Lecture 9: Large-sample properties of OLS and robust standard errors

Economics 527: Econometric Methods

Vadim Marmer, UBC

Consistency of OLS

OLS in terms of sample averages

- Regression, $\{(Y_i, X_i) : i = 1, \dots, n\}$ iid:

$$Y_i = X_i^\top \beta + U_i, \qquad \underset{k\times 1}{X_i} = \begin{pmatrix} X_{i,1} \\ \vdots \\ X_{i,k} \end{pmatrix}$$

- OLS from n observations, $\text{rank}\,(X) = k$ (Lecture 1); substitute $Y_i = X_i^\top \beta + U_i$:

$$\begin{aligned}
\hat{\beta}_n &= (X^\top X)^{-1} X^\top Y \\
&= \Big(\underbrace{\sum_{i=1}^n X_i X_i^\top}_{k\times k} \Big)^{-1} \sum_{i=1}^n X_i Y_i \\
&= \Big(\sum_{i=1}^n X_i X_i^\top \Big)^{-1} \sum_{i=1}^n X_i (X_i^\top \beta + U_i) \\
&= \Big(\sum_{i=1}^n X_i X_i^\top \Big)^{-1} \Big(\sum_{i=1}^n X_i X_i^\top \Big) \beta + \Big(\sum_{i=1}^n X_i X_i^\top \Big)^{-1} \sum_{i=1}^n X_i U_i \\
&= \beta + \Big(\sum_{i=1}^n X_i X_i^\top \Big)^{-1} \sum_{i=1}^n X_i U_i \\
&= \beta + \Big(\frac{1}{n} \sum_{i=1}^n X_i X_i^\top \Big)^{-1} \frac{1}{n} \sum_{i=1}^n X_i U_i
\end{aligned}$$

- Squares and cross-products:

$$\underset{k\times k}{X_i X_i^\top} = \begin{pmatrix} X_{i,1}^2 & X_{i,1}X_{i,2} & \cdots & X_{i,1}X_{i,k} \\ X_{i,2}X_{i,1} & X_{i,2}^2 & \cdots & X_{i,2}X_{i,k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{i,k}X_{i,1} & X_{i,k}X_{i,2} & \cdots & X_{i,k}^2 \end{pmatrix}$$

- Element (r, s):

$$\Big[\sum_{i=1}^n X_i X_i^\top \Big]_{rs} = \sum_{i=1}^n X_{i,r} X_{i,s}$$

LLN for $\frac{1}{n}\sum X_i U_i$

- $X_1 U_1, \dots, X_n U_n$: iid; weak LLN, element by element:

$$\frac{1}{n} \sum_{i=1}^n X_i U_i \to_p \mathrm{E}\,[X_i U_i]$$

- **Assume:**

$$\mathrm{E}\left[X_i U_i\right] = 0$$

- Weak LLN for element j requires:

$$\mathrm{E}\left[\left|X_{i,j} U_i\right|\right] < \infty, \qquad j = 1, \ldots, k$$

- Cauchy-Schwarz inequality:

$$\mathrm{E}\left[\left|X_{i,j} U_i\right|\right] \leq \sqrt{\mathrm{E}\left[X_{i,j}^2\right] \mathrm{E}\left[U_i^2\right]}$$

- **Assume:**

$$\mathrm{E}\left[U_i^2\right] < \infty, \qquad \mathrm{E}\left[X_{i,j}^2\right] < \infty, \quad j = 1, \ldots, k$$

LLN for $\frac{1}{n}\sum X_i X_i^\top$

- Element (r, s) of $\mathrm{E}\left[X_i X_i^\top\right]$ well defined, Cauchy-Schwarz:

$$\mathrm{E}\left[\left|X_{i,r} X_{i,s}\right|\right] \leq \sqrt{\mathrm{E}\left[X_{i,r}^2\right] \mathrm{E}\left[X_{i,s}^2\right]} < \infty$$

- Weak LLN, element by element:

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top \to_p \mathrm{E}\left[X_i X_i^\top\right]$$

- $\mathrm{E}\left[X_i X_i^\top\right]$: symmetric, p.s.d.
- **Assume:**

$$\mathrm{E}\left[X_i X_i^\top\right] \quad \text{p.d.}$$

- Slutsky, $h(\cdot) = (\cdot)^{-1}$ continuous at a p.d. matrix (Lecture 8):

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top\right)^{-1} \to_p \left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1}$$

Assumptions

- Assumptions for consistency:

$\{(Y_i, X_i) : i = 1, \ldots, n\}$ iid
$Y_i = X_i^\top \beta + U_i$
$\mathrm{E}\left[X_i U_i\right] = 0$
$\mathrm{E}\left[U_i^2\right] < \infty$
$\mathrm{E}\left[X_{i,j}^2\right] < \infty, \quad j = 1, \ldots, k$
$\mathrm{E}\left[X_i X_i^\top\right]$ p.d. (no perfect multicollinearity)

- $U_i = Y_i - X_i^\top \beta$:

$$\{(Y_i, X_i)\} \text{ iid} \quad \Longrightarrow \quad \{(U_i, X_i)\} \text{ iid}$$

- (Y_1, X_1) independent of (Y_2, X_2); Y_1 and X_1 need not be independent.
- Weaker than $\mathrm{E}\left[U_i \mid X_i\right] = 0$:

$$\mathrm{E}\left[X_i U_i\right] = 0 \quad \not\Rightarrow \quad \mathrm{E}\left[Y_i \mid X_i\right] = X_i^\top \beta \quad \text{(in general)}$$

Consistency

- **Definition.** $\hat{\beta}_n$ is a consistent estimator of β if $\hat{\beta}_n \to_p \beta$.
- Element by element, by definition (Lecture 8):

$$\hat{\beta}_n \to_p \beta \quad \iff \quad \hat{\beta}_{n,j} \to_p \beta_j, \quad j = 1, \dots, k$$

- **Theorem.** Under the assumptions for consistency, $\hat{\beta}_n \to_p \beta$.
- The two limits, then Slutsky for the product:

$$\begin{aligned}\hat{\beta}_n &= \beta + \Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top}_{\to_p \mathrm{E}[X_i X_i^\top]} \Big)^{-1} \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i U_i}_{\to_p 0} \\ &\to_p \beta + \left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1} \cdot 0 \\ &= \beta\end{aligned}$$

- Therefore:

$$\boxed{\hat{\beta}_n \to_p \beta}$$

Asymptotic normality of OLS

Assumptions I and II

- $\hat{\beta}_n - \beta \to_p 0$. Approximate the distribution of $\hat{\beta}_n - \beta$: scale by $\sqrt{n}$ (Lecture 8).
- **Assumption I.**

$$\begin{aligned}&\{(Y_i, X_i) : i = 1, \dots, n\} \ \text{ iid} \\ &Y_i = X_i^\top \beta + U_i, \qquad \beta \in \mathbb{R}^k\end{aligned}$$

- OLS:

$$\begin{aligned}\hat{\beta}_n &= \Big(\sum_{i=1}^{n} X_i X_i^\top\Big)^{-1} \sum_{i=1}^{n} X_i Y_i \\ &= \beta + \Big(\sum_{i=1}^{n} X_i X_i^\top\Big)^{-1} \sum_{i=1}^{n} X_i U_i \\ \implies \quad \hat{\beta}_n - \beta &= \Big(\sum_{i=1}^{n} X_i X_i^\top\Big)^{-1} \sum_{i=1}^{n} X_i U_i\end{aligned}$$

- **Assumption II.**

$$\underset{k\times k}{Q} = \mathrm{E}\left[X_i X_i^\top\right] \quad \text{p.d. and finite}$$

- Weak LLN, then Slutsky:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top &\to_p \mathrm{E}\left[X_i X_i^\top\right] = Q \\ \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top\Big)^{-1} &\to_p Q^{-1} = \left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1}\end{aligned}$$

Scaling by $\sqrt{n}$

- Multiply by $\sqrt{n}$:

$$\begin{aligned}\sqrt{n}(\hat{\beta}_n - \beta) &= \sqrt{n}\Big(\sum_{i=1}^n X_i X_i^\top\Big)^{-1} \sum_{i=1}^n X_i U_i \\ &= \sqrt{n}\Big(n \cdot \frac{1}{n}\sum_{i=1}^n X_i X_i^\top\Big)^{-1} \sum_{i=1}^n X_i U_i \\ &= \frac{\sqrt{n}}{n}\Big(\frac{1}{n}\sum_{i=1}^n X_i X_i^\top\Big)^{-1} \sum_{i=1}^n X_i U_i \\ &= \Big(\frac{1}{n}\sum_{i=1}^n X_i X_i^\top\Big)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i U_i\end{aligned}$$

Assumption III and the CLT

- **Assumption III.**

$$\begin{aligned}&\mathrm{E}\left[X_i U_i\right] = 0 \quad (X\text{'s and } U\text{'s uncorrelated}) \\ &\underset{k\times k}{\Omega} = \mathrm{Var}\left(X_i U_i\right) \quad \text{finite and p.d.}\end{aligned}$$

- Variance of $X_i U_i$, with $\mathrm{E}\left[X_i U_i\right] = 0$:

$$\begin{aligned}\Omega &= \mathrm{Var}\left(X_i U_i\right) \\ &= \mathrm{E}\left[X_i U_i (X_i U_i)^\top\right] \\ &= \mathrm{E}\left[\underset{1\times 1}{X_i U_i}\ \underset{1\times 1}{U_i} X_i^\top\right] \\ &= \mathrm{E}\left[U_i^2 X_i X_i^\top\right]\end{aligned}$$

- $X_1 U_1, \ldots, X_n U_n$: iid $k \times 1$ (Assumption I), mean 0, variance Ω finite and p.d. (Assumption III).
- Multivariate CLT (Lecture 8):

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i U_i \to_d N(0,\ \underbrace{\mathrm{E}\left[U_i^2 X_i X_i^\top\right]}_{\Omega})$$

Limit of $\sqrt{n}(\hat{\beta}_n - \beta)$

- We will use, without proof (Cramér convergence theorem): $A_n \to_p A$, A non-random, and $Z_n \to_d Z$:

$$A_n Z_n \to_d AZ$$

- The two limits:

$$\sqrt{n}(\hat{\beta}_n - \beta) = \underbrace{\Big(\frac{1}{n}\sum_{i=1}^n X_i X_i^\top\Big)^{-1}}_{\to_p Q^{-1}} \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i U_i}_{\to_d N(0,\Omega)} \to_d Q^{-1} N(0,\Omega)$$

- Linear map of a normal vector (Lecture 4):

$$W \sim N(\mu, \Sigma) \quad \Longrightarrow \quad \Gamma W \sim N(\Gamma\mu,\ \Gamma\Sigma\Gamma^\top)$$

- $\Gamma = Q^{-1}$; $Q = \mathrm{E}\left[X_i X_i^\top\right] = Q^\top$, so $(Q^{-1})^\top = Q^{-1}$:

$$\begin{aligned}Q^{-1} N(0,\Omega) &\overset{d}{=} N(Q^{-1}\cdot 0,\ Q^{-1}\Omega(Q^{-1})^\top) \\ &= N(0,\ Q^{-1}\Omega Q^{-1})\end{aligned}$$

Asymptotic variance

- **Theorem.** Under Assumptions I to III:

$$\boxed{\sqrt{n}(\hat{\beta}_n - \beta) \to_d N\Big(0,\ \underbrace{\left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1} \mathrm{E}\left[U_i^2 X_i X_i^\top\right] \left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1}}_{\mathrm{AsyVar}(\hat{\beta}_n)}\Big)}$$

$$\mathrm{AsyVar}(\hat{\beta}_n) = Q^{-1}\Omega Q^{-1} \quad \text{(sandwich)}$$

- CDF form, for all $a \neq 0$ and $t \in \mathbb{R}$:

$$\mathrm{P}(a^\top \sqrt{n}(\hat{\beta}_n - \beta) \leq t) \to \Phi\Big(\frac{t}{\sqrt{a^\top Q^{-1}\Omega Q^{-1} a}}\Big)$$

- $a = e_j$, element j:

$$\sqrt{n}(\hat{\beta}_{n,j} - \beta_j) \to_d N(0,\ [Q^{-1}\Omega Q^{-1}]_{jj})$$

Homoskedasticity

- Suppose:

$$\mathrm{E}\left[U_i^2 \mid X_i\right] = \sigma^2 \quad \text{(homoskedasticity)}$$

- Law of iterated expectations:

$$\begin{aligned}
\Omega &= \mathrm{E}\left[U_i^2 X_i X_i^\top\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[U_i^2 X_i X_i^\top \mid X_i\right]\right] \\
&= \mathrm{E}\left[\underbrace{\mathrm{E}\left[U_i^2 \mid X_i\right]}_{=\sigma^2} X_i X_i^\top\right] \\
&= \sigma^2 \mathrm{E}\left[X_i X_i^\top\right] \\
&= \sigma^2 Q
\end{aligned}$$

- The sandwich collapses:

$$\begin{aligned}
\mathrm{AsyVar}(\hat{\beta}_n) &= Q^{-1}\Omega Q^{-1} \\
&= Q^{-1}(\sigma^2 Q)Q^{-1} \\
&= \sigma^2 Q^{-1}
\end{aligned}$$

- Therefore:

$$\boxed{\sqrt{n}(\hat{\beta}_n - \beta) \to_d N\Big(0,\ \sigma^2 \left(\mathrm{E}\left[X_i X_i^\top\right]\right)^{-1}\Big)}$$

Finite sample against large sample

- Normal errors (Lecture 4, Assumptions 1 to 5), exact for every n (requires stronger assumptions), $a \neq 0$:

$$\begin{aligned}
& Y = X\beta + U, \qquad U \mid X \sim N(0, \sigma^2 I_n) \\
\implies\ & \hat{\beta}_n \mid X \sim N(\beta,\ \sigma^2 (X^\top X)^{-1}) \\
\implies\ & \mathrm{P}(a^\top(\hat{\beta}_n - \beta) \leq t \mid X) = \Phi\Big(\frac{t}{\sqrt{\sigma^2 a^\top (X^\top X)^{-1} a}}\Big)
\end{aligned}$$

- Homoskedastic errors, large sample, $a \neq 0$:

$$\begin{aligned} & \text{Assumptions I and II,} \quad \mathrm{E}\left[X_i U_i\right] = 0, \quad \mathrm{E}\left[U_i^2 \mid X_i\right] = \sigma^2 > 0 \\ \implies \quad & \mathrm{P}(a^\top \sqrt{n}(\hat{\beta}_n - \beta) \leq t) \to \Phi\Big(\frac{t}{\sqrt{\sigma^2 a^\top (\mathrm{E}\left[X_i X_i^\top\right])^{-1} a}}\Big) \end{aligned}$$

- General, large sample, $a \neq 0$:

$$\begin{aligned} & \text{Assumptions I to III;} \quad Q = \mathrm{E}\left[X_i X_i^\top\right], \quad \Omega = \mathrm{E}\left[U_i^2 X_i X_i^\top\right] \\ \implies \quad & \mathrm{P}(a^\top \sqrt{n}(\hat{\beta}_n - \beta) \leq t) \to \Phi\Big(\frac{t}{\sqrt{a^\top Q^{-1} \Omega Q^{-1} a}}\Big) \end{aligned}$$

Estimation of the asymptotic variance

The estimator $\hat{V}_n$

- Unknown:

$$\begin{aligned} & V = \mathrm{AsyVar}(\hat{\beta}_n) = Q^{-1} \Omega Q^{-1} \\ & Q = \mathrm{E}\left[X_i X_i^\top\right], \qquad \Omega = \mathrm{E}\left[U_i^2 X_i X_i^\top\right] \end{aligned}$$

- Estimator, sample averages in place of expectations:

$$\hat{V}_n = \hat{Q}_n^{-1} \hat{\Omega}_n \hat{Q}_n^{-1}$$

- $\hat{Q}_n$: weak LLN, then Slutsky (Assumption II):

$$\begin{aligned} & \hat{Q}_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top \to_p \mathrm{E}\left[X_i X_i^\top\right] = Q \\ & \hat{Q}_n^{-1} \to_p Q^{-1} \end{aligned}$$

- $\hat{\Omega}_n$, residuals $\hat{U}_i$ in place of U_i:

$$\hat{\Omega}_n = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top$$

- Residual and error, i fixed (continuous mapping theorem, without proof):

$$\begin{aligned} \hat{U}_i &= Y_i - X_i^\top \hat{\beta}_n \\ &= X_i^\top \beta + U_i - X_i^\top \hat{\beta}_n \\ &= U_i + X_i^\top \underbrace{(\beta - \hat{\beta}_n)}_{\to_p 0} \\ \implies \quad & \hat{U}_i \to_p U_i \\ \implies \quad & \hat{U}_1^2 + \hat{U}_2^2 \to_p U_1^2 + U_2^2 \end{aligned}$$

Can we use the LLN on $\hat{\Omega}_n$?

- **Definition.** Small o_p:

$$Z_n = o_p(1) \quad \iff \quad Z_n \to_p 0$$

- One $o_p(1)$ term for each i, $\hat{U}_i = U_i + o_p^i(1)$:

$$\begin{aligned}\hat{\Omega}_n &= \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^n \left(U_i + o_p^i(1)\right)^2 X_i X_i^\top \\ &\stackrel{?}{=} \frac{1}{n}\sum_{i=1}^n U_i^2 X_i X_i^\top + o_p(1)\end{aligned}$$

- $\sum_{i=1}^n o_p^i(1)$ can be large.
- Weak LLN? Summands not iid:

$$\hat{U}_i = U_i - X_i^\top \underbrace{(\hat{\beta}_n - \beta)}_{\text{random}}$$

- $\hat{U}_1$ and $\hat{U}_2$: correlated via $\hat{\beta}_n$.

Matrix norm

- To show:

$$\hat{\Omega}_n = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top \to_p \mathrm{E}\left[U_i^2 X_i X_i^\top\right] = \Omega, \qquad \underset{k\times k}{\Omega}, \quad \underset{k\times k}{X_i X_i^\top}$$

- Frobenius norm, $\underset{k\times k}{A}$ (Euclidean norm of the k^2 elements):

$$\begin{aligned}\|A\| &= \sqrt{\sum_{r=1}^k \sum_{s=1}^k A_{rs}^2} \\ &= \sqrt{\mathrm{tr}(A^\top A)}\end{aligned}$$

- Element by element $\Longleftrightarrow$ in norm (Lecture 8); to show:

$$\| \underbrace{\hat{\Omega}_n - \Omega}_{k\times k} \| \to_p 0$$

Expansion of $\hat{\Omega}_n$

- Substitute $\hat{U}_i = U_i - X_i^\top(\hat{\beta}_n - \beta)$; square, $(a-b)^2 = a^2 + b^2 - 2ab$:

$$\begin{aligned}\hat{\Omega}_n &= \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^n \big(\underbrace{Y_i - X_i^\top \hat{\beta}_n}_{\hat{U}_i}\big)^2 X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^n \big(U_i - \underbrace{X_i^\top(\hat{\beta}_n - \beta)}_{1\times 1}\big)^2 X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^n \Big(U_i^2 + \big(X_i^\top(\hat{\beta}_n - \beta)\big)^2 - 2\,U_i X_i^\top(\hat{\beta}_n - \beta)\Big) X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^n U_i^2 X_i X_i^\top + \frac{1}{n}\sum_{i=1}^n \big(X_i^\top(\hat{\beta}_n - \beta)\big)^2 X_i X_i^\top \\ &\quad - 2\,\frac{1}{n}\sum_{i=1}^n (U_i X_i^\top(\hat{\beta}_n - \beta)) X_i X_i^\top\end{aligned}$$

- First term, weak LLN, Ω finite (Assumption III):

$$\frac{1}{n}\sum_{i=1}^{n} U_i^2 X_i X_i^\top \to_p \mathrm{E}\left[U_i^2 X_i X_i^\top\right] = \Omega$$

- To show: the other two terms $\to_p 0$.

Cross term: triangle inequality

- Triangle inequality, $\|a+b\| \le \|a\| + \|b\|$; then $\|cA\| = |c|\,\|A\|$ for 1×1 c; then $|ab| = |a|\,|b|$ for 1×1 a, b:

$$\begin{aligned}
\Big\|\frac{1}{n}\sum_{i=1}^{n} (U_i X_i^\top(\hat\beta_n - \beta)) \underset{k\times k}{X_i X_i^\top}\Big\| &\le \frac{1}{n}\sum_{i=1}^{n} \Big\| \underbrace{(U_i X_i^\top(\hat\beta_n - \beta))}_{1\times 1} X_i X_i^\top\Big\| \\
&= \frac{1}{n}\sum_{i=1}^{n} \left|U_i X_i^\top(\hat\beta_n - \beta)\right| \|X_i X_i^\top\| \\
&= \frac{1}{n}\sum_{i=1}^{n} |U_i| \left|X_i^\top(\hat\beta_n - \beta)\right| \|X_i X_i^\top\|
\end{aligned}$$

Cauchy-Schwarz and $\|X_i X_i^\top\|$

- Cauchy-Schwarz:

$$\left|X_i^\top(\hat\beta_n - \beta)\right| \le \|X_i\|\,\|\hat\beta_n - \beta\|$$

- $\underset{k\times k}{A} = X_i X_i^\top = A^\top$; $\mathrm{tr}(BC) = \mathrm{tr}(CB)$:

$$\begin{aligned}
\|X_i X_i^\top\| &= \sqrt{\mathrm{tr}\left(\underbrace{X_i X_i^\top}_{A^\top}\ \underbrace{X_i X_i^\top}_{A}\right)} \\
&= \sqrt{\mathrm{tr}\left(X_i \underbrace{X_i^\top X_i}_{=\|X_i\|^2} X_i^\top\right)} \\
&= \sqrt{\mathrm{tr}\left(X_i \underbrace{\|X_i\|^2}_{1\times 1} X_i^\top\right)} \\
&= \sqrt{\|X_i\|^2 \,\mathrm{tr}\left(X_i X_i^\top\right)} \\
&= \|X_i\| \sqrt{\mathrm{tr}\left(X_i X_i^\top\right)} \\
&= \|X_i\| \sqrt{\mathrm{tr}\left(X_i^\top X_i\right)} \\
&= \|X_i\| \sqrt{\|X_i\|^2} \\
&= \|X_i\|^2
\end{aligned}$$

Cross term $\to_p 0$

- Cauchy-Schwarz and $\|X_i X_i^\top\| = \|X_i\|^2$:

$$\begin{aligned}
\underbrace{\Big\|\frac{1}{n}\sum_{i=1}^{n} (U_i X_i^\top(\hat\beta_n - \beta)) X_i X_i^\top\Big\|}_{1\times 1} &\le \frac{1}{n}\sum_{i=1}^{n} |U_i| \left|X_i^\top(\hat\beta_n - \beta)\right| \|X_i X_i^\top\| \\
&\le \frac{1}{n}\sum_{i=1}^{n} \underset{1\times 1}{|U_i|}\ \underset{1\times 1}{\|X_i\|}\ \underset{1\times 1}{\|\hat\beta_n - \beta\|}\ \underset{1\times 1}{\|X_i\|^2} \\
&= \Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n} |U_i|\,\|X_i\|^3}_{\to_p \mathrm{E}[|U_i|\,\|X_i\|^3]}\Big) \cdot \underbrace{\|\hat\beta_n - \beta\|}_{\to_p 0}
\end{aligned}$$

- Weak LLN, provided $\mathrm{E}\left[|U_i| \, \|X_i\|^3\right] < \infty$. Cauchy-Schwarz twice:

$$\begin{aligned}\mathrm{E}\left[|U_i| \, \|X_i\|^3\right] &= \mathrm{E}\left[|U_i| \, \|X_i\| \cdot \|X_i\|^2\right] \\ &\leq \sqrt{\mathrm{E}\left[U_i^2 \|X_i\|^2\right] \, \mathrm{E}\left[\|X_i\|^4\right]} \\ &\leq \sqrt{\sqrt{\mathrm{E}\left[U_i^4\right] \, \mathrm{E}\left[\|X_i\|^4\right]} \cdot \mathrm{E}\left[\|X_i\|^4\right]}\end{aligned}$$

- **Assume:**

$$\mathrm{E}\left[U_i^4\right] < \infty, \qquad \mathrm{E}\left[\|X_i\|^4\right] < \infty$$

- Slutsky, bound $\to_p 0$; non-negative norm:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(U_i X_i^\top(\hat{\beta}_n - \beta))X_i X_i^\top\right\| \to_p 0$$

Squared term $\to_p 0$

- Triangle inequality and $\|cA\| = |c| \, \|A\|$, then Cauchy-Schwarz and $\|X_i X_i^\top\| = \|X_i\|^2$:

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\Big(\underbrace{X_i^\top(\hat{\beta}_n - \beta)}_{1\times 1}\Big)^2 X_i X_i^\top\right\| &\leq \frac{1}{n}\sum_{i=1}^{n}\left(X_i^\top(\hat{\beta}_n - \beta)\right)^2 \|X_i X_i^\top\| \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\|X_i\|^2 \, \|\hat{\beta}_n - \beta\|^2 \, \|X_i\|^2 \\ &= \Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n}\|X_i\|^4}_{\to_p \mathrm{E}[\|X_i\|^4] < \infty}\Big) \cdot \underbrace{\|\hat{\beta}_n - \beta\|^2}_{\to_p 0}\end{aligned}$$

- Slutsky, bound $\to_p 0$; non-negative norm:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left(X_i^\top(\hat{\beta}_n - \beta)\right)^2 X_i X_i^\top\right\| \to_p 0$$

$\hat{\Omega}_n \to_p \Omega$ and $\hat{V}_n \to_p V$

- Three terms:

$$\begin{aligned}\hat{\Omega}_n &= \frac{1}{n}\sum_{i=1}^{n}\hat{U}_i^2 X_i X_i^\top \\ &= \frac{1}{n}\sum_{i=1}^{n}U_i^2 X_i X_i^\top + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(X_i^\top(\hat{\beta}_n - \beta)\right)^2 X_i X_i^\top}_{\to_p 0} \\ &\quad - 2\,\underbrace{\frac{1}{n}\sum_{i=1}^{n}(U_i X_i^\top(\hat{\beta}_n - \beta))X_i X_i^\top}_{\to_p 0} \\ &= \underbrace{\frac{1}{n}\sum_{i=1}^{n}U_i^2 X_i X_i^\top}_{\to_p \mathrm{E}[U_i^2 X_i X_i^\top] = \Omega} + o_p(1) \\ &\to_p \Omega\end{aligned}$$

- **Theorem.** Under Assumptions I to III, $\mathrm{E}\left[U_i^4\right] < \infty$ and $\mathrm{E}\left[\|X_i\|^4\right] < \infty$, with or without homoskedasticity:

$$\boxed{\hat{V}_n = \hat{Q}_n^{-1}\hat{\Omega}_n\hat{Q}_n^{-1} \to_p Q^{-1}\Omega Q^{-1} = V} \quad (\text{Slutsky, } \hat{Q}_n^{-1} \to_p Q^{-1})$$

Robust and usual standard errors

Robust standard errors

- $\overset{a}{\sim}$: approximately distributed, for large n:

$$\sqrt{n}(\hat{\beta}_n - \beta) \overset{a}{\sim} N(0, V) \quad \Longrightarrow \quad \hat{\beta}_{n,j} \overset{a}{\sim} N\Big(\beta_j,\ \frac{[V]_{jj}}{n}\Big)$$

- **Definition.** Robust (heteroskedasticity-consistent, White) standard error; $\widehat{Q}_n = \frac{1}{n} X^\top X$:

$$\begin{aligned}
&\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j}) \\
&= \sqrt{[\hat{V}_n]_{jj}/n} \\
&= \sqrt{\Big[\frac{1}{n}\Big(\frac{1}{n}X^\top X\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top\Big)\Big(\frac{1}{n}X^\top X\Big)^{-1}\Big]_{jj}} \\
&= \sqrt{\Big[\frac{1}{n}\cdot n(X^\top X)^{-1}\cdot\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top \cdot n(X^\top X)^{-1}\Big]_{jj}} \\
&= \sqrt{\Big[(X^\top X)^{-1}\Big(\sum_{i=1}^n \hat{U}_i^2 X_i X_i^\top\Big)(X^\top X)^{-1}\Big]_{jj}}
\end{aligned}$$

- Lecture 2, heteroskedastic errors, $\text{Var}\,(U_i \mid X) = \sigma_i^2$; $\hat{U}_i^2$ in place of σ_i^2:

$$\text{Var}\,\big(\hat{\beta} \mid X\big) = (X^\top X)^{-1}\Big(\sum_{i=1}^n \sigma_i^2 X_i X_i^\top\Big)(X^\top X)^{-1}$$

t-ratio with robust standard errors

- We will use, without proof (Cramér convergence theorem): $Z_n \to_d Z$, $c_n \to_p c \neq 0$, c non-random:

$$\frac{Z_n}{c_n} \to_d \frac{Z}{c}$$

- Slutsky, $\sqrt{[\hat{V}_n]_{jj}} \to_p \sqrt{[V]_{jj}} > 0$ (V p.d.); $\sqrt{n}(\hat{\beta}_{n,j} - \beta_j) \to_d N(0, [V]_{jj})$:

$$\begin{aligned}
\frac{\hat{\beta}_{n,j} - \beta_j}{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})} &= \frac{\hat{\beta}_{n,j} - \beta_j}{\sqrt{[\hat{V}_n]_{jj}/n}} \\
&= \frac{\sqrt{n}(\hat{\beta}_{n,j} - \beta_j)}{\sqrt{[\hat{V}_n]_{jj}}} \\
&\to_d \frac{N(0, [V]_{jj})}{\sqrt{[V]_{jj}}} \\
&\overset{d}{=} N(0,1)
\end{aligned}$$

- Asymptotic confidence interval:

$$\begin{aligned}
CI_{1-\alpha} = \Big[&\hat{\beta}_{n,j} - z_{1-\alpha/2}\,\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j}), \\
&\hat{\beta}_{n,j} + z_{1-\alpha/2}\,\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})\Big]
\end{aligned}$$

- Coverage, t-ratio $\to_d Z \sim N(0,1)$:

$$\begin{aligned}\mathrm{P}(\beta_j \in CI_{1-\alpha}) &= \mathrm{P}\Big(\Big|\frac{\hat{\beta}_{n,j} - \beta_j}{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})}\Big| \le z_{1-\alpha/2}\Big) \\ &\to \mathrm{P}(|Z| \le z_{1-\alpha/2}) \\ &= 1-\alpha\end{aligned}$$

Usual standard errors

- Lecture 3:

$$\text{std. err}(\hat{\beta}_{n,j}) = \sqrt{s^2\big[(X^\top X)^{-1}\big]_{jj}}, \qquad s^2 = \frac{1}{n-k}\sum_{i=1}^n \hat{U}_i^2$$

- In terms of $\hat{Q}_n = \frac{1}{n}X^\top X$:

$$\begin{aligned}s^2(X^\top X)^{-1} &= \frac{1}{n}\, s^2\Big(\frac{1}{n}X^\top X\Big)^{-1} \\ &= \frac{1}{n}\, s^2 \hat{Q}_n^{-1}\end{aligned}$$

- Usual: $s^2\hat{Q}_n^{-1}$, built for $\sigma^2 Q^{-1}$ (homoskedasticity); robust: $\hat{V}_n$, for V:

$$\text{std. err}(\hat{\beta}_{n,j}) = \sqrt{[s^2\hat{Q}_n^{-1}]_{jj}/n}, \qquad \text{std. err}_{\text{rob}}(\hat{\beta}_{n,j}) = \sqrt{[\hat{V}_n]_{jj}/n}$$

$s^2 \to_p \mathrm{E}\left[U_i^2\right]$

- $\hat{U}_i = U_i - X_i^\top(\hat{\beta}_n - \beta)$; $U_i X_i^\top(\hat{\beta}_n - \beta) = (\hat{\beta}_n - \beta)^\top X_i U_i$; $\mathrm{E}\left[U_i^2\right] \le \sqrt{\mathrm{E}\left[U_i^4\right]} < \infty$; weak LLN, consistency, Slutsky:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 &= \frac{1}{n}\sum_{i=1}^n \left(U_i - X_i^\top(\hat{\beta}_n - \beta)\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^n U_i^2 - 2\,\frac{1}{n}\sum_{i=1}^n U_i X_i^\top(\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{n}\sum_{i=1}^n \left(X_i^\top(\hat{\beta}_n - \beta)\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^n U_i^2 - 2\,(\hat{\beta}_n - \beta)^\top \frac{1}{n}\sum_{i=1}^n X_i U_i \\ &\quad + (\hat{\beta}_n - \beta)^\top\Big(\frac{1}{n}\sum_{i=1}^n X_i X_i^\top\Big)(\hat{\beta}_n - \beta) \\ &\to_p \mathrm{E}\left[U_i^2\right] - 2\cdot 0^\top 0 + 0^\top Q\, 0 \\ &= \mathrm{E}\left[U_i^2\right]\end{aligned}$$

- $\frac{n}{n-k} \to 1$:

$$\begin{aligned}s^2 &= \frac{n}{n-k}\cdot\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 \\ &\to_p 1\cdot \mathrm{E}\left[U_i^2\right] = \mathrm{E}\left[U_i^2\right]\end{aligned}$$

- Homoskedasticity, law of iterated expectations:

$$\mathrm{E}\left[U_i^2\right] = \mathrm{E}\left[\mathrm{E}\left[U_i^2 \mid X_i\right]\right] = \sigma^2$$

Homoskedasticity: robust and usual agree

- Both estimators of $\text{AsyVar}(\hat{\beta}_n) = \sigma^2 Q^{-1}$ consistent:

$$\begin{aligned} \hat{V}_n &\to_p Q^{-1}\Omega Q^{-1} = \sigma^2 Q^{-1} \\ s^2\hat{Q}_n^{-1} &\to_p \sigma^2 Q^{-1} \end{aligned}$$

- Ratio of the two standard errors, Slutsky, $\sigma^2[Q^{-1}]_{jj} = [V]_{jj} > 0$:

$$\begin{aligned} \frac{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})}{\text{std. err}(\hat{\beta}_{n,j})} &= \sqrt{\frac{[\hat{V}_n]_{jj}/n}{[s^2\hat{Q}_n^{-1}]_{jj}/n}} \\ &= \sqrt{\frac{[\hat{V}_n]_{jj}}{[s^2\hat{Q}_n^{-1}]_{jj}}} \\ &\to_p \sqrt{\frac{\sigma^2[Q^{-1}]_{jj}}{\sigma^2[Q^{-1}]_{jj}}} \\ &= 1 \end{aligned}$$

- Therefore:

$$\boxed{\frac{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})}{\text{std. err}(\hat{\beta}_{n,j})} \to_p 1}$$

- Usual t-ratio, $A_n Z_n \to_d AZ$:

$$\begin{aligned} \frac{\hat{\beta}_{n,j} - \beta_j}{\text{std. err}(\hat{\beta}_{n,j})} &= \underbrace{\frac{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})}{\text{std. err}(\hat{\beta}_{n,j})}}_{\to_p 1} \cdot \underbrace{\frac{\hat{\beta}_{n,j} - \beta_j}{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})}}_{\to_d N(0,1)} \\ &\to_d N(0,1) \end{aligned}$$

Heteroskedasticity: usual standard errors fail

- Without homoskedasticity:

$$s^2\hat{Q}_n^{-1} \to_p \text{E}\left[U_i^2\right] Q^{-1} \neq Q^{-1}\Omega Q^{-1} = V \quad \text{(in general)}$$

- Example, $k = 1$, no intercept: $Q = \text{E}\left[X_i^2\right]$, $\Omega = \text{E}\left[U_i^2 X_i^2\right]$:

$$\begin{aligned} V &= \frac{\text{E}\left[U_i^2 X_i^2\right]}{\left(\text{E}\left[X_i^2\right]\right)^2}, \qquad \text{usual: } \frac{\text{E}\left[U_i^2\right]}{\text{E}\left[X_i^2\right]} \\ V - \frac{\text{E}\left[U_i^2\right]}{\text{E}\left[X_i^2\right]} &= \frac{\text{E}\left[U_i^2 X_i^2\right] - \text{E}\left[U_i^2\right]\text{E}\left[X_i^2\right]}{\left(\text{E}\left[X_i^2\right]\right)^2} \\ &= \frac{\text{Cov}\left(U_i^2, X_i^2\right)}{\left(\text{E}\left[X_i^2\right]\right)^2} \end{aligned}$$

- $\text{Cov}\left(U_i^2, X_i^2\right) > 0$, errors more variable where $|X_i|$ is large: usual standard error too small.
- Robust: $\hat{V}_n \to_p V$ with or without homoskedasticity.

Normality: exact and asymptotic

- Normal errors (Lecture 4, Assumptions 1 to 5):

$$\begin{aligned} & U \mid X \sim N(0, \sigma^2 I_n) \\ \implies \quad & \hat{\beta}_n \mid X \sim N(\beta,\ \sigma^2 (X^\top X)^{-1}) \quad \text{for every } n \end{aligned}$$

- Variance of $\sqrt{n}(\hat{\beta}_n - \beta)$ given X, and its limit under Assumptions I and II:

$$\begin{aligned} \operatorname{Var}\left(\sqrt{n}(\hat{\beta}_n - \beta) \mid X\right) &= n \operatorname{Var}\left(\hat{\beta}_n \mid X\right) \\ &= n\, \sigma^2 (X^\top X)^{-1} \\ &= \sigma^2 \Big(\frac{1}{n} X^\top X\Big)^{-1} \\ &= \sigma^2 \hat{Q}_n^{-1} \\ &\to_p \sigma^2 Q^{-1} = \operatorname{AsyVar}(\hat{\beta}_n) \end{aligned}$$

- σ^2 estimated by s^2 (typed notes, Lecture 4):

$$\begin{aligned} & \frac{\hat{\beta}_{n,j} - \beta_j}{\text{std. err}(\hat{\beta}_{n,j})} \,\Big|\, X \sim t_{n-k} \quad \text{for every } n \\ & t_{n-k,\,1-\alpha/2} \to z_{1-\alpha/2} \end{aligned}$$

- Normal errors are homoskedastic, so robust and usual agree in large samples ($\mathrm{E}\left[U_i^4\right] = 3\sigma^4$, $\mathrm{E}\left[\|X_i\|^4\right] < \infty$):

$$\begin{aligned} & \mathrm{E}\left[U_i^2 \mid X\right] = \left[\sigma^2 I_n\right]_{ii} = \sigma^2 \\ \implies \quad & \mathrm{E}\left[U_i^2 \mid X_i\right] = \mathrm{E}\left[\mathrm{E}\left[U_i^2 \mid X\right] \mid X_i\right] = \sigma^2 \\ \implies \quad & \frac{\text{std. err}_{\text{rob}}(\hat{\beta}_{n,j})}{\text{std. err}(\hat{\beta}_{n,j})} \to_p 1 \end{aligned}$$

Usual and robust standard errors compared

- Validity of the t-ratio (large n: conditions for $\hat{V}_n \to_p V$):

Errors	Usual std. err	Robust std. err
$U \mid X \sim N(0, \sigma^2 I_n)$	exact, t_{n-k}	large n, $N(0,1)$
$\mathrm{E}\left[U_i^2 \mid X_i\right] = \sigma^2$	large n, $N(0,1)$	large n, $N(0,1)$
$\mathrm{E}\left[U_i^2 \mid X_i\right]$ depends on X_i	invalid in general	large n, $N(0,1)$

- Rows 1 and 2, $t_{n-k,\,1-\alpha/2} \to z_{1-\alpha/2}$ and std. err_{rob}/std. err $\to_p 1$; the three intervals agree in large samples:

$$\begin{aligned} \text{usual, exact (row 1):} \quad & \hat{\beta}_{n,j} \pm t_{n-k,\,1-\alpha/2}\, \text{std. err}(\hat{\beta}_{n,j}) \\ \text{usual, large } n\text{:} \quad & \hat{\beta}_{n,j} \pm z_{1-\alpha/2}\, \text{std. err}(\hat{\beta}_{n,j}) \\ \text{robust, large } n\text{:} \quad & \hat{\beta}_{n,j} \pm z_{1-\alpha/2}\, \text{std. err}_{\text{rob}}(\hat{\beta}_{n,j}) \end{aligned}$$

Summary

Consistency and asymptotic normality

- $\hat{\beta}_n = \beta + \left(\frac{1}{n}\sum X_i X_i^\top\right)^{-1} \frac{1}{n}\sum X_i U_i$; iid, $\mathrm{E}\left[X_i U_i\right] = 0$, $\mathrm{E}\left[U_i^2\right] < \infty$, $\mathrm{E}\left[X_{i,j}^2\right] < \infty$, $\mathrm{E}\left[X_i X_i^\top\right]$ p.d. $\implies \hat{\beta}_n \to_p \beta$.

- $\sqrt{n}(\hat{\beta}_n - \beta) = \left(\frac{1}{n}\sum X_i X_i^\top\right)^{-1} \frac{1}{\sqrt{n}} \sum X_i U_i$.
- Assumptions I to III $\implies \sqrt{n}(\hat{\beta}_n - \beta) \to_d N(0, Q^{-1}\Omega Q^{-1})$, $Q = \mathrm{E}\left[X_i X_i^\top\right]$, $\Omega = \mathrm{E}\left[U_i^2 X_i X_i^\top\right]$.
- $\mathrm{E}\left[U_i^2 \mid X_i\right] = \sigma^2 \implies \Omega = \sigma^2 Q$, $\mathrm{AsyVar}(\hat{\beta}_n) = \sigma^2 Q^{-1}$.

Variance estimation and standard errors

- $\hat{V}_n = \hat{Q}_n^{-1}\hat{\Omega}_n\hat{Q}_n^{-1}$, $\hat{Q}_n = \frac{1}{n}\sum X_i X_i^\top$, $\hat{\Omega}_n = \frac{1}{n}\sum \hat{U}_i^2 X_i X_i^\top$.
- $\hat{\Omega}_n = \frac{1}{n}\sum U_i^2 X_i X_i^\top + o_p(1)$ under Assumptions I to III, $\mathrm{E}\left[U_i^4\right] < \infty$, $\mathrm{E}\left[\|X_i\|^4\right] < \infty$; $\hat{V}_n \to_p V$.
- std. $\mathrm{err}_{\mathrm{rob}}(\hat{\beta}_{n,j}) = \sqrt{\left[(X^\top X)^{-1}(\sum \hat{U}_i^2 X_i X_i^\top)(X^\top X)^{-1}\right]_{jj}}$; t-ratio $\to_d N(0,1)$; $\hat{\beta}_{n,j} \pm z_{1-\alpha/2}$ std. $\mathrm{err}_{\mathrm{rob}}$: coverage $\to 1-\alpha$.
- std. $\mathrm{err}(\hat{\beta}_{n,j}) = \sqrt{[s^2\hat{Q}_n^{-1}]_{jj}/n}$, $s^2 \to_p \mathrm{E}\left[U_i^2\right]$; homoskedasticity: std. $\mathrm{err}_{\mathrm{rob}}$/std. err $\to_p 1$.
- Heteroskedasticity: $s^2\hat{Q}_n^{-1} \to_p \mathrm{E}\left[U_i^2\right] Q^{-1} \neq V$ in general; only the robust standard error is always valid.
- Normality: $\mathrm{Var}\left(\sqrt{n}(\hat{\beta}_n - \beta) \mid X\right) = \sigma^2\hat{Q}_n^{-1} \to_p \sigma^2 Q^{-1}$; usual t-ratio exactly t_{n-k}, $t_{n-k,\,1-\alpha/2} \to z_{1-\alpha/2}$; homoskedastic, so std. $\mathrm{err}_{\mathrm{rob}}$/std. err $\to_p 1$.