Advanced Econometrics

Lecture 4: Least Squares Regression (Hansen Chapter 4)

Instructor: Ma, Jun

Renmin University of China

October 23, 2018

随机样本

Random Sampling

- The simplest context is when the observations are mutually independent, in which case we say that they are **independent and identically distributed**, or **i.i.d.** It is also common to describe iid observations as a **random sample**.

← 随机抽样假设

独立同分布

Assumption

The observations $\{(Y_1, \boldsymbol{X}_1), ..., (Y_i, \boldsymbol{X}_i), ..., (Y_n, \boldsymbol{X}_n)\}$ *are independent and identically distributed.*

- If you take any two individuals $i \neq j$ in a sample, the values $(Y_i, \boldsymbol{X}_i)$ are independent of the values $(Y_j, \boldsymbol{X}_j)$ yet have the same distribution.

总体均值

Sample Mean

- Suppose we have a random sample $\{Y_1, ..., Y_n\}$ and we want to estimate the population mean $\mu = \mathbb{E}Y_i$. The sample mean is $\overline{Y} = n^{-1} \sum_{i=1}^{n} Y_i$.
- The sample mean is a linear function of the observations. Its expectation is:

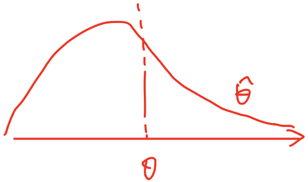
$$\mathbb{E}\overline{Y} = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}Y_i = \mu.$$

- An estimator with the property that its expectation equals the parameter it is estimating is called unbiased.

Definition
An estimator $\widehat{\theta}$ for θ is unbiased if $\mathbb{E}\left(\widehat{\theta}\right) = \theta$.

估计量的期望等于真值.
称为无偏的.

是一个有限样本性质



若真值θ. 那么观测到离θ很远的一个观测值的概率很小.

Sample Mean

- Write $Y_i = \mu + e_i$ with $\sigma^2 = \mathbb{E}\left(e_i^2\right)$. Then

$$
\begin{aligned}
\operatorname{Var}\left(\overline{Y}\right) &= \mathbb{E}\left(\overline{Y}-\mu\right)^2 \\
&= \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} e_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} e_j\right)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(e_i e_j\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \\
&= \frac{1}{n}\sigma^2.
\end{aligned}
$$

$E(e_i e_j) = 0 \quad \forall i \neq j$

$E(e_i^2) = \sigma^2$

线性回归模型

Linear Regression Model

Assumption (*Linear Regression Model*)

The observations $(Y_i, \boldsymbol{X}_i)$ satisfy the linear regression equation

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + e_i$$

$$\mathbb{E}\left(e_i \mid \boldsymbol{X}_i\right) = 0$$

The variables have finite second moments

$$\mathbb{E}\left(Y_i^2\right) < \infty,$$

$$\mathbb{E}\left\|\boldsymbol{X}_i\right\|^2 < \infty,$$

and an invertible design matrix

$$\boldsymbol{Q}_{\boldsymbol{XX}} = \mathbb{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right) > 0.$$

回归模型的几个假设：

① $Y_i = X_i'\beta + \underset{\sim}{e_i}$ ← 观测不到的解释因素.

② $E(e_i \mid X_i) = 0$

$\Downarrow$

$\mathrm{cov}(e_i, X_i) = 0$. e_i与X_i不相关

①② $\Rightarrow E(Y_i \mid X_i) = X_i'\beta$

Linear Regression Model

- Heteroskedastic regression:

$$\mathbb{E}\left(e_i^2 \mid \boldsymbol{X}_i\right) = \sigma^2\left(\boldsymbol{X}_i\right) = \sigma_i^2.$$

异方差

- Homoskedastic regression: the conditional variance is constant.

同方差 $E(e_i^2 \mid X_i) = \sigma^2$ ← 常数

Assumption

The conditional variance of the error

$$\mathbb{E}\left(e_i^2 \mid \boldsymbol{X}_i\right) = \sigma^2\left(\boldsymbol{X}_i\right) = \sigma^2$$

is independent of $\boldsymbol{X}_i$.

首先假设同方差.

即 σ^2 与 X_i 独立.

最小二乘估计均值

Mean of Least-Squares Estimator

- Since the observations are assumed to be iid, then

$$\mathbb{E}\left(Y_i \mid \boldsymbol{X}\right) \overset{iid}{=} \mathbb{E}\left(Y_i \mid \boldsymbol{X}_i\right) = \boldsymbol{X}_i'\boldsymbol{\beta}.$$

- By the conditioning theorem and the linearity of expectations,

$$\begin{aligned}
\mathbb{E}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) &= \mathbb{E}\left(\left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right) \mid \boldsymbol{X}\right) \\
&= \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \mathbb{E}\left(\left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right) \mid \boldsymbol{X}\right) \\
&= \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \sum_{i=1}^{n} \mathbb{E}\left(\boldsymbol{X}_i Y_i \mid \boldsymbol{X}\right) \\
&= \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \mathbb{E}\left(Y_i \mid \boldsymbol{X}\right) \\
&= \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\boldsymbol{\beta} = \boldsymbol{\beta}.
\end{aligned}$$

$$X = \begin{pmatrix} X_{1,1} & \cdots & X_{1,k} \\ X_{2,1} & \cdots & X_{2,k} \\ \vdots & & \vdots \\ X_{n,1} & \cdots & X_{n,k} \end{pmatrix}_{n\times k}$$

$$E(Y_i \mid X) \overset{iid}{=} E(Y_i \mid X_i) = X_i'\beta$$

Mean of Least-Squares Estimator

- Using matrix notation,

$$\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{X}) = \begin{pmatrix} \vdots \\ \mathbb{E}(Y_i \mid \boldsymbol{X}) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \boldsymbol{X}_i' \boldsymbol{\beta} \\ \vdots \end{pmatrix} = \boldsymbol{X}\boldsymbol{\beta}$$

$$\mathbb{E}(\boldsymbol{e} \mid \boldsymbol{X}) = \begin{pmatrix} \vdots \\ \mathbb{E}(e_i \mid \boldsymbol{X}) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \mathbb{E}(e_i \mid \boldsymbol{X}_i) \\ \vdots \end{pmatrix} = \mathbf{0}.$$

- By the conditioning theorem and the linearity of expectations,

$$\begin{aligned} \mathbb{E}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) &= \mathbb{E}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Y} \mid \boldsymbol{X}\right) \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{X}) \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}. \end{aligned}$$

Mean of Least-Squares Estimator

- Since $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$,

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}\right)\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{e}\right) \\
&= \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e}.
\end{aligned}$$

- By the conditioning theorem and the linearity of expectations,

$$\begin{aligned}
\mathbb{E}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X}\right) &= \mathbb{E}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e} \mid \boldsymbol{X}\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}\left(\boldsymbol{e} \mid \boldsymbol{X}\right) \\
&= \mathbf{0}.
\end{aligned}$$

Mean of Least-Squares Estimator

Theorem
In the linear regression model and i.i.d. sampling

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \boldsymbol{\beta}$$

- The conditional distribution of $\widehat{\boldsymbol{\beta}}$ is centered at $\boldsymbol{\beta}$.
- Applying the law of iterated expectations,

$$\mathbb{E}\left(\hat{\boldsymbol{\beta}}\right) = \mathbb{E}\left(\mathbb{E}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right)\right) = \boldsymbol{\beta}$$

最小二乘估计量的方差 Variance of Least Squares Estimator

- For any $r \times 1$ random vector $\boldsymbol{Z}$, define the $r \times r$ covariance matrix

$$\begin{aligned}\operatorname{Var}(\boldsymbol{Z}) &= \mathbb{E}\left((\boldsymbol{Z} - \mathbb{E}(\boldsymbol{Z}))(\boldsymbol{Z} - \mathbb{E}(\boldsymbol{Z}))'\right) \\ &= \mathbb{E}\left(\boldsymbol{Z}\boldsymbol{Z}'\right) - (\mathbb{E}(\boldsymbol{Z}))(\mathbb{E}(\boldsymbol{Z}))'.\end{aligned}$$

- For any pair $(\boldsymbol{Z}, \boldsymbol{X})$, define the conditional covariance matrix

$$\operatorname{Var}(\boldsymbol{Z} \mid \boldsymbol{X}) = \mathbb{E}\left((\boldsymbol{Z} - \mathbb{E}(\boldsymbol{Z} \mid \boldsymbol{X}))(\boldsymbol{Z} - \mathbb{E}(\boldsymbol{Z} \mid \boldsymbol{X}))' \mid \boldsymbol{X}\right).$$

- Define

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \operatorname{Var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right),$$

  the conditional covariance matrix of the LS estimators.

Variance of Least Squares Estimator

$$D = E(ee'|X) = \begin{pmatrix} h(x_1) & & & 0 \\ & h(x_2) & & \\ & & \ddots & \\ 0 & & & h(x_n) \end{pmatrix}$$

$$h(x_i) = E(e_i^2|x_i) = \sigma_i^2 \text{ (异方差)}$$

$$= \sigma^2 \text{ (同方差)} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \sigma^2 = \sigma^2 I_n$$

- The conditional covariance matrix of the error $\boldsymbol{e}$ is

$$\operatorname{Var}(\boldsymbol{e} \mid \boldsymbol{X}) = \mathbb{E}(\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}) = \boldsymbol{D}.$$

  The i^{th} diagonal element of $\boldsymbol{D}$ is

$$\mathbb{E}(e_i^2 \mid \boldsymbol{X}) = \mathbb{E}(e_i^2 \mid \boldsymbol{X}_i) = \sigma_i^2$$

  while the ij^{th} off-diagonal element of $\boldsymbol{D}$ is

$$\mathbb{E}(e_i e_j \mid \boldsymbol{X}) \overset{iid}{=} \mathbb{E}(e_i \mid \boldsymbol{X}_i)\,\mathbb{E}(e_j \mid \boldsymbol{X}_j) = 0.$$

  The first equality holds because of independence of the observations.

- Thus $\boldsymbol{D}$ is a diagonal matrix with i^{th} diagonal element σ_i^2:

$$\boldsymbol{D} = diag(\sigma_1^2, \ldots, \sigma_n^2) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

⇒ 误差项的方差,斜方差矩阵

如果同方差，$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

Variance of Least Squares Estimator

- In the special case of homoskedasticity,

$$\mathbb{E}\left(e_i^2 \mid \boldsymbol{X}_i\right) = \sigma_i^2 = \sigma^2$$

  and we have $\boldsymbol{D} = \boldsymbol{I}_n \sigma^2$.
- For any $n \times r$ matrix $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{X})$,

$$\operatorname{Var}\left(\boldsymbol{A}'\boldsymbol{Y} \mid \boldsymbol{X}\right) = \operatorname{Var}\left(\boldsymbol{A}'\boldsymbol{e} \mid \boldsymbol{X}\right) = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}.$$

- We write $\hat{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{Y}$ where $\boldsymbol{A} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ and thus

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \operatorname{Var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

$$\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \sigma_i^2 .$$

- In the special case of homoskedasticity, $\boldsymbol{D} = \boldsymbol{I}_n \sigma^2$, so

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2.$$

$$Var(A'Y \mid X) = Var(A'X\beta + A'e \mid X)$$

$$= Var(A'e \mid X)$$

$$= A' \underbrace{Var(e \mid X)}_{D} A$$

$$X = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix}$$

$$X'DX = [X_1, \cdots, X_n] \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix}$$

$$= [X_1, \cdots, X_n] \begin{bmatrix} \sigma_1^2 X_1' \\ \vdots \\ \sigma_n^2 X_n' \end{bmatrix}$$

$$= \sigma_1^2 X_1 X_1' + \cdots + \sigma_n^2 X_n X_n'$$

Variance of Least Squares Estimator

Theorem

In the linear regression model and i.i.d. sampling

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \operatorname{var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

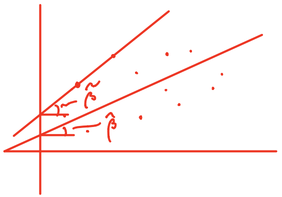
$$= (X'X)^{-1} X'DX(X'X)^{-1} = \sigma^2 (X'X)^{-1} \underbrace{X' I_n X}_{X'X} (X'X)^{-1} = \sigma^2 (X'X)^{-1}$$

In the homoskedastic linear regression model and i.i.d. sampling

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

高斯马科夫定理

Gauss-Markov Theorem



- Now consider the class of estimators that can be written as $\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{Y}$, where $\boldsymbol{A}$ is an $n \times k$ matrix depending only on $\boldsymbol{X}$.
- The LS estimator is the special case: $\boldsymbol{A} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$.
- The Gauss-Markov theorem says that the LS estimator is the best choice among linear unbiased estimators when the errors are homoskedastic, in the sense that the least-squares estimator has the smallest variance.

$E(\tilde{\beta}) = \beta$ 两个都是无偏估计.

$E(\hat{\beta}) = \beta$

↳ OLS

$Var(\tilde{\beta}|X) > Var(\hat{\beta}|X)$

Gauss-Markov Theorem

- For any linear estimator $\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{Y}$ we have

$$\mathbb{E}\left(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \boldsymbol{A}'\mathbb{E}\left(\boldsymbol{Y} \mid \boldsymbol{X}\right) = \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta}$$

  so that $\widetilde{\boldsymbol{\beta}}$ is unbiased if and only if $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$. Furthermore,

$$\operatorname{Var}\left(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) = \operatorname{Var}\left(\boldsymbol{A}'\boldsymbol{Y} \mid \boldsymbol{X}\right) = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{A}\sigma^2.$$

- The best unbiased linear estimator is obtained by finding the matrix $\boldsymbol{A}_0$ satisfying $\boldsymbol{A}_0'\boldsymbol{X} = \boldsymbol{I}_k$ such that for any other matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$ then $\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{A}_0'\boldsymbol{A}_0$ is positive semi-definite.

① $E(\tilde\beta \mid X) = E(A'Y \mid X)$

$= E(A'(X\beta + e) \mid X)$

$= E(A'X\beta \mid X) + E(A'e \mid X)$

$= A'X\beta + A'\underbrace{E(e \mid X)}_{=0}$

$= A'X\beta = \beta$

$\therefore A'X = I_k$ 才有 $\tilde\beta$ 是无偏的.

最小二乘的 $A' = (X'X)^{-1}X'$

显然满足 $A'X = I_k$

$\Rightarrow$ 最小二乘 $\hat\beta$ 是无偏的.

$\hat\beta$ 是最小二乘估计. $\tilde\beta$ 是任意一个线性估计

Gauss-Markov Theorem

③ $\operatorname{Var}(\hat\beta - \tilde\beta \mid X) \geq 0$

$= \operatorname{Var}(\hat\beta \mid X) + \operatorname{Var}(\tilde\beta \mid X) - \underbrace{\operatorname{cov}(\hat\beta, \tilde\beta \mid X)}_{= \operatorname{Var}(\hat\beta \mid X)} - \underbrace{\operatorname{cov}(\tilde\beta, \hat\beta \mid X)}_{= \operatorname{Var}(\hat\beta \mid X)}$

$= \operatorname{Var}(\tilde\beta \mid X) - \underbrace{\operatorname{Var}(\hat\beta \mid X)}_{= \sigma^2 (X'X)^{-1}} \geq 0$

$\Rightarrow \operatorname{Var}(\tilde\beta \mid X) \geq \sigma^2 (X'X)^{-1}$

② $\operatorname{cov}(\hat\beta, \tilde\beta \mid X) = \operatorname{Var}(\hat\beta \mid X)$

$= E\left((\hat\beta - E(\hat\beta \mid X))(\tilde\beta - E(\tilde\beta \mid X))' \mid X\right)$

$= E\left((\hat\beta - \beta)(\tilde\beta - \beta)' \mid X\right)$

$= E\left((X'X)^{-1} X' e e' A \mid X\right)$

$= (X'X)^{-1} X' \underbrace{E(ee' \mid X)}_{\sigma^2 I_n} A$

$= \sigma^2 (X'X)^{-1} \underbrace{X'A}_{I_n}$

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

a. $\tilde\beta = A'Y = A'(X\beta + e)$

$= \underbrace{A'X\beta}_{=\beta} + A'e$

b. $\hat\beta = (X'X)^{-1} X'Y$

$= \beta + (X'X)^{-1} X'e$

Theorem

In the homoskedastic linear regression model and i.i.d sampling, if $\tilde{\boldsymbol\beta}$ is a linear unbiased estimator of $\boldsymbol\beta$ then

$$\operatorname{var}\left(\tilde{\boldsymbol\beta} \mid \boldsymbol X\right) \geq \sigma^2 \left(\boldsymbol X' \boldsymbol X\right)^{-1}.$$

- The theorem is limited because the class of models is restricted to homoskedastic linear regression and the class of potential estimators is restricted to linear unbiased estimators.

$\operatorname{Var}(\tilde\beta_1 \mid X)$ vs. $\operatorname{Var}(\hat\beta_1 \mid X)$

$= \operatorname{Var}(e_1' \tilde\beta \mid X)$ $\qquad = \operatorname{Var}(e_1' \hat\beta \mid X)$

$= e_1' \operatorname{Var}(\tilde\beta \mid X) e_1$ $\qquad = e_1' \operatorname{Var}(\hat\beta \mid X) e_1$

$\therefore \operatorname{Var}(\tilde\beta_1 \mid X) - \operatorname{Var}(\hat\beta_1 \mid X) = e_1' \left(\operatorname{Var}(\tilde\beta \mid X) - \operatorname{Var}(\hat\beta \mid X)\right) e_1 \geq 0 \Rightarrow$ 所有线性估计中最小二乘估

回归残差

Residuals

$Y = \hat{Y} + \hat{e}$

$= PY + MY$

$MY = M(X\beta + e)$

$= \underbrace{MX\beta}_{=0} + Me$

$= Me$

$\therefore \hat{e} = Me$

- The residuals:

$$\hat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e}$$

where $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$.

- We compute

$$\mathbb{E}\left(\hat{\boldsymbol{e}} \mid \boldsymbol{X}\right) = \mathbb{E}\left(\boldsymbol{M}\boldsymbol{e} \mid \boldsymbol{X}\right) = \boldsymbol{M}\mathbb{E}\left(\boldsymbol{e} \mid \boldsymbol{X}\right) = \mathbf{0}$$

$$\operatorname{Var}\left(\hat{\boldsymbol{e}} \mid \boldsymbol{X}\right) = \operatorname{Var}\left(\boldsymbol{M}\boldsymbol{e} \mid \boldsymbol{X}\right) = \boldsymbol{M}\operatorname{Var}\left(\boldsymbol{M}\boldsymbol{e} \mid \boldsymbol{X}\right)\boldsymbol{M} = \boldsymbol{M}\boldsymbol{D}\boldsymbol{M}.$$

- Under the assumption of conditional homoskedasticity,

$$\mathbb{E}\left(e_i^2 \mid \boldsymbol{X}_i\right) = \sigma^2 \text{ and } \operatorname{Var}\left(\hat{\boldsymbol{e}} \mid \boldsymbol{X}\right) = \boldsymbol{M}\sigma^2.$$

计一定是最小的.

广义最小二乘估计是最小的、最优的

$D^{-\frac{1}{2}}Y$ on $D^{-\frac{1}{2}}X = \operatorname{Var}(\tilde{\beta} \mid X) - \operatorname{Var}(\hat{\beta} \mid X)$

$(D = E(ee' \mid X))$ 需要假设D是已知的.

误差估计

Estimation of Error Variance

- The method of moments estimator (MME) of $\sigma^2 = \mathbb{E}\left(e_i^2\right)$ is the sample average of the squared residuals:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 .$$

$\sigma^2 = E(e_i^2) = Var(e_i)$

$\downarrow$ 矩估计

$\frac{1}{n}\sum_{i=1}^{n} e_i^2$

估计不出来用 $\hat{e}_i$ 替换 e_i

$\downarrow$

$\frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2$

- Observe

$$\hat{\sigma}^2 = \frac{1}{n}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}\right) = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{e}\boldsymbol{e}'\right)$$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 = \frac{1}{n}\hat{e}'\hat{e}$

$= \frac{1}{n}(Me)'(Me)$

$= \frac{1}{n} e'M'Me$ M是对称且幂等 $M'M=M$

$= \frac{1}{n} e'Me$

$= \frac{1}{n} tr(e'Me)$

$= \frac{1}{n} tr(Mee')$ [$tr(AB)=tr(BA)$]

and

$$\begin{aligned}
\mathbb{E}\left(\hat{\sigma}^2 \mid \boldsymbol{X}\right) &= \frac{1}{n}\operatorname{tr}\left(\mathbb{E}\left(\boldsymbol{M}\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right)\right) \\
&= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right)\right) \\
&= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{D}\right).
\end{aligned}$$

[$E(tr(X)) = tr(E(X))$]

Estimation of Error Variance

$\operatorname{tr}(M) = \operatorname{rank}(M) = n-k$

- Under the homoskedasticity assumption $\mathbb{E}\left(e_i^2 | \boldsymbol{X}_i\right) = \sigma^2$ so that $\boldsymbol{D} = \boldsymbol{I}_n \sigma^2$,

$$\begin{aligned} \mathbb{E}\left(\hat{\sigma}^2 \mid \boldsymbol{X}\right) &= \frac{1}{n} \operatorname{tr}\left(\boldsymbol{M} \sigma^2\right) \\ &= \sigma^2 \left(\frac{n-k}{n}\right). \end{aligned}$$

- To obtain an unbiased estimator is by rescaling the estimator:

$$s^2 = \frac{1}{n-k} \sum_{i=1}^{n} \hat{e}_i^2 .$$

Now $\mathbb{E}\left(s^2 \mid \boldsymbol{X}\right) = \sigma^2$ and $\mathbb{E}\left(s^2\right) = \sigma^2$.

同方差协方差矩阵估计

Covariance Matrix Estimation Under Homoskedasticity

- Under homoskedasticity, the covariance matrix takes the relatively simple form $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}^{0} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \sigma^2$ which is known up to the unknown scale σ^2.
- The classic covariance matrix estimator:

$$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{0} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} s^2.$$

- $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{0}$ is conditionally unbiased for $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ under homoskedasticity:

$$\begin{aligned}
\mathbb{E}\left(\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{0} \mid \boldsymbol{X}\right) &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \mathbb{E}\left(s^2 \mid \boldsymbol{X}\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \sigma^2 \\
&= \boldsymbol{V}_{\hat{\boldsymbol{\beta}}}^{0}.
\end{aligned}$$

- This was the dominant covariance matrix estimator in applied econometrics for many years, and is still the default method in most regression packages.

Covariance Matrix Estimation Under Homoskedasticity

- If the estimator $\hat{\boldsymbol{V}}^0_{\hat{\boldsymbol{\beta}}}$ is used, but the regression error is heteroskedastic, it is possible for $\hat{\boldsymbol{V}}^0_{\hat{\boldsymbol{\beta}}}$ to be quite biased for

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

⇐真实的方差、斜方差矩阵.

eg.

- Suppose $k = 1$ and $\sigma_i^2 = X_i^2$ with $\mathbb{E}\left(X_i\right) = 0$.

假设了 $D = \operatorname{diag}(X_1^2, \cdots, X_n^2)$

$$\frac{\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}}{\mathbb{E}\left(\hat{\boldsymbol{V}}^0_{\hat{\boldsymbol{\beta}}} \mid \boldsymbol{X}\right)} = \frac{\sum_{i=1}^n X_i^4}{\sigma^2 \sum_{i=1}^n X_i^2} \simeq \frac{\mathbb{E}\left(X_i^4\right)}{\left(\mathbb{E}\left(X_i^2\right)\right)^2}.$$

⇒如果异方差，用原来的公式是错误的.

- If X_i is normally distributed, this ratio is 3. The true variance $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ is three times larger than the expectation of the homoskedastic estimator $\hat{\boldsymbol{V}}^0_{\hat{\boldsymbol{\beta}}}$.

异方差协方差矩阵估计

Covariance Matrix Estimation Under Heteroskedasticity

- The general form is

Stata默认的 $\longrightarrow$

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

where

$$\begin{aligned}\boldsymbol{D} &= \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right)\\ &= \mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right).\end{aligned}$$

$$D = \begin{pmatrix} \underbrace{E(e_1^2 \mid X_1)}_{=\sigma_1^2} & & & 0 \\ & \underbrace{E(e_2^2 \mid X_2)}_{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \underbrace{E(e_n^2 \mid X_n)}_{\sigma_n^2} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix}$$

- If e_i^2, $i = 1, \ldots, n$ are observed, we can construct an unbiased estimator for $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$:

$$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{ideal} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'e_i^2\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}.$$

Covariance Matrix Estimation Under Heteroskedasticity

- Compute

$$
\begin{aligned}
\mathbb{E}\left(\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{ideal} \mid \boldsymbol{X}\right) &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \mathbb{E}\left(e_i^2 \mid \boldsymbol{X}\right)\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i' \sigma_i^2\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \boldsymbol{V}_{\hat{\boldsymbol{\beta}}}.
\end{aligned}
$$

Covariance Matrix Estimation Under Heteroskedasticity

- A feasible version:

怀特 $\longrightarrow$

$$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{W} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\hat{e}_i^2\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

e_i 换成 $\hat{e}_i$ 就可以算异方差.

This is known as the White covariance matrix estimator.

$$E(\hat{e}_i^2|X) = E((\xi_i' M e)^2|X)$$
$$= \xi_i' M E(ee'|X) M \xi_i$$
$$= \sigma^2(\xi_i' M \xi_i)$$
$$= \sigma^2(\xi_i'\xi_i - h_{ii})$$
$$= \sigma^2(1 - h_{ii})$$

- Under homoskedasticity,

$$\begin{aligned}
\mathbb{E}\left(\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{W} \mid \boldsymbol{X}\right) &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\mathbb{E}\left(\hat{e}_i^2 \mid \boldsymbol{X}\right)\right)(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\left(1-h_{ii}\right)\sigma^2\right)(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2 - (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'h_{ii}\sigma^2\right)(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&< (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2 = \boldsymbol{V}_{\hat{\boldsymbol{\beta}}}.
\end{aligned}$$

$\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}^{W}$ is biased towards 0.

怀特估计量一定是有偏的. 且一定是向0单方向有偏.

Measures of Fit

$R^2 > 0.2$ 都算非常好的.

不小于 0-1

- A commonly reported measure of regression fit is the regression R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{e}_i^2}{\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_Y^2}.$$

分母分子都有偏.
没有调整过自由度.

- R^2 can be viewed as an estimator of the population parameter

$$\rho^2 = \frac{\operatorname{Var}\left(\boldsymbol{X}_i'\boldsymbol{\beta}\right)}{\operatorname{Var}\left(Y_i\right)} = 1 - \frac{\sigma^2}{\sigma_Y^2}.$$

- $\hat{\sigma}^2$ and $\hat{\sigma}_Y^2$ are biased estimators. The adjusted R^2 uses unbiased versions:

$$\bar{R}^2 = 1 - \frac{s^2}{\tilde{\sigma}_Y^2} = 1 - \frac{(n-k)^{-1}\sum_{i=1}^n \widehat{e}_i^2}{(n-1)^{-1}\sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2}.$$

调整后 R^2, 调整了分子分母的自由度.
调整后分子分母都无偏了.

- R^2 cannot be used for model selection, as it necessarily increases when regressors are added to a regression model.