STAT 34300: Lecture 8

$X \in \mathbb{R}^{n \times p}$ fixed.

$Y \sim (X\beta, \Sigma)$ $\quad [\Sigma \neq \sigma^2 I_n]$

$\hat{\beta}^{OLS} = (X^T X)^{-1} X^T Y$

$\mathbb{E}[\hat{\beta}^{OLS}] = \beta$

$\mathrm{Var}[\hat{\beta}^{OLS}] = \mathrm{Var}\big(\underbrace{(X^T X)^{-1} X^T}\, \underbrace{Y}_{\sim (0, \Sigma)}\big)$

$$= (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1}.$$ "sandwich"

[Homoscedasticity: $\Sigma = \sigma^2 I$.

$\rightsquigarrow \sigma^2 (X^T X)^{-1}$]

Inference pretending homoscedasticity holds $\longrightarrow$ misleading.

Approach I: Come up with new estimator.

Suppose $\Sigma = \underbrace{\sigma^2}_{>0 \text{ unknown}} \Gamma > 0$ known.

$\Gamma^{-1/2}$

$$\left(\begin{aligned} \Gamma &= V \Lambda V^T \\ \Gamma^{\alpha} &= V \Lambda^{\alpha} V^T \end{aligned} \right).$$

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim (0, \sigma^2 \Gamma)$$

$$\underbrace{\Gamma^{-1/2} y}_{\tilde{y}} = \underbrace{\Gamma^{-\frac{1}{2}} X}_{\tilde{X}} \beta + \underbrace{\Gamma^{-\frac{1}{2}} \varepsilon}_{\tilde{\varepsilon}}.$$

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

OLS with $\tilde{X}$ and $\tilde{Y}$.

$\rightarrow$ GLS : Generalized LS.

$$\hat{\theta}^{GLS} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}$$

$$= (X^T \Gamma^{-1} X)^{-1} X^T \Gamma^{-1} Y.$$

$$\hat{\theta}^{GLS} \sim (\theta, \sigma^2 (\tilde{X}^T \tilde{X})^{-1})$$

$$\Leftrightarrow \hat{\theta}^{GLS} \sim (\theta, \sigma^2 (X^T \Gamma^{-1} X)^{-1}).$$

Any $c \in \mathbb{R}^p$: $\mathrm{Var}(c^T \hat{\theta}^{GLS}) \leq \mathrm{Var}(c^T \hat{\theta}^{OLS})$

Γ is diagonal.

$\Gamma_{ii} = \frac{1}{w_i}$ $\rightsquigarrow$ weighted least squares.

<u>why would we know Γ?</u>

- Y_i is the avg. of $Y_{i1}, \ldots, Y_{ik_i}$

Y_{ij} were homoscedastic

$\rightarrow \mathrm{Var}(Y_i) = \frac{\sigma^2}{k_i}$, & know k_i

$\rightarrow$ weighted least squares with $w_i = K_i$.

- We know $\Gamma = \Gamma(\vartheta)$, ϑ low-dim. parameter.

[time-series, spatial dependence].

Two-step approach:

① $\hat{\vartheta}$

② Do GLS with $\Gamma(\hat{\vartheta})$.

$\rightarrow$ Asymptotically valid inference as $n \to \infty$.

Approach II: Keep working with $\hat{\theta}^{OLS}$ & adjust inference.

Need some structure (but less than for GLS)

Assume: Σ diagonal.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{pmatrix}$$

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

$$\mathbb{E}[\hat{\beta}^{OLS}] = \beta$$

$$\mathrm{Var}[\hat{\beta}^{OLS}] = \mathrm{Var}(\underbrace{(X^T X)^{-1} X^T}\, \underset{\sim (0, \Sigma)}{y}) = (X^T X)^{-1} X^T \underline{\Sigma} X (X^T X)^{-1}$$

$$(\text{if} \quad \Sigma = \sigma^2 I_n)$$

$$\underbrace{X^T \Sigma X} = \sum_i^n x_i x_i^T \underline{\sigma_i^2}$$

Estimate σ_i^2 by $\underline{\hat{\sigma}_i^2} = \hat{\varepsilon}_i^2$

$$= (y_i - \hat{\beta}^{OLS\,T} x_i)^2.$$

$$\sigma_i^2 = \mathrm{Var}(y_i)$$

$$\widehat{X^T \Sigma X} = \sum_{i=1}^{n} x_i x_i^T \hat{\varepsilon}_i^2$$

Any $c \in \mathbb{R}^p$

$$\frac{c^T (X^T X)^{-1} \widehat{X^T \Sigma X} (X^T X)^{-1} c}{c^T (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} c} \xrightarrow{\mathbb{P}} 1 .$$

$n \to \infty$.

$$\widehat{Var}[\hat{\beta}^{OLS}] = (X^T X)^{-1} \left(\sum_{i=1}^{n} x_i x_i^T \hat{\varepsilon}_i^2 \right) (X^T X)^{-1}$$

- Heterosk. robust s.e.
- Robust Standard Errors
- Eicker – Huber – White standard errors.

Assumption lean regression

→ X_i random.

$(X_i, Y_i) \overset{iid}{\sim} \mathbb{P}$ on $\mathbb{R}^p \times \mathbb{R}$.

(works even when: $\mathbb{E}(Y_i | X_i) \neq \theta^T X_i$

$\mathrm{Var}(Y_i | X_i) \neq \sigma^2 I$)

$$\hat{\theta}^{OLS} \in \underset{\theta \in \mathbb{R}^p}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \theta^T X_i)^2 \right\}$$

$$\theta^*(\mathbb{P}) \in \operatorname{argmin} \{ \mathbb{E}[(Y_i - \theta^\top X_i)^2] \}.$$

$$D_\theta \overset{!}{=} 0 \implies \boxed{\mathbb{E}_{\mathbb{P}}[\underbrace{X_i}_{\in \mathbb{R}^p} (\underbrace{Y_i - \theta^*(\mathbb{P})^\top X_i}_{\in \mathbb{R}})] = 0.}$$

If $\mathbb{E}[X_i X_i^\top] \succ 0$

$$\implies \boxed{\theta^*(\mathbb{P}) = \mathbb{E}[X_i X_i^\top]^{-1} \, \mathbb{E}[X_i Y_i]}$$

- $\hat{\theta}^{OLS} \approx \theta^*(\mathbb{P})$.
- Inference for $\theta^*(\mathbb{P})$.