10.2 Misspecified regression space

DATA: $(Y_i, Z_i^T)^T$, $i = 1, \dots, n$, $Z_i = (Z_{i1}, \dots, Z_{ip})^T \in \mathcal{Z}$

$\mathbb{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, the following model matrices will now be assumed

$$\underset{n\times}{}\Big(\underbrace{\overbrace{X^0, X^1, \dots, X^{k-1}}^{t_X(Z)}}_{\mathbb{X}}, \underbrace{\overbrace{V^1, \dots, V^l}^{t_V(Z)}}_{\mathbb{V}}\Big) = \mathbb{G}$$

ASSUMPTION (everything of full-rank)!

$\operatorname{rank}(\mathbb{X}) = k$, $\operatorname{rank}(\mathbb{V}) = l$,

$\operatorname{rank}(\mathbb{G}) = \operatorname{rank}(\mathbb{X}, \mathbb{V}) = k + l < n$

---

typical situation:

$$\Big(\underbrace{\mathbb{1}, Z^1, \dots, Z^{k-1}}_{\mathbb{X}}, \underbrace{Z^k, \dots, Z^p}_{\mathbb{V}}\Big) \qquad (l = p - k + 1)$$

→ two nested models

M_X: $\mathbb{Y} | \mathbb{Z} \sim (\mathbb{X}\beta, \sigma^2 I_n)$

M_{XV}: $\sim (\underbrace{\mathbb{X}\beta + \mathbb{V}\gamma}_{\mathbb{G}\binom{\beta}{\gamma}}, \sigma^2 I_n)$

typical situation: $\mathbb{X}$ = covariates included in model M_X

$\mathbb{V}$ = covariates included in model M_{XV} but not in model M_X

10.2.1 Omitted and irrelevant regressors [21]

omitted important regressors (covariates)

- M_{XV} is correct (with $\gamma \neq 0_e$) but inference is based on model M_X
  - β estimated using M_X
  - σ^2 estimated using M_X
  - prediction based on fitted M_X

irrelevant regressors included in the model (covariates)

- M_X is correct but inference based on M_{XV} (which is then correct as well)
  - β estimated using M_{XV}
  - σ^2 estimated using M_{XV}
  - prediction based on fitted M_{XV}

---

summary of (obvious) notation [22-23]

Consequence of Lemma 9.1 (Relationship between the quantities derived while assuming the two models)

$$\hat{Y}_{XV} - \hat{Y}_X = M_X V (V^T M_X V)^{-1} V^T U_X$$
$$= X(\hat{\beta}_{XV} - \hat{\beta}_X) + V \hat{\gamma}_{XV}$$

$$\hat{\gamma}_{XV} = (V^T M_X V)^{-1} V^T U_X$$

$$\hat{\beta}_{XV} - \hat{\beta}_X = -(X^T X)^{-1} X^T V \hat{\gamma}_{XV}$$

$$SS_{e,X} - SS_{e,XV} = \| M_X V \hat{\gamma}_{XV} \|^2$$

$$H_{XV} = H_X + M_X V (V^T M_X V)^{-1} V^T M_X$$

→ direct use of formulas from Lemma 9.1

Lemma 10.3 Variance of LSE in the two models

Irrespective of whether M_X or M_{XV} holds, the covariance matrices of the fitted values and the LSE of the regression coefficients satisfy the following:

$$\mathrm{var}(\hat{Y}_{XV} | Z) - \mathrm{var}(\hat{Y}_X | Z) \geq 0,$$
$$\mathrm{var}(\hat{\beta}_{XV} | Z) - \mathrm{var}(\hat{\beta}_X | Z) \geq 0.$$

Proof: • $\mathrm{var}(\hat{Y}_X | Z) = \mathrm{var}(H_X Y | Z) = H_X \underbrace{\mathrm{var}(Y|Z)}_{\sigma^2 I \text{ under both } M_X \text{ and } M_{XV}} H_X^T =$
$= \sigma^2 H_X$

• $\mathrm{var}(\hat{Y}_{XV} | Z) = \mathrm{var}(H_{XV} Y | Z) = \sigma^2 H_{XV}$

$$\overset{\text{see above}}{=} \sigma^2 H_X + \sigma^2 \underbrace{M_X V (V^T M_X V)^{-1} V^T M_X}_{\geq 0}$$

- proof for $var(\hat{\beta}_0 | \underline{x})$
  - see notes (if interested)
  - inverse of a matrix divided in blocks is needed

CONCLUSION (by now):

- In terms of variability of LSE, bigger model is always worse (or the same).

10.2.2 Prediction quality of the fitted model

We will now assume $(Y_i, Z_i^T)^T \overset{iid}{\sim} (Y, Z^T)^T$

$+$ $\boxed{\begin{aligned} \mathbb{E}(Y|Z) &= m(Z) \\ var(Y|Z) &= \sigma^2 \end{aligned}}$ $\leftarrow$ some regression funct.

$\equiv$ MODEL

replicated response

$z_1, .., z_n$: values of $Z_1, .., Z_n$ in data

Consider new data $(Y_{n+i}, Z_{n+i}^T)^T \overset{iid}{\sim} (Y, Z^T)^T$

$\perp\!\!\!\perp (Y_i, Z_i^T)^T$, $i = 1, .., n$

- new data independent of old data
- new data satisfy the same model as old data, in particular

$$\mathbb{E}(Y_{n+i} | Z_{n+i} = z) = m(z)$$
$$var(Y_{n+i} | Z_{n+i} = z) = \sigma^2$$

AIM: PREDICTION of Y_{n+i} given $Z_{n+i} = z_i$, $i = 1, .., n$

$Y_{new} := (Y_{n+1}, .., Y_{n+n})^T$ = new (not yet observed) response vector obtained with the same covariate values as the old data

$\equiv$ replicated response vector

Y_{n+i} generated by the conditional distribution $Y | Z = z_i$

$Y_{new} \perp\!\!\!\perp Y = (Y_1, .., Y_n)^T$

= old (observed) response vector

We have

$$\mathbb{E}(Y | \overbrace{Z_1 = z_1, .., Z_n = z_n}^{Z}) =: \mu = \mathbb{E}(Y_{new} | Z)$$

$$var(Y | Z) = \sigma^2 I_n = var(Y_{new} | Z)$$

$$\mu = (m(z_1), \dots, m(z_n))^T$$

PREDICTION of Y_{new}:

$= \hat{Y}_{new} := (\hat{Y}_{n+1}, \dots, \hat{Y}_{n+n})^T$

= prediction of Y_{new} based on the <u>assumed model</u> fitted using the original data Y with $Z_1 = z_1, \dots, Z_n = z_n$

$\Rightarrow \hat{Y}_{new}$ = some statistic (measurable function) of Y and $\mathbb{Z}$

→ BOX on slide, discussion

29

Def 10.2 Quantification of a prediction quality of the fitted regression model

Prediction quality of the fitted regression model will be evaluated by the <u>mean squared error of prediction</u> (MSEP) defined as $MSEP(\hat{Y}_{new}) = \sum_{i=1}^{n} E\{(\hat{Y}_{n+i} - Y_{n+i})^2 / \mathbb{Z}\}$,

$E\{\sum_{i=1}^{n}(\hat{Y}_{n+i} - Y_{n+i})^2 / \mathbb{Z}\}$

w.r.t. joint distribution $(Y^T, Y_{n+i})^T / \mathbb{Z}$

$g(Y, \mathbb{Z})$

random

where the expectation is with respect to the $(n+n)$-dimensional conditional distribution of the vector $(Y^T, Y_{new}^T)^T$ given $\mathbb{Z} = \begin{pmatrix} Z_1^T \\ \vdots \\ Z_n^T \end{pmatrix} = \begin{pmatrix} Z_{n+1}^T \\ \vdots \\ Z_{n+n}^T \end{pmatrix}$.

30

Additionally, we define the <u>averaged mean squared error of prediction</u> (AMSEP) as

$$AMSEP(\hat{Y}_{new}) = \frac{1}{n} MSEP(\hat{Y}_{new}).$$

"Special" case of linear model 31

MODEL is

$$\mathbb{E}(Y/Z) = \underbrace{X\beta}_{\mu} = \mathbb{E}(Y_{new}/Z)$$

i.e. $m(z) = x^T\beta$, where

$x = t_X(z)$ for some t_X

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} t_X^T(z_1) \\ \vdots \\ t_X^T(z_n) \end{pmatrix}$$

Best linear unbiased prediction in a linear model 32

Just another version of Gauss-Markov theorem:

$\hat{Y}_{new}$ which minimizes $MSEP(\hat{Y}_{new})$ subject to (i) linearity ($\hat{Y}_{new} = a + AY$ for some a, A (that may depend on Z)),

(ii) unbiasedness ($\mathbb{E}(\hat{Y}_{new}/Z) = \mu = X\beta$ for any β)

is $\hat{Y}_{new} = Xb$, $b = (X^TX)^- X^TY$ (any solution to normal eqs.)

$$= \underbrace{X(X^TX)^-X^T}_{H} Y = HY$$

$$= \hat{Y} \;(=: \hat{\mu})$$

Proof (sketch):

- linearity (seen directly)
- unbiasedness → easy
- optimality → same steps as in proof of Gauss-Markov

Lemma 10.4 Mean squared error of the BLUP in a linear model

In a linear model the mean squared error of the best linear unbiased prediction can be expressed as

$$MSEP(\hat{Y}_{new}) = n\sigma^2 + \sum_{i=1}^{n} MSE(\hat{Y}_i),$$

where $MSE(\hat{Y}_i) = \mathbb{E}\{(\hat{Y}_i - \underset{x_i^T\beta}{\overset{\shortparallel}{\mu_i}})^2 / Z\}$, $i=1,\dots,n$,

is the mean squared error of $\hat{Y}_i$ if this is viewed as estimator of μ_i, $i=1,\dots,n$.

NOTE! In a correct model (correctly specified regr. fun.)

$\mathbb{E}(\hat{Y}_i/Z) = X_i^T\beta = \mu_i$ and $MSE(\hat{Y}_i) = var(\hat{Y}_i/Z) = \sigma^2 h_{ii}$

But in an incorrect model

$$MSE(\hat{Y}_i) = \mathbb{E}\{(\hat{Y}_i - \mu_i)^2/Z\} = \mathbb{E}\{(\hat{Y}_i - \mathbb{E}(\hat{Y}_i/Z) + \mathbb{E}(\hat{Y}_i/Z) - \mu_i)^2/Z\} =$$

$$= \mathbb{E}((\hat{Y}_i - \mathbb{E}(\hat{Y}_i/Z))^2/Z) + \{\mathbb{E}(\hat{Y}_i/Z) - \mu_i\}^2 =$$

$$= var(\hat{Y}_i/Z) + \{bias(\hat{Y}_i)\}^2$$

Proof: all expectations are given Z:

$$\mathrm{MSEP}(\hat{Y}_{new}) = \sum_{i=1}^{n} \mathbb{E}(\hat{Y}_{n+i} - Y_{n+i})^2$$

$$\mathbb{E}(\underbrace{\hat{Y}_{n+i}}_{=\hat{Y}_i} - Y_{n+i})^2 = \mathbb{E}(\hat{Y}_i - \mu_i + \mu_i - Y_{n+i})^2 =$$

$$= \mathbb{E}(\hat{Y}_i - \mu_i)^2 + \mathbb{E}(\mu_i - Y_{n+i})^2$$

$$\underbrace{+2\,\mathbb{E}(\hat{Y}_i - \mu_i)(\mu_i - Y_{n+i})}_{} =$$

$\hat{Y}_i \perp\!\!\!\perp Y_{n+i} \quad \forall i$

$$= 2\,\underbrace{\mathbb{E}(\hat{Y}_i - \mu_i)}_{=?} \cdot \underbrace{\mathbb{E}(\mu_i - Y_{n+i})}_{=0}$$

$$= \underbrace{\mathbb{E}(\hat{Y}_i - \mu_i)^2}_{\mathrm{MSE}(\hat{Y}_i)} + \underbrace{\mathbb{E}(\mu_i - Y_{n+i})^2}_{\mathrm{var}(Y_{n+i}) = \sigma^2}$$

Hence $\mathrm{MSEP}(\hat{Y}_{new}) = \sum_{i=1}^{n} \mathrm{MSE}(\hat{Y}_i) + n\sigma^2$ ∎

Remark: $\sum_{i=1}^{n} \mathrm{MSE}(\hat{Y}_i) = \sum_i \mathbb{E}(\hat{Y}_i - \mu_i)^2 = \mathbb{E}\|\hat{Y} - \mu\|^2$

as in proof of Lemma 10.1

$$\underbrace{\mathbb{E}\|\hat{Y}\|^2}_{} - \|\mu\|^2$$

not problematic w.r.t. multicollinearity

Hence we have

$$MSEP(\hat{Y}_{new}) = n\sigma^2 + \sum_{i=1}^{n} MSE(\hat{Y}_i)$$

$$= \underline{n\sigma^2} + \underbrace{\mathbb{E}\|\hat{Y} - \mu\|^2}_{\sum_{i=1}^{n} MSE(\hat{Y}_i)}.$$

does not depend on specification of the regression space (comes from variability of response)

depends on a quality of the model for $\mathbb{E}(Y|Z)$ ($\equiv$ quality component)

As shown above (p. 20):

$$MSE(\hat{Y}_i) = \mathbb{E}\{(\hat{Y}_i - \mu_i)^2 | Z\}$$

$$= var(\hat{Y}_i | Z) + \{\mathbb{E}(\hat{Y}_i - \mu_i | Z)\}^2 =$$

$$= var(\hat{Y}_i | Z) + \underline{(bias\ \hat{Y}_i)^2}$$

$= 0$ if correctly specified regres. space

$\neq 0$ if wrong regress. space

10.2.3 Omitted regressors

Correct model now: $\mathsf{M}_{XV}: \ Y \mid \mathbb{Z} \sim (X\beta + V\gamma, \sigma^2 I_n)$
with $\gamma \neq 0_l$

(Standard) properties of LSE derived under the correct model M_{XV}:

- $\mathbb{E}(\hat{\beta}_{XV} \mid \mathbb{Z}) = \beta$
- $\mathbb{E}(\hat{Y}_{XV} \mid \mathbb{Z}) = X\beta + V\gamma =: \mu$
- $\sum_{i=1}^{n} \mathrm{MSE}(\hat{Y}_{XV,i}) = \sum_{i=1}^{n} \mathrm{var}(\hat{Y}_{XV,i} \mid \mathbb{Z}) = \mathrm{tr}\big(\underbrace{\mathrm{var}(\hat{Y}_{XV} \mid \mathbb{Z})}_{\sigma^2 H_{XV}}\big) = \sigma^2(k+l)$
- $\mathbb{E}(MS_{e,XV} \mid \mathbb{Z}) = \sigma^2$

Lemma 10.5 Properties of the LSE in a model with omitted regressors

Let $\mathsf{M}_{XV}: \ Y \mid \mathbb{Z} \sim (X\beta + V\gamma, \sigma^2 I_n)$ hold, i.e. $\mu := \mathbb{E}(Y \mid \mathbb{Z})$ satisfies $\mu = X\beta + V\gamma$ for some $\beta \in \mathbb{R}^k$, $\gamma \in \mathbb{R}^l$. Then the least squares estimators derived while assuming model $\mathsf{M}_X: \ Y \mid \mathbb{Z} \sim (X\beta, \sigma^2 I_n)$ attain the following properties:

$$\mathbb{E}(\hat{\beta}_X \mid \mathbb{Z}) = \beta + (X^\top X)^{-1} X^\top V\gamma,$$

$$\mathbb{E}(\hat{Y}_X \mid \mathbb{Z}) = \mu - M_X V\gamma,$$

$$\sum_{i=1}^{n} \mathrm{MSE}(\hat{Y}_{X,i}) = k\sigma^2 + \|M_X V\gamma\|^2,$$

$$\mathbb{E}(MS_{e,X} \mid \mathbb{Z}) = \sigma^2 + \frac{\|M_X V\gamma\|^2}{n-k}.$$

Proof: All expectations given Z.

$\hat\beta$: Lemma 9.1: $\hat\beta_{XV} - \hat\beta_X = -(X^TX)^{-1}X^TV\hat\gamma_{XV}$

$$E\hat\beta_X = \underbrace{E\hat\beta_{XV}}_{\beta} + (X^TX)^{-1}X^TV\underbrace{E\hat\gamma_{XV}}_{\gamma} =$$

(correct model)

$$= \beta + (X^TX)^{-1}X^TV\gamma .$$

$$\Rightarrow \text{bias}\,\hat\beta_X = (X^TX)^{-1}X^TV\gamma .$$

Additional discussion (also slide 36)

$\text{bias}\,\hat\beta_X = 0$ if $X^TV = 0$ $\equiv$ omitted regressors are orthogonal (uncorrelated) to included regressors)

bias otherwise

BUT remember:

$$MSE(\hat\beta_X) = \text{var}\,\hat\beta_X + \text{bias}(\hat\beta_X)\,\text{bias}^T(\hat\beta_X)$$

$\wedge$|| Lemma 10.3

$$MSE(\hat\beta_{XV}) = \text{var}\,\hat\beta_{XV}$$

$\rightarrow$ Even (biased) estimator $\hat\beta_X$ might still be better than (unbiased) estimator $\hat\beta_{XV}$ in terms of MSE.

(= strength of collinearity)

$\hat{Y}$: Lemma 9.1 $\quad \hat{Y}_{XV} - \hat{Y}_X = X(\hat{\beta}_{XV} - \hat{\beta}_X) + V\hat{\gamma}_{XV}$

$$\Rightarrow \mathbb{E}\hat{Y}_X = \mathbb{E}\left(\hat{Y}_{XV} - X(\hat{\beta}_{XV} - \hat{\beta}_X) - V\hat{\gamma}_{XV}\right) =$$

$$= \mu - X\beta + X\underbrace{(\beta + (X^TX)^{-1}X^TV\gamma)}_{\mathbb{E}\hat{\beta}_X} - V\gamma =$$

$$= \mu + (\underbrace{X(X^TX)^{-1}X^T - I_n}_{-M_X})V\gamma = \underline{\underline{\mu - M_X V\gamma}}$$

$$\Rightarrow \text{bias}\,\hat{Y}_X = -M_X V\gamma \neq 0 \quad (\text{always unless } \mathcal{M}(V) \subseteq \mathcal{M}(X))$$

not allowed by our assumptions

$\sum_{i=1}^{n} MSE(\hat{Y}_{X,i})$: $\quad \sum_{i=1}^{n} MSE(\hat{Y}_{X,i}) = tr(MSE(\hat{Y}_X))$ **

$$^{**}MSE(\hat{Y}_X) = var\,\hat{Y}_X + bias(\hat{Y}_X)\cdot bias^T(\hat{Y}_X) =$$

$$= \sigma^2 H_X + M_X V\gamma\gamma^T V M_X$$

$$\Rightarrow \sum_{i=1}^{n} MSE(\hat{Y}_{X,i}) = tr\left(\sigma^2 H_X + M_X V\gamma\gamma^T V M_X\right) =$$

$$= \sigma^2 \underbrace{tr(H_X)}_{k} + tr(\boxed{M_X V\gamma}\gamma^T V M_X) =$$

$$= \sigma^2 k + \cancel{tr}(\|M_X V\gamma\|^2) =$$

$$= \underline{\underline{\sigma^2 k + \|M_X V\gamma\|^2}}$$

$MSe_{,X}$: (can be skipped during lecture)

First calculate $\mathbb{E}(SSe_{,X}) := \mathbb{E}(SSe_{,X} \mid Z)$.

M_{XV} written using the error terms:

$$Y = X\beta + V\gamma + \varepsilon, \quad \mathbb{E}(\varepsilon \mid Z) = 0$$
$$\operatorname{var}(\varepsilon \mid Z) = \sigma^2 I_n$$

$$\mathbb{E}(SSe_{,X}) = \mathbb{E}\|M_X Y\|^2 = \mathbb{E}\|M_X(X\beta + V\gamma + \varepsilon)\|^2 =$$
$$= \mathbb{E}\|M_X V\gamma + M_X\varepsilon\|^2 =$$
$$= \mathbb{E}\|M_X V\gamma\|^2 + \mathbb{E}\|M_X\varepsilon\|^2 + 2\underbrace{\mathbb{E}(\gamma^T V^T M_X M_X \varepsilon)}_{\gamma^T V^T M_X \underbrace{\mathbb{E}(\varepsilon \mid Z)}_{0}}$$
$$= \|M_X V\gamma\|^2 + \underbrace{\mathbb{E}(\varepsilon^T M_X \varepsilon)}_{(n-k)\sigma^2}$$

$$\Rightarrow \mathbb{E}(MSe_{,X}) = \mathbb{E}\left(\frac{SSe_{,X}}{n-k}\right) = \underline{\underline{\sigma^2 + \frac{\|M_X V\gamma\|^2}{n-k}}}.$$

$\boxed{\square}$

was discussed before

Prediction

- compare $\hat{Y}_{new,X} = \hat{Y}_X$ (prediction using WRONG model)

and $\hat{Y}_{new,XV} = \hat{Y}_{XV}$ (prediction using correct model)

Remember: $MSEP(\hat{Y}_{new}) = n\sigma^2 + \sum_{i=1}^{n} MSE(\hat{Y}_i)$

M_{XV}: $MSEP(\hat{Y}_{new,XV}) = n\sigma^2 + \underbrace{\sum_{i=1}^{n} var(\hat{Y}_i \mid Z)}_{\sigma^2 tr(H_{XV}) = \sigma^2(k+l)}$

M_X: $MSEP(\hat{Y}_{new,X}) = n\sigma^2 + k \cdot \sigma^2 + \|M_X V \gamma\|^2$

$AMSEP(\hat{Y}_{new,XV}) = \sigma^2 + \underbrace{\sigma^2 \frac{k+l}{n}}_{\to 0 \text{ as } n \to \infty}$

$AMSEP(\hat{Y}_{new,X}) = \sigma^2 + \underbrace{\sigma^2 \frac{k}{n}}_{\to 0 \text{ as } n\to\infty} + \underbrace{\frac{1}{n}\|M_X V \gamma\|^2}_{\text{not necessarily tends to } 0 \text{ as } n\to\infty\ !}$

Estimator of the residual variance

38

$bias(MS_{e,X}) = E(MS_{e,X} - \sigma^2 \mid Z) = \frac{\|M_X V \gamma\|^2}{n-k}$

10.2.4 Irrelevant regressors

Correct models now are both

M_X: $Y|Z \sim (X\beta, \sigma^2 I_n)$

and M_{XV}: $Y|Z \sim (X\beta + V\gamma, \sigma^2 I_n)$

(this with $\gamma = 0_\ell$)

(Standard) properties of LSE derived under either of two correct models:

- $E(\hat\beta_X|Z) = E(\hat\beta_{XV}|Z) = \beta$
- $E(\hat Y_X|Z) = E(\hat Y_{XV}|Z) = X\beta =: \mu$
- $\sum_{i=1}^n MSE(\hat Y_{X,i}) = \sum_{i=1}^n var(\hat Y_{X,i}|Z) = tr(var(\hat Y_X|Z))$
  $= tr(\sigma^2 H_X) = \sigma^2 k$

  $\sum_{i=1}^n MSE(\hat Y_{XV,i}) = \sum_{i=1}^n var(\hat Y_{XV,i}|Z) = tr(var(\hat Y_{XV}|Z))$
  $= tr(\sigma^2 H_{XV}) = \sigma^2(k+\ell)$

- $E(MSe,_X|Z) = E(MSe,_{XV}|Z) = \sigma^2.$

Interest in β

Remember $\hat{\beta}_X = \hat{\beta}_{XV} + (X^TX)^{-1} X^T V \hat{\gamma}_{XV}$

(now $\mathbb{E}(\hat{\gamma}_{XV} | Z) = \gamma = 0$)

- $\mathbb{E}(\hat{\beta}_X | Z) = \mathbb{E}(\hat{\beta}_{XV} | Z) = \beta$

$\equiv$ both estimators are unbiased

- $MSE(\hat{\beta}_{XV}) - MSE(\hat{\beta}_X) = var(\hat{\beta}_{XV} | Z) - var(\hat{\beta}_X | Z) =$

$\overset{\text{Lemma 90.3}}{=}$ ~~var~~ $\sigma^2 [\{X^TX - X^TV(V^TV)^{-1}V^TX\}^{-1} - (X^TX)^{-1}]$

≥ 0

(i) $X^TV = 0_{k \times l}$ (irrelevant regressors are orthogonal to those included in both models)

$\equiv$ uncorrelated

$\Rightarrow \hat{\beta}_X = \hat{\beta}_{XV}$ and $var(\hat{\beta}_{XV} | Z) = var(\hat{\beta}_X | Z)$

$\rightarrow$ irrelevant regressors do not influence quality of LSE of β

(ii) $X^TV \neq 0$

- $\hat{\beta}_{XV}$ is worse than $\hat{\beta}_X$ in terms of MSE and variability

- remember (multicollinearity)

"$\sum_j var(\hat{\beta}_j | Z) = \sigma^2 tr(X^TX)^{-1}$"

more regressors $\equiv$ higher chance that "$tr(X^TX)^{-1}$" is huge

$\equiv$ higher chance on multicollinearity

Prediction

- compare $\hat{Y}_{new,X} = \hat{Y}_X$ (prediction using simpler correct model)

and $\hat{Y}_{new,XV} = \hat{Y}_{XV}$ (prediction using also correct model which however is unnecessarily complex)

Remember $MSEP(\hat{Y}_{new}) = n\sigma^2 + \underbrace{\sum_{i=1}^{n} MSE(\hat{Y}_i)}$

$= \sum_{i=1}^{n} var(\hat{Y}_i | Z)$ now

$= tr(\sigma^2 H)$

A M_{XV}: $MSEP(\hat{Y}_{new,XV}) = n\sigma^2 + (k+l)\sigma^2$

M_X: $MSEP(\hat{Y}_{new,X}) = n\sigma^2 + k\sigma^2$

$AMSEP(\hat{Y}_{new,XV}) = \sigma^2 + \underbrace{\sigma^2 \frac{k+l}{n}}_{\to 0 \text{ as } n \to \infty}$

$AMSEP(\hat{Y}_{new,X}) = \sigma^2 + \underbrace{\sigma^2 \frac{k}{n}}_{\to 0 \text{ as } n \to \infty}$

+ comments on slide

10.2.5 Summary

Interest in estimation of the regression coefficients and inference on them

→ see slide

‼ conclusions for observational studies ‼

– see also model building

Interest in prediction

→ see slide