$$F = \frac{\text{Increase in Reg S.S.} / \text{d.f.}}{\hat{\sigma}^2}$$

$$= \frac{(\text{Reg S.S.}|_{H_\phi} - \text{Reg S.S.}|_{H_0}) / df}{\hat{\sigma}^2}$$

$$= \frac{(\text{Res S.S.}|_{H_0} - \text{Res S.S.}|_{H_\phi}) / \boxed{df}}{\boxed{\hat{\sigma}^2}} \quad 1$$

$$\boxed{\hat{\sigma}^2} = \text{Res S.S.}|_{H_1} / (n - p')$$

$$= \left(\frac{\text{Res S.S.}|_{H_0}}{\text{Res S.S.}|_{H_1}} - 1 \right) * (n - p')$$

Best subset selection methods

C_p method $\qquad C_p = \text{est. of } \sum_{i=1}^{n} \frac{MSE(\hat{y}(x_i))}{\sigma^2}$

MSE $\qquad \hat{\theta} \xrightarrow{est.} \theta \leftarrow$ true value

$$MSE = E(\hat{\theta} - \theta)^2$$

$$= E(\underline{\hat{\theta} - E(\hat{\theta})} + \underline{E(\hat{\theta}) - \theta})^2$$

$$= E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta)^2 + \cancel{2E(\hat{\theta} - E(\hat{\theta}))(E(\theta}$$

$$+ \underbrace{2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)]}_{\shortparallel}$$

$$\cancel{(E(\hat{\theta}) - \theta)E[\hat{\theta} - E(\hat{\theta})]}$$

constant

$$(E(\hat{\theta}) - \theta)\, \underbrace{E[\hat{\theta} - \boxed{E(\hat{\theta})}]}_{E(\hat{\theta}) - E(\hat{\theta}) = 0}$$

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

unbiased

If $\hat{\theta}$ is not unbiased, then $E(\hat{\theta}) \neq \theta$

$$= \underbrace{E[(\hat{\theta} - E(\hat{\theta}))^2]}_{Var(\hat{\theta})\text{ by definition}} + (\underbrace{E(\hat{\theta}) - \theta}_{bias})^2$$

$$E[\underbrace{E(\hat{\theta}) - \theta}_{\text{constant}}] = E(\hat{\theta}) - \theta \stackrel{?}{=} 0$$

Underfitting

True model : $\underset{\sim}{Y} = \underset{\sim}{X_1}\underset{\sim}{\beta_1} + \boxed{\underset{\sim}{X_2}}\underset{\sim}{\beta_2} + \underset{\sim}{e}^*$ $\quad m'$

Fitted model : $\underset{\sim}{Y} = \underset{\sim}{X_1}\underset{\sim}{\beta_1} + e$ $\quad p'$ $\Rightarrow \hat{\underset{\sim}{\beta_1}} = (\underset{\sim}{X_1}^T \underset{\sim}{X_1})^{-1} \underset{\sim}{X_1}^T \underset{\sim}{Y}$

$m' > p'$

e.g. True model : $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i^*$

Fitted model : $y_i = \beta_0 + \beta_1 X_{i1} + e_i$

p.37 of Chapter 1 (Section 6 - lack of fit)

$$E(\text{Res S.S.}\,|\,\text{fitted model}) = (n-p')\sigma^2 + \underset{\sim}{\beta_2}^T(\underset{\sim}{X_2}^T \underset{\sim}{X_2} - \underset{\sim}{X_2}^T \underset{\sim}{X_1}(\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{X_2})\beta_2$$

$$\hat{\sigma}^2_{\text{fitted model}} = \frac{\text{Res S.S.}\,|\,\text{fitted model}}{n-p'}$$

$$\Rightarrow E(\hat{\sigma}^2_{\text{fitted model}}) = \frac{(n-p')\sigma^2 + \underset{\sim}{\beta_2}^T(\underset{\sim}{X_2}^T \underset{\sim}{X_2} - \underset{\sim}{X_2}^T \underset{\sim}{X_1}(\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{X_2})\beta_2}{n-p'}$$

$$= \sigma^2 + \frac{\underset{\sim}{\beta_2}^T(\underset{\sim}{X_2}^T \underset{\sim}{X_2} - \underset{\sim}{X_2}^T \underset{\sim}{X_1}(\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{X_2})\beta_2}{n-p'}$$

$$\neq \sigma^2$$

– biased est.

$E(\underset{\sim}{Y}) = \underset{\sim}{X_1}\underset{\sim}{\beta_1} + \underset{\sim}{X_2}\underset{\sim}{\beta_2}$

$$E(\hat{\underset{\sim}{\beta_1}}) = E((\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{Y})$$

$$= (\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T \underbrace{E(\underset{\sim}{Y})}_{\text{under the true model}}$$

$$= (\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T(\underset{\sim}{X_1}\underset{\sim}{\beta_1} + \underset{\sim}{X_2}\underset{\sim}{\beta_2})$$

$$= \underline{(\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{X_1}}\,\underset{\sim}{\beta_1} + \underbrace{(\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{X_2}}_{\neq 0 \text{ except } \underset{\sim}{X_2} = \underset{\sim}{0}}\,\underset{\sim}{\beta_2}$$

$$= \underset{\sim}{\beta_1} + (\underset{\sim}{X_1}^T\underset{\sim}{X_1})^{-1}\underset{\sim}{X_1}^T\underset{\sim}{X_2}\underset{\sim}{\beta_2}$$

$$\neq \underset{\sim}{\beta_1} \text{ (biased est)}$$

e.g. true model : $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i^*$

Fitted model : $y_i = \beta_0 + \beta_1 X_{i1} + e_i$

$$\hat{\beta}_1 = \frac{S_{x_1 y}}{S_{x_1 x_1}}$$

under the true model

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)E(y_i)}{\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)^2} = \frac{\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2})}{\sum_{i=1}^{n}(X_{i1}-\bar{X}_1)^2}$$

$$= \frac{\beta_0 \sum_{i=1}^{n}(X_{i1}-\bar{X}_1) + \beta_1 \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)X_{i1} + \beta_2 \sum_{i=1}^{n}(X_{i1}-\bar{X}_1)X_2}{S_{x_1 x_1}}$$

$$= \frac{\beta_1 S_{x_1 x_1} + \beta_2 S_{x_1 x_2}}{S_{x_1 x_1}}$$

$$= \beta_1 + \beta_2 \frac{\boxed{S_{x_1 x_2}}}{S_{x_1 x_1}}$$

If $\boxed{S_{x_1 x_2}} = 0$, $\Rightarrow$ unbiased est.

$\neq \beta_1$ (biased est.)

New observation $\underset{\sim}{x}_0 = (\underset{\sim}{x}_{10}, \boxed{\underset{\sim}{x}_{20}})$

can't be observed.

By true model : $y_{x_0} = \underset{\sim}{x}_{10}^T \underset{\sim}{\beta}_1 + \boxed{\underset{\sim}{x}_{20}^T} \underset{\sim}{\beta}_2 + e_{x_0}^*$

mean prediction : $E(y_{x_0}) = \underset{\sim}{x}_{10}^T \underset{\sim}{\beta}_1 + \boxed{\underset{\sim}{x}_{20}^T} \underset{\sim}{\beta}_2 \leftarrow \theta$

By Fitted model $\hat{y}_{x_0} = \underset{\sim}{x}_{10}^T \hat{\underset{\sim}{\beta}}_1 \leftarrow \hat{\theta}$

$$E(\hat{\theta}) = E(\hat{y}_{x_0}) = E(\underset{\sim}{x}_{10}^T \hat{\underset{\sim}{\beta}}_1)$$

$$= \underset{\sim}{x}_{10}^T E(\hat{\underset{\sim}{\beta}}_1)$$

$$= \underset{\sim}{x}_{10}^T (\underset{\sim}{\beta}_1 + (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{X}_2 \underset{\sim}{\beta}_2)$$

$$= \underset{\sim}{x}_{10}^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{10}^T (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{X}_2 \underset{\sim}{\beta}_2 \neq \overbrace{\boxed{\underset{\sim}{x}_{10}^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{20}^T \underset{\sim}{\beta}_2}}^{\theta}$$

— biased est.

<u>Overfitting</u>

True model : $\underset{\sim}{Y} = \underset{\sim}{X_1}\underset{\sim}{\beta_1} + \underset{\sim}{e}$ $\qquad p'$

Fitted model : $\underset{\sim}{Y} = \underset{\sim}{X_1}\underset{\sim}{\beta_1} + \underset{\sim}{X_2}\underset{\sim}{\beta_2} + \underset{\sim}{e}^*$ $\qquad m' \qquad m' > p'$

e.g. True model : $y_i = \beta_0 + \beta_1 x_{i1} + e_i$

Fitted model : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i^*$

$$(\underset{\sim}{X_c}^T \underset{\sim}{X_c}) = \begin{pmatrix} n & 0 & 0 \\ 0 & S_{x_1x_1} & S_{x_1x_2} \\ 0 & S_{x_1x_2} & S_{x_2x_2} \end{pmatrix} \qquad \underset{\sim}{X_c}^T \underset{\sim}{y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ S_{x_1y} \\ S_{x_2y} \end{pmatrix}$$

$$\hat{\beta}_1 = \frac{S_{x_2x_2}S_{x_1y} - S_{x_1x_2}S_{x_2y}}{S_{x_1x_1}S_{x_2x_2} - S_{x_1x_2}^2}$$

under the true model $E(y_i) = \beta_0 + \beta_1 x_{i1}$

$$E(\hat{\beta}_1) = \frac{S_{x_2x_2}\sum_{i=1}^n (x_{i1} - \bar{x}_1)E(y_i) - S_{x_1x_2}\sum_{i=1}^n (x_{i2} - \bar{x}_2)E(y_i)}{S_{x_1x_1}S_{x_2x_2} - S_{x_1x_2}^2}$$

$$= \frac{S_{x_2x_2}(\beta_1 S_{x_1x_1}) - S_{x_1x_2}(\beta_1 S_{x_1x_2})}{S_{x_1x_1}S_{x_2x_2} - S_{x_1x_2}^2}$$

$= \beta_1$ (unbiased)

$$\text{Res S.S}\big|_{\text{fitted model}} = \underset{\sim}{Y}^T(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{Y} \qquad \underset{\sim}{X} = (\underset{\sim}{X_1}, \underset{\sim}{X_2})$$

<u>Theorem 3.3 in p.16 of chapter 1</u> $\quad E(\underset{\sim}{Y}) = \underset{\sim}{\mu} \quad Var(\underset{\sim}{Y}) = \underset{\sim}{\Sigma}$

Then $E(\underset{\sim}{Y}^T \underset{\sim}{A} \underset{\sim}{Y}) = \text{trace}(\underset{\sim}{A}\underset{\sim}{\Sigma}) + \underset{\sim}{\mu}^T \underset{\sim}{A} \underset{\sim}{\mu}$

$E(\text{Res S.S} \mid \text{fitted model})$

$$= E(\underset{\sim}{Y}^T \underbrace{(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)}_{\underset{\sim}{A}}\underset{\sim}{Y})$$

under the true model, $E(\underset{\sim}{Y}) = \underset{\sim}{X_1}\underset{\sim}{\beta_1}$

$$= \sigma^2 \text{trace}(\underset{n\times n}{\underset{\sim}{I}} - \underset{n \times m'}{\underset{\sim}{X}}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T) + \underbrace{\underset{\sim}{\beta_1}^T \underset{\sim}{X_1}^T(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{X_1}\underset{\sim}{\beta_1}}$$

$$= \sigma^2(n - \underbrace{\text{trace}(\underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)}_{= \text{trace}(\underset{\sim}{I}_{m'\times m'})}) \quad \underbrace{\underset{\sim}{\beta_1}^T(\underset{\sim}{X_1}^T\underset{\sim}{X_1} - \underset{\sim}{X_1}^T\underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{X_1})\underset{\sim}{\beta_1}}_{=0}$$

$$\underset{\sim}{X}^T\underset{\sim}{X} = \left(\begin{array}{c|c} \underset{\sim}{X_1}^T\underset{\sim}{X_1} & \underset{\sim}{X_1}^T\underset{\sim}{X_2} \\ \hline \underset{\sim}{X_2}^T\underset{\sim}{X_1} & \underset{\sim}{X_2}^T\underset{\sim}{X_2} \end{array}\right)$$

$= \sigma^2 (m - m')$

$$\Rightarrow \hat{\sigma}^2_{\text{fitted model}} = \frac{\text{Res S.S} \mid \text{fitted model}}{n - m'}$$

$E(\hat{\sigma}^2_{\text{fitted model}}) = \sigma^2$ — unbiased

$$\frac{\text{Res S.S.}}{\sigma^2} \sim \chi^2 (n-m')$$

$$\Rightarrow Var\left(\frac{\text{Res S.S.}}{\sigma^2}\right) = 2(n-m')$$

$$\Rightarrow Var(\hat{\sigma}^2_{\text{fitted model}}) = \frac{2\sigma^4 (n-m')}{(n-m')^2}$$

$$= \frac{2\sigma^4}{n-m'}$$

$m' > p'$

$\Rightarrow n - m' < n - p'$

$$\Rightarrow Var(\hat{\sigma}^2_{\text{fitted model}}) > \frac{2\sigma^4}{n-p'} \leftarrow Var(\hat{\sigma}^2_{\text{true model}})$$

larger variance

At $\underset{\sim}{x}_0$ ~~$\hat{y}_{\underset{\sim}{x}_0} = \underset{\sim}{x}_{10}^T \hat{\beta}_1 + \underset{\sim}{x}_{20}^T \hat{\underset{\sim}{\beta}}_2 \leftarrow \hat{\theta}$~~

By ~~the~~ true model, $E(y_{\underset{\sim}{x}_0}) = \underset{\sim}{x}_{10}^T \beta_1$ — θ

By Fitted model, $\hat{y}_{\underset{\sim}{x}_0} = \underset{\sim}{x}_{10}^T \hat{\underset{\sim}{\beta}}_1 + \underset{\sim}{x}_{20}^T \hat{\underset{\sim}{\beta}}_2$ — $\hat{\theta}$

$\hat{y}_{x_0}$ — unbiased $\leftarrow E(\hat{y}_{\underset{\sim}{x}_0}) = \underset{\sim}{x}_{10}^T E(\hat{\underset{\sim}{\beta}}_1) + \underset{\sim}{x}_{20}^T E(\hat{\underset{\sim}{\beta}}_2)$

under the true model

$Var(\hat{y}_{x_0}) > Var(\hat{y}^*_{x_0})$

$E(\hat{\underset{\sim}{\beta}}_1) = \underset{\sim}{\beta}_1$

$E(\hat{\underset{\sim}{\beta}}_2) = \underset{\sim}{0}$

C_p $\qquad \sum_{i=1}^{n} \frac{MSE(\hat{y}(\underset{\sim}{x}_i))}{\sigma^2} = \sum_{i=1}^{n} \frac{Var(\hat{y}(\underset{\sim}{x}_i))}{\sigma^2} + \sum_{i=1}^{n} \frac{(bias(\hat{y}(\underset{\sim}{x}_i))^2}{\sigma^2}$

underfitting

Consider the model is underfitted

Fitted model : $\underset{\sim}{Y} = \underset{\sim}{X}_1 \underset{\sim}{\beta}_1 + \underset{\sim}{e}$ — p'

True model : $\underset{\sim}{Y} = \underset{\sim}{X}_1 \underset{\sim}{\beta}_1 + \boxed{\underset{\sim}{X}_2} \underset{\sim}{\beta}_2 + \underset{\sim}{e}^*$ — m'

~~Not~~ Observation

$\underset{\sim}{x}_i$ $\qquad E(y_i) = \underset{\sim}{x}_{1i}^T \underset{\sim}{\beta}_1$ — from fitted model

$\Rightarrow \underset{\sim}{x}_{1i}^T \hat{\underset{\sim}{\beta}}_1 \rightarrow$ est. $\hat{y}(\underset{\sim}{x}_i)$

$$E(\hat{y}(\underset{\sim}{x}_i)) = E(\underset{\sim}{x}_{1i}^T \hat{\underset{\sim}{\beta}}_1)$$
$$= \underset{\sim}{x}_{1i}^T E(\hat{\underset{\sim}{\beta}}_1)$$

under true model

$$Var(\hat{y}(\underset{\sim}{x}_i)) = Var(\underset{\sim}{x}_{1i}^T \hat{\underset{\sim}{\beta}}_1)$$
$$= \underset{\sim}{x}_{1i}^T \underbrace{Var(\hat{\underset{\sim}{\beta}}_1)}_{(\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1}\sigma^2} \underset{\sim}{x}_{1i}$$

p.14 of Chapter 1

$Var(\underset{\sim}{C}\underset{\sim}{z}) = \underset{\sim}{C}\, Var(\underset{\sim}{z})\, \underset{\sim}{C}^T$

$$\sum_{i=1}^{n} \frac{Var(\hat{y}(\underset{\sim}{x}_i))}{\sigma^2} = \sum_{i=1}^{n} \boxed{\underset{\sim}{x}_{1i}^T} (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{x}_{1i}$$

ith row of $\underset{\sim}{X}_1$ i.e. $\underset{\sim}{x}_{1i} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$

$$= \sum_{i=1}^{n} trace(\underset{\sim}{x}_{1i}^T (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{x}_{1i})$$
$$= \sum_{i=1}^{n} trace((\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{x}_{1i} \underset{\sim}{x}_{1i}^T)$$
$$\cancel{= \sum_{i=1}^{n} trace(I)}$$
$$= trace\Big(\underbrace{\sum_{i=1}^{n} \underset{\sim}{x}_{1i} \underset{\sim}{x}_{1i}^T}_{\underset{\sim}{X}_1^T \underset{\sim}{X}_1} (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \Big)$$

$p' \times p'$ matrix

$$= p'$$

$$\sum_{i=1}^{n} \frac{\left(\boxed{\text{bias}(\hat{y}(\underset{\sim}{x}_i))}\right)^2}{\sigma^2}$$

$\theta = E(y_i)$ at $\underset{\sim}{x}_i$

$$= \underset{\sim}{x}_{1i}^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{2i}^T \underset{\sim}{\beta}_2$$

$$\hat{\theta} = \underset{\sim}{x}_{1i}^T \hat{\underset{\sim}{\beta}}_1$$

Under the true model

$$E(\hat{\theta}) = E(\underset{\sim}{x}_{1i}^T \hat{\underset{\sim}{\beta}}_1) = \underset{\sim}{x}_{1i}^T \underbrace{E(\hat{\underset{\sim}{\beta}}_1)}$$

$$= (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T E(\underset{\sim}{Y})$$

$$= (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T (\underset{\sim}{X}_1 \underset{\sim}{\beta}_1 + \underset{\sim}{X}_2 \underset{\sim}{\beta}_2)$$

$$= \underset{\sim}{x}_{1i}^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{1i}^T (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{X}_2 \underset{\sim}{\beta}_2$$

bias $= E(\hat{\theta}) - \theta$

$$= \underset{\sim}{x}_{1i}^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{1i}^T \underbrace{(\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{X}_2}_{\underset{\sim}{A}} \underset{\sim}{\beta}_2$$

$$- (\underset{\sim}{x}_{1i}^T \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{2i}^T \underset{\sim}{\beta}_2)$$

$$= (\underset{\sim}{x}_{1i}^T \underset{\sim}{A} - \underset{\sim}{x}_{2i}^T) \underset{\sim}{\beta}_2$$

$$\text{bias}^2 = \text{bias}^T \text{bias}$$

$$= \underset{\sim}{\beta}_2^T (\underset{\sim}{A}^T \underset{\sim}{x}_{1i} - \underset{\sim}{x}_{2i})(\underset{\sim}{x}_{1i}^T \underset{\sim}{A} - \underset{\sim}{x}_{2i}^T) \underset{\sim}{\beta}_2$$

$$= \underset{\sim}{\beta}_2^T (\underset{\sim}{x}_{2i} - \underset{\sim}{A}^T \underset{\sim}{x}_{1i})(\underset{\sim}{x}_{2i}^T - \underset{\sim}{x}_{1i}^T \underset{\sim}{A}) \underset{\sim}{\beta}_2$$

$$\sum_{i=1}^{n} \frac{\text{bias}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \underset{\sim}{\beta}_2^T (\underset{\sim}{x}_{2i} - \underset{\sim}{A}^T \underset{\sim}{x}_{1i})(\underset{\sim}{x}_{2i}^T - \underset{\sim}{x}_{1i}^T \underset{\sim}{A}) \underset{\sim}{\beta}_2$$

$$= \frac{1}{\sigma^2} \underset{\sim}{\beta}_2^T \boxed{\sum_{i=1}^{n} (\underset{\sim}{x}_{2i} - \underset{\sim}{A}^T \underset{\sim}{x}_{1i})(\underset{\sim}{x}_{2i}^T - \underset{\sim}{x}_{1i}^T \underset{\sim}{A})} \underset{\sim}{\beta}_2$$

$$\|$$

$$\boxed{\underset{\sim}{X}_2^T \underset{\sim}{X}_2 - \underset{\sim}{X}_2^T \underset{\sim}{X}_1 (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{X}_2}$$

From p. 37 of Chapter 1

$$E(\hat{\sigma}^2_{\text{fitted model}}) = \frac{(n-p')\sigma^2 + \boxed{\underset{\sim}{\beta}_2^T (\underset{\sim}{X}_2^T \underset{\sim}{X}_2 - \underset{\sim}{X}_2^T \underset{\sim}{X}_1 (\underset{\sim}{X}_1^T \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{X}_2) \underset{\sim}{\beta}_2}}{n-p'}$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\text{bias}^2}{\sigma^2} = \frac{(E(\hat{\sigma}^2_{p'}) - \sigma^2) * (n-p')}{\sigma^2}$$

R. $\sum_{i=1}^{n} \frac{MSE(\hat{y}(x_i))}{\sigma^2} = \sum_{i=1}^{n} \frac{Var(\hat{y}(x_i))}{\sigma^2} + \sum_{i=1}^{n} \frac{bias^2}{\sigma^2}$

$$\Downarrow \qquad\qquad \Downarrow$$

$$p' \qquad\qquad \frac{(E(\hat{\sigma}^2_{p'}) - \sigma^2) + (n-p')}{\sigma^2}$$

$$= p' + \frac{1}{\sigma^2}(n-p')\,(\underbrace{E(\hat{\sigma}^2_{p'})}_{\text{est. by } \hat{\sigma}^2_{p'}} - \underbrace{\boxed{\sigma^2}}_{\text{est. by } \hat{\sigma}^2_{\text{full model}}})$$

$$C_p = p' + \frac{(n-p')(\hat{\sigma}^2_{p'} - \hat{\sigma}^2_{\text{full model}})}{\hat{\sigma}^2_{\text{full model}}}$$

$$= p' + (n-p')\left(\frac{\hat{\sigma}^2_{p'}}{\hat{\sigma}^2_{\text{full model}}} - 1\right)$$

$$= 2p' - n + \frac{Res\,S.S.|p'}{\hat{\sigma}^2_{\text{full model}}}$$

Choose a model with smallest C_p stat.

For each p, choose a model with $\underbrace{\text{smallest } C_p}$

$$\Downarrow$$

smallest $Res\,S.S|p'$

e.g.

$$\left.\begin{array}{l} p=1 \\ p=2 \\ \vdots \\ p=5 \text{ — full model} \end{array}\right\} \Longrightarrow \text{best model}$$