Chapt 3

Properties of Least Squares

Classical Linear Regression Model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u} \qquad \underset{\sim}{u} \text{ iid } (\underset{\sim}{0}, \sigma^2 I_n)$$

CNLRM add the assumption ~~u iid~~ $\underset{\sim}{u} \sim N(0, \sigma^2 I_n)$

$\underset{\sim}{y}$ $n \times 1$ $\qquad$ $\underset{\sim}{\beta}$ $k \times 1$

X $n \times k$ $\qquad$ $\underset{\sim}{u}$ $n \times 1$

$$E(\underset{\sim}{u}) = 0 \Rightarrow \begin{matrix} E(u_1) = 0 \\ E(u_2) = 0 \\ \vdots \\ E(u_n) = 0 \end{matrix} = \underset{\sim}{0}$$

$$Cov(\underset{\sim}{u}) = E\left[(\underset{\sim}{u} - E(\underset{\sim}{u}))(u - E(\underset{\sim}{u}))^T\right] = E\,\underset{\sim}{u}\underset{\sim}{u}^T;$$

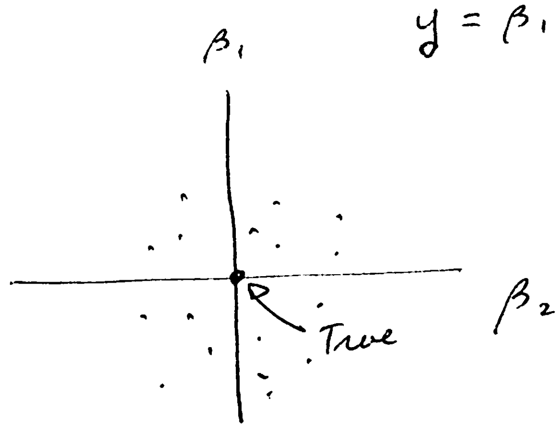
$$= \begin{matrix} E(u_1 u_1) & E(u_1 u_2) & \cdots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2 u_2) & \cdots & E(u_2 u_n) \\ \vdots & & & \\ E(u_n u_1) & \cdots & \cdots & E(u_n u_n) \end{matrix} = \sigma^2 I_n$$

The OLS estimate is $\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y}$

$$E(\hat{\beta}) = E\,(X^T X)^{-1} X^T y$$
$$= E\,(X^T X)^{-1} X^T (X\beta + u)$$
$$= \underset{\sim}{\beta} + E(X^T X)^{-1} X^T \underset{\sim}{u}$$

If $E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}$ then LS is unbiased for $\underset{\sim}{\beta}$.

$y = \beta_1 + \beta_2 x + u$



on average $\hat{\beta}_1, \hat{\beta}_2 = \beta_1, \beta_2$

For OLS to be unbiased

$$E\,(X^T X)^{-1} X^T \underset{\sim}{u} = 0$$

(1) if X fixed

$E(X^T X)^{-1} X^T u = \cancel{E}(X^T X)^{-1} X^T E(u)$

so if $E(u) = 0$, OLS unbiased

(2) If X is stochastic then $E(u|X) = 0$ is required

$$E\left[(X^TX)^{-1}X^Tu \mid X\right] = 0$$

(see eq. 1.17)

here X is not fixed, but is <u>exogenous</u> in the sense that the randomness that generates X is <u>no</u>t related to that of $\underset{\sim}{u}$ in the specified model.

$$E(\underset{\sim}{u} \mid X) = 0 \qquad (3.08)$$

is referred to as an <u>exogeneity assumption</u>.

Each u indep. of all x's = Exog.

NOTE: If $E(u|X) = 0 \Rightarrow E(\underset{\sim}{u}) = 0$

By the Law of Iterated Expectations.

This is a fairly strong assumption. ~~A weak~~ It implies that each u_i is independent of all X_i.

It is sufficient to weaken this to this

$$E(u_t \mid \underset{\sim}{X}_t) = 0 \qquad (3.10)$$

mean of u_t does not depend on the current OBS. OLS Biased here, but consistent.

Bottom line

If $E(\underset{\sim}{u} \mid X) = 0 \qquad (3.08)$

then OLS is unbiased. If this is not true, OLS is biased.

Example: LDV

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t \quad u_t \text{ iid}(0,\sigma^2)$$

In this case $E(u_t \mid y_{t-1}) = 0$
But lagged values of u_t will be correlated with y_{t-1}. Exogeneity is violated. OLS is Biased in this model.

Consistency

An estimator gets close to the unknown parameter with high prob as $n \to \infty$.

Biased estimators can be consistent.

The consistency property uses the idea of a probability limit.

A probability limit generalizes the limit concept to random variables.

Let $\alpha(\underset{\sim}{y}^n)$ be a function of a the $n\times 1$ vector $\underset{\sim}{y}$. The superscript tells us how long $\underset{\sim}{y}$ is.

For all $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\left(\| \alpha(y^n) - \alpha_0 \| < \varepsilon \right) = 1$$

If this is true, then

$$\underset{n\to\infty}{\text{plim}}\ \alpha(y^n) = \alpha_0$$

and $\alpha(y)^n$ is said to be consistent for α_0.

(also, $\alpha(y)^n \xrightarrow{p} \alpha_0$).

When the context of the plim is unambiguous ($n\to\infty$), then it is conventional to drop it from the notation: $\text{plim}\ \alpha(y) = \alpha_0$

The prob limit can be random or nonstochastic

LLN: under various conditions, sample means are consistent for popl means!

$X \overset{iid}{\sim} (\mu_x, \sigma^2)$

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$\operatorname{plim}_{n \to \infty} (\bar{X}) = \mu_x$

Convergence in quadratic mean.

If $\lim_{n\to\infty} E(X) = \mu_x$ AND

$\lim_{n\to\infty} Var(X) = 0$

then $\bar{X} \xrightarrow{q.m.} \mu_x$

This is ~~nec~~ sufficient (But not nec) for consistency.

Properties of plim

suppose $\{x^n\}$ $n = 1, 2, \ldots, \infty$

is a sequence of R.V. with a

nonstoch plim of x_0

$$\text{plim}(x^n) = x_0$$

then for smooth functions $n(x^n)$

$$\text{plim } n(x^n) = n(x_0).$$

$$\text{plim}\big(n(x^n)\big) = n(\text{plim } x^n) = n(x_0)$$

Certainly not true of expectations.

where $E(n(x)) \neq n(E(x))$.

Some quantities in econometrics only have prob limits if they are properly scaled by sample size or n.

For instance plim $(X^T X)$ has no plim, but

$$\text{plim}\left(\frac{X^T X}{n}\right) = S_{x^T x} \text{ does.}$$

Another example

$$\text{plim } X^T \underset{\sim}{u} / \sqrt{n}$$

So, regressor cross products matrix can't grow faster or slower than n without causing inconsistency.

Consistency of OCS

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

$$\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y} = (X^T X)^{-1} X^T (X\beta + u)$$

$$= \beta + (X^T X)^{-1} X^T \underset{\sim}{u}.$$

To take plim we have to do the following

$$\text{plim}(\hat{\beta}) = \text{plim}(\beta) + \text{plim}\left(\frac{X^T X}{n}\right)^{-1} \frac{X^T u}{n}$$

$$= \text{plim}\,\beta + \text{plim}\left(\frac{X^T X}{n}\right)^{-1} \text{plim}\left(\frac{X^T u}{n}\right)$$

$$= \beta + S_{X^T X}^{-1} \cdot \text{plim}\left(\frac{X^T u}{n}\right)$$

By convergence in q.m.

$$\lim_{n \to \infty} E\left(\frac{X^T u}{n}\right) = 0 \quad \text{Provided } E(u|X) = 0$$

$$\lim_{n \to \infty} Var\left(\frac{X^T u}{n}\right) =$$

X^T

If

$\text{plim}\ \frac{1}{n}(X^T \underset{\sim}{u}) = 0$ Then OLS is consistent. When is this so?

(1) $E(\underset{\sim}{u}|X) = 0$

(2) $E(u_t | x_t) = 0$

since this implies.

If $E(u_t|x_t) = 0$ then

$E(\underset{\sim}{x}_t^T u_t | x_t) = 0.$

Then By the Law of Iterated expectations

$E(x_t^T u_t) = 0.$

Application of a law of large Numbers

$$\underset{n\to\infty}{\text{plim}}\ \underbrace{\frac{1}{n}\sum_{t=1}^{n} x_t^T u_t}_{} = 0$$

sample mean converges in prob. to the pop'le mean $= 0$.

$\therefore$ OLS is consistent.

Two things can cause inconsistency.

(1) plims don't exist.

(2) plims converge to the wrong thing.

In the first instance, the estimator could still be unbiased.

Avoid the second possibility at all <u>costs</u>.

Covariance Matrix

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t \qquad t = 1, 2, \ldots, n$$

or.

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{u} \qquad \underset{\sim}{u} \sim (\underset{\sim}{0}, \sigma^2 I_n)$$

$\Longrightarrow$ next page.

$$\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y}$$

$$\underset{k \times k}{\text{Cov}(\hat{\beta})} \equiv E\Big[\underset{k \times 1}{(\underset{\sim}{\beta} - E(\underset{\sim}{\beta}))} \underset{1 \times k}{(\underset{\sim}{\beta} - E(\underset{\sim}{\beta}))^T} \Big]$$

$$= E\left[\begin{pmatrix} \beta_1 - E(\beta_1) \\ \beta_2 - E(\beta_2) \\ \vdots \\ \beta_k - E(\beta_k) \end{pmatrix} \begin{pmatrix} \beta_1 - E(\beta_1) & \cdots & \beta_k - E(\beta_k) \end{pmatrix} \right]$$

$$= \begin{matrix} E(\beta_1 - E(\beta_1))(\beta_1 - E(\beta_1)) & \cdots \\ E(\beta_2 - E(\beta_2))(\beta_1 - E(\beta_1)) & \\ \vdots & \ddots \end{matrix}$$

$$Cov(\underset{\sim}{u}) = E\left[(u - E(u))(u - E(u))^T\right]$$

$$E(\underset{\sim}{u}) = 0$$

$$= E(\underset{\sim}{u}\underset{\sim}{u}^T)$$

$$= \begin{bmatrix} Var(u_1) & Cov(u_1, u_2) & \cdots & Cov(u_1, u_n) \\ & Var(u_2) & & \\ & & \ddots & \\ & & & Var(u_n) \end{bmatrix}$$

If R.S. drawn (iid sample)

$cov(u_i, u_j) = 0 \qquad Var(u_i) = \sigma^2$

$$= \sigma^2 I_n$$

$$= \begin{bmatrix} Var(\hat\beta_1) & Cov(\hat\beta_1,\hat\beta_2) & \cdot & Cov(\hat\beta_1,\hat\beta_k) \\ & Var(\hat\beta_2) & & \\ & & \ddots & \\ & & & Var(\hat\beta_k) \end{bmatrix}$$

Properties

Cov matrices are symmetric

" " are pos. def.

<u>Notes on Pos Def Matrices.</u>

Defn: A $k \times k$ symmetric matrix is pos def if for all nonzero k-vectors $\underset{\sim}{x}$

$$\underset{\sim}{x}^T A \underset{\sim}{x} > 0$$

$\underset{\sim}{x}^T A \underset{\sim}{x}$ is a <u>quadratic form</u>.

A matrix is pos semidefinite if $\underset{\sim}{x}^T A \underset{\sim}{x} = 0$ for some nonzero $\underset{\sim}{x}$.

Rule: Take any matrix $B_{n \times m}$

$B^T B$ is pos. <u>semidefinite</u> and symmetric.

For nonzero $\underset{\sim}{x}$

$$\underset{\sim}{x}^T B^T B \underset{\sim}{x} = (B\underset{\sim}{x})^T (B\underset{\sim}{x}) = \| B x \|^2 \geq 0$$

If this holds with equality, then $B\underset{\sim}{x} = 0$ which implies that B is <u>not</u> <u>of</u> <u>full</u> <u>column</u> <u>rank</u>.

When $B_{m \times n}$ has full column rank then $B^T B$ is pos def.

Also, if A is pos def.

$B^T A B$ is also pos. def. if B is full column rank.

Examples: let $\underset{n \times k}{X}$ $n \geq k$ and Rank$(X) = k$

and $\Omega_{n \times n}$ (sym and Rank$(\Omega) = n$)

$X^T X$ is pos definite. and has Rank of k.

Suppose $\underset{n \times n}{\Omega}$ is pos def. ($\Rightarrow$ Rank$(\Omega) = n$)

Then $\underset{k \times k}{X^T \Omega X}$ is pos def AND has Rank k.

Pos definite matrices are non-singular and can be inverted.

OLS cov.

$$\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y} = (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T u$$
$$= \beta + (X^T X)^{-1} X^T \underset{\sim}{u}$$

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T \underset{\sim}{u}$$

$$E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^T | X] =$$

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T | X]$$ since OLS unbiased.

$$\Rightarrow E[(X^T X)^{-1} X^T \underset{\sim}{u}\, \underset{\sim}{u}^T X (X^T X)^{-1} | X]$$

since $(X^T X)^{-1} = [(X^T X)^{-1}]^T$ sym.

Taking Expectation conditional on $\underset{\sim}{X}$

$$= (X^T X)^{-1} X^T \overbrace{\sigma^2 I_n}^{E(\underset{\sim}{u}\underset{\sim}{u}^T)} X (X^T X)^{-1}$$
$$= \sigma^2 (X^T X)^{-1}$$

E

Precision of LS

The <u>smaller</u> the Cov Matrix the <u>more</u> precise our estimate is.

If Cov $(\hat{\beta})$ = Cov (β) + A [illegible] then Cov β is more precise

$\Rightarrow$ As precision increases Variance gets smaller.

$\Rightarrow$ As precision increases variance of linear combs of estimates gets smaller:

3 things improve precision of LS

(1) smaller σ^2 (true variance of errors themselves)

(2) Larger Samples.

$$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$
$$= \frac{\sigma^2}{n} \left(\frac{X^T X}{n} \right)^{-1}$$

For consistency, $\left(\frac{X^T X}{n}\right)$ converges to a finite limit. $\therefore$ $n \to \infty$ reduces σ^2/n

(3) The third thing that affects precision is

$$X^T X$$

To see this, consider

$$y = x_1 \beta_1 + X_2 \beta_2 + u$$

x_1 is a single regressor
X_2 contains the rest.

Create the X_2 based residual maker matrix

$$M_2 = I - X_2 (X_2^T X_2)^{-1} X_2^T$$

and premultiply the model to remove β_2

$$M_2 y = M_2 x_1 \beta_1 + \text{residuals}$$

$M_2 u$

OLS yields

$$\hat{\beta}_1 = (x_1^T M_2^T M_2 x_1)^{-1} x_1^T M_2' M_2 y$$
$$= (x_1^T M_2 x_1)^{-1} x_1^T M_2 y$$

$\hat{\beta}_1$ = A Scalar

$$\text{Var}(\hat{\beta}_1) = \sigma^2 (x_1^T M_2 x_1)^{-1}$$

$$= \frac{\sigma^2}{x_1^T M_2^T M_2 x_1}$$

$$= \frac{\sigma^2}{\| M_2 x_1 \|^2}$$

As $M_2 x_1$ gets longer, Variance shrinks

$$M_2 x_1 = \left(I - X_2 (X_2^T X_2)^{-1} X_2\right) x_1$$

$$= x_1 - X_2 \underbrace{(X_2^T X_2)^{-1} \cancel{X_2} x_1}_{\text{Regression } \hat{c}}$$

$$= x_1 - X_2 \hat{c}$$

$$x_1 = \cancel{a} X_2 c + \text{res.}$$

$$\hat{c} = (X_2^T X_2)^{-1} X_2^T x_1$$

So Basically $\| M_2 x_1 \|^2$ are SSE from the Regression.

The more highly correlated X_1 is with X_2, the smaller this is and the less precise OLS is for β_1.

If X_1 shows lots of independent variation (not in common with X_2) much of its variation is unexplained by the other vars in model and OLS will be more precise for β_1!

Linear Functions of parameters

Let $\underset{\sim}{\gamma} = \omega^T \hat{\beta}$ where $\underset{\sim}{\omega}$ $k \times 1$ vector of known coeffs.

Be a linear comb of OLS EST.

Example

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 x_{t3} + \hat{\beta}_4 x_{t4} + e_t$$

$$3\hat{\beta}_2 = 2\hat{\beta}_3$$

$$3\hat{\beta}_2 - 2\hat{\beta}_3 = 0$$

$$\omega^T = \begin{matrix} 0 & 3 & -2 & 0 \end{matrix}$$

$$\omega^T \hat{\underset{\sim}{\beta}} = 3\hat{\beta}_2 - 2\hat{\beta}_3$$

$$\begin{aligned} \text{Var}(\omega^T\hat{\beta}) &= E\left[\left(\omega^T\hat{\beta} - E(\omega^T\hat{\beta})\right)\left(\omega^T\hat{\beta} - E(\omega^T\hat{\beta})\right)^T\right] \\ &= E\left[\omega^T(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^T\omega\right] \\ &= \underset{\sim}{\omega}^T \text{Cov}(\hat{\beta})\, \underset{\sim}{\omega} \end{aligned}$$

GAUSS Markov Theorem

Let $E(\underset{\sim}{u} \mid X) = 0$ and $E(\underset{\sim}{u}\underset{\sim}{u}^T \mid X) = \sigma^2 I_n$

Then $\hat{\beta} = (X^TX)^{-1}X^Ty$ is ~~unbiased~~ BLUE of $\underset{\sim}{\beta}$ in $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$.

① Linear estimator

$$\tilde{\beta} = A\underset{\sim}{y}$$

$$\text{let } A = (X^TX)^{-1}X^T + C$$

$$\tilde{\beta} = \left[(X^TX)^{-1}X^T + C\right](X\beta + \underset{\sim}{u})$$

$$= (X^TX)^{-1}X^TX\beta + CX\beta + (X^TX)^{-1}X^T\underset{\sim}{u} + C\underset{\sim}{u}$$

② ~~B~~ Unbiased

$E(\tilde{\beta}) = \beta$ if $CX\beta = 0$ and $E(\underset{\sim}{u} \mid X) = 0$

$\therefore$ Require $CX = 0$

$$\mathrm{Cov}(\tilde{\beta}) \equiv E\left((\tilde{\beta} - E(\tilde{\beta}))(\tilde{\beta} - E(\tilde{\beta}))^T\right]$$

$$= E\left((X^TX)^{-1}X^T\underset{\sim}{u} + C\underset{\sim}{u}\right)\left((X^TX)^{-1}X^T\underset{\sim}{u} + C\underset{\sim}{u}\right)^T$$

$$= E(X^TX)^{-1}X^T\underset{\sim}{u}\underset{\sim}{u}^TX(X^TX)^{-1} + E\,C\underset{\sim}{u}\underset{\sim}{u}^TX(X^TX)^{-1}$$

$$+ E\,(X^TX)^{-1}X^T\underset{\sim}{u}\underset{\sim}{u}^TC^T + C\underset{\sim}{u}\underset{\sim}{u}^TC^T$$

$$E(\underset{\sim}{u}\underset{\sim}{u}^T | X) = \sigma^2 I_n$$

$$\therefore$$

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

$$+ \sigma^2(X^TX)^{-1}X^TC^T + \sigma^2(CC^T)$$

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

$$\mathrm{Cov}(\tilde{\beta}) - \mathrm{Cov}(\hat{\beta}) = \sigma^2 CC^T \quad \text{psd.}$$

Least Square Residuals

$$M_x y = M_x X\beta + M_x \underset{\sim}{u}$$
$$= 0 + M_x u = \hat{u}$$

$$M_x \underset{\sim}{u} = (I - X(X^TX)^{-1}X^T)\underset{\sim}{y}$$
$$= y - X(X^TX)^{-1}X^Ty$$
$$= \tilde{y} - X\hat{\beta} = \hat{u}$$

$$\boxed{M_x \underset{\sim}{u} = \hat{u}}$$

$$Var(\hat{\underset{\sim}{u}}) = Var(M_x \underset{\sim}{u}) = M_x Var(\underset{\sim}{u}) M_x^T$$
$$= M_x \sigma^2 I_n M_x^T$$
$$= \sigma^2 M_x \neq \sigma^2 I_n$$

$$\boxed{Var(\hat{\underset{\sim}{u}}) \neq Var(\underset{\sim}{u})}$$

The variance of an individual error, u_i is found on the i^{th} diagonal element of $\sigma^2 M_x \Longrightarrow \sigma^2(1 - \underset{1\times k}{x_i^T}\underset{k\times k}{(X^TX)^{-1}}\underset{k\times 1}{\underset{\sim}{x}_t}) \equiv \sigma^2(1-h_t)$

It can be shown (sec. 2.6)

$$0 < h_t \le 1.$$

$$\sigma^2(1-h_t) < \sigma^2$$

$$\text{Var}(\hat{u}) < \text{Var}(\underset{\sim}{u})$$

Avg value of $h_t = \frac{k}{n}$

Estimating Error Variance, σ^2.

MOM if $\underset{\sim}{u}$ were observed, then we could estimate σ^2 using

$$\frac{1}{n}\underset{\sim}{u}^T\underset{\sim}{u} = \frac{1}{n}\sum_{i=1}^{n} u_i^2$$

i.e., $E(\underset{\sim}{u}_i^2) = \sigma^2. = \text{Var}(u_i)$

There are not observed, so we replace $\underset{\sim}{u}$ with consistent estimates $\underset{\sim}{\hat{u}}$ from LS. According to the above result we expect to underestimate σ^2 using

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2 \qquad \text{since Var}(\hat{u}_i) < \text{Var}(u_i)$$

$$E\left(\frac{\hat{u}^T\hat{u}}{n}\right) = E\left(tr(\hat{u}^T\hat{u})/n\right)$$

- trace of a scalar is itself.
- trace is sum of diag. elements of a matrix
- $tr(A \cdot B) = tr(B \cdot A)$ for suitably conformable matrices.
- tr and E are linear operators and their order can be switched.

$$E\left(tr(\hat{u}^T\hat{u})/n\right) = E\left[tr(\underset{\sim}{u}^T M_x \underset{\sim}{u})/n\right]$$

$$E\left[tr(M_x u u^T/n)\right] =$$

$$E\left[tr\left(M_x \underset{\sim}{u}\underset{\sim}{u}^T/n\right)\right] =$$

$$tr\left[E\left(M_x \underset{\sim}{u}\underset{\sim}{u}^T/n\right)\right] = tr\left(\tfrac{1}{n}M_x \sigma^2 I_n\right)$$

$$= \frac{1}{n}\sigma^2 tr(M_x)$$

$$= \frac{1}{n} \sigma^2 tr(M_x)$$

$$= \frac{1}{n} \sigma^2 tr(I - X(X^TX)^{-1}X^T)$$

$$= \frac{1}{n} \sigma^2 \left[tr(I) - tr(X(X^TX)^{-1}X^T)\right]$$

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

$$= \frac{1}{n} \sigma^2 (n - k) = \frac{n-k}{n} \sigma^2$$

$$\therefore E\left(\frac{1}{n} \sum_{i=1}^{n} \hat{u}_i^2\right) = \frac{n-k}{n} \sigma^2$$

to make unbiased

$$\frac{1}{n} \cdot \frac{n}{n-k} \sum_{i=1}^{n} \hat{u}_i^2 = \frac{\hat{u}^T\hat{u}}{n-k} = \hat{\sigma}^2 \text{ or } s^2$$

unbiased estimator of cs cov matrix

$$\widehat{Cov}(\hat{\beta}) = s^2 (X^TX)^{-1}$$