066	434	2.04	103.0
2008.11	439394	439	1077	533	2.06	102.1
2008.12	969449	1076	1711	1231	3.24	101.7
2009.01	443337	479	962	636	1.85	101.1
2009.02	464665	417	849	705	1.64	100.8
2009.03	443429	444	1009	478	1.88	101.1
2009.04	473779	354	958	428	1.87	101.2
2009.05	436123	370	1180	495	2.34	101.0
2009.06	700239	343	1126	386	2.14	100.8
2009.07	573821	287	1478	327	2.78	100.5
2009.08	466393	300	1519	345	2.90	100.8
2009.09	422120	263	974	363	1.86	100.8
2009.10	459704	349	941	435	1.81	100.4
2009.11	428219	432	941	588	1.81	100.2
2009.12	906884	943	1546	1019	2.98	100.0
2010.01	434344	420	800	464	1.46	100.1

2010.02	464866	347	751	500	1.49	100.0
2010.03	439410	386	885	578	1.74	100.3
2010.04	474616	317	926	404	1.79	100.4
2010.05	421413	316	1040	455	1.99	100.3
2010.06	733886	316	1236	375	2.36	100.1
2010.07	562094	362	1600	382	3.05	99.5
2010.08	470717	314	1571	397	3.03	99.7
2010.09	425771	255	1028	361	1.93	99.9
2010.10	494398	337	1017	520	1.95	100.2
2010.11	431281	374	870	485	1.67	99.9
2010.12	895511	943	1456	912	2.80	99.6
2011.01	419728	418	693	495	1.33	99.5
2011.02	470071	345	650	552	1.23	99.5
2011.03	419862	366	703	472	1.34	99.8
2011.04	454433	371	814	485	1.53	99.9
2011.05	413506	345	888	432	1.67	99.9
2011.06	687212	317	1025	327	1.95	99.7
2011.07	572662	267	1407	367	2.68	99.7
2011.08	463760	277	1378	345	2.66	99.9
2011.09	422720	276	917	419	1.77	99.9
2011.10	479749	345	789	433	1.52	100.0
2011.11	424272	329	848	426	1.64	99.4
2011.12	893811	884	1398	907	2.73	99.4
2012.01	430477	432	711	619	1.43	99.6
2012.02	483625	394	721	495	1.39	99.8
2012.03	441015	397	787	592	1.50	100.3
2012.04	469381	381	833	466	1.56	100.4
2012.05	417723	309	845	411	1.67	100.1
2012.06	712592	337	1015	417	1.96	99.6

2012.07	557032	284	1242	375	2.36	99.3
2012.08	470470	288	1374	357	2.61	99.4
2012.09	422046	294	903	337	1.76	99.6
2012.10	482101	282	752	361	1.45	99.6
2012.11	432681	361	756	510	1.40	99.2
2012.12	902928	859	1347	863	2.56	99.3
2013.01	433858	377	743	467	1.43	99.3
2013.02	476256	325	656	410	1.26	99.2
2013.03	444379	384	815	467	1.52	99.4
2013.04	479854	323	680	359	1.21	99.7
2013.05	422724	322	853	402	1.61	99.8
2013.06	728678	433	973	541	1.86	99.8
2013.07	569174	281	1104	315	2.04	100.0
2013.08	471411	298	1200	324	2.32	100.3
2013.09	431931	258	848	311	1.59	100.6
2013.10	482684	282	805	296	1.47	100.7
2013.11	436293	377	725	447	1.32	100.8
2013.12	905822	835	1351	933	2.62	100.9
2014.01	438646	431	703	530	1.38	100.7
2014.02	479268	365	612	446	1.21	100.7
2014.03	438145	397	891	476	1.66	101.0
2014.04	463964	304	630	401	1.15	103.1
2014.05	421117	348	846	432	1.56	103.5
2014.06	710375	356	933	394	1.72	103.4
2014.07	555276	304	1182	361	2.18	103.4
2014.08	463810	325	1159	356	2.12	103.7
2014.09	421809	319	840	383	1.51	103.9
2014.10	488273	345	707	391	1.33	103.6
2014.11	431543	398	716	519	1.33	103.2

```
2014.12  924911  892 1324  901 2.47 103.3
2015.01  440226  400  622  494 1.14 103.1
2015.02  488519  356  600  469 1.12 102.9
2015.03  449243  353  712  416 1.28 103.3
2015.04  476880  353  739  413 1.35 103.7
2015.05  430325  331  909  377 1.64 104.0
2015.06  733589  350  928  231 1.66 103.8
2015.07  587156  331 1105  347 2.02 103.7
2015.08  475369  339 1165  462 2.18 103.9
2015.09  415467  346  797  354 1.47 103.9
```

```
. gen t=_n                               <---- make data

. tsset t                               <---- set t as time series data
        time variable:  t, 1 to 189
                delta:  1 unit

. gen ry=y/(cpi/100)                    <---- real income

. gen rsp=(s/sml)/(cpi/100)             <---- real sake price per 1ml

. gen rbp=(b/bl)/(cpi/100)              <---- real beer price per 1l

. gen ds=0                              <---- dfault data

. replace ds=1 if f.rsp>rsp             <---- set ds=1 when excess demand exists
(94 real changes made)
```

```
. probit ds ry rsp rbp, if t<188.5 <---- Estimate probit
                                         during the period from 1 to 188

Iteration 0:   log likelihood = -130.30103
Iteration 1:   log likelihood = -95.883766
Iteration 2:   log likelihood = -95.419505
Iteration 3:   log likelihood = -95.419207
Iteration 4:   log likelihood = -95.419207

Probit regression                                 Number of obs   =        188
                                                  LR chi2(3)      =      69.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -95.419207                       Pseudo R2       =     0.2677

------------------------------------------------------------------------------
          ds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ry |   8.04e-07   9.42e-07     0.85   0.393    -1.04e-06    2.65e-06
         rsp |   -13.8574   2.166711    -6.40   0.000    -18.10408   -9.610729
         rbp |   .0026681   .0067234     0.40   0.691    -.0105094    .0158457
       _cons |    9.44494   3.697318     2.55   0.011     2.198331    16.69155
------------------------------------------------------------------------------
Note: 1 failure and 0 successes completely determined.

. logit ds ry rsp rbp if t<188.5 <---- Estimate logit
                                       during the period from 1 to 188

Iteration 0:   log likelihood = -130.30103
```

```
Iteration 1:   log likelihood = -96.132508
Iteration 2:   log likelihood =  -95.65503
Iteration 3:   log likelihood = -95.653538
Iteration 4:   log likelihood = -95.653538

Logistic regression                               Number of obs   =        188
                                                  LR chi2(3)      =      69.29
                                                  Prob > chi2     =     0.0000
Log likelihood = -95.653538                       Pseudo R2       =     0.2659

------------------------------------------------------------------------------
          ds |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ry |   1.41e-06   1.56e-06     0.90   0.368    -1.65e-06    4.47e-06
         rsp |  -23.36485   3.941076    -5.93   0.000    -31.08922   -15.64048
         rbp |   .0046687   .0113128     0.41   0.680    -.0175041    .0268414
       _cons |   15.82621   6.301665     2.51   0.012     3.475172    28.17724
------------------------------------------------------------------------------
```

$D_t - S_t = \beta_0 + \beta_1 ry_t + \beta_2 rsp_t + \beta_3 rbp_t$

D_t is observed, but S_t is not observed. Therefore, $D_t - S_t$ is unobserved.

$rsp_{t+1} > rsp_t \implies D_t - S_t > 0 \implies ds_t = 1.$

$rsp_{t+1} \leq rsp_t \implies D_t - S_t \leq 0 \implies ds_t = 0.$

Example 2: Consider the two utility functions: $U_{1i} = X_i\beta_1 + \epsilon_{1i}$ and $U_{2i} = X_i\beta_2 + \epsilon_{2i}$. A linear utility function is problematic, but we consider the linear function for simplicity of discussion.

We purchase a good when $U_{1i} > U_{2i}$ and do not purchase it when $U_{1i} < U_{2i}$.

We can observe $y_i = 1$ when we purchase the good, i.e., when $U_{1i} > U_{2i}$, and $y_i = 0$ otherwise.

$$
\begin{aligned}
P(y_i = 1) &= P(U_{1i} > U_{2i}) = P(X_i(\beta_1 - \beta_2) > -\epsilon_{1i} + \epsilon_{2i}) \\
&= P(-X_i\beta^* > \epsilon_i^*) = P(-X_i\beta^{**} > \epsilon_i^{**}) = 1 - F(-X_i\beta^{**}) = F(X_i\beta^{**})
\end{aligned}
$$

where $\beta^* = \beta_1 - \beta_2$, $\quad \epsilon_i^* = \epsilon_{1i} - \epsilon_{2i}$, $\quad \beta^{**} = \dfrac{\beta^*}{\sigma^*}$ $\quad$ and $\quad \epsilon_i^{**} = \dfrac{\epsilon_i^*}{\sigma^*}$.

We can estimate β^{**}, but we cannot estimate ϵ_i^* and σ^*, separately.

Mean and variance of ϵ_i^{**} are normalized to be zero and one, respectively.

If the distribution of ϵ_i^{**} is symmetric, the last equality holds.

We can estimate β^{**} by MLE as in Example 1.

Example 3: Consider the questionnaire:

$$y_i = \begin{cases} 1, & \text{if the } i\text{th person answers YES,} \\ 0, & \text{if the } i\text{th person answers NO.} \end{cases}$$

Consider estimating the following linear regression model:

$$y_i = X_i\beta + u_i.$$

When $\mathrm{E}(u_i) = 0$, the expectation of y_i is given by:

$$\mathrm{E}(y_i) = X_i\beta.$$

Because of the linear function, $X_i\beta$ takes the value from $-\infty$ to ∞.

However, $\mathrm{E}(y_i)$ indicates the ratio of the people who answer YES out of all the people, because of $\mathrm{E}(y_i) = 1 \times P(y_i = 1) + 0 \times P(y_i = 0) = P(y_i = 1)$.

That is, $\mathrm{E}(y_i)$ has to be between zero and one.

Therefore, it is not appropriate that $\mathrm{E}(y_i)$ is approximated as $X_i\beta$.

The model is written as:

$$y_i = P(y_i = 1) + u_i,$$

where u_i is a discrete type of random variable, i.e., u_i takes $1 - P(y_i = 1)$ with probability $P(y_i = 1)$ and $-P(y_i = 1)$ with probability $1 - P(y_i = 1) = P(y_i = 0)$.

Consider that $P(y_i)$ is connected with the distribution function $F(X_i\beta)$ as follows:

$$P(y_i = 1) = F(X_i\beta),$$

where $F(\cdot)$ denotes a distribution function such as normal dist., logistic dist., and so on. $\longrightarrow$ probit model or logit model.

The probability function of y_i is:

$$f(y_i) = F(X_i\beta)^{y_i}(1 - F(X_i\beta))^{1-y_i} \equiv F_i^{y_i}(1 - F_i)^{1-y_i}, \qquad y_i = 0, 1.$$

The joint distribution of $y_1, y_2, \cdots, y_n$ is:

$$f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} F_i^{y_i}(1 - F_i)^{1-y_i} \equiv L(\beta),$$

which corresponds to the likelihood function. $\longrightarrow$ MLE

Example 4: Ordered probit or logit model:

Consider the regression model:

$$y_i^* = X_i\beta + u_i, \qquad u_i \sim (0, 1), \qquad i = 1, 2, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as $1, 2, \cdots, m$, i.e.,

$$y_i = \begin{cases} 1, & \text{if } -\infty < y_i^* \leq a_1, \\ 2, & \text{if } a_1 < y_i^* \leq a_2, \\ \vdots, & \\ m, & \text{if } a_{m-1} < y_i^* < \infty, \end{cases}$$

where a_1, a_2, $\cdots$, a_{m-1} are assumed to be known.

Consider the probability that y_i takes 1, 2, $\cdots$, m, i.e.,

$$
\begin{aligned}
P(y_i = 1) &= P(y_i^* \le a_1) = P(u_i \le a_1 - X_i\beta) \\
&= F(a_1 - X_i\beta), \\
P(y_i = 2) &= P(a_1 < y_i^* \le a_2) = P(a_1 - X_i\beta < u_i \le a_2 - X_i\beta) \\
&= F(a_2 - X_i\beta) - F(a_1 - X_i\beta), \\
P(y_i = 3) &= P(a_2 < y_i^* \le a_3) = P(a_2 - X_i\beta < u_i \le a_3 - X_i\beta) \\
&= F(a_3 - X_i\beta) - F(a_2 - X_i\beta), \\
&\vdots \\
P(y_i = m) &= P(a_{m-1} < y_i^*) = P(a_{m-1} - X_i\beta < u_i) \\
&= 1 - F(a_{m-1} - X_i\beta).
\end{aligned}
$$

Define the following indicator functions:

$$I_{i1} = \begin{cases} 1, & \text{if } y_i = 1, \\ 0, & \text{otherwise.} \end{cases} \qquad I_{i2} = \begin{cases} 1, & \text{if } y_i = 2, \\ 0, & \text{otherwise.} \end{cases} \qquad \cdots \qquad I_{im} = \begin{cases} 1, & \text{if } y_i = m, \\ 0, & \text{otherwise.} \end{cases}$$

More compactly,

$$P(y_i = j) = F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta),$$

for $j = 1, 2, \cdots, m$, where $a_0 = -\infty$ and $a_m = \infty$.

$$I_{ij} = \begin{cases} 1, & \text{if } y_i = j, \\ 0, & \text{otherwise,} \end{cases}$$

for $j = 1, 2, \cdots, m$.

Then, the likelihood function is:

$$L(\beta) = \prod_{i=1}^{n} \Big(F(a_1 - X_i\beta)\Big)^{I_{i1}} \Big(F(a_2 - X_i\beta) - F(a_1 - X_i\beta)\Big)^{I_{i2}} \cdots \Big(1 - F(a_{m-1} - X_i\beta)\Big)^{I_{im}}$$
$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \Big(F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)\Big)^{I_{ij}},$$

where $a_0 = -\infty$ and $a_m = \infty$. Remember that $F(-\infty) = 0$ and $F(\infty) = 1$.

The log-likelihood function is:

$$\log L(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} \log\Big(F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)\Big).$$

The first derivative of $\log L(\beta)$ with respect to β is:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{-I_{ij} X_i' \Big(f(a_j - X_i\beta) - f(a_{j-1} - X_i\beta)\Big)}{F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)} = 0.$$

Usually, normal distribution or logistic distribution is chosen for $F(\cdot)$.

Example 5: Multinomial logit model:

The ith individual has $m+1$ choices, i.e., $j = 0, 1, \cdots, m$.

$$P(y_i = j) = \frac{\exp(X_i\beta_j)}{\sum_{j=0}^{m} \exp(X_i\beta_j)} \equiv P_{ij},$$

for $\beta_0 = 0$. The case of $m = 1$ corresponds to the bivariate logit model (binary choice).

Note that

$$\log \frac{P_{ij}}{P_{i0}} = X_i\beta_j$$

The log-likelihood function is:

$$\log L(\beta_1, \cdots, \beta_m) = \sum_{i=1}^{n} \sum_{j=0}^{m} d_{ij} \ln P_{ij},$$

where $d_{ij} = 1$ when the ith individual chooses jth choice, and $d_{ij} = 0$ otherwise.

Example 6: Nested logit model:

(i) In the 1st step, choose YES or NO. Each probability is P_Y and $P_N = 1 - P_Y$.

(ii) Stop if NO is chosen in the 1st step. Go to the next if YES is chosen in the 1st step.

(iii) In the 2nd step, choose A or B if YES is chosen in the 1st step. Each probability is $P_{A|Y}$ and $P_{B|Y}$.

For simplicity, usually we assume the logistic distribution.

So, we call the nested logit model.

The probability that the ith individual chooses NO is:

$$P_{N,i} = \frac{1}{1 + \exp(X_i\beta)}.$$

The probability that the ith individual chooses YES and A is:

$$P_{A|Y,i}P_{Y,i} = P_{A|Y,i}(1 - P_{N,i}) = \frac{\exp(Z_i\alpha)}{1 + \exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

The probability that the ith individual chooses YES and B is:

$$P_{B|Y,i}P_{Y,i} = (1 - P_{A|Y,i})(1 - P_{N,i}) = \frac{1}{1 + \exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

In the 1st step, decide if the ith individual buys a car or not.

In the 2nd step, choose A or B.

X_i includes annual income, distance from the nearest station, and so on.

Z_i are speed, fuel-efficiency, car company, color, and so on.

The likelihood function is:

$$\begin{aligned} L(\alpha,\beta) &= \prod_{i=1}^{n} P_{N,i}^{I_{1i}} \Big(((1 - P_{N,i})P_{A|Y,i})^{I_{2i}} ((1 - P_{N,i})(1 - P_{A|Y,i}))^{1-I_{2i}} \Big)^{1-I_{1i}} \\ &= \prod_{i=1}^{n} P_{N,i}^{I_{1i}} (1 - P_{N,i})^{1-I_{1i}} \Big(P_{A|Y,i}^{I_{2i}} (1 - P_{A|Y,i})^{1-I_{2i}} \Big)^{1-I_{1i}}, \end{aligned}$$

where

$$I_{1i} = \begin{cases} 1, & \text{if the } i\text{th individual decides not to buy a car in the 1st step,} \\ 0, & \text{if the } i\text{th individual decides to buy a car in the 1st step,} \end{cases}$$

$$I_{2i} = \begin{cases} 1, & \text{if the } i\text{th individual chooses A in the 2nd step,} \\ 0, & \text{if the } i\text{th individual chooses B in the 2nd step,} \end{cases}$$

Remember that $\mathrm{E}(y_i) = F(X_i\beta^*)$, where $\beta^* = \dfrac{\beta}{\sigma}$.
Therefore, size of β^* does not mean anything.

The marginal effect is given by:

$$\frac{\partial \mathrm{E}(y_i)}{\partial X_i} = f(X_i\beta^*)\beta^*.$$

Thus, the marginal effect depends on the height of the density function $f(X_i\beta^*)$.

7.2 Limited Dependent Variable Model (制限従属変数モデル)

Truncated Regression Model: Consider the following model:

$$y_i = X_i\beta + u_i, \qquad u_i \sim N(0, \sigma^2) \text{ when } y_i > a, \text{ where } a \text{ is a constant,}$$

for $i = 1, 2, \cdots, n$.

Consider the case of $y_i > a$ (i.e., in the case of $y_i \leq a$, y_i is not observed).

$$\mathrm{E}(u_i | X_i\beta + u_i > a) = \int_{a - X_i\beta}^{\infty} u_i \frac{f(u_i)}{1 - F(a - X_i\beta)} \mathrm{d}u_i.$$

Suppose that $u_i \sim N(0, \sigma^2)$, i.e., $\dfrac{u_i}{\sigma} \sim N(0, 1)$.

Using the following standard normal density and distribution functions:

$$\phi(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2),$$

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-\frac{1}{2}z^2)\mathrm{d}z = \int_{-\infty}^{x} \phi(z)\mathrm{d}z,$$

$f(x)$ and $F(x)$ are given by:

$$f(x) = (2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}x^2) = \frac{1}{\sigma}\phi(\frac{x}{\sigma}),$$

$$F(x) = \int_{-\infty}^{x}(2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}z^2)\mathrm{d}z = \Phi(\frac{x}{\sigma}).$$

[Review — Mean of Truncated Normal Random Variable:]

Let X be a normal random variable with mean μ and variance σ^2.

Consider $\mathrm{E}(X|X > a)$, where a is known.

The truncated distribution of X given $X > a$ is:

$$f(x|x > a) = \frac{(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)}{\displaystyle\int_a^{\infty}(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\mathrm{d}x} = \frac{\frac{1}{\sigma}\phi(\frac{x}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})}.$$

$$\mathrm{E}(X|X>a)=\int_a^\infty xf(x|x>a)\mathrm{d}x=\frac{\displaystyle\int_a^\infty x(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\mathrm{d}x}{\displaystyle\int_a^\infty(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\mathrm{d}x}$$

$$=\frac{\sigma\phi(\frac{a-\mu}{\sigma})+\mu\Big(1-\Phi(\frac{a-\mu}{\sigma})\Big)}{1-\Phi(\frac{a-\mu}{\sigma})}=\frac{\sigma\phi(\frac{a-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})}+\mu,$$

which are shown below. The denominator is:

$$\int_a^\infty(2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)\mathrm{d}x=\int_{\frac{a-\mu}{\sigma}}^\infty(2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z$$

$$=1-\int_{-\infty}^{\frac{a-\mu}{\sigma}}(2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z$$

$$=1-\Phi(\frac{a-\mu}{\sigma}),$$

where x is transformed into $z=\dfrac{x-\mu}{\sigma}$. $\qquad x>a\implies z=\dfrac{x-\mu}{\sigma}>\dfrac{a-\mu}{\sigma}$.

The numerator is:

$$
\begin{aligned}
&\int_a^\infty x(2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)\mathrm{d}x \\
&= \int_{\frac{a-\mu}{\sigma}}^\infty (\sigma z+\mu)(2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z \\
&= \sigma\int_{\frac{a-\mu}{\sigma}}^\infty z(2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z + \mu\int_{\frac{a-\mu}{\sigma}}^\infty (2\pi)^{-1/2}\exp(-\frac{1}{2\sigma^2}z^2)\mathrm{d}z \\
&= \sigma\int_{\frac{1}{2}(\frac{a-\mu}{\sigma})^2}^\infty (2\pi)^{-1/2}\exp(-t)\mathrm{d}t + \mu\Big(1-\Phi(\frac{a-\mu}{\sigma})\Big) \\
&= \sigma\phi(\frac{a-\mu}{\sigma}) + \mu\Big(1-\Phi(\frac{a-\mu}{\sigma})\Big),
\end{aligned}
$$

where z is transformed into $t=\frac{1}{2}z^2$. $\quad z>\frac{a-\mu}{\sigma} \implies t=\frac{1}{2}z^2>\frac{1}{2}(\frac{a-\mu}{\sigma})^2$.

[End of Review]

Therefore, the conditional expectation of u_i given $X_i\beta + u_i > a$ is:

$$\mathrm{E}(u_i|X_i\beta + u_i > a) = \int_{a-X_i\beta}^{\infty} u_i \frac{f(u_i)}{1 - F(a - X_i\beta)} \mathrm{d}u_i = \int_{a-X_i\beta}^{\infty} \frac{u_i}{\sigma} \frac{\phi(\dfrac{u_i}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})} \mathrm{d}u_i$$

$$= \frac{\sigma\phi(\dfrac{a - X_i\beta}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})}.$$

Accordingly, the conditional expectation of y_i given $y_i > a$ is given by:

$$\mathrm{E}(y_i|y_i > a) = \mathrm{E}(y_i|X_i\beta + u_i > a) = \mathrm{E}(X_i\beta + u_i|X_i\beta + u_i > a)$$

$$= X_i\beta + \mathrm{E}(u_i|X_i\beta + u_i > a) = X_i\beta + \frac{\sigma\phi(\dfrac{a - X_i\beta}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})},$$

for $i = 1, 2, \cdots, n$.

Estimation:

MLE:

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{f(y_i - X_i\beta)}{1 - F(a - X_i\beta)} = \prod_{i=1}^{n} \frac{1}{\sigma} \frac{\phi(\dfrac{y_i - X_i\beta}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})}$$

is maximized with respect to β and σ^2.

Some Examples:

1. Buying a Car:

   $y_i = x_i\beta + u_i$, where y_i denotes expenditure for a car, and x_i includes income, price of the car, etc.

   Data on people who bought a car are observed.

   People who did not buy a car are ignored.

2. Working-hours of Wife:

   y_i represents working-hours of wife, and x_i includes the number of children, age, education, income of husband, etc.

3. Stochastic Frontier Model:

   $y_i = f(K_i, L_i) + u_i$, where y_i denotes production, K_i is stock, and L_i is amount of labor.

   We always have $y_i \leq f(K_i, L_i)$, i.e., $u_i \leq 0$.

   $f(K_i, L_i)$ is a maximum value when we input K_i and L_i.

Censored Regression Model or Tobit Model:

$$y_i = \begin{cases} X_i\beta + u_i, & \text{if } y_i > a, \\ a, & \text{otherwise.} \end{cases}$$

The probability which y_i takes a is given by:

$$P(y_i = a) = P(y_i \le a) = F(a) \equiv \int_{-\infty}^{a} f(x)\mathrm{d}x,$$

where $f(\cdot)$ and $F(\cdot)$ denote the density function and cumulative distribution function of y_i, respectively.

Therefore, the likelihood function is:

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} F(a)^{I(y_i=a)} \times f(y_i)^{1-I(y_i=a)},$$

where $I(y_i = a)$ denotes the indicator function which takes one when $y_i = a$ or zero otherwise.

When $u_i \sim N(0, \sigma^2)$, the likelihood function is:

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \Big(\int_{-\infty}^{a} (2\pi\sigma^2)^{-1/2} \exp(-\frac{1}{2\sigma^2}(y_i - X_i\beta)^2) \mathrm{d}y_i \Big)^{I(y_i=a)} \times \Big((2\pi\sigma^2)^{-1/2} \exp(-\frac{1}{2\sigma^2}(y_i - X_i\beta)^2) \Big)^{1-I(y_i=a)},$$

which is maximized with respect to β and σ^2.

Example of Truncated Regression Model:

Demand Function of Watermelon

二人以上の世帯のうち勤労者世帯（2000 年〜）

```
y    実収入【円】
a    りんご【円】
g    ぶどう【円】
w    すいか【円】
ag   りんご【1g】
gg   ぶどう【1g】
wg   すいか【1g】
CPI  CPI
```

year	y	a	g	w	ag	gg	wg	CPI
2000.01	458911	371	6	3	1093	9	3	102.8
2000.02	486601	416	4	4	1285	5	4	102.5
2000.03	494395	388	8	8	1145	12	10	102.7
2000.04	505409	350	19	46	899	28	85	102.9
2000.05	460116	258	46	243	598	45	638	103.0
2000.06	772611	191	153	352	446	169	1163	102.8
2000.07	640258	139	317	571	306	293	2152	102.5
2000.08	506757	144	1032	397	282	1073	1558	102.8
2000.09	446405	354	826	30	884	1002	109	102.7
2000.10	488921	501	292	3	1460	360	8	102.7
2000.11	457054	739	37	1	2024	43	2	102.4

2000.12	1035616	938	16	5	2230	27	11	102.5
2001.01	453748	329	11	1	905	16	2	102.5
2001.02	475556	350	5	1	920	6	3	102.1
2001.03	481198	321	7	3	835	11	5	101.9
2001.04	498080	287	17	52	713	26	92	102.1
2001.05	447510	255	43	236	582	43	602	102.2
2001.06	766471	169	138	355	352	120	1167	101.9
2001.07	614715	108	301	616	203	278	2403	101.6
2001.08	496482	129	827	400	265	916	1577	102.0
2001.09	447397	449	661	26	1087	823	90	101.8
2001.10	489834	598	241	1	1581	308	2	101.8
2001.11	461094	673	27	1	2026	34	2	101.3
2001.12	1000728	961	16	3	2622	16	6	101.2
2002.01	462389	331	4	2	997	4	3	101.0
2002.02	477622	343	2	1	1327	2	1	100.5
2002.03	496351	326	8	8	1114	10	22	100.7
2002.04	485770	273	14	50	826	21	90	101.0
2002.05	444612	243	57	208	726	55	517	101.3
2002.06	745480	194	170	353	524	157	1225	101.2
2002.07	583862	126	324	499	313	341	2075	100.8
2002.08	488257	151	722	335	312	813	1406	101.1
2002.09	440319	376	730	24	939	853	88	101.1
2002.10	475494	506	366	1	1504	462	3	100.9
2002.11	439186	733	36	3	2056	52	3	100.9
2002.12	939747	847	24	2	2599	38	2	100.9
2003.01	435989	303	7	1	900	12	0	100.6
2003.02	455309	305	3	2	1148	5	1	100.3
2003.03	456873	326	11	2	1094	22	8	100.6
2003.04	475037	273	18	36	815	28	63	100.9

2003.05	429669	221	40	171	583	58	422	101.1
2003.06	730617	157	177	294	368	150	967	100.8
2003.07	574574	153	244	379	326	242	1412	100.6
2003.08	474973	128	683	293	264	873	1110	100.8
2003.09	429301	333	636	33	938	738	88	100.9
2003.10	467408	506	258	5	1193	346	5	100.9
2003.11	435079	618	39	1	2105	46	0	100.4
2003.12	932887	757	12	3	1856	13	2	100.5
2004.01	445133	327	5	1	995	6	0	100.3
2004.02	474143	348	3	2	1044	4	3	100.3
2004.03	456288	287	7	5	829	9	6	100.5
2004.04	488217	221	13	52	640	26	114	100.5
2004.05	446758	192	52	168	487	61	542	100.6
2004.06	723370	141	133	289	362	123	725	100.8
2004.07	599045	94	313	462	223	307	1689	100.5
2004.08	476264	115	675	276	260	761	892	100.6
2004.09	440187	328	583	25	859	814	82	100.9
2004.10	467895	482	156	1	1192	204	4	101.4
2004.11	442885	563	48	2	1613	58	7	101.2
2004.12	920100	673	15	3	1686	24	0	100.7
2005.01	448635	310	6	4	785	9	3	100.5
2005.02	469673	340	4	6	911	4	17	100.2
2005.03	451360	360	11	7	933	13	9	100.5
2005.04	495036	294	18	23	787	30	37	100.6
2005.05	440388	226	47	149	485	50	416	100.7
2005.06	720667	152	126	337	335	111	1088	100.3
2005.07	576129	105	217	402	216	226	1546	100.2
2005.08	463034	104	582	328	234	652	1225	100.3
2005.09	427753	277	644	30	771	838	93	100.6

2005.10	463838	404	363	1	1147	544	4	100.6
2005.11	433036	540	45	1	1594	67	1	100.2
2005.12	905473	631	13	2	1519	20	2	100.3
2006.01	437787	294	7	1	994	10	1	100.4
2006.02	461368	310	4	0	950	6	0	100.1
2006.03	429948	302	7	7	920	12	0	100.3
2006.04	472583	256	17	25	728	26	40	100.5
2006.05	426680	202	32	141	515	44	332	100.8
2006.06	684632	148	114	240	338	97	720	100.8
2006.07	613269	105	209	361	228	205	1413	100.5
2006.08	475866	82	595	324	163	634	1034	101.2
2006.09	429017	263	628	32	647	716	108	101.2
2006.10	467163	455	263	4	1144	359	4	101.0
2006.11	442147	605	23	0	1556	22	1	100.5
2006.12	968162	719	18	1	1949	13	0	100.6
2007.01	441039	309	5	1	858	4	0	100.4
2007.02	471681	319	3	8	950	5	0	99.9
2007.03	445076	346	6	2	1012	9	0	100.2
2007.04	472446	304	15	35	770	23	75	100.5
2007.05	431013	233	35	159	539	37	355	100.8
2007.06	735579	177	122	320	369	110	926	100.6
2007.07	592452	110	201	360	258	212	1322	100.5
2007.08	467786	103	581	341	211	639	1126	101.0
2007.09	431793	291	717	28	735	745	77	101.0
2007.10	469981	443	261	1	1185	331	3	101.3
2007.11	435640	574	45	0	1423	29	0	101.1
2007.12	950654	748	17	1	1873	27	0	101.3
2008.01	438998	302	4	2	835	5	0	101.1
2008.02	476282	309	4	0	884	5	0	100.9

2008.03	453482	291	5	4	905	6	0	101.4
2008.04	469774	232	12	28	676	18	43	101.3
2008.05	435076	192	30	148	471	39	293	102.1
2008.06	737166	150	102	222	358	95	661	102.6
2008.07	587732	103	236	400	227	245	1212	102.8
2008.08	488216	88	615	307	197	670	1012	103.1
2008.09	433502	278	625	28	827	693	125	103.1
2008.10	481746	445	241	2	1336	337	7	103.0
2008.11	439394	526	36	0	1601	39	0	102.1
2008.12	969449	661	10	1	1949	13	2	101.7
2009.01	443337	268	5	0	865	17	0	101.1
2009.02	464665	277	3	1	1084	3	0	100.8
2009.03	443429	265	5	0	861	6	2	101.1
2009.04	473779	210	15	32	648	15	56	101.2
2009.05	436123	167	33	141	478	31	301	101.0
2009.06	700239	129	110	243	351	97	735	100.8
2009.07	573821	84	209	329	219	232	1253	100.5
2009.08	466393	80	493	303	193	494	1054	100.8
2009.09	422120	259	522	27	774	686	80	100.8
2009.10	459704	366	204	3	1129	248	5	100.4
2009.11	428219	558	41	1	1732	48	3	100.2
2009.12	906884	525	16	2	1561	17	2	100.0
2010.01	434344	256	7	0	804	5	0	100.1
2010.02	464866	265	2	0	917	3	0	100.0
2010.03	439410	264	5	4	829	8	10	100.3
2010.04	474616	208	12	11	578	21	19	100.4
2010.05	421413	167	31	102	391	31	219	100.3
2010.06	733886	129	96	205	285	93	513	100.1
2010.07	562094	78	183	339	168	161	1054	99.5

2010.08	470717	67	543	327	141	566	935	99.7
2010.09	425771	245	608	22	567	624	36	99.9
2010.10	494398	371	237	2	955	271	5	100.2
2010.11	431281	541	44	1	1538	47	3	99.9
2010.12	895511	533	17	1	1511	23	0	99.6
2011.01	419728	239	6	0	666	6	0	99.5
2011.02	470071	257	6	0	732	6	0	99.5
2011.03	419862	250	8	0	758	13	0	99.8
2011.04	454433	210	16	19	634	27	25	99.9
2011.05	413506	177	37	115	508	54	281	99.9
2011.06	687212	158	84	206	416	70	606	99.7
2011.07	572662	97	162	351	257	138	849	99.7
2011.08	463760	94	487	285	204	508	909	99.9
2011.09	422720	230	453	35	621	517	136	99.9
2011.10	479749	350	215	3	932	220	0	100.0
2011.11	424272	410	41	1	1105	43	2	99.4
2011.12	893811	546	51	0	1487	67	0	99.4
2012.01	430477	252	7	0	574	12	0	99.6
2012.02	483625	268	7	0	647	8	0	99.8
2012.03	441015	257	16	2	505	21	1	100.3
2012.04	469381	199	25	19	355	42	30	100.4
2012.05	417723	158	38	99	312	51	145	100.1
2012.06	712592	129	90	181	208	87	567	99.6
2012.07	557032	97	166	326	179	129	1279	99.3
2012.08	470470	74	519	307	166	503	1087	99.4
2012.09	422046	211	491	44	513	567	118	99.6
2012.10	482101	355	295	2	945	342	3	99.6
2012.11	432681	482	50	1	1572	72	1	99.2
2012.12	902928	508	21	1	1404	29	0	99.3

2013.01	433858	264	8	0	753	13	0	99.3
2013.02	476256	264	9	1	743	13	0	99.2
2013.03	444379	276	16	1	781	22	1	99.4
2013.04	479854	229	30	17	643	42	28	99.7
2013.05	422724	168	41	113	454	58	250	99.8
2013.06	728678	136	82	205	307	99	634	99.8
2013.07	569174	99	169	370	218	154	1204	100.0
2013.08	471411	75	480	284	182	506	862	100.3
2013.09	431931	217	566	27	544	621	85	100.6
2013.10	482684	374	231	2	1045	294	9	100.7
2013.11	436293	417	47	1	1080	56	0	100.8
2013.12	905822	574	25	0	1377	31	1	100.9
2014.01	438646	270	8	2	674	11	6	100.7
2014.02	479268	278	7	0	734	15	0	100.7
2014.03	438145	256	15	1	655	21	4	101.0
2014.04	463964	216	35	20	536	43	27	103.1
2014.05	421117	177	46	114	355	52	273	103.5
2014.06	710375	135	86	190	267	68	453	103.4
2014.07	555276	89	180	315	163	155	1067	103.4
2014.08	463810	82	511	224	147	431	704	103.7
2014.09	421809	236	528	34	574	551	147	103.9
2014.10	488273	379	250	2	1042	247	2	103.6
2014.11	431543	504	57	0	1397	57	1	103.2
2014.12	924911	692	22	0	1555	28	1	103.3
2015.01	440226	301	8	0	847	11	0	103.1
2015.02	488519	307	9	1	820	7	0	102.9
2015.03	449243	327	21	2	842	22	6	103.3
2015.04	476880	262	49	23	604	64	38	103.7
2015.05	430325	186	54	99	364	63	180	104.0

```
2015.06  733589 140   96 203  235   80  529 103.8
2015.07  587156 101  189 297  158  146 1124 103.7
2015.08  475369 103  548 279  212  520  889 103.9
2015.09  415467 272  599  41  655  606  159 103.9
2015.10  485330 397  246   4 1107  252   19 103.9


-------------------
. gen t=_n                      <--- Make data t=1,2,...,190.

. tsset t
        time variable:  t, 1 to 190
                delta:  1 unit

. gen ry=log(y/(cpi/100))       <--- log of real income

. gen rap=log((a/ag)/(cpi/100))  <--- log of real price of apple (yen/1g)

. gen rgp=log((g/gg)/(cpi/100))  <--- log of real price of grape (yen/1g)

. gen rwp=log((w/wg)/(cpi/100))  <--- log of real price of watermelon (yen/1g)
(40 missing values generated)

. gen lwg=log(wg)               <--- log of demand of watermelon (1g)
(35 missing values generated)

. reg lwg ry rwp rap rgp if lwg>log(10)  <--- OLS using the data for wg>10

      Source |       SS       df       MS              Number of obs =     102
```

```
-------------+------------------------------           F(  4,    97) =   40.23
       Model |  138.187919     4  34.5469797           Prob > F      =  0.0000
    Residual |  83.2907458    97  .858667482           R-squared     =  0.6239
-------------+------------------------------           Adj R-squared =  0.6084
       Total |  221.478665   101  2.19285807           Root MSE      =  .92664

------------------------------------------------------------------------------
         lwg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ry |   .8102279   .5131024     1.58   0.118    -.2081383    1.828594
         rwp |  -1.575157   .3569839    -4.41   0.000    -2.283671   -.8666422
         rap |   2.854632   .6983476     4.09   0.000     1.468606    4.240659
         rgp |   2.158679   .6110691     3.53   0.001     .9458762    3.371482
       _cons |  -3.826122   6.894227    -0.55   0.580    -17.50925    9.857011
------------------------------------------------------------------------------

. truncreg lwg ry rwp rap rgp if lwg>log(10)  <--- This is equivalent to OLS
(note: 0 obs. truncated)

Fitting full model:

Iteration 0:   log likelihood = -134.46068
Iteration 1:   log likelihood = -134.39761
Iteration 2:   log likelihood = -134.39733
Iteration 3:   log likelihood = -134.39733

Truncated regression
Limit:   lower =       -inf                             Number of obs =    102
         upper =       +inf                             Wald chi2(4)  = 169.23
```

```
Log likelihood = -134.39733                            Prob > chi2     = 0.0000

------------------------------------------------------------------------------
         lwg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ry |   .8102279   .5003683     1.62   0.105     -.170476    1.790932
         rwp |  -1.575157   .3481244    -4.52   0.000    -2.257468   -.8928453
         rap |   2.854632   .6810162     4.19   0.000     1.519865    4.189399
         rgp |   2.158679   .5959037     3.62   0.000     .9907293    3.326629
       _cons |  -3.826122   6.723128    -0.57   0.569    -17.00321    9.350967
-------------+----------------------------------------------------------------
      /sigma |   .9036459   .0632679    14.28   0.000     .7796432    1.027649
------------------------------------------------------------------------------

. truncreg lwg ry rwp rap rgp if lwg>log(10), ll(log(10))   <--- truncated reg
(note: 0 obs. truncated)

Fitting full model:

Iteration 0:   log likelihood = -132.93358
Iteration 1:   log likelihood = -132.70871
Iteration 2:   log likelihood = -132.70789
Iteration 3:   log likelihood = -132.70789

Truncated regression
Limit:   lower =  2.3025851                            Number of obs =     102
         upper =       +inf                            Wald chi2(4)  = 145.68
Log likelihood = -132.70789                            Prob > chi2   = 0.0000
```

lwg	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ry	.760959	.5179994	1.47	0.142	-.2543011	1.776219
rwp	-1.682078	.3724194	-4.52	0.000	-2.412006	-.952149
rap	2.958551	.7114935	4.16	0.000	1.564049	4.353053
rgp	2.299172	.6349926	3.62	0.000	1.054609	3.543734
_cons	-3.212068	6.960658	-0.46	0.644	-16.85471	10.43057
/sigma	.9260598	.0686138	13.50	0.000	.7915792	1.06054

$$\log(wg_t) = \beta_0 + \beta_1 \log(ry_t) + \beta_2 \log(rwp_t) + \beta_3 \log(rap_t) + \beta_4 \log(rgp_t)$$

Pick up the cases of $wg_t > 10$.

7.3 Count Data Model (計数データモデル)

Poisson distribution:

$$\mathrm{P}(X = x) = f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!},$$

for $x = 0, 1, 2, \cdots$.

In the case of Poisson random variable X, the expectation of X is:

$$\mathrm{E}(X) = \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=1}^{\infty} \lambda\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda\sum_{x'=0}^{\infty} \frac{e^{-\lambda}\lambda^{x'}}{x'!} = \lambda.$$

Remember that $\sum_x f(x) = 1$, i.e., $\sum_{x=0}^{\infty} e^{-\lambda}\lambda^{x}/x! = 1$.

Therefore, the probability function of the count data y_i is taken as the Poisson distribution with parameter λ_i.

In the case where the explained variable y_i takes 0, 1, 2, $\cdots$ (discrete numbers), assuming that the distribution of y_i is Poisson, the logarithm of λ_i is specified as a linear function, i.e.,

$$\mathrm{E}(y_i) = \lambda_i = \exp(X_i\beta).$$

Note that λ_i should be positive.

Therefore, it is better to avoid the specification: $\lambda = X_i\beta$.

The joint distribution of $y_1, y_2, \cdots, y_n$ is:

$$f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} = L(\beta),$$

where $\lambda_i = \exp(X_i\beta)$.

The log-likelihood function is:

$$\begin{aligned}\log L(\beta) &= -\sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} y_i! \\ &= -\sum_{i=1}^{n} \exp(X_i\beta) + \sum_{i=1}^{n} y_i X_i\beta - \sum_{i=1}^{n} y_i!.\end{aligned}$$

The first-order condition is:

$$\frac{\partial \log L(\beta)}{\partial \beta} = -\sum_{i=1}^{n} X_i' \exp(X_i\beta) + \sum_{i=1}^{n} X_i' y_i = 0.$$

$\Longrightarrow$ Nonlinear optimization procedure

[Review] Nonlinear Optimization Procedures:

Note that the Newton-Raphson method (one of the nonlinear optimization procedures) is:

$$\beta^{(j+1)} = \beta^{(j)} - \left(\frac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'}\right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial \beta},$$

which comes from the first-order Taylor series expansion around $\beta = \beta^*$:

$$0 = \frac{\partial \log L(\beta)}{\partial \beta} \approx \frac{\partial \log L(\beta^*)}{\partial \beta} + \frac{\partial^2 \log L(\beta^*)}{\partial\beta\partial\beta'}(\beta - \beta^*),$$

and β and β^* are replaced by $\beta^{(j+1)}$ and $\beta^{(j)}$, respectively.

An alternative nonlinear optimization procedure is known as the method of scoring, which is shown as:

$$\beta^{(j+1)} = \beta^{(j)} - \left(\mathrm{E}\Big(\frac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'}\Big)\right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial\beta},$$

where $\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'}\right)$ is replaced by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'}\right)$.

[End of Review]

In this case, we have the following iterative procedure:

$$\beta^{(j+1)} = \beta^{(j)} - \left(-\sum_{i=1}^{n} X_i' X_i \exp(X_i\beta^{(j)})\right)^{-1} \left(-\sum_{i=1}^{n} X_i' \exp(X_i\beta^{(j)}) + \sum_{i=1}^{n} X_i' y_i\right).$$

The Newton-Raphson method is equivalent to the scoring method in this count model, because any random variable is not included in the expectation.

Zero-Inflated Poisson Count Data Model: In the case of too many zeros, we have to modify the estimation procedure.

Suppose that the probability of $y_i = j$ is decomposed of two regimes.

$\longrightarrow$ We have the case of $y_i = j$ and Regime 1, and that of $y_i = j$ and Regime 2.

Consider $P(y_i = 0)$ and $P(y_i = j)$ separately as follows:

$$P(y_i = 0) = P(y_i = 0|\text{Regime 1})P(\text{Regime 1}) + P(y_i = 0|\text{Regime 2})P(\text{Regime 2})$$

$$P(y_i = j) = P(y_i = j|\text{Regime 1})P(\text{Regime 1}) + P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 1, 2, \cdots$.

Assume:

- $P(y_i = 0|\text{Regime 1}) = 1$ and $P(y_i = j|\text{Regime 1}) = 0$ for $j = 1, 2, \cdots,$

- $P(\text{Regime 1}) = F_i$ and $P(\text{Regime 2}) = 1 - F_i,$

- $P(y_i = j|\text{Regime 2}) = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ for $j = 0, 1, 2, \cdots,$

where $F_i = F(Z_i\alpha)$ and $\lambda_i = \exp(X_i\beta)$. $\Longrightarrow$ w_i and X_i are exogenous variables.

Under the first assumption, we have the following equations:

$$P(y_i = 0) = P(\text{Regime 1}) + P(y_i = 0|\text{Regime 2})P(\text{Regime 2})$$
$$P(y_i = j) = P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 1, 2, \cdots.$

Combining the above two equations, we obtain the following:

$$P(y_i = j) = P(\text{Regime 1})I_i + P(y_i = j|\text{Regime 2})P(\text{Regime 2}),$$

for $j = 0, 1, 2, \cdots,$

where the indicator function I_i is given by $I_i = 1$ for $y_i = 0$ and $I_i = 0$ for $y_i \neq 0$.

F_i denotes a cumulative distribution function of $Z_i\alpha$. $\Longrightarrow$ We have to assume F_i.

Including the other two assumptions, we obtain the distribution of y_i as follows:

$$P(y_i = j) = F_i I_i + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F_i), \qquad j = 0, 1, 2, \cdots$$

where $F_i \equiv F(Z_i\alpha)$, $\lambda_i = \exp(X_i\beta)$, and the indicator function I_i is given by $I_i = 1$ for $y_i = 0$ and $I_i = 0$ for $y_i \neq 0$.

Therefore, the log-likelihood function is:

$$\log L(\alpha,\beta) = \sum_{i=1}^{n} \log P(y_i = j) = \sum_{i=1}^{n} \log\left(F_i I_i + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F_i)\right),$$

where $F_i \equiv F(Z_i\alpha)$ and $\lambda_i = \exp(X_i\beta)$.

$\log L(\alpha,\beta)$ is maximized with respect to α and β.

$\Longrightarrow$ The Newton-Raphson method or the method of scoring is utilized for maximization.

Example of Poisson Regression:

bike	自転車事故死者数（2012 年）
lowland	低地面積（平方キロ，2012 年）
dwellings	居住用宅地面積（平方キロ，2012 年）
pop	人口（2010 年）

	pref	bike	lowland	dwellings	pop
北海道	1	11	9794	543	5504
青森	2	6	1237	193	1374
岩手	3	7	1261	216	1326
宮城	4	4	1757	259	2352
秋田	5	2	2453	170	1085
山形	6	5	1393	163	1167
福島	7	5	1437	255	2021
茨城	8	20	1647	454	2887
栃木	9	17	752	289	1990
群馬	10	17	585	272	2005
埼玉	11	42	1414	487	6373
千葉	12	30	1452	489	5560
東京	13	34	274	421	15576
神奈川	14	17	575	418	8254
新潟	19	5	2775	274	2375
富山	20	4	987	145	1091
石川	15	5	656	116	1172
福井	16	2	932	93	807
山梨	17	4	343	115	855

長野	21	7	751	307	2149
岐阜	22	12	1174	226	1998
静岡	23	22	1155	338	3760
愛知	24	44	1148	521	7521
三重	18	8	1031	207	1820
滋賀	25	6	935	132	1363
京都	26	15	820	149	2668
大阪	27	47	610	318	9281
兵庫	28	23	1604	346	5348
奈良	29	4	273	110	1260
和歌山	30	7	316	93	983
鳥取	31	4	411	70	589
島根	32	3	495	94	718
岡山	33	14	1141	216	1943
広島	34	12	559	232	2869
山口	35	2	461	173	1444
徳島	36	7	551	88	783
香川	37	17	474	117	998
愛媛	38	9	557	146	1433
高知	39	6	327	70	763
福岡	40	18	1224	400	5078
佐賀	41	6	645	103	852
長崎	42	1	339	141	1423
熊本	43	14	958	225	1810
大分	44	6	595	140	1197
宮崎	45	6	764	163	1136
鹿児島	46	5	771	258	1704
沖縄	47	1	151	98	1392

```
----

. poisson bike lowland dwellings pop

Iteration 0:   log likelihood = -156.83031
Iteration 1:   log likelihood = -153.97721
Iteration 2:   log likelihood = -153.97403
Iteration 3:   log likelihood = -153.97403

Poisson regression                                Number of obs   =         47
                                                  LR chi2(3)      =     286.85
                                                  Prob > chi2     =     0.0000
Log likelihood = -153.97403                       Pseudo R2       =     0.4823

------------------------------------------------------------------------------
        bike |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lowland |  -.0001559   .0000368    -4.23   0.000    -.0002281   -.0000837
   dwellings |   .0042478    .000447     9.50   0.000     .0033716    .0051239
         pop |   .0000519   .0000146     3.56   0.000     .0000234    .0000804
       _cons |   1.309844   .1051302    12.46   0.000     1.103793    1.515896
------------------------------------------------------------------------------

. gen llland=log(lowland)

. gen ldwellings=log(dwellings)

. gen lpop=log(pop)
```

```
. poisson bike llland ldwellings lpop

Iteration 0:   log likelihood = -156.15686
Iteration 1:   log likelihood =  -155.6255
Iteration 2:   log likelihood = -155.62489
Iteration 3:   log likelihood = -155.62489

Poisson regression                                Number of obs   =         47
                                                  LR chi2(3)      =     283.54
                                                  Prob > chi2     =     0.0000
Log likelihood = -155.62489                       Pseudo R2       =     0.4767

------------------------------------------------------------------------------
        bike |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      llland |  -.1028579   .0800629    -1.28   0.199    -.2597784    .0540625
  ldwellings |   .4817018   .2171779     2.22   0.027      .056041    .9073626
        lpop |   .5715923   .1220733     4.68   0.000      .332333    .8108517
       _cons |   -3.93974    .559487    -7.04   0.000    -5.036315   -2.843166
------------------------------------------------------------------------------
```

8 Panel Data

8.1 Some Formulas of Matrix Algebra — Review

1. Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lk} \end{pmatrix} = [a_{ij}]$,

   which is a $l \times k$ matrix, where a_{ij} denotes ith row and jth column of A.

   The **transposed matrix (転置行列)** of A, denoted by A', is defined as:

   $$A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{l1} \\ a_{12} & a_{22} & \cdots & a_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{lk} \end{pmatrix} = [a_{ji}],$$

where the ith row of A' is the ith column of A.

2. $(Ax)' = x'A'$,

   where A and x are a $l \times k$ matrix and a $k \times 1$ vector, respectively.

3. $a' = a$,

   where a denotes a scalar.

4. $\dfrac{\partial a'x}{\partial x} = a$,

   where a and x are $k \times 1$ vectors.

5. $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$,

   where A and x are a $k \times k$ matrix and a $k \times 1$ vector, respectively.

   Especially, when A is symmetric,

$$\frac{\partial x'Ax}{\partial x} = 2Ax.$$

6. Let A and B be $k \times k$ matrices, and I_k be a $k \times k$ **identity matrix (単位行列)** (one in the diagonal elements and zero in the other elements).

   When $AB = I_k$, B is called the **inverse matrix (逆行列)** of A, denoted by $B = A^{-1}$.

   That is, $AA^{-1} = A^{-1}A = I_k$.

7. Let A be a $k \times k$ matrix and x be a $k \times 1$ vector.

   If A is a **positive definite matrix (正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax > 0.$$

   If A is a **positive semidefinite matrix (非負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \geq 0.$$

If A is a **negative definite matrix (負値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax < 0.$$

If A is a **negative semidefinite matrix (非正値定符号行列)**, for any x except for $x = 0$ we have:

$$x'Ax \leq 0.$$

Trace, Rank and etc.: $A : k \times k, \qquad B : n \times k, \qquad C : k \times n.$

1. The **trace (トレース)** of A is: $\mathrm{tr}(A) = \sum_{i=1}^{k} a_{ii}$, where $A = [a_{ij}]$.

2. The **rank (ランク，階数)** of A is the maximum number of linearly independent column (or row) vectors of A, which is denoted by rank(A).

3. If A is an **idempotent matrix (べき等行列)**, $A = A^2$.

4. If A is an idempotent and symmetric matrix, $A = A^2 = A'A$.

5. A is idempotent if and only if the eigen values of A consist of 1 and 0.

6. If A is idempotent, rank(A) =tr(A) .

7. tr(BC) =tr(CB)

Distributions in Matrix Form:

1. Let X, μ and Σ be $k \times 1$, $k \times 1$ and $k \times k$ matrices.

When $X \sim N(\mu, \Sigma)$, the density function of X is given by:

$$f(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big).$$

$\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) = \mathrm{E}\Big((X-\mu)(X-\mu)'\Big) = \Sigma$

The moment-generating function: $\phi(\theta) = \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta)$

(*) In the univariate case, when $X \sim N(\mu, \sigma^2)$, the density function of X is:

$$f(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big).$$

2. If $X \sim N(\mu, \Sigma)$, then $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(k)$.

   Note that $X'X \sim \chi^2(k)$ when $X \sim N(0, I_k)$.

3. X: $n \times 1$, Y: $m \times 1$, $X \sim N(\mu_x, \Sigma_x)$, $Y \sim N(\mu_y, \Sigma_y)$

X is independent of Y, i.e., $\mathrm{E}\Big((X-\mu_x)(Y-\mu_y)'\Big) = 0$ in the case of normal random variables.

$$\frac{(X-\mu_x)'\Sigma_x^{-1}(X-\mu_x)/n}{(Y-\mu_y)'\Sigma_y^{-1}(Y-\mu_y)/m} \sim F(n,m)$$

4. If $X \sim N(0,\sigma^2 I_n)$ and A is a symmetric idempotent $n\times n$ matrix of rank G, then $X'AX/\sigma^2 \sim \chi^2(G)$.

   Note that $X'AX = (AX)'(AX)$ and $\mathrm{rank}(A) = \mathrm{tr}(A)$ because A is idempotent.

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

$$\frac{X'AX}{G\sigma^2}\Big/\frac{X'BX}{K\sigma^2} = \frac{X'AX/G}{X'BX/K} \sim F(G,K).$$

8.2 OLS — Review

We consider the following regression model:

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k} + u_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,k}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + u_i = x_i \beta + u_i,$$

for $i = 1, 2, \cdots, n$, where x_i and β denote a $1 \times k$ vector of the independent variables and a $k \times 1$ vector of the unknown parameters to be estimated, which are given by:

$$x_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,k}), \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

$x_{i,j}$ denotes the ith observation of the jth independent variable.

The case of $k = 2$ and $x_{i,1} = 1$ for all i is exactly equivalent to (4).

Therefore, the matrix form above is a generalization of (4).

Writing all the equations for $i = 1, 2, \cdots, n$, we have:

$$\begin{aligned}
y_1 &= \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_k x_{1,k} + u_1 = x_1\beta + u_1, \\
y_2 &= \beta_1 x_{2,1} + \beta_2 x_{2,2} + \cdots + \beta_k x_{2,k} + u_2 = x_2\beta + u_2, \\
&\vdots \\
y_n &= \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_k x_{n,k} + u_n = x_n\beta + u_n,
\end{aligned}$$

which is rewritten as:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

$$= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Again, the above equation is compactly rewritten as:

$$y = X\beta + u, \tag{19}$$

where y, X and u are denoted by:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Utilizing the matrix form (19), we derive the ordinary least squares estimator of β, denoted by $\hat{\beta}$.

In (19), replacing β by $\hat{\beta}$, we have the following equation:

$$y = X\hat{\beta} + e,$$

where e denotes a $n \times 1$ vector of the residuals.

The ith element of e is given by e_i.

The sum of squared residuals is written as follows:

$$\begin{aligned} S(\hat{\beta}) &= \sum_{i=1}^{n} e_i^2 = e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) = (y' - \hat{\beta}'X')(y - X\hat{\beta}) \\ &= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} = y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}. \end{aligned}$$

In the last equality, note that $\hat{\beta}'X'y = y'X\hat{\beta}$ because both are scalars.

To minimize $S(\hat{\beta})$ with respect to $\hat{\beta}$, we set the first derivative of $S(\hat{\beta})$ equal to zero, i.e.,

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0.$$

Solving the equation above with respect to $\hat{\beta}$, the **ordinary least squares estimator (OLS, 最小自乗推定量)** of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{20}$$

Thus, the ordinary least squares estimator is derived in the matrix form.

(*) Remark

The second order condition for minimization:

$$\frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta} \partial \hat{\beta}'} = 2X'X$$

is a positive definite matrix.

Set $c = Xd$.

For any $d \neq 0$, we have $c'c = d'X'Xd > 0$.

Now, in order to obtain the properties of $\hat{\beta}$ such as mean, variance, distribution and so on, (20) is rewritten as follows:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u. \end{aligned} \qquad (21)$$

Taking the expectation on both sides of (21), we have the following:

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'\mathrm{E}(u) = \beta,$$

because of $\mathrm{E}(u) = 0$ by the assumption of the error term u_i.

Thus, unbiasedness of $\hat{\beta}$ is shown.

The variance of $\hat{\beta}$ is obtained as:

$$\begin{aligned}
\mathrm{V}(\hat{\beta}) &= \mathrm{E}((\hat{\beta}-\beta)(\hat{\beta}-\beta)') = \mathrm{E}\Big((X'X)^{-1}X'u((X'X)^{-1}X'u)'\Big) \\
&= \mathrm{E}((X'X)^{-1}X'uu'X(X'X)^{-1}) = (X'X)^{-1}X'\mathrm{E}(uu')X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}.
\end{aligned}$$

The first equality is the definition of variance in the case of vector.

In the fifth equality, $\mathrm{E}(uu') = \sigma^2 I_n$ is used, which implies that $\mathrm{E}(u_i^2) = \sigma^2$ for all i and $\mathrm{E}(u_i u_j) = 0$ for $i \neq j$.

Remember that $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2.

Under normality assumption on the error term u, it is known that the distribution of $\hat{\beta}$ is given by:

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

Proof:

First, when $X \sim N(\mu, \Sigma)$, the moment-generating function, i.e., $\phi(\theta)$, is given by:

$$\phi(\theta) \equiv \mathrm{E}\Big(\exp(\theta' X)\Big) = \exp\Big(\theta'\mu + \frac{1}{2}\theta'\Sigma\theta\Big)$$

θ_u: $n \times 1$, u: $n \times 1$, θ_β: $k \times 1$, $\hat{\beta}$: $k \times 1$

The moment-generating function of u, i.e., $\phi_u(\theta_u)$, is:

$$\phi_u(\theta_u) \equiv \mathrm{E}\Big(\exp(\theta_u' u)\Big) = \exp\Big(\frac{\sigma^2}{2}\theta_u'\theta_u\Big),$$

which is $N(0, \sigma^2 I_n)$.

The moment-generating function of $\hat{\beta}$, i.e., $\phi_\beta(\theta_\beta)$, is:

$$\begin{aligned}
\phi_\beta(\theta_\beta) &\equiv \mathrm{E}\Big(\exp(\theta'_\beta\hat{\beta})\Big) = \mathrm{E}\Big(\exp(\theta'_\beta\beta + \theta'_\beta(X'X)^{-1}X'u)\Big) \\
&= \exp(\theta'_\beta\beta)\mathrm{E}\Big(\exp(\theta'_\beta(X'X)^{-1}X'u)\Big) = \exp(\theta'_\beta\beta)\phi_u\Big(\theta'_\beta(X'X)^{-1}X'\Big) \\
&= \exp(\theta'_\beta\beta)\exp\Big(\frac{\sigma^2}{2}\theta'_\beta(X'X)^{-1}\theta_\beta\Big) = \exp\Big(\theta'_\beta\beta + \frac{\sigma^2}{2}\theta'_\beta(X'X)^{-1}\theta_\beta\Big),
\end{aligned}$$

which is equivalent to the normal distribution with mean β and variance $\sigma^2(X'X)^{-1}$.

Note that $\theta_u = X(X'X)^{-1}\theta_\beta$. QED

Taking the jth element of $\hat{\beta}$, its distribution is given by:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 a_{jj}), \qquad \text{i.e.,} \qquad \frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{a_{jj}}} \sim N(0, 1),$$

where a_{jj} denotes the jth diagonal element of $(X'X)^{-1}$.

Replacing σ^2 by its estimator s^2, we have the following t distribution:

$$\frac{\hat{\beta}_j - \beta_j}{s \sqrt{a_{jj}}} \sim t(n-k),$$

where $t(n-k)$ denotes the t distribution with $n-k$ degrees of freedom.

● When $\mathrm{Cov}(u, X) \neq 0$, i.e., $\mathrm{E}(X'u) \neq 0$, what happens?

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + (X'X)^{-1}X'u) = \beta + \mathrm{E}((X'X)^{-1}X'u) \neq \beta,$$

because of $\mathrm{E}(X'u) \neq 0$ by the assumption of the error term u_i.

Thus, $\hat{\beta}$ is biased.

[Review] Trace (トレース):

1. A: $n \times n$, $\quad \mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$, where a_{ij} denotes an element in the ith row and the jth column of a matrix A.

2. a: scalar (1×1), $\quad \mathrm{tr}(a) = a$

3. A: $n \times k$, B: $k \times n$, $\quad \mathrm{tr}(AB) = \mathrm{tr}(BA)$

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

5. When X is a square matrix of random variables, $\mathrm{E}(\mathrm{tr}(AX)) = \mathrm{tr}(A\mathrm{E}(X))$

End of Review

8.3 GLS — Review

Regression model:

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

where y, X, β, u, 0 and Ω are $n \times 1$, $n \times k$, $k \times 1$, $n \times 1$, $n \times 1$, and $n \times n$, respectively.

We solve the following minimization problem:

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

Let b be a solution of the above minimization problem.

GLS estimator of β is given by:

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

In general, when Ω is symmetric, Ω is decomposed as follows.

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

Λ is a diagonal matrix, where the diagonal elements of Λ are given by the eigen values.

A is a matrix consisting of eigen vectors.

When Ω is a positive definite matrix, all the diagonal elements of Λ are positive.

There exists P such that $\Omega = PP'$ (i.e., take $P = A'\Lambda^{1/2}$). $\implies P^{-1}\Omega P'^{-1} = I_n$

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

We have:

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

where $\quad y^\star = P^{-1}y, \quad X^\star = P^{-1}X, \quad$ and $\quad u^\star = P^{-1}u.$

The variance of $u^\star$ is:

$$\mathrm{V}(u^\star) = \mathrm{V}(P^{-1}u) = P^{-1}\mathrm{V}(u)P'^{-1} = \sigma^2 P^{-1}\Omega P'^{-1} = \sigma^2 I_n.$$

because $\Omega = PP'$, i.e., $P^{-1}\Omega P'^{-1} = I_n$.

Accordingly, the regression model is rewritten as:

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

Apply OLS to the above model.

Let b be as estimator of β from the above model.

That is, the minimization problem is given by:

$$\min_{b} \; (y^{\star} - X^{\star}b)'(y^{\star} - X^{\star}b),$$

which is equivalent to:

$$\min_{b} \; (y - Xb)'\Omega^{-1}(y - Xb).$$

Solving the minimization problem above, we have the following estimator:

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

which is called GLS (Generalized Least Squares) estimator.

b is rewritten as follows:

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

The mean and variance of b are given by:

$$\mathrm{E}(b) = \beta,$$
$$\mathrm{V}(b) = \sigma^2(X^{\star\prime}X^{\star})^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}.$$

Suppose that the regression model is given by:

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

In this case, when we use OLS, what happens?

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

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

● Compare GLS and OLS.

Expectation:

$$\mathrm{E}(\hat{\beta}) = \beta, \quad \text{and} \quad \mathrm{E}(b) = \beta$$

Thus, both $\hat{\beta}$ and b are unbiased estimator.

Variance:

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

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

Which is more efficient, OLS or GLS?.

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

Ω is the variance-covariance matrix of u, which is a positive definite matrix.

Therefore, except for $\Omega = I_n$, $A\Omega A'$ is also a positive definite matrix.

This implies that $\mathrm{V}(\hat{\beta}_i) - \mathrm{V}(b_i) > 0$ for the ith element of β.

Accordingly, b is more efficient than $\hat{\beta}$.

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

● Maximum Likelihood Estimation (MLE):

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

$\mathrm{E}(y) = X\beta$ and $\mathrm{V}(y) = \sigma^2\Omega$

$$f(y) = (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(y - X\beta)\Omega^{-1}(y - X\beta)'\Big) = L(\beta, \sigma^2)$$

b is equivalent to MLE.

8.4 Panel Model Basic

Model:

$$y_{it} = X_{it}\beta + v_i + u_{it}, \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T$$

where i indicates individual and t denotes time.

There are n observations for each t.

u_{it} indicates the error term, assuming that $\mathrm{E}(u_{it}) = 0$, $\mathrm{V}(u_{it}) = \sigma_u^2$ and $\mathrm{Cov}(u_{it}, u_{js}) = 0$ for $i \neq j$ and $t \neq s$.

v_i denotes the individual effect, which is fixed or random.

8.4.1 Fixed Effect Model (固定効果モデル)

In the case where v_i is fixed, the case of $v_i = z_i\alpha$ is included.

$$y_{it} = X_{it}\beta + v_i + u_{it}, \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T,$$

$$\overline{y}_i = \overline{X}_i\beta + v_i + \overline{u}_i, \qquad i = 1, 2, \cdots, n,$$

where $\overline{y}_i = \dfrac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{X}_i = \dfrac{1}{T}\sum_{t=1}^{T} X_{it}$, and $\overline{u}_i = \dfrac{1}{T}\sum_{t=1}^{T} u_{it}$.

$$(y_{it} - \overline{y}_i) = (X_{it} - \overline{X}_i)\beta + (u_{it} - \overline{u}_i), \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T,$$

Taking an example of y, the left-hand side of the above equation is rewritten as:

$$y_{it} - \overline{y}_i = y_{it} - \frac{1}{T}1_T' y_i,$$

where $1_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, which is a $T \times 1$ vector, and $y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}$.

$$\begin{pmatrix} y_{i1} - \overline{y}_i \\ y_{i2} - \overline{y}_i \\ \vdots \\ y_{iT} - \overline{y}_i \end{pmatrix} = I_T y_i - 1_T \overline{y}_i = I_T y_i - \frac{1}{T} 1_T 1_T' y_i = (I_T - \frac{1}{T} 1_T 1_T') y_i$$

Thus,

$$\begin{pmatrix} y_{i1} - \overline{y}_i \\ y_{i2} - \overline{y}_i \\ \vdots \\ y_{iT} - \overline{y}_i \end{pmatrix} = \begin{pmatrix} X_{i1} - \overline{X}_i \\ X_{i2} - \overline{X}_i \\ \vdots \\ X_{iT} - \overline{X}_i \end{pmatrix} \beta + \begin{pmatrix} u_{i1} - \overline{u}_i \\ u_{i2} - \overline{u}_i \\ \vdots \\ u_{iT} - \overline{u}_i \end{pmatrix}, \qquad i = 1, 2, \cdots, n,$$

which is re-written as:

$$(I_T - \frac{1}{T}1_T 1_T')y_i = (I_T - \frac{1}{T}1_T 1_T')X_i\beta + (I_T - \frac{1}{T}1_T 1_T')u_i, \qquad i = 1, 2, \cdots, n,$$

i.e.,

$$D_T y_i = D_T X_i\beta + D_T u_i, \qquad i = 1, 2, \cdots, n,$$

where $D_T = (I_T - \frac{1}{T}1_T 1_T')$, which is a $T \times T$ matrix.

Note that $D_T D_T' = D_T$, i.e., D_T is a symmetric and idempotent matrix.

Using the matrix form for $i = 1, 2, \cdots, n$, we have:

$$\begin{pmatrix} D_T y_1 \\ D_T y_2 \\ \vdots \\ D_T y_n \end{pmatrix} = \begin{pmatrix} D_T X_1 \\ D_T X_2 \\ \vdots \\ D_T X_n \end{pmatrix}\beta + \begin{pmatrix} D_T u_1 \\ D_T u_2 \\ \vdots \\ D_T u_n \end{pmatrix},$$

i.e.,

$$\begin{pmatrix} D_T & 0 & \cdots & 0 \\ 0 & D_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_T \end{pmatrix} y = \begin{pmatrix} D_T & 0 & \cdots & 0 \\ 0 & D_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_T \end{pmatrix} X\beta + \begin{pmatrix} D_T & 0 & \cdots & 0 \\ 0 & D_T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_T \end{pmatrix} u,$$

where $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $X \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, and $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$, which are $Tn \times 1$, $Tn \times k$ and $Tn \times 1$ matrices, respectively

Using the Kronecker product, we obtain the following expression:

$$(I_n \otimes D_T)y = (I_n \otimes D_T)X\beta + (I_n \otimes D_T)u,$$

where $(I_n \otimes D_T)$, y, X, and u are $nT \times nT$, $nT \times 1$, $nT \times k$, and $nT \times 1$, respectively.

Kronecker Product — Review:

1. A: $n \times m$, B: $T \times k$

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nm}B \end{pmatrix}, \text{ which is a } nT \times mk \text{ matrix.}$$

2. A: $n \times n$, B: $m \times m$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}, \qquad |A \otimes B| = |A|^m |B|^n,$$

$$(A \otimes B)' = A' \otimes B', \qquad \mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B).$$

3. For A, B, C and D such that the products are defined,

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

End of Review

Going back to the previous slide, using the Kronecker product, we obtain the following expression:

$$(I_n \otimes D_T)y = (I_n \otimes D_T)X\beta + (I_n \otimes D_T)u,$$

where $(I_n \otimes D_T)$, y, X, and u are $nT \times nT$, $nT \times 1$, $nT \times k$, and $nT \times 1$, respectively.

Apply OLS to the above regression model.

$$\begin{aligned}
\hat{\beta} &= \Big(((I_n \otimes D_T)X)'(I_n \otimes D_T)X\Big)^{-1}((I_n \otimes D_T)X)'(I_n \otimes D_T)y \\
&= \Big(X'(I_n \otimes D_T'D_T)X\Big)^{-1}X'(I_n \otimes D_T'D_T)y \\
&= \Big(X'(I_n \otimes D_T)X\Big)^{-1}X'(I_n \otimes D_T)y.
\end{aligned}$$

Note that the inverse matrix of D_T is not available, because the rank of D_T is $T-1$, not T (full rank).

The rank of a symmetric and idempotent matrix is equal to its trace.

The fixed effect v_i is estimated as:

$$\hat{v}_i = \bar{y}_i - \overline{X}_i \hat{\beta}.$$

Possibly, we can estimate the following regression:

$$\hat{v}_i = Z_i \alpha + \epsilon_i,$$

where it is assumed that the individual-specific effect depends on Z_i.

The estimator of σ_u^2 is given by:

$$\hat{\sigma}_u^2 = \frac{1}{nT - k - n} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - X_{it}\hat{\beta} - \hat{v}_i)^2.$$

[Remark]

More than ten years ago, “fixed” indicates that v_i is nonstochastic.

Recently, however, “fixed” does not mean anything.

“fixed” indicates that OLS is applied and that v_i may be correlated with X_{it}.

Possibly, $\mathrm{E}(v_i|X) = \alpha_i(X)$, where $\alpha_i(X)$ is a function of X_{it} for $i = 1, 2, \cdots, n$ and $t = 1, 2, \cdots, T$, and it is normalized to $\sum_{i=1}^{n} \alpha_i(X) = 0$.

8.4.2 Random Effect Model (ランダム効果モデル)

Model:

$$y_{it} = X_{it}\beta + v_i + u_{it}, \qquad i = 1, 2, \cdots, n, \quad t = 1, 2, \cdots, T$$

where i indicates individual and t denotes time.

The assumptions on the error terms v_i and u_{it} are:

$$\mathrm{E}(v_i|X) = \mathrm{E}(u_{it}|X) = 0 \ \text{ for all } i,$$

$$\mathrm{V}(v_i|X) = \sigma_v^2 \ \text{ for all } i, \qquad \mathrm{V}(u_{it}|X) = \sigma_u^2 \ \text{ for all } i \text{ and } t,$$

$$\mathrm{Cov}(v_i, v_j|X) = 0 \ \text{ for } i \neq j, \qquad \mathrm{Cov}(u_{it}, u_{js}|X) = 0 \ \text{ for } i \neq j \text{ and } t \neq s,$$

$$\mathrm{Cov}(v_i, u_{jt}|X) = 0 \ \text{ for all } i,\ j \text{ and } t.$$

Note that X includes X_{it} for $i = 1, 2, \cdots, n$ and $t = 1, 2, \cdots, T$.

In a matrix form with respect to $t = 1, 2, \cdots, T$, we have the following:

$$y_i = X_i\beta + v_i 1_T + u_i, \qquad i = 1, 2, \cdots, n,$$

where $y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}$, $X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iT} \end{pmatrix}$ and $u_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}$ are $T \times 1$, $T \times k$ and $T \times 1$, respectively.

$$u_i \sim N(0, \sigma_u^2 I_T) \text{ and } v_i 1_T \sim N(0, \sigma_v^2) \implies v_i 1_T + u_i \sim N(0, \sigma_v^2 1_T 1_T' + \sigma_u^2 I_T).$$

Again, in a matrix form with respect to i, we have the following:

$$y = X\beta + v + u,$$

where $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, $v = \begin{pmatrix} v_1 1_T \\ v_2 1_T \\ \vdots \\ v_n 1_T \end{pmatrix}$ and $u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ are $nT \times 1$, $nT \times k$, $nT \times 1$ and

$nT \times 1$, respectively.

The distribution of $u + v$ is given by:

$$v + u \sim N\Big(0, I_n \otimes (\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)\Big)$$

The likelihood function is given by:

$$\begin{aligned} L(\beta, \sigma_v^2, \sigma_u^2) &= (2\pi)^{-nT/2} \Big| I_n \otimes (\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T) \Big|^{-1/2} \\ &\quad \times \exp\Big(-\frac{1}{2} (y - X\beta)' \Big(I_n \otimes (\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T) \Big)^{-1} (y - X\beta) \Big). \end{aligned}$$

Remember that $f(x) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big)$ when $X \sim N(\mu, \Sigma)$, where X denotes a k-variate random variable.

The estimators of β, σ_v^2 and σ_u^2 are given by maximizing the following log-likelihood function:

$$\log L(\beta,\sigma_v^2,\sigma_u^2) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\log\Big|I_n\otimes(\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)\Big|$$
$$-\frac{1}{2}(y - X\beta)'\Big(I_n\otimes(\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)\Big)^{-1}(y - X\beta).$$

MLE of β, denoted by $\tilde{\beta}$, is given by:

$$\begin{aligned}\tilde{\beta} &= \Big(X'\Big(I_n\otimes(\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)\Big)^{-1}X\Big)^{-1}X'\Big(I_n\otimes(\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)\Big)^{-1}y \\ &= \Big(\sum_{i=1}^{n} X_i'(\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)^{-1}X_i\Big)^{-1}\Big(\sum_{i=1}^{n} X_i'(\sigma_v^2 1_T 1_T' + \sigma_u^2 I_T)^{-1}y_i\Big),\end{aligned}$$

which is equivalent to GLS.

Note that $\tilde{\beta}$ is not operational, because $\hat{\beta}$ depends on σ_v^2 and σ_u^2.

8.5 Consistency and Asymptotic Normality of OLSE — Review

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

Consistency:

1. Let $\hat{\beta}_n = (X'X)^{-1}X'y$ be the OLS with sample size n.

   Consistency: As n is large, $\hat{\beta}_n$ converges to β.

2. Assume the stationarity assumption for X, i.e.,

$$\frac{1}{n}X'X \longrightarrow M_{xx}.$$

   Then, we have the following result:

$$\frac{1}{n}X'u \longrightarrow 0.$$

Proof:

According to Chebyshev's inequality, for $g(Z) \geq 0$,

$$P(g(Z) \geq k) \leq \frac{\mathrm{E}(g(Z))}{k},$$

where k is a positive constant.

Set $g(Z) = Z'Z$, and $Z = \frac{1}{n}X'u$.

Apply Chebyshev's inequality.

$$\begin{aligned}\mathrm{E}\Big((\frac{1}{n}X'u)'\frac{1}{n}X'u\Big) &= \frac{1}{n^2}\mathrm{E}\Big(u'XX'u\Big) = \frac{1}{n^2}\mathrm{E}\Big(\mathrm{tr}(u'XX'u)\Big) = \frac{1}{n^2}\mathrm{E}\Big(\mathrm{tr}(XX'uu')\Big) \\ &= \frac{1}{n^2}\mathrm{tr}\Big(XX'\mathrm{E}(uu')\Big) = \frac{\sigma^2}{n^2}\mathrm{tr}(XX') = \frac{\sigma^2}{n^2}\mathrm{tr}(X'X) = \frac{\sigma^2}{n}\mathrm{tr}(\frac{1}{n}X'X).\end{aligned}$$

Therefore,

$$P\Big((\frac{1}{n}X'u)'\frac{1}{n}X'u \geq k\Big) \leq \frac{\sigma^2}{nk}\mathrm{tr}(\frac{1}{n}X'X) \longrightarrow 0 \times \mathrm{tr}(M_{xx}) = 0.$$

Note that from the assumption,

$$\frac{1}{n}X'X \longrightarrow M_{xx}.$$

Therefore, we have:

$$(\frac{1}{n}X'u)'\frac{1}{n}X'u \longrightarrow 0,$$

which implies:

$$\frac{1}{n}X'u \longrightarrow 0,$$

because $(\frac{1}{n}X'u)'\frac{1}{n}X'u$ indicates a quadratic form.

3. Note that $\frac{1}{n}X'X \longrightarrow M_{xx}$ results in $(\frac{1}{n}X'X)^{-1} \longrightarrow M_{xx}^{-1}$.

$\Longrightarrow$ Slutsky's Theorem

(*) **Slutsky's Theorem** $g(\hat{\theta}) \longrightarrow g(\theta)$, when $\hat{\theta} \longrightarrow \theta$.

4. OLS is given by:

$$\hat{\beta}_n = \beta + (X'X)^{-1}X'u = \beta + (\frac{1}{n}X'X)^{-1}(\frac{1}{n}X'u).$$

Therefore,

$$\hat{\beta}_n \longrightarrow \beta + M_{xx}^{-1} \times 0 = \beta$$

Thus, OLSE is a consitent estimator.

Asymptotic Normality:

1. Asymptotic Normality of OLSE

$$\sqrt{n}(\hat{\beta}_n - \beta) \ \longrightarrow \ N(0,\ \sigma^2 M_{xx}^{-1}), \quad \text{when } n \ \longrightarrow \ \infty.$$

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

   $Z_1, Z_2, \cdots, Z_n$ are mutually indelendently distributed with mean μ and variance Σ_i.

   Then, we have the following result:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i - \mu) \longrightarrow N(0, \Sigma),$$

   where

$$\Sigma = \lim_{n \to \infty} \left(\frac{1}{n} \sum_{i=1}^{n} \Sigma_i \right).$$

   The distribution of Z_i is not assumed.

3. Define $Z_i = x_i' u_i$. Then, $\Sigma_i = \mathrm{Var}(Z_i) = \sigma^2 x_i' x_i$.

4. Σ is defined as:

$$\Sigma = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma^2 x_i' x_i\right) = \sigma^2 \lim_{n\to\infty}\left(\frac{1}{n}X'X\right) = \sigma^2 M_{xx},$$

where

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

5. Applying Central Limit Theorem (Greenberg and Webster (1983), we obtain the following:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i' u_i = \frac{1}{\sqrt{n}}X'u \longrightarrow N(0, \sigma^2 M_{xx}).$$

On the other hand, from $\hat{\beta}_n = \beta + (X'X)^{-1}X'u$, we can rewrite as:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}X'X\right)^{-1}\frac{1}{\sqrt{n}}X'u.$$

$$
\begin{aligned}
\mathrm{Var}\Big(\big(\frac{1}{n}X'X\big)^{-1}\frac{1}{\sqrt{n}}X'u\Big) &= \mathrm{E}\Big(\big(\frac{1}{n}X'X\big)^{-1}\frac{1}{\sqrt{n}}X'u\big(\big(\frac{1}{n}X'X\big)^{-1}\frac{1}{\sqrt{n}}X'u\big)'\Big) \\
&= \big(\frac{1}{n}X'X\big)^{-1}\big(\frac{1}{n}X'\mathrm{E}(uu')X\big)\big(\frac{1}{n}X'X\big)^{-1} \\
&= \sigma^2\big(\frac{1}{n}X'X\big)^{-1} \longrightarrow \sigma^2 M_{xx}^{-1}.
\end{aligned}
$$

Therefore,

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

$\Longrightarrow$ Asymptotic normality (漸近的正規性) of OLSE

The distribution of u_i is not assumed.

8.6 Instrumental Variable (IV) Method (操作変数法 or IV 法) — Review

Instrumental Variable (IV)

1. Consider the regression model: $y = X\beta + u$ and $u \sim N(0, \sigma^2 I_n)$.

   In the case of $\mathrm{E}(X'u) \neq 0$, OLSE of β is inconsistent.

2. **Proof:**

$$\hat{\beta} = \beta + (\frac{1}{n}X'X)^{-1}\frac{1}{n}X'u \longrightarrow \beta + M_{xx}^{-1}M_{xu},$$

   where

$$\frac{1}{n}X'X \longrightarrow M_{xx}, \qquad \frac{1}{n}X'u \longrightarrow M_{xu} \neq 0$$

3. Find the Z which satisfies $\frac{1}{n}Z'u \longrightarrow M_{zu} = 0$.

Multiplying Z' on both sides of the regression model: $y = X\beta + u$,

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

Dividing n on both sides of the above equation, we take plim on both sides.

Then, we obtain the following:

$$\text{plim}\left(\frac{1}{n}Z'y\right) = \text{plim}\left(\frac{1}{n}Z'X\right)\beta + \text{plim}\left(\frac{1}{n}Z'u\right) = \text{plim}\left(\frac{1}{n}Z'X\right)\beta.$$

Accordingly, we obtain:

$$\beta = \left(\text{plim}\left(\frac{1}{n}Z'X\right)\right)^{-1} \text{plim}\left(\frac{1}{n}Z'y\right).$$

Therefore, we consider the following estimator:

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

which is taken as an estimator of β.

$\Longrightarrow$ **Instrumental Variable Method (操作変数法 or IV 法)**

4. Assume the followings:

$$\frac{1}{n}Z'X \longrightarrow M_{zx}, \qquad \frac{1}{n}Z'Z \longrightarrow M_{zz}, \qquad \frac{1}{n}Z'u \longrightarrow 0$$

5. **Asymptotic Distribution of β_{IV}:**

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

which is rewritten as:

$$\sqrt{n}(\beta_{IV} - \beta) = \Big(\frac{1}{n}Z'X\Big)^{-1}\Big(\frac{1}{\sqrt{n}}Z'u\Big)$$

Applying the Central Limit Theorem to $\Big(\frac{1}{\sqrt{n}}Z'u\Big)$, we have the following result:

$$\frac{1}{\sqrt{n}}Z'u \longrightarrow N(0, \sigma^2 M_{zz}).$$

Therefore,

$$\sqrt{n}(\beta_{IV} - \beta) = \Big(\frac{1}{n}Z'X\Big)^{-1}\Big(\frac{1}{\sqrt{n}}Z'u\Big) \longrightarrow N(0, \sigma^2 M_{zx}^{-1} M_{zz} M_{zx}'^{\,-1})$$

$\Longrightarrow$ Consistency and Asymptotic Normality

6. The variance of β_{IV} is given by:

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

where

$$s^2 = \frac{(y - X\beta_{IV})'(y - X\beta_{IV})}{n - k}.$$

8.7 Two-Stage Least Squares Method (2段階最小二乗法, 2SLS or TSLS) — Review

1. Regression Model:

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

   In the case of $\mathrm{E}(X'u) \neq 0$, OLSE is not consistent.

2. Find the variable Z which satisfies $\dfrac{1}{n}Z'u \longrightarrow M_{zu} = 0$.

3. Use $Z = \hat{X}$ for the instrumental variable.

   $\hat{X}$ is the predicted value which regresses X on the other exogenous variables, say W.

   That is, consider the following regression model:

$$X = WB + V.$$

Estimate B by OLS.

Then, we obtain the prediction:

$$\hat{X} = W\hat{B},$$

where $\hat{B} = (W'W)^{-1}W'X$.

Or, equivalently,

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

$\hat{X}$ is used for the instrumental variable of X.

4. The IV method is rewritten as:

$$\beta_{IV} = (\hat{X}'X)^{-1}\hat{X}'y = (X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'y.$$

Furthermore, β_{IV} is written as follows:

$$\beta_{IV} = \beta + (X'W(W'W)^{-1}W'X)^{-1}X'W(W'W)^{-1}W'u.$$

Therefore, we obtain the following expression:

$$\sqrt{n}(\beta_{IV}-\beta) = \Big(\big(\frac{1}{n}X'W\big)\big(\frac{1}{n}W'W\big)^{-1}\big(\frac{1}{n}XW'\big)'\Big)^{-1}\big(\frac{1}{n}X'W\big)\big(\frac{1}{n}W'W\big)^{-1}\big(\frac{1}{\sqrt{n}}W'u\big)$$
$$\longrightarrow N\Big(0, \sigma^2(M_{xw}M_{ww}^{-1}M_{xw}')^{-1}\Big).$$

5. Clearly, there is no correlation between W and u at least in the limit, i.e.,

$$\text{plim}\big(\frac{1}{n}W'u\big) = 0.$$

6. **Remark:**

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

Therefore,

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

which implies the OLS estimator of β in the regression model: $y = \hat{X}\beta + u$ and $u \sim N(0, \sigma^2 I_n)$.

Example:

$$y_t = \alpha x_t + \beta z_t + u_t, \qquad u_t \sim (0, \sigma^2).$$

Suppose that x_t is correlated with u_t but z_t is not correlated with u_t.

- 1st Step:

Estimate the following regression model:

$$x_t = \gamma w_t + \delta z_t + \cdots + v_t,$$

by OLS. $\Longrightarrow$ Obtain $\hat{x}_t$ through OLS.

- 2nd Step:

Estimate the following regression model:

$$y_t = \alpha \hat{x}_t + \beta z_t + u_t,$$

by OLS. $\Longrightarrow$ α_{iv} and β_{iv}

Note as follows. Estimate the following regression model:

$$z_t = \gamma_2 w_t + \delta_2 z_t + \cdots + v_{2t},$$

by OLS.

$\Longrightarrow$ $\hat{\gamma}_2 = 0$, $\hat{\delta}_2 = 1$, and the other coefficient estimates are zeros. i.e., $\hat{z}_t = z_t$.

Eviews Command:

```
tsls y x z @ w z ...
```

Going back to Panel Data:

8.8 Hausman's Specification Error (特定化誤差) Test

Regression model:

$$y = X\beta + u, \qquad y: n\times 1, \quad X: n\times k, \quad \beta: k\times 1, \quad u: n\times 1.$$

Suppose that X is stochastic.

If $\mathrm{E}(u|X) = 0$, OLSE $\hat{\beta}$ is unbiased because of $\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$ and $\mathrm{E}((X'X)^{-1}X'u) = 0$.

However, If $\mathrm{E}(u|X) \neq 0$, OLSE $\hat{\beta}$ is biased and inconsistent.

Therefore, we need to check if X is correlated with u or not.

$\Longrightarrow$ **Hausman's Specification Error Test**

The null and alternative hypotheses are:

- H_0 : X and u are independent, i.e., $\text{Cov}(X, u) = 0$,
- H_1 : X and u are not independent.

Suppose that we have two estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, which have the following properties:

- $\hat{\beta}_0$ is consistent and efficient under H_0, but is not consistent under H_1,
- $\hat{\beta}_1$ is consistent under both H_0 and H_1, but is not efficient under H_0.

Under the conditions above, we have the following test statistic:

$$(\hat{\beta}_1 - \hat{\beta}_0)'\Big(\text{V}(\hat{\beta}_1) - \text{V}(\hat{\beta}_0)\Big)^{-1}(\hat{\beta}_1 - \hat{\beta}_0) \longrightarrow \chi^2(k).$$

Example: $\hat{\beta}_0$ is OLS, while $\hat{\beta}_1$ is IV such as 2SLS.

Hausman, J.A. (1978) "Specification Tests in Econometrics," *Econometrica*, Vol.46, No.6, pp.1251–1271.

8.9 Choice of Fixed Effect Model or Random Effect Model

8.9.1 The Case where X is Correlated with u — Review

The standard regression model is given by:

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

OLS is:

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

If X is not correlated with u, i.e., $\mathrm{E}(X'u) = 0$, we have the result: $\mathrm{E}(\hat{\beta}) = \beta$.

However, if X is correlated with u, i.e., $\mathrm{E}(X'u) \neq 0$, we have the result: $\mathrm{E}(\hat{\beta}) \neq \beta$.
$\Longrightarrow$ $\hat{\beta}$ is biased.

Assume that in the limit we have the followings:

$$
\begin{aligned}
(\frac{1}{n}X'X)^{-1} &\longrightarrow M_{xx}^{-1}, \\
\frac{1}{n}X'u &\longrightarrow M_{xu} \neq 0 \text{ when } X \text{ is correlated with } u.
\end{aligned}
$$

Therefore, even in the limit,

$$
\text{plim}\,\hat{\beta} = \beta + M_{xx}^{-1}M_{xu} \neq \beta,
$$

which implies that $\hat{\beta}$ is not a consistent estimator of β.

Thus, in the case where X is correlated with u, OLSE $\hat{\beta}$ is neither unbiased nor consistent.

8.9.2 Fixed Effect Model or Random Effect Model

Usually, in the random effect model, we can consider that v_i is correlated with X_{it}.

[Reason:]

v_i includes the unobserved variables in the ith individual, i.e., ability, intelligence, and so on.

X_{it} represents the observed variables in the ith individual, i.e., income, assets, and so on.

The unobserved variables v_i are related to the observed variables X_{it}.

Therefore, we consider that v_i is correlated with X_{it}.

Thus, in the case of the random effect model, usually we cannot use OLS or GLS.

In order to use the random effect model, we need to test whether v_i is uncorrelated with X_{it}.

Apply Hausman's test.

- H_0 : X_{it} and e_{it} are independent ($\longrightarrow$ Use the random effect model),
- H_1 : X_{it} and e_{it} are not independent ($\longrightarrow$ Use the fixed effect model),

where $e_{it} = v_i + u_{it}$.

Example of Panel Data:

Production Function of Prefectures from 2001 to 2010.

pref: 都道府県（通し番号 1〜47）

year: 年度（2001〜2010 年）

y : 県内総生産（支出側、実質：固定基準年方式），出所：県民経済計算（平成 13 年度 - 平成 24 年度）（93SNA，平成 17 年基準計数）

k : 都道府県別民間資本ストック（平成 12 暦年価格，年度末，国民経済計算ベース　平成 23 年 3 月時点）一期前（2000〜2009 年）

l : 県内就業者数，出所:県民経済計算（平成 13 年度 - 平成 24 年度）（93SNA，平成 17 年基準計数）

```
. tsset pref year
       panel variable:  pref (strongly balanced)
        time variable:  year, 2001 to 2010
                delta:  1 unit
```

```
. gen ly=log(y)

. gen lk=log(k)

. gen ll=log(l)

. reg ly lk ll

      Source |       SS       df       MS              Number of obs =     470
-------------+------------------------------           F(  2,   467) =19374.95
       Model |  316.479302     2  158.239651           Prob > F      =  0.0000
    Residual |  3.81409572   467  .008167229           R-squared     =  0.9881
-------------+------------------------------           Adj R-squared =  0.9880
       Total |  320.293398   469  .682928354           Root MSE      =  .09037

------------------------------------------------------------------------------
          ly |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |   .0941587   .0081273    11.59   0.000     .0781881    .1101294
          ll |   .9976399   .0102641    97.20   0.000     .9774703    1.017809
       _cons |   .5970719   .0773137     7.72   0.000     .4451461    .7489978
------------------------------------------------------------------------------

. xtreg ly lk ll,fe

Fixed-effects (within) regression               Number of obs      =       470
Group variable: pref                            Number of groups   =        47
```

```
R-sq:  within  = 0.1721                         Obs per group: min =        10
       between = 0.9456                                        avg =      10.0
       overall = 0.9439                                        max =        10

                                                F(2,421)           =     43.77
corr(u_i, Xb)  = 0.8803                         Prob > F           =    0.0000

------------------------------------------------------------------------------
          ly |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |   .2329208   .0252321     9.23   0.000     .1833242    .2825175
          ll |   .3268537   .0810662     4.03   0.000     .1675088    .4861987
       _cons |   7.691145   1.376677     5.59   0.000     4.985128    10.39716
-------------+----------------------------------------------------------------
     sigma_u |  .41045507
     sigma_e |  .03561437
         rho |  .99252757   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(46, 421) =    56.22             Prob > F = 0.0000

. est store fixed

. xtreg ly lk ll,re

Random-effects GLS regression                   Number of obs      =       470
Group variable: pref                            Number of groups   =        47

R-sq:  within  = 0.1058                         Obs per group: min =        10
       between = 0.9805                                        avg =      10.0
```

```
     overall = 0.9787                                        max =         10

                                                Wald chi2(2)       =   3875.75
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          ly |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |   .2457767   .0153094    16.05   0.000     .2157708    .2757827
          ll |   .8105099   .0220256    36.80   0.000     .7673406    .8536793
       _cons |   .8332015   .2411141     3.46   0.001     .3606265    1.305776
-------------+----------------------------------------------------------------
     sigma_u |    .081609
     sigma_e |  .03561437
         rho |   .8400205   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. hausman fixed

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
          lk |    .2329208     .2457767       -.0128559         .020057
          ll |    .3268537     .8105099       -.4836562        .0780167
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg
```

```
Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      144.66
                Prob>chi2 =      0.0000
```

9　ノンパラメトリック回帰

回帰モデル

$$y_i = X_i\beta + u_i$$

通常，線形を仮定 → この仮定を緩める

$$y_i = m(x_i) + u_i$$

$m(\cdot)$ が未知 → $m(\cdot)$ を推定 → ノンパラメトリック回帰

準備として，密度関数が未知で，密度関数を推定する

9.1　密度関数のノンパラメトリック推定

分布を仮定しない → 分布関数自体を推定 → ノンパラメトリック推定

ノンパラメトリック推定に関するテキスト：

- Pagan, A. and Ullah, A., (1999),
  Nonparametric Econometrics, Cambridge University Press.
- Prakasa Rao, B.L.S., (1983),
  Nonparametric Functional Estimation, Academic Press, Inc.
- Silverman, B.W., (1986),
  Density Estimation for Statistics and Data Analysis (Monographs on Statistics and Applied Probability, No.26), Chapman & Hall.

同じ分布からの n 個の観測値 x_i, $i = 1, 2, \cdots, n$, があるとする。

x_i の密度関数を $f(x)$ とする。

密度関数 $f(x)$ が未知であるとする。

密度関数 $f(x)$ の推定値を $\hat{f}(x)$ とする。

n 個のデータから密度関数を次のように推定することができる。

密度関数 $f(x)$ と分布関数 $F(x)$ の関係：

$$f(x) = \lim_{h \to 0} \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h} \approx \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h}$$

$F(x) = \mathrm{P}(X \leq x) = x$ 以下の確率

$$\hat{f}(x) = \frac{F(x + \frac{h}{2}) - F(x - \frac{h}{2})}{h}$$

$$\approx \frac{1}{nh} [\ n \text{ 個の観測値 } x_1, \cdots, x_n \text{ の中で } (x - \frac{h}{2},\ x + \frac{h}{2}) \text{ の範囲に入っている個数 }]$$

$$= \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h}) \tag{22}$$

ただし，h はある小さな数で，バンド幅 (Band Width)，窓幅 (Windows Width)，平滑化定数 (Smoothing Parameter) などと様々な呼び名があるが，以降では，h のことをバンド幅と呼ぶことにする。

また，(22) に含まれる $K(\cdot)$ は，

$$K(z) = \begin{cases} 1 & |z| < \dfrac{1}{2} \text{ のとき} \\ 0 & \text{その他} \end{cases}$$

と表される。→ この場合，$K(\cdot)$ は一様分布 → Rectangle Kernel と呼ばれる。

$K(\cdot)$ はカーネル（Kernel）と呼ばれる。

しかし，より一般的には，$\int K(t)\,\mathrm{d}t = 1$，かつ，すべての t について $K(t) \geq 0$ となるような $K(\cdot)$ を選べばよい。

$K(\cdot)$ の選択として，代表的なものは標準正規分布である。

問題は h の選び方である。

h を大きくすると，分布関数は滑らかに近似され真の分布関数とは異なったものとなってしまう。

逆に h を小さくすると，分布関数は必要以上に凸凹になり，真の分布関数とはかけ離れたものになる。

よって，適切な h が選ばれる必要がある。

密度関数 $f(x)$ の推定量は実現値 x_i をその確率変数 X_i で置き換えて，

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{23}$$

と書くことができる。

特に区別を必要としない限り，密度関数 $f(x)$ の推定値と推定量は同じ記号 $\hat{f}(x)$ を使うことにする。

さらに，x が p 次元の場合は x の密度関数の推定値は次のように表される。

$$\hat{f}(x) = \frac{1}{nh^p}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

しかし，ここでは簡単化のため，$p = 1$ の場合のみを扱うことにする。

9.2　バンド幅 h の選択

準備として，平均自乗誤差の積分 (Integrated Mean Square Error，IMSE) を考える。

それは次のように変形することができる。

$$\begin{aligned} \mathrm{IMSE}(\hat{f}(x)) &\equiv \int \mathrm{MSE}(\hat{f}(x))\,\mathrm{d}x \equiv \int \mathrm{E}(\hat{f}(x) - f(x))^2\,\mathrm{d}x \\ &= \int (\mathrm{E}(\hat{f}(x)) - f(x))^2\,\mathrm{d}x + \int \mathrm{Var}(\hat{f}(x))\,\mathrm{d}x \end{aligned} \tag{24}$$

(24) に含まれる期待値，分散は，密度関数 $f(x)$ の推定量 $\hat{f}(x)$ に含まれる確率変数 X_i，$i = 1, 2, \cdots, n$ に関する期待値，分散であることに注意せよ。

また，X_i は密度関数 $f(x)$ を持つ確率変数である。

よって，(23) を用いて，$\mathrm{E}(\hat{f}(x))$，$\mathrm{Var}(\hat{f}(x))$ は次のように書き換えられる。

$$\begin{aligned}\mathrm{E}(\hat{f}(x)) &= \mathrm{E}\Big(\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h})\Big) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big(\frac{1}{h}K(\frac{x-X_i}{h})\Big) = \mathrm{E}\Big(\frac{1}{h}K(\frac{x-X}{h})\Big) \\ &= \int \frac{1}{h}K(\frac{x-y}{h})f(y)\,\mathrm{d}y \end{aligned} \tag{25}$$

$$\begin{aligned}\mathrm{Var}(\hat{f}(x)) &= \mathrm{Var}\Big(\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h})\Big) = \sum_{i=1}^{n}\mathrm{Var}\Big(\frac{1}{nh}K(\frac{x-X_i}{h})\Big) \\ &= \frac{1}{n}\mathrm{Var}\Big(\frac{1}{h}K(\frac{x-X}{h})\Big) = \frac{1}{n}\mathrm{E}\Big(\frac{1}{h}K(\frac{x-X}{h})\Big)^2 - \frac{1}{n}\Big(\mathrm{E}\Big(\frac{1}{h}K(\frac{x-X}{h})\Big)\Big)^2 \\ &= \frac{1}{n}\int \frac{1}{h^2}K(\frac{x-y}{h})^2 f(y)\,\mathrm{d}y - \frac{1}{n}\Big(\int \frac{1}{h}K(\frac{x-y}{h})f(y)\,\mathrm{d}y\Big)^2 \end{aligned} \tag{26}$$

簡単化のために，カーネル $K(\cdot)$ は，

$$\int K(t)\,\mathrm{d}t = 1 \qquad \int tK(t)\,\mathrm{d}t = 0 \qquad \int t^2K(t)\,\mathrm{d}t = k_2 \neq 0 \tag{27}$$

を満たす左右対称な関数とする。

また，未知の密度関数 $f(x)$ はすべての次数で微分可能な連続関数とする。

$k_2 = 1$ と置くこともできる。

$y = x - ht$ と変数変換を行い，(25) を次のように書き直す。

$$\mathrm{E}(\hat{f}(x)) = \int \frac{1}{h}K(\frac{x-y}{h})f(y)\,\mathrm{d}y = \int K(t)f(x-ht)\,\mathrm{d}t \tag{28}$$

$ht = 0$ の近傍で $f(x - ht)$ をテーラー展開すると，次のようになる。

$$f(x-ht) = f(x) - htf'(x) + \frac{1}{2}h^2t^2f''(x) + \frac{1}{6}h^3t^3f'''(x) + O(h^4)$$

(28) に代入すると，

$$\begin{aligned} \mathrm{E}(\hat{f}(x)) &= \int K(t)\Big(f(x) - htf'(x) + \frac{1}{2}h^2t^2f''(x) + \frac{1}{6}h^3t^3f'''(x) + O(h^4)\Big)\,\mathrm{d}t \\ &= f(x) + \frac{1}{2}h^2f''(x)k_2 + O(h^4) \end{aligned} \tag{29}$$

が得られる (カーネル $K(\cdot)$ は左右対称と仮定しているので，1 行目の式の右辺の第 1 項と第 3 項はゼロになることに注意)。

同様に，$y = x - ht$ と変数変換を行うと，(26) の第 1 項，第 2 項は，それぞれ，

$$\begin{aligned} \frac{1}{n}\int \frac{1}{h^2}K\Big(\frac{x-y}{h}\Big)^2 f(y)\,\mathrm{d}y &= \frac{1}{n}\int \frac{1}{h}K(t)^2 f(x-ht)\,\mathrm{d}t \\ &= \frac{1}{n}\int \frac{1}{h}K(t)^2\Big(f(x) - htf'(x) + O(h^2)\Big)\,\mathrm{d}t \\ &= \frac{1}{nh}f(x)\int K(t)^2\,\mathrm{d}t + O(n^{-1}) \end{aligned} \tag{30}$$

$$\frac{1}{n}\Big(\int \frac{1}{h}K(\frac{x-y}{h})f(y)\,\mathrm{d}y\Big)^2 = \frac{1}{n}(f(x) + \frac{1}{2}h^2 f''(x)k_2 + O(h^4))^2 = \frac{1}{n}(f(x)^2 + O(h^2))$$
$$= O(n^{-1}) \tag{31}$$

となる (n が大きくなるにつれて，h は小さくなると考える)。

よって，$\mathrm{Var}(\hat{f}(x))$ は，(26) に (30) と (31) を代入して，

$$\mathrm{Var}(\hat{f}(x)) = \frac{1}{nh}f(x)\int K(t)^2\,\mathrm{d}t + O(n^{-1}) = \frac{1}{nh}\Big(f(x)\int K(t)^2\,\mathrm{d}t + O(h)\Big) \tag{32}$$

と表される。(29) と (32) を用いて，$\mathrm{MSE}(\hat{f}(x))$ は次のように近似される。

$$\mathrm{MSE}(\hat{f}(x)) = \mathrm{Var}(\hat{f}(x)) + (\mathrm{E}(\hat{f}(x)) - f(x))^2$$
$$\approx \frac{1}{nh}f(x)\int K(t)^2\,\mathrm{d}t + \frac{1}{4}h^4 k_2^2 f''(x)^2 \tag{33}$$

また，(24) の $\mathrm{IMSE}(\hat{f}(x))$ は，(33) の $\mathrm{MSE}(\hat{f}(x))$ の近似を x について積分する

ことにより，

$$\mathrm{IMSE}(\hat{f}(x)) = \int \mathrm{MSE}(\hat{f}(x))\,\mathrm{d}x$$
$$\approx \frac{1}{nh}\int K(t)^2\,\mathrm{d}t + \frac{1}{4}h^4 k_2^2 \int f''(x)^2\,\mathrm{d}x \tag{34}$$

が得られる。

(34) が最小になる h を $\hat{h}$ とすると，

$$\hat{h} = k_2^{-2/5}\Big(\int K(t)^2\,\mathrm{d}t\Big)^{1/5}\Big(\int f''(x)^2\,\mathrm{d}x\Big)^{-1/5} n^{-1/5} \tag{35}$$

となる。

また，(27) の条件を満たし，(34) の IMSE($\hat{f}(x)$) を最小にするカーネル $K(\cdot)$ は，

$$K(t) = \begin{cases} \frac{3}{4}(1-t^2) & -1 \le t \le 1 \\ 0 & \text{その他} \end{cases} \tag{36}$$

となり，Epanechnikov カーネルと呼ばれる (Epanechnikov (1969) を参照せよ)。ただし，(36) は IMSE($\hat{f}(x)$) の近似 (34) を最小にしたものであり，必ずしも最適なカーネルとは限らないことに注意せよ。

以上をもとにして，バンド幅 h の選択として考えられるものとしては，大別して，(24) や尤度関数に基づくものと (35) に基づくものの 2 つがあげられる。前者はクロス・バリデーション，後者はプラグ・イン法として知られている。

ここでは，クロス・バリデーションに基づく方法を 2 つ，プラグ・イン法に基づく方法を 2 つの合計 4 つのバンド幅 h の推定を紹介する。

9.2.1　クロス・バリデーション (Cross-Validation)

クロス・バリデーション (Cross-Validation) には 2 つの種類がある。一つは (24) を最小にする h を求めるという最小自乗クロス・バリデーション (Least-Squares

Cross-Validation) と呼ばれるものであり，もう一つは尤度関数を最大にする h を求めるという尤度クロス・バリデーション (Likelihood Cross-Validation) である。2 つを以下に簡単に説明しておく。

最小自乗クロス・バリデーション (Least-Squares Cross-Validation): (24) の IMSE($\hat{f}(x)$) は，

$$\mathrm{IMSE}(\hat{f}(x)) = \mathrm{E}\int \hat{f}(x)^2\,\mathrm{d}x - 2\mathrm{E}\int \hat{f}(x)f(x)\,\mathrm{d}x + \int f(x)^2\,\mathrm{d}x \tag{37}$$

と分解される。(37) の第 3 項は h に依存しないため，第 2 項までを最小にする h を求めればよい。期待値 $E(\cdot)$ を無視すると，第 1 項，第 2 項は次のようになる。

$$\mathrm{CV}(h) \equiv \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int K(\frac{x_i - x_j}{h} - t)K(t)\,\mathrm{d}t - 2\frac{1}{n}\sum_{i=1}^{n}\hat{f}_{-i}(x_i) \tag{38}$$

ただし，$\hat{f}_{-i}(x_i)$ は，

$$\hat{f}_{-i}(x_i) \equiv \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x_i - x_j}{h}\right)$$

である。

(38) の第 1 項の評価は，モンテ・カルロ積分や数値積分が考えられる。モンテ・カルロ積分を利用すると，$K(\cdot)$ から m 個の乱数 $t_1, t_2, \cdots, t_m$ を発生させて，次のように CV(h) が計算される。

$$\mathrm{CV}(h) = \frac{1}{n^2h^2}\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} K\left(\frac{x_i - x_j}{h} - t_k\right) - \frac{2}{n(n-1)h}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x_i - x_j}{h}\right) \quad (39)$$

$K(\cdot)$ は，分析者によって特定される密度関数であるので，乱数生成の簡単な関数を選べばよい。

尤度クロス・バリデーション (Likelihood Cross-Validation): 対数尤度関数 $\log L$ は次のように近似される。

$$\log L = \sum_{i=1}^{n} \log f(x_i) \approx \sum_{i=1}^{n} \log \hat{f}_{-i}(x_i) = \sum_{i=1}^{n} \log\Big(\frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^{n} K(\frac{x_i - x_j}{h})\Big) \tag{40}$$

(40) が最大になるような h を選べばよい。(39) や (40) に基づいて h の推定値 $\hat{h}$ を求める方法は，単純探査法 (Simle Grid Search) 等を用いる必要がある。

9.2.2　プラグ・イン法 (Plug-In Method)

(35) は，近似された IMSE($\hat{f}(x)$) を最小にするような h である。問題は，(35) において未知の部分は $f''(x)$ であるということである。最も簡単な方法は，未知の密度関数 $f(\cdot)$ を正規分布 (分散を σ^2 とする) で近似してしまうことである。すなわち，

$$\hat{h} = k_2^{-2/5}\Big(\int K(t)^2\, \mathrm{d}t\Big)^{1/5}\Big(\frac{3}{8\sqrt{\pi}}\sigma^{-5}\Big)^{-1/5} n^{-1/5} \tag{41}$$

となる。ただし，$k_2 \equiv \int t^2 K(t)\,\mathrm{d}t = 1$ とする。σ は $x_1, x_2, \cdots, x_n$ から得られる標本不偏分散の平方根 (すなわち，標本標準偏差) に置き換えられる。$f(x)$ が分散 σ^2 の正規分布のとき，$\int f''(x)^2\,\mathrm{d}x = \frac{3}{8\sqrt{\pi}}\sigma^{-5}$ と計算されることに注意せよ。

また，$f(\cdot)$ を $\hat{f}(\cdot)$ で置き換えると，次のようになる。

$$\hat{h} = k_2^{-2/5}\Big(\int K(t)^2\,\mathrm{d}t\Big)^{1/5}\Big(\int \hat{f}''(x)^2\,\mathrm{d}x\Big)^{-1/5} n^{-1/5} \tag{42}$$

(42) の $\int \hat{f}''(x)^2\,\mathrm{d}x$ の評価について，カーネル $K(\cdot)$ を標準正規分布としたとき，

$$\begin{aligned}\int \hat{f}''(x)^2\,\mathrm{d}x &= \frac{1}{n^2h^5}\sum_{i=1}^{n}\sum_{j=1}^{n}\int\Big(\big(\frac{x_i - x_j}{h} - t\big)^2 - 1\Big)(t^2 - 1)K\big(\frac{x_i - x_j}{h} - t\big)K(t)\,\mathrm{d}t \\ &\approx \frac{1}{n^2h^5m}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\Big(\big(\frac{x_i - x_j}{h} - t_k\big)^2 - 1\Big)(t_k^2 - 1)K\big(\frac{x_i - x_j}{h} - t_k\big)\end{aligned} \tag{43}$$

となる。ただし，$t_1, t_2, \cdots, t_m$ は $K(t)$ から生成された乱数とする。問題は，(43) が h に依存することである。次節では，$K(\cdot)$ を標準正規分布とおいて，(41) を用いて，h を求めると，$1.06\hat{\sigma}n^{-1/5}$ となるので，これを (43) の中の h とした。

(39) や (43) と同様に，$\int K(t)^2\,\mathrm{d}t$ は，モンテ・カルロ積分や数値積分で評価される。$K(t)$ から生成された $t_1, t_2, \cdots, t_m$ をもとにして，

$$\int K(t)^2\,\mathrm{d}t = \mathrm{E}(K(X)) \approx \frac{1}{m}\sum_{i=1}^{n} K(t_k) \tag{44}$$

と計算される。ただし，X は密度関数 $K(\cdot)$ を持つ確率変数とする。次節では。カーネルには (36) と標準正規分布を用いるので，$\int K(t)^2\,\mathrm{d}t$ を明示的に求めることができる。(44) のようなモンテ・カルロ積分による積分の近似は必要としない。すなわち，$K(\cdot)$ を (36) とすると，$\int K(t)^2\,\mathrm{d}t = 0.6$ となり，$K(\cdot)$ を標準正規分布とすると，$\int K(t)^2\,\mathrm{d}t = (2\sqrt{\pi})^{-1}$ となる。

(39)，(43)，(44) の積分値を求める場合に，誤差を減らすためのテクニックを以下に記しておく。$K(t)$ から生成された乱数 $t_1, t_2, \cdots, t_m$ は，既に，小さい順に並べ替えられているものとする。$t_1, t_2, \cdots, t_m$ は乱数なので，等確率で現れるもの

と離散近似することが出来る。すなわち，初期値 t_1 は

$$\int_{-\infty}^{t_1} K(t)\,\mathrm{d}t = \frac{1}{2m}$$

を満たす値として求め，$t_2, t_3, \cdots, t_m$ は

$$\int_{t_{k-1}}^{t_k} K(t)\,\mathrm{d}t = \frac{1}{m} \qquad k = 2, 3, \cdots, m$$

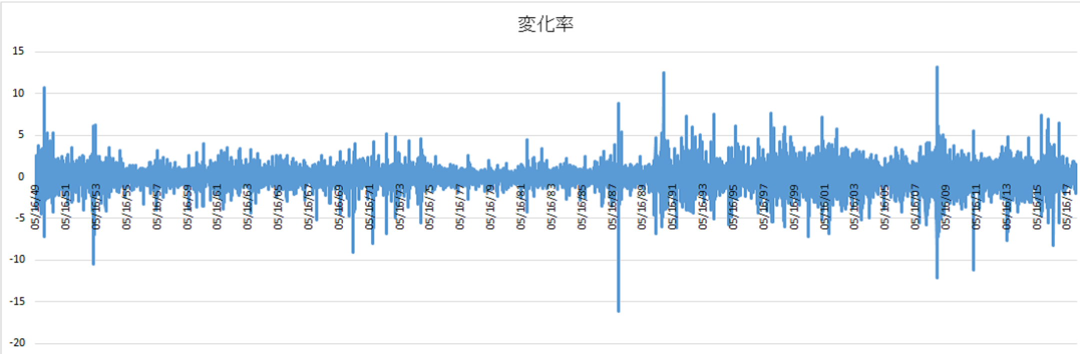
を満たす値として順次求められる。言い換えると，t_k は，密度関数 $K(\cdot)$ の $100\times(k-0.5)/m$ % 点に相当する。このようにして $t_1, t_2, \cdots, t_m$ の値を定めることによって，(39)，(43)，(44) の積分値を求めるためのシミュレーション誤差を減らすことができる。

以上のように，h の推定としては，(39) を最小にする h，(40) を最大にする h，(41)，(43) を用いた (42) の 4 通りを紹介した。(39) や (40) に基づいて h の推定値 $\hat{h}$ を求める方法は，単純探査法 (Simle Grid Search) を用いる必要があるため，

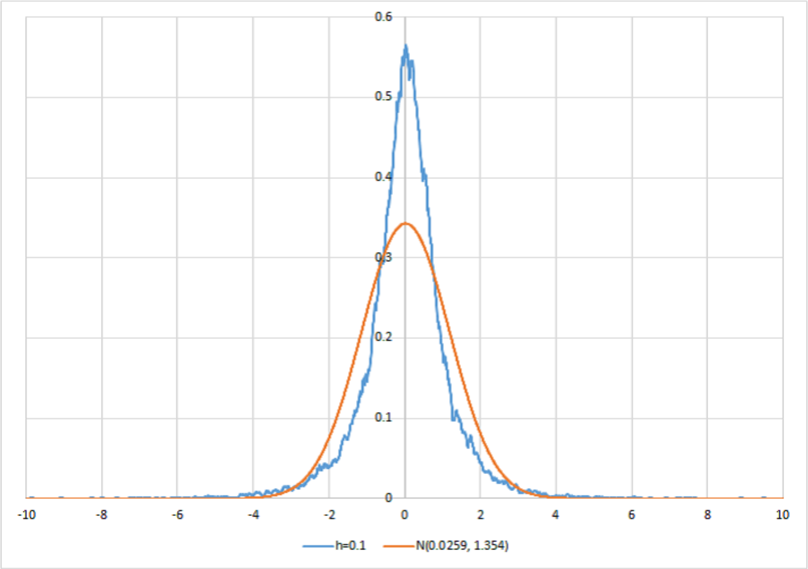
計算時間が非常にかかる。特に，(39) には，$\sum$ が 3 つも含まれているため，単純探査法による h の最適点を求めることはほとんど不可能に近い。

グラフの例：

日経平均株価指数　1949 年 5 月 16 日～2017 年 12 月 29 日の日時データ

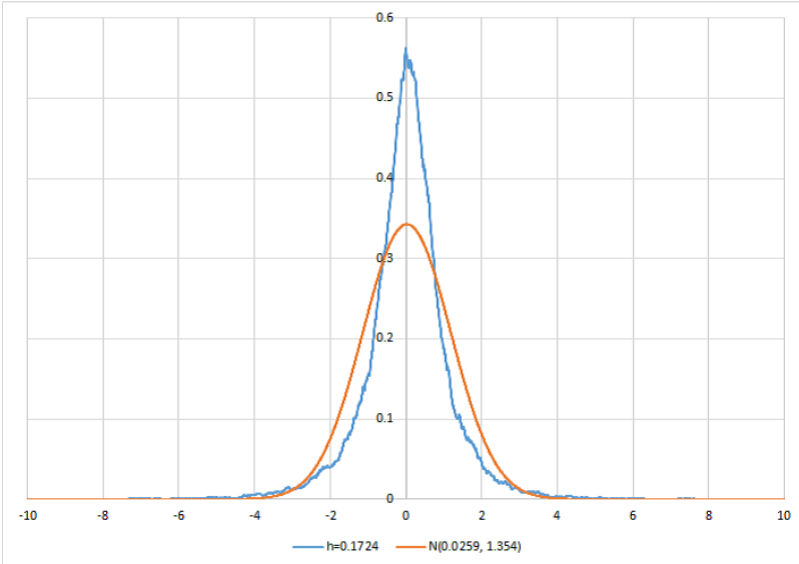


上昇率　$100\ln(x_i/x_{i-1})$,　$i = 1, 2, \cdots, n$,　$n = 18767$



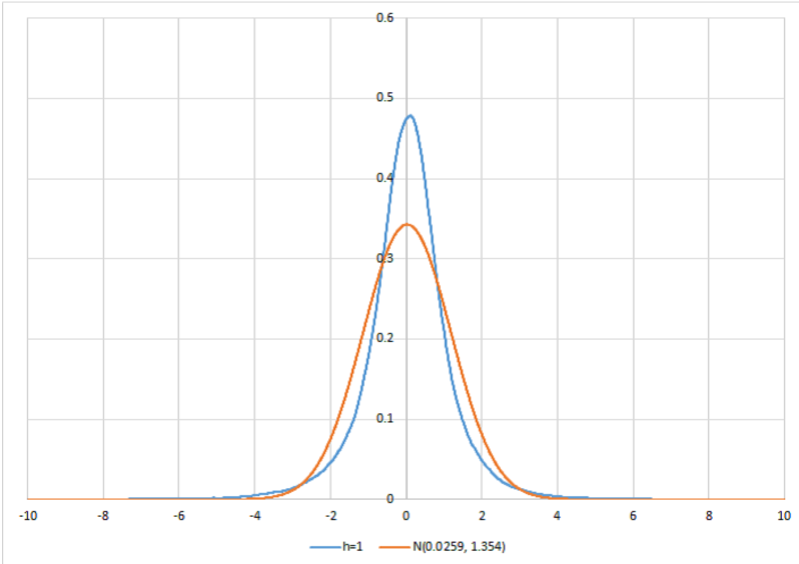
青線： $h = 0.1$ で密度関数を推定

赤線： $N(0.0259,\ 1.354)$ $\longleftarrow$ データからの平均と分散による正規分布



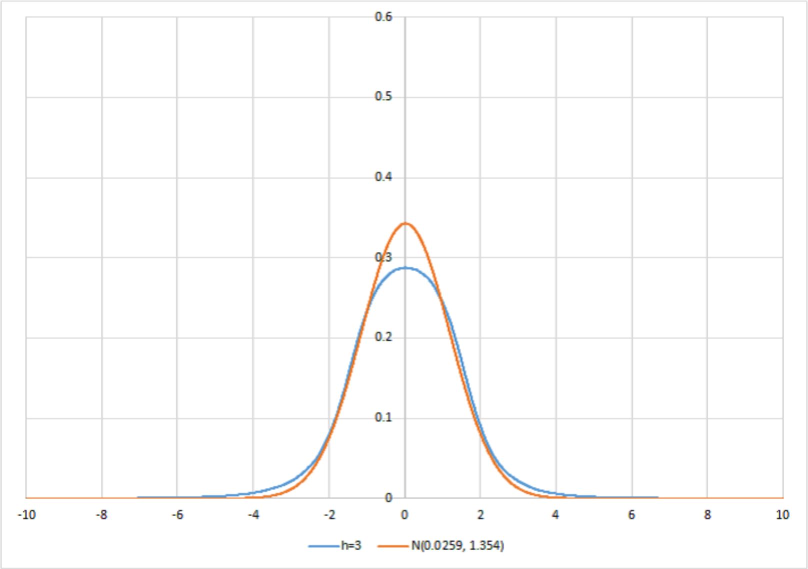
青線： $h = 0.17$ で密度関数を推定 $\longleftarrow$ $h = 1.06\hat{\sigma}n^{-1/5} \approx 0.17$，$\hat{\sigma}^2 = 1.354$

赤線： $N(0.0259,\ 1.354)$ $\longleftarrow$ データからの平均と分散による正規分布



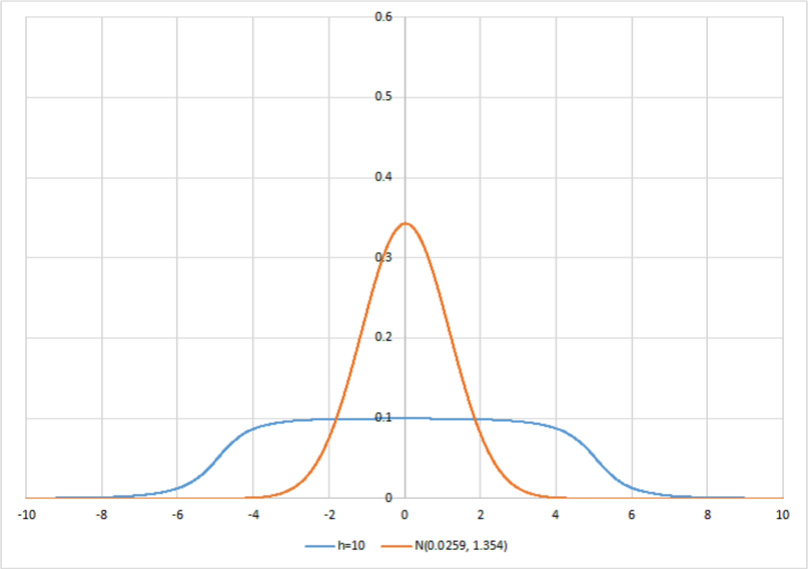
青線： $h = 1$ で密度関数を推定

赤線： $N(0.0259,\ 1.354) \longleftarrow$ データからの平均と分散による正規分布



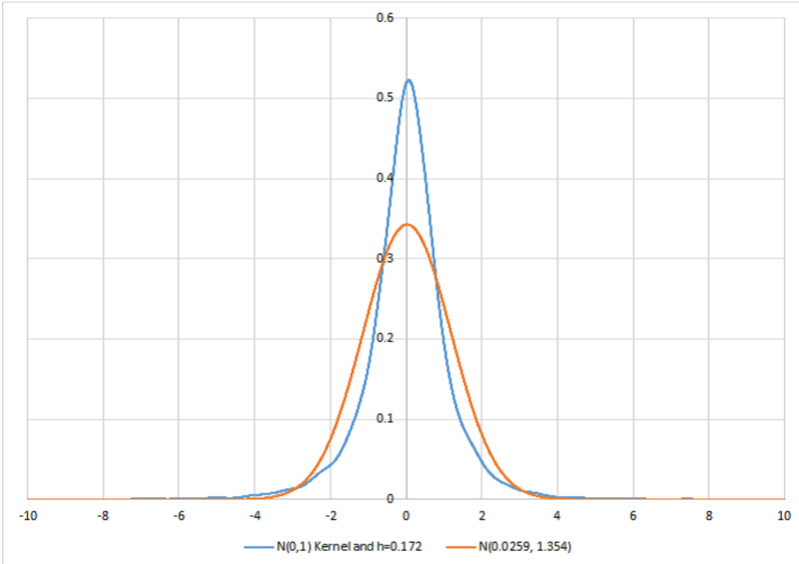
青線：　$h = 3$ で密度関数を推定

赤線：　$N(0.0259,\ 1.354) \longleftarrow$ データからの平均と分散による正規分布



青線：　$h = 10$ で密度関数を推定

赤線：　$N(0.0259,\ 1.354) \longleftarrow$ データからの平均と分散による正規分布



青線： $N(0,1)$ Kernel，$h = 0.17$ で密度関数を推定 $\longleftarrow$ $h = 1.06\hat{\sigma}n^{-1/5} \approx 0.17$, $\hat{\sigma}^2 = 1.354$

赤線： $N(0.0259,\ 1.354)$ $\longleftarrow$ データからの平均と分散による正規分布

赤線，青線のプログラム (Fortran 77)

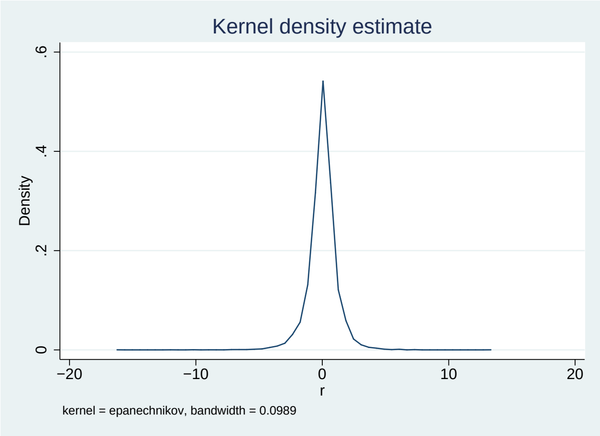
```
      implicit real*8 (a-h,o-z)
      dimension y(100000)
      open(1,file='r.txt')
      read(1,*)
        do 1 i=1,100000
    1 read(1,*,end=2) y(i)
    2 close(1)
      n=i-1

      a=0.0
      v=0.0
        do 3 i=1,n
      a=a+y(i)/float(n)
    3 v=v+y(i)*y(i)/float(n)
      se=sqrt(v-a*a)

      h=1.06*se*( float(n)**(-0.2) )
c     h=1
```

```
      write(1,4) h,a,se*se
    4 format(3f15.10)

        do 5 i=-1000,1000
      x=float(i)/100.
      sum=0.0
         do 6 j=1,n
      z=(x-y(j))/h
      if( abs(z).le.0.5 ) sum=sum+1.
    6 continue
      f=sum/(float(n)*h)
      fn=exp(-0.5*(x-a)*(x-a)/(se*se))/sqrt(2.*3.141592*se*se)
    5 write(1,7) x,f,fn
    7 format(f7.2,2f15.10)

      end
```

do 6 j=1,n から 6 continue の部分が，Rectangle Kernel に対応する。

Stata による結果：

```
. kdensity r
```



デフォルトで (36) の Epanechnikov カーネルが選択され，$h = 0.0989$

9.3 回帰分析への応用

y_i を被説明変数 (スカラー)，x_i を説明変数 ($1 \times k$ ベクトル) とする。

回帰モデルは，関数形を特定化しないで，次のように表される。

$$y_i = m(x_i) + u_i, \qquad i = 1, 2, \cdots, n \tag{45}$$

$m(\cdot)$ が未知 → $m(\cdot)$ を推定 → ノンパラメトリック回帰

ただし，誤差項 u_i は平均 0，分散 $\sigma^2(x)$ のある分布に従うものとする。

y と x の密度関数を $f_{yx}(y, x)$，x の密度関数を $f(x)$ とするとき，y の条件付期待値は,

$$\mathrm{E}(y|x) = m(x) = \int \frac{y f_{yx}(y, x)}{f(x)} \mathrm{d}y$$

となり，$m(x)$ によって表される。

$f(x)$ は,

$$f(x) = \int f_{yx}(y, x) \mathrm{d}y$$

となる。$\longrightarrow$ $f(x)$ は周辺分布

2 つの密度関数 $f_{yx}(y,x)$，$f(x)$ の推定量を $\hat{f}_{yx}(y,x)$，$\hat{f}(x)$ として，

$$\hat{f}_{yx}(y,x) = \frac{1}{nh^{k+1}}\sum_{i=1}^{n} K_{yx}\Big(\frac{y-y_i}{h},\frac{x-x_i}{h}\Big)$$

$$\begin{aligned}
\hat{f}(x) &\equiv \int \hat{f}_{yx}(y,x)\mathrm{d}y = \int \frac{1}{nh^{k+1}}\sum_{i=1}^{n} K_{yx}\Big(\frac{y-y_i}{h},\frac{x-x_i}{h}\Big)\mathrm{d}y \\
&= \frac{1}{nh^{k+1}}\sum_{i=1}^{n}\int K_{yx}\Big(\frac{y-y_i}{h},\frac{x-x_i}{h}\Big)\mathrm{d}y = \frac{1}{nh^{k+1}}\sum_{i=1}^{n} h\int K_{yx}\Big(u,\frac{x-x_i}{h}\Big)\mathrm{d}u \\
&= \frac{1}{nh^{k}}\sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big) \qquad (46)
\end{aligned}$$

とする。

$\hat{f}(x)$ の 4 番目の等式では，$u = \dfrac{y-y_i}{h}$ として変数変換。

$\hat{f}(x)$ の 5 番目の等式では，$\displaystyle\int K_{yx}(z)\mathrm{d}y = K(x)$ を利用。

$z=(y,x)$ として，$K_{yx}(z)$ と $K(x)$ の性質は,

$$\int K_{yx}(z)\mathrm{d}z=1, \qquad \int zK_{yx}(z)\mathrm{d}z=0, \qquad \int zz'K_{yx}(z)\mathrm{d}z=\Omega_*,$$
$$\int K(x)\mathrm{d}x=1, \qquad \int xK(x)\mathrm{d}x=0, \qquad \int xx'K(x)\mathrm{d}x=\Omega,$$
$$\int K_{yx}(z)\mathrm{d}y=K(x)$$

として表される。

Ω_* は $(k+1)\times(k+1)$ 行列，Ω は $k\times k$ 行列である。Ω_* と Ω との関係は,

$$\Omega_*=\begin{pmatrix}\omega_y^2 & \cdots \\ \vdots & \Omega\end{pmatrix}$$

となるものとする。

実証分析では，ω_y^2 は $y_i,\ i=1,2,\cdots,n$ から得られた標本分散，Ω は $x_1,x_2,\cdots,x_n$ から計算された $(1/n)\sum_{i=1}^n x_i'x_i$ とする。

このとき，$m(x)$ の推定量 $\hat{m}(x)$ は，$f_{yx}(y,x)$ と $f(x)$ を $\hat{f}_{yx}(y,x)$ と $\hat{f}(x)$ で置き換えると，

$$\hat{m}(x) = \int \frac{y\hat{f}_{yx}(y,x)}{\hat{f}(x)}\mathrm{d}y = \frac{\frac{1}{nh^{k+1}}\sum_{i=1}^{n}\int yK_{yx}\left(\frac{y-y_i}{h}, \frac{x-x_i}{h}\right)\mathrm{d}y}{\frac{1}{nh^k}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)}$$

$$= \frac{\frac{1}{nh^k}\sum_{i=1}^{n}\int (y_i + hu)K_{yx}\left(u, \frac{x-x_i}{h}\right)\mathrm{d}u}{\frac{1}{nh^k}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)}$$

$$= \frac{\frac{1}{nh^k}\sum_{i=1}^{n}y_iK\left(\frac{x-x_i}{h}\right)}{\frac{1}{nh^k}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right)} = \frac{\hat{g}(x)}{\hat{f}(x)} = \frac{\sum_{i=1}^{n}y_iK\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{x-x_j}{h}\right)} = \sum_{i=1}^{n}y_i\omega_i(x) \qquad (47)$$

となる。

2 行目では $(y-y_i)/h=u$ として変数変換を行っている。

3 行目では $\int uK_{yx}(u,v)\mathrm{d}u=0$ が用いられている。

ただし，分子の $\hat{g}(x)$ を

$$\hat{g}(x)=\frac{1}{nh^k}\sum_{i=1}^{n}y_iK\left(\frac{x-x_i}{h}\right)$$

と定義する。

また，

$$\omega_i(x)=\frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^{n}K\left(\frac{x-x_j}{h}\right)}$$

とする。

$\hat{m}(x)$ の漸近的性質は，$nh^k \longrightarrow \infty$ のとき，次のようになることが知られている。

$$(nh^k)^{1/2}(\hat{m}(x) - m(x)) \sim N\left(0, \frac{\sigma^2(x)}{f(x)} \int K^2(u)\mathrm{d}u\right) \tag{48}$$

ただし，

$$\sigma^2(x) = \int \frac{(y - m(x))^2 f(y, x)}{f(x)} \mathrm{d}y$$

とする。

このように，$\hat{m}(x)$ は $m(x)$ の一致推定量となり，漸近的に正規分布に従うことになる。

言い換えると，n が大きいとき，近似的に，

$$\hat{m}(x) \sim N\Bigl(m(x),\ \frac{\sigma^2(x)}{f(x)}\frac{1}{nh^k}\int K^2(u)\mathrm{d}u\Bigr)$$

となる。

実証分析において，$\hat{m}(x)$ の漸近分散は，(48) 式の $\sigma^2(x)$ と $f(x)$ をその推定量 $\hat{\sigma}^2(x)$，$\hat{f}(x)$ で置き換えて，求められる。

すなわち，

$$\hat{m}(x) \sim N\Bigl(m(x),\ \frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\frac{1}{n\hat{h}^k}\int K^2(u)\mathrm{d}u\Bigr)$$

が利用される。

実践では，$m(x)$ の $100(1-\alpha)\%$ 信頼区間は，

$$\left(\hat{m}(x) - z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\int K^2(u)\mathrm{d}u\Big)^{1/2},\quad \hat{m}(x) + z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\int K^2(u)\mathrm{d}u\Big)^{1/2}\right)$$

となる。

$z_{\alpha/2}$ は $N(0,1)$ の上側 $100\alpha/2\%$ 点とする ($\alpha = 0.05$ のとき，$z_{\alpha/2} = 1.96$)。

ただし，$\hat{f}(x) = \frac{1}{nh^k}\sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big)$，

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} y_i K(\frac{x-x_i}{\hat{h}})}{\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}$$

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n} (y_i - \hat{m}(x_i))^2 K(\frac{x-x_i}{\hat{h}})}{\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}, \qquad \hat{h} = 1.06 s_{x_i} n^{-1/5}$$

とする (上の $\hat{h}$ は 1 つの例，s_{x_i} は i 番目の x 変数の標準偏差)。

さらに，$K(\cdot)$ を $N(0,1)$ と仮定すれば,

$$\int K^2(u)\mathrm{d}u = \int\Big(\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)\Big)^2\mathrm{d}u = \int\frac{1}{2\pi}\exp(-u^2)\mathrm{d}u$$
$$= \frac{1}{2\sqrt{\pi}}\int\frac{1}{\sqrt{2\pi(1/2)}}\exp(-\frac{1}{2(1/2)}u^2)\Big)^2\mathrm{d}u = \frac{1}{2\sqrt{\pi}}$$

となる。

最後の等式は $N(0,1/2)$ を利用。

(*) $N(\mu,\sigma^2)$ の密度関数は,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$$

である。

よって，この場合，$m(x)$ の $100(1-\alpha)\%$ 信頼区間は，

$$\left(\hat{m}(x) - z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\frac{1}{2\sqrt{\pi}}\Big)^{1/2},\quad \hat{m}(x) + z_{\alpha/2}\Big(\frac{1}{n\hat{h}^k}\frac{\hat{\sigma}^2(x)}{\hat{f}(x)}\frac{1}{2\sqrt{\pi}}\Big)^{1/2}\right) \tag{49}$$

再度，まとめると，$\hat{f}(x)$，$\hat{m}(x)$，$\hat{\sigma}^2(x)$，$\hat{h}$ は，

$$\hat{f}(x) = \frac{1}{nh^k}\sum_{i=1}^{n} K\Big(\frac{x-x_i}{h}\Big), \qquad \hat{m}(x) = \frac{\sum_{i=1}^{n} y_i K(\frac{x-x_i}{\hat{h}})}{\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}$$

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n}(y_i - \hat{m}(x_i))^2 K(\frac{x-x_i}{\hat{h}})}{\sum_{j=1}^{n} K(\frac{x-x_j}{\hat{h}})}, \qquad \hat{h} = 1.06 s_{x_i} n^{-1/5}$$

上の $\hat{h}$ は 1 つの例，s_{x_i} は i 番目の x 変数の標準偏差

さらに，$K(u)$ を次のように $N(0,\Omega)$ の多変数正規分布を仮定する ($\Omega = I_k$ でも構わない)。

$$K(u) = (2\pi)^{-k/2}|\Omega|^{-1/2}\exp\Bigl(-\frac{1}{2}u'\Omega^{-1}u\Bigr) \tag{50}$$

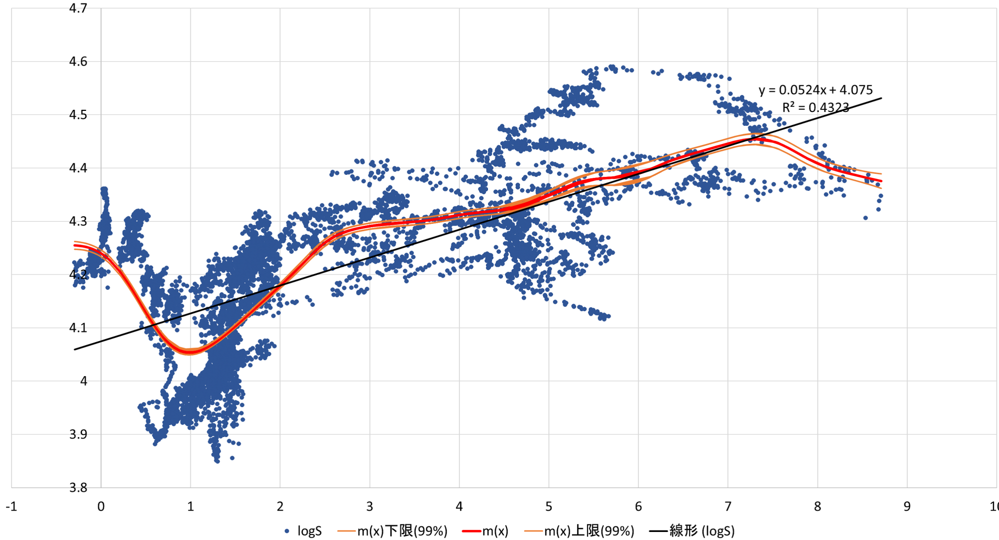
このとき，(48) 式の積分値は,

$$\begin{aligned}\int K^2(u)\mathrm{d}u &= \int (2\pi)^{-k}|\Omega|^{-1}\exp\bigl(-u'\Omega^{-1}u\bigr)\mathrm{d}u \\ &= (2\pi)^{-k/2}|2\Omega|^{-1/2}\int (2\pi)^{-k/2}\Bigl|\frac{1}{2}\Omega\Bigr|^{-1/2}\exp\Bigl(-\frac{1}{2}u'\Bigl(\frac{1}{2}\Omega\Bigr)^{-1}u\Bigr)\mathrm{d}u \\ &= 2^{-k}\pi^{-k/2}|\Omega|^{-1/2}\end{aligned}$$

として計算される。

日経平均株価指数の対数 vs 国債利回り：　推定期間：1986.2.1〜2017.12.29,

データ数：$n = 7932$，$h = 1.06\hat{\sigma}n^{-1/5} \approx 0.343565$，$\hat{\sigma}^2 = 1.95245$



横軸：国債利回り，縦軸：日経平均株価指数の対数

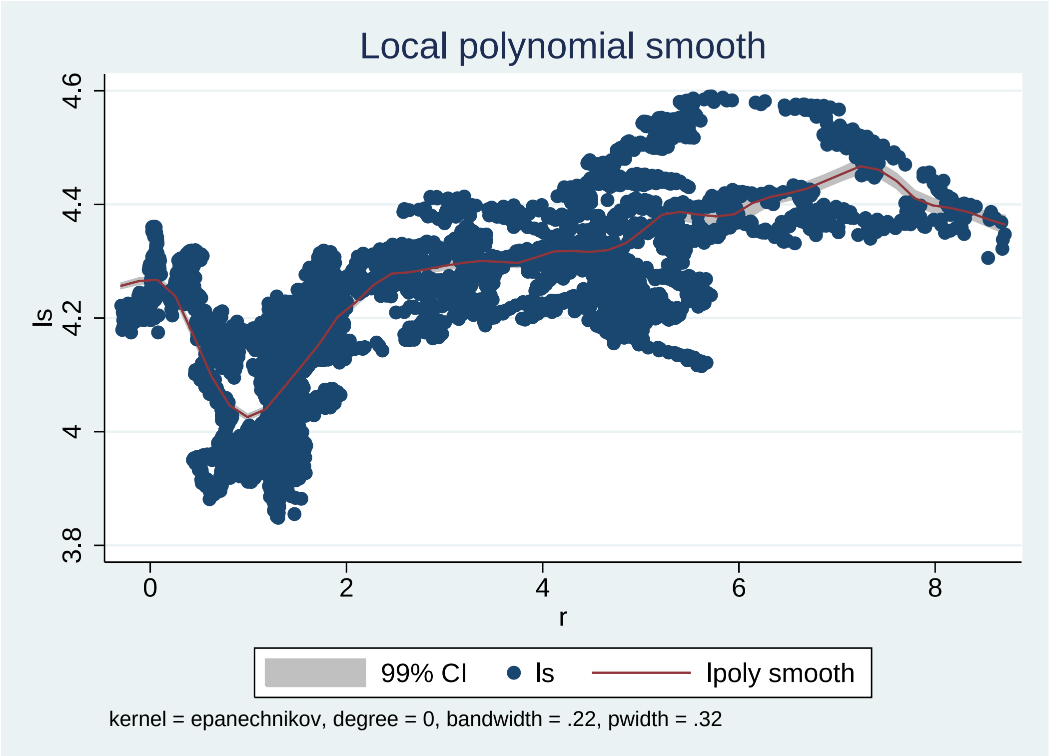
Stata による結果：

```
. lpoly ls r, ci l(99)
```

`ls` は日経平均株価椎の対数，`r` は国債利回り

`ci` はオプションで，95 %信頼区間 (Confidence Interval)

`l(99)` は `ci` のオプションで，99 %信頼区間 (無ければ，デフォルトで 95 %信頼区間)



横軸：国債利回り，縦軸：日経平均株価指数の対数