LS: Least Squares 线性回归模型

y	$N \times 1$
X	$N \times k$
β	$k \times 1$
u	$N \times 1$

例:

y	x_0	x_1	x_2
4	1	1	2
5	1	3	3
6	1	2	7

$$4 = \beta_0 + \beta_1 + 2\beta_2$$
$$5 = \beta_0 + 3\beta_1 + 3\beta_2$$
$$6 = \beta_0 + 2\beta_1 + 7\beta_2$$

3个未知量, 3个等式 => 直接求解.

例:

y	x_0	x_1	x_2
4	1	1	2
5	1	3	3
6	1	2	7
7	3	4	5

3个未知量, 4个等式 => 没法求出唯一解.

找到 $\beta_0, \beta_1, \beta_2$, 使得每行误差之和最小

(Least Squares)

因为总是有 $N > K$，必有残差.

$$\underset{N\times 1}{y} = \underset{N\times K}{X}\,\underset{K\times 1}{\beta} + \underset{N\times 1}{u}$$

- y 是一个结果，由 X, β, u 组成
- X 是观测到的值，本身已知
- β 表示真实的参数，即通过总体数据 y 与 X，能拟合出的结果
- u 表示真实的残差 Disturbance
  也是总体的产物.

目标：找到 b 作为 β 的 estimator

使得 $(y-\hat{y})^T(y-\hat{y})$ 最小.

$$y = X\beta + u$$
$$\hat{y} = Xb$$

目标是寻找 b，但在这之前要做假设，否则不好找.

线性 $y = X\beta + u$ — [X与y是线性 / X内部无共线性 Multi colinearity]

X不随机 $E[X] = X$

u {
正态分布

零残差 $E[u|X] = 0$ — [$E[u] = 0$ / u与X不相关 $u^TX = 0$]

方差 $\sigma^2 I$ — [同方差 constant Homoskedasticity / 无自相关 non-auto correlation]

*
$$\begin{array}{c|ccc} & u_1 & u_2 & u_3 \\ \hline u_1 & \sigma^2 & 0 & 0 \\ u_2 & 0 & \sigma^2 & 0 \\ u_3 & 0 & 0 & \sigma^2 \end{array}$$

线性 $\boxed{y = X\beta + u}$ —
- X与y是线性
- X内部无共线性 Multi colinearity

X不随机 $\boxed{E[X] = X}$

u
- 正态分布
- 零残差 $E[u|X] = 0$ —
  - $E[u] = 0$
  - u与X不相关
- 方差 $\sigma^2 I$ —
  - 同方差 constant Homoskedasticity
  - 无自相关 non-autocorrelation

- LS 模型必备
- OLS 的简化

Ordinary	OLS	→ PLS 允许共线性.
Weighted	WLS	允许方差不 constant $\sigma^2 \begin{bmatrix} a & & \\ & b & \\ & & c \end{bmatrix}$
Generalized	GLS	允许方差不 constant 且有自关性 $\sigma^2 \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

Fact:

$$y = X\beta + u$$
$$\hat{y} = Xb$$
$$\varepsilon = y - \hat{y}$$

$$E[X] = X$$
$$E[u|X] = 0$$
$$Var(u) = \sigma^2 I$$

求 b. 因为只有样本.

注意: 可变量是 b, 它是函数 f(b)

$\Sigma(y-\hat{y})^2$ 最小

y $N\times 1$

X $N\times K$

b $K\times 1$

y^TXb 1×1

b^TX^Ty 1×1

$$y^TXb = (y^TXb)^T = b^TX^Ty$$

$$\min_b (y-Xb)^T(y-Xb)$$
$$= \min_b \left(y^Ty - y^TXb - b^TX^Ty + b^TX^TXb \right)$$
$$= \min_b \left(y^Ty - 2y^TXb + b^TX^TXb \right)$$
$$= \min_b f(b)$$

$$\frac{\partial f}{\partial b} = \frac{标量}{列向量\ K\times 1} = 列向量\ K\times 1 = \begin{bmatrix} \frac{\partial f}{\partial b_1} \\ \frac{\partial f}{\partial b_2} \\ \vdots \end{bmatrix}$$

$$\frac{\partial y^Ty}{\partial b} = 0$$

$$\frac{\partial\, -2y^TXb}{\partial b} = \Big(-2\underset{1\times N}{y^T}\underset{N\times K}{X} \Big)^T = -2\underset{K\times N}{X^T}\underset{N\times 1}{y}$$

$$\frac{\partial\, b^TX^TXb}{\partial b} = 2\underset{K\times N}{X^T}\underset{N\times K}{X}\underset{K\times 1}{b}$$

$$\frac{\partial f}{\partial b} = -2X^Ty + 2X^TXb = 0$$

$$X^Ty = X^TXb$$

$$b = (X^TX)^{-1}X^Ty, \quad \text{if } X^TX \text{ invertible.}$$

满秩. $N>K$

不存在 $X_1=2X_2$

共线性

找最值: 一阶导为0, 二阶导 >0

$$\frac{\partial^2 f}{\partial b \partial b^T} = \frac{\partial}{\partial b^T}\left(\frac{\partial f}{\partial b}\right) = \frac{\text{列向量 } k\times 1}{\text{行向量 } 1\times k} = k\times k$$

$$= \frac{\partial \begin{bmatrix} \frac{\partial f}{\partial b_1} \\ \frac{\partial f}{\partial b_2} \end{bmatrix}}{\partial \text{行}} \overset{\text{例}}{=} \begin{bmatrix} \frac{\partial^2 f}{\partial b_1 \partial b_1} & \frac{\partial^2 f}{\partial b_1 \partial b_2} \\ \frac{\partial^2 f}{\partial b_2 \partial b_1} & \frac{\partial^2 f}{\partial b_2 \partial b_2} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial b \partial b^T} = \frac{\partial -2X^Ty + 2X^TXb}{\partial b^T} = 2X^TX$$

$$\frac{\partial(-2X^Ty)}{\partial b^T} = 0$$

$$\frac{\partial\, 2X^TXb}{\partial b^T} = 2\underset{k\times N}{X^T}\underset{N\times k}{X}$$

X^TX 正定 $\longrightarrow$ 大于 0

Positie Definite

所以，$b = (X^TX)^{-1}X^Ty$ 是最小值的解.

b 作为 estimator，不是完全精确的. 有方差.

每次采样，都会有一个 b 值.

b $K\times 1$

$var(b)$ $K\times K$ 矩阵

协方差矩阵

$$var(b) = var\left((X^TX)^{-1}X^Ty \right)$$

$$= var\left((X^TX)^{-1}X^T(X\beta + u) \right)$$

$$= var\left(\underbrace{\beta}_{\text{确定的未知量}} + \underbrace{(X^TX)^{-1}X^T}_{\text{不随机}} \underbrace{u}_{var(b)\text{ 来源于 } var(u)} \right)$$

$var(Ax)$

$= A\, var(x)\, A^T$

$$= (X^TX)^{-1}X^T \underbrace{var(u)}_{=\sigma^2 I} \left((X^TX)^{-1}X^T \right)^T$$

$$= \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1}$$

b 是什么, $\mathrm{var}(b)$

还剩下 $E[b]$

高斯—马尔可夫定理.

在线性回归模型中,

如果误差满足零均差、同方差、不相关,

则 $b = (X^TX)^{-1}X^Ty$ 是 BLUE.

→ 不要求 u 服从正态分布

→ B : Best 方差最小

L : Linear

U : Unbiased $E[b] = \beta$

E : Estimator

证明 BLUE.

先写下已知事实

$$y = X\beta + u \qquad E[u|X] = 0$$

$$\hat{y} = Xb \qquad var(u) = \sigma^2 I$$

$$b = (X^T X)^{-1} X^T y$$

$$var(b) = \sigma^2 (X^T X)^{-1}$$

(1) 证明 unbiased $E[b] = \beta$.

$$
\begin{aligned}
& E[b] \\
&= E[(X^T X)^{-1} X^T y] \\
&= E[(X^T X)^{-1} X^T (X\beta + u)] \\
&= E[\underbrace{\beta}_{\text{已知}} + \underbrace{(X^T X)^{-1} X^T}_{\text{已知}} \underbrace{u}_{E[u]=0}] \\
&= E[\beta] \\
&= \beta.
\end{aligned}
$$

(2) 证明 Best.

反证: $\exists\ \tilde{\beta} = ((X^T X)^{-1} X^T + D)\, y$

$var(\tilde{\beta}) < var(b)$

$E[\tilde{\beta}] = \beta.$

$E[((X^T X)^{-1} X^T + D)\, y]$

$= E[((X^T X)^{-1} X^T + D)(X\beta + u)]$

$= E[\beta + (X^T X)^{-1} X^T u + DX\beta + Du]$

$= \beta + DX\beta$

$= (I + DX)\beta$

$\therefore\ I + DX = I$

$DX = 0$

$$\mathrm{var}(\tilde{\beta}) = \mathrm{var}\left(\left((X^TX)^{-1}X^T + D\right)(X\beta + u) \right)$$

$$= \mathrm{var}\left(\underbrace{\beta}_{\text{常量}} + (X^TX)^{-1}X^Tu + \underbrace{DX}_{=0}\beta + Du \right)$$

$$= \mathrm{var}\left((X^TX)^{-1}X^Tu + Du \right)$$

$$(\text{Want}) \quad < \mathrm{var}\left((X^TX)^{-1}X^Tu \right)$$

加一个Du绝不可能让

var 变更小

$*$ $\mathrm{var}(X-Y) = \mathrm{var}X + \mathrm{var}Y$

补充:

X^TX 何时 invertible : 满秩, $N>K$

$var(AX) = A\, varX\, A^T$

X^TX (半)正定

二阶导正定,即为 >0

$varX = E[(X-EX)(X-EX)^T]$

GLS

$\mathrm{var}(b) \neq \sigma^2 I$ $\begin{cases} \text{unequal variance} \\ \text{auto correlated} \end{cases}$

If so, $(X^T X)^{-1} X^T y$ is still unbiased, but not best.

Transform into a new set of observations that satisfy assumptions, then use OLS.

Given $\mathrm{var}(b) = \sigma^2 \Sigma$, where $\Sigma = K^T K = KK$

Define
$$y^* = K^{-1} y$$
$$X^* = K^{-1} X$$
$$u^* = K^{-1} u$$

$$E[u^*] = K^{-1} E[u] = 0$$
$$\mathrm{Var}(u^*) = K^{-1} \mathrm{Var}(u) K^{-1} = K^{-1} \sigma^2 V K^{-1}$$
$$= \sigma^2 K^{-1} K K K^{-1}$$
$$= \underline{\sigma^2 I}$$

Proceed with OLS

$$\min_b (y^* - X^* b)^T (y^* - X^* b)$$
$$= \min_b (y^T K^{-1} - b^T X^T K^{-1})(K^{-1} y - K^{-1} X b)$$
$$= \min_b (y^T - b^T X^T) K^{-1} K^{-1} (y - Xb)$$
$$= \min_b (y^T - b^T X^T) \Sigma^{-1} (y - Xb)$$
$$= \min_b y^T \Sigma^{-1} y - y^T \Sigma^{-1} X b - b^T X^T \Sigma^{-1} y + b^T X^T \Sigma^{-1} X b$$

$$= \min_b \; y^T\Sigma^{-1}y - 2y^T\Sigma^{-1}Xb + b^TX^T\Sigma^{-1}Xb$$

$$= \min_b f(b)$$

$$\frac{\partial f}{\partial b} = -2X^T\Sigma^{-1}y + 2X^T\Sigma^{-1}Xb = 0$$

$$X^T\Sigma^{-1}Xb = X^T\Sigma^{-1}y$$

$$b = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y$$

BLUE again

(1) unbiased : $E[b] = \beta$.

$$E[b] = E[(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}(X\beta+u)]$$

$$= E[\beta + (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}u]$$

$$= \beta.$$

(2) best :

first find variance, then 反推 (略)

$$Var(b) = Var((X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}u)$$

$$= (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\underbrace{Var(u)}_{\sigma^2\Sigma}\Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}$$

$$= \sigma^2(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}$$

$$= \sigma^2(X^T\Sigma^{-1}X)^{-1}$$

GLS 我们须假设已知 Σ，而这不现实

退而求其次，只允许 var(u) 变化 $\sigma^2\begin{bmatrix} a & & \\ & b & \\ & & c \end{bmatrix}$

If so,

$$b = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y$$

例 $\Sigma = \begin{bmatrix} 1 & & \\ & 2 & \\ & & 3 \end{bmatrix}$

$$\Sigma^{-1} = \begin{bmatrix} 1 & & \\ & \frac{1}{2} & \\ & & \frac{1}{3} \end{bmatrix}$$

$$X_{3\times 2} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

本来 $X^TX = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$

$$= \begin{bmatrix} 1^2+3^2+5^2 & 1\times 2+3\times 4+5\times 6 \\ 1\times 2+3\times 4+5\times 6 & 2^2+4^2+6^2 \end{bmatrix}$$

现在 $X^T\Sigma^{-1}X$

$$= \begin{bmatrix} 1^2\times\frac{1}{1}+3^2\times\frac{1}{2}+5^2\times\frac{1}{3} & 1\times 2\times\frac{1}{1}+3\times 4\times\frac{1}{2}+5\times 6\times\frac{1}{3} \\ \cdots & 2^2\times\frac{1}{1}+4^2\times\frac{1}{2}+6^2\times\frac{1}{3} \end{bmatrix}$$

Observations with smaller variances will receive larger weights in Σ.

Usually, $w_i \propto e_i$

第一步：OLS，求得残差 e

第二步：按照 e 构建 Σ