

Lecture 7: Omitted and irrelevant regressors

Economics 527: Econometric Methods

Vadim Marmer, UBC

Long and short regressions

Model and two estimators of β_1

- Two blocks of regressors, $k_1 + k_2 = k$:

$$Y = \underset{n \times k_1}{X_1} \underset{k_1 \times 1}{\beta_1} + \underset{n \times k_2}{X_2} \underset{k_2 \times 1}{\beta_2} + U, \quad X = [X_1 \ X_2]$$

- Assumptions 2, 3 and 4:

$$E[U \mid X_1, X_2] = 0, \quad \text{Var}(U \mid X_1, X_2) = \sigma^2 I_n, \quad \text{rank}(X) = k$$

- “Long” regression $Y \mid X_1, X_2$, controlling for X_2 (FWL):

$$\hat{\beta}_1 = (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 Y, \quad M_2 = I_n - X_2 (X_2^\top X_2)^{-1} X_2^\top$$

- “Short” regression $Y \mid X_1$, not controlling for X_2 :

$$\tilde{\beta}_1 = (X_1^\top X_1)^{-1} X_1^\top Y$$

Short against long in the sample

- Long regression’s fit (third lecture):

$$Y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{U}$$

- Substitute into $\tilde{\beta}_1$:

$$\begin{aligned} \tilde{\beta}_1 &= (X_1^\top X_1)^{-1} X_1^\top Y \\ &= (X_1^\top X_1)^{-1} X_1^\top (X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{U}) \\ &= \hat{\beta}_1 + (X_1^\top X_1)^{-1} X_1^\top X_2 \hat{\beta}_2 + \underbrace{(X_1^\top X_1)^{-1} X_1^\top \hat{U}}_{=0} \\ &= \boxed{\hat{\beta}_1 + (X_1^\top X_1)^{-1} X_1^\top X_2 \hat{\beta}_2} \\ &\neq \hat{\beta}_1 \quad \text{in general} \\ \tilde{\beta}_1 &= \hat{\beta}_1 \quad \text{if } X_1^\top X_2 = 0 \end{aligned}$$

Omitting a relevant regressor

Signal, bias and noise

- True model, estimated by the short regression $Y \mid X_1$:

$$Y = X_1 \beta_1 + X_2 \beta_2 + U, \quad \beta_2 \neq 0$$

- Substitute the true model into $\tilde{\beta}_1$:

$$\begin{aligned}\tilde{\beta}_1 &= (X_1^\top X_1)^{-1} X_1^\top Y \\ &= (X_1^\top X_1)^{-1} X_1^\top (X_1 \beta_1 + X_2 \beta_2 + U) \\ &= \underbrace{\beta_1}_{\text{signal}} + \underbrace{(X_1^\top X_1)^{-1} X_1^\top X_2 \beta_2}_{\text{bias}} + \underbrace{(X_1^\top X_1)^{-1} X_1^\top U}_{\text{noise}}\end{aligned}$$

- Conditional mean of $\tilde{\beta}_1$, by Assumption 2:

$$\begin{aligned}\mathbb{E} [\tilde{\beta}_1 \mid X_1, X_2] &= \beta_1 + (X_1^\top X_1)^{-1} X_1^\top X_2 \beta_2 + (X_1^\top X_1)^{-1} X_1^\top \underbrace{\mathbb{E} [U \mid X_1, X_2]}_{=0} \\ &= \boxed{\beta_1 + \underbrace{(X_1^\top X_1)^{-1} X_1^\top X_2 \beta_2}_{\text{bias}}} \\ &\neq \beta_1 \quad \text{in general}\end{aligned}$$

- Bias = 0 if:
 1. $\beta_2 = 0$
 2. $X_1^\top X_2 = 0$

Example: return to education, ability omitted

- Wage equation, $k_1 = k_2 = 1$:

$$\ln \text{Wage}_i = \underbrace{\text{Education}_i}_{X_{i,1}} \beta_1 + \underbrace{\text{Ability}_i}_{X_{i,2}} \beta_2 + U_i$$

- β_1 : return to education; $\beta_2 > 0$.
- Suppose $X_1^\top X_2 > 0$:

$$X_1^\top X_2 = \sum_{i=1}^n \text{Education}_i \cdot \text{Ability}_i > 0$$

- Sign of the bias, three positive factors:

$$\begin{aligned}\text{Bias} &= \underbrace{(X_1^\top X_1)^{-1}}_{+} \underbrace{X_1^\top X_2}_{+} \overset{+}{\beta_2} > 0 \\ \implies \mathbb{E} [\tilde{\beta}_1 \mid X_1, X_2] &> \beta_1\end{aligned}$$

- Return to education over-estimated.

Variances of the long and short estimators

Two conditional variances

- We will use, for Γ a function of X_1, X_2 :

$$\text{Var} (\Gamma U \mid X_1, X_2) = \Gamma \text{Var} (U \mid X_1, X_2) \Gamma^\top$$

- Short regression, by Assumption 3:

$$\begin{aligned}\text{Var} (\tilde{\beta}_1 \mid X_1, X_2) &= \text{Var} (\beta_1 + (X_1^\top X_1)^{-1} X_1^\top X_2 \beta_2 + (X_1^\top X_1)^{-1} X_1^\top U \mid X_1, X_2) \\ &= \text{Var} ((X_1^\top X_1)^{-1} X_1^\top U \mid X_1, X_2) \\ &= (X_1^\top X_1)^{-1} X_1^\top \underbrace{\text{Var} (U \mid X_1, X_2)}_{=\sigma^2 I_n} X_1 (X_1^\top X_1)^{-1} \\ &= \sigma^2 (X_1^\top X_1)^{-1}\end{aligned}$$

- Long regression:

$$\begin{aligned}
\text{Var}(\hat{\beta}_1 | X_1, X_2) &= \text{Var}(\beta_1 + (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 U | X_1, X_2) \\
&= \text{Var}((X_1^\top M_2 X_1)^{-1} X_1^\top M_2 U | X_1, X_2) \\
&= (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 \sigma^2 I_n M_2 X_1 (X_1^\top M_2 X_1)^{-1} \quad (M_2^\top = M_2) \\
&= \sigma^2 (X_1^\top M_2 X_1)^{-1} \underbrace{X_1^\top M_2 M_2 X_1}_{= X_1^\top M_2 X_1} (X_1^\top M_2 X_1)^{-1} \\
&= \sigma^2 (X_1^\top M_2 X_1)^{-1}
\end{aligned}$$

Larger variance for the long regression

- **Claim.** For $\beta_2 \neq 0$:

$$\text{Var}(\hat{\beta}_1 | X_1, X_2) - \text{Var}(\tilde{\beta}_1 | X_1, X_2) \geq 0 \quad (\text{p.s.d.})$$

- Difference of the two variances:

$$\begin{aligned}
\text{Var}(\hat{\beta}_1 | X_1, X_2) - \text{Var}(\tilde{\beta}_1 | X_1, X_2) &= \sigma^2 (X_1^\top M_2 X_1)^{-1} - \sigma^2 (X_1^\top X_1)^{-1} \\
&= \sigma^2 [(X_1^\top M_2 X_1)^{-1} - (X_1^\top X_1)^{-1}]
\end{aligned}$$

- Need to show:

$$(X_1^\top M_2 X_1)^{-1} - (X_1^\top X_1)^{-1} \geq 0$$

Proof, step 1: inverses to matrices

- Scalars $a, b > 0$:

$$b - a \geq 0 \iff \frac{1}{a} - \frac{1}{b} \geq 0$$

- **Claim** (homework). For A, B p.d.:

$$B - A \geq 0 \iff A^{-1} - B^{-1} \geq 0$$

- Apply the homework claim, A and B p.d. by Assumption 4:

$$\begin{aligned}
&\underbrace{(X_1^\top M_2 X_1)^{-1}}_{A^{-1}} - \underbrace{(X_1^\top X_1)^{-1}}_{B^{-1}} \geq 0 \\
\iff &\underbrace{X_1^\top X_1}_B - \underbrace{X_1^\top M_2 X_1}_A \geq 0
\end{aligned}$$

Proof, step 2: $X_1^\top P_2 X_1$ p.s.d.

- Rewrite the difference, $I_n - M_2 = P_2$:

$$\begin{aligned}
X_1^\top X_1 - X_1^\top M_2 X_1 &= X_1^\top (I_n - M_2) X_1 \\
&= X_1^\top P_2 X_1, \quad P_2 = X_2 (X_2^\top X_2)^{-1} X_2^\top
\end{aligned}$$

- P_2 p.s.d. $\implies$ quadratic form of $X_1^\top P_2 X_1$ p.s.d., for $c \in \mathbb{R}^{k_1}$:

$$c^\top X_1^\top P_2 X_1 c = w^\top P_2 w \geq 0, \quad w = X_1 c$$

- Therefore:

$$\begin{aligned}
&X_1^\top X_1 - X_1^\top M_2 X_1 = X_1^\top P_2 X_1 \geq 0 \\
\implies &(X_1^\top M_2 X_1)^{-1} - (X_1^\top X_1)^{-1} \geq 0 \quad (\text{step 1}) \\
\implies &\boxed{\text{Var}(\hat{\beta}_1 | X_1, X_2) - \text{Var}(\tilde{\beta}_1 | X_1, X_2) \geq 0}
\end{aligned}$$

Why controlling for X_2 increases the variance

- $X_1^\top M_2 X_1$ in terms of $\tilde{X}_1$:

$$\begin{aligned} X_1^\top M_2 X_1 &= \underbrace{X_1^\top M_2^\top}_{\tilde{X}_1^\top} \underbrace{M_2 X_1}_{\tilde{X}_1} \quad (\text{symmetric, idempotent}) \\ &= \tilde{X}_1^\top \tilde{X}_1 \end{aligned}$$

- $\tilde{X}_1 = M_2 X_1$: residual from regressing X_1 on X_2 .
- Conditional variances of $\hat{\beta}_1$ and $\tilde{\beta}_1$:

$$\begin{aligned} \text{Var}(\hat{\beta}_1 \mid X_1, X_2) &= \sigma^2 (X_1^\top M_2 X_1)^{-1} \\ &= \sigma^2 (\tilde{X}_1^\top \tilde{X}_1)^{-1} \\ \text{Var}(\tilde{\beta}_1 \mid X_1, X_2) &= \sigma^2 (X_1^\top X_1)^{-1} \end{aligned}$$

- $X_1^\top X_1$ against $\tilde{X}_1^\top \tilde{X}_1$, from step 2:

$$X_1^\top X_1 - \tilde{X}_1^\top \tilde{X}_1 \geq 0$$

- $\tilde{X}_1^\top \tilde{X}_1$: variation of X_1 not explained by X_2 .
- Adding $X_2 \implies$ variance $\uparrow$.

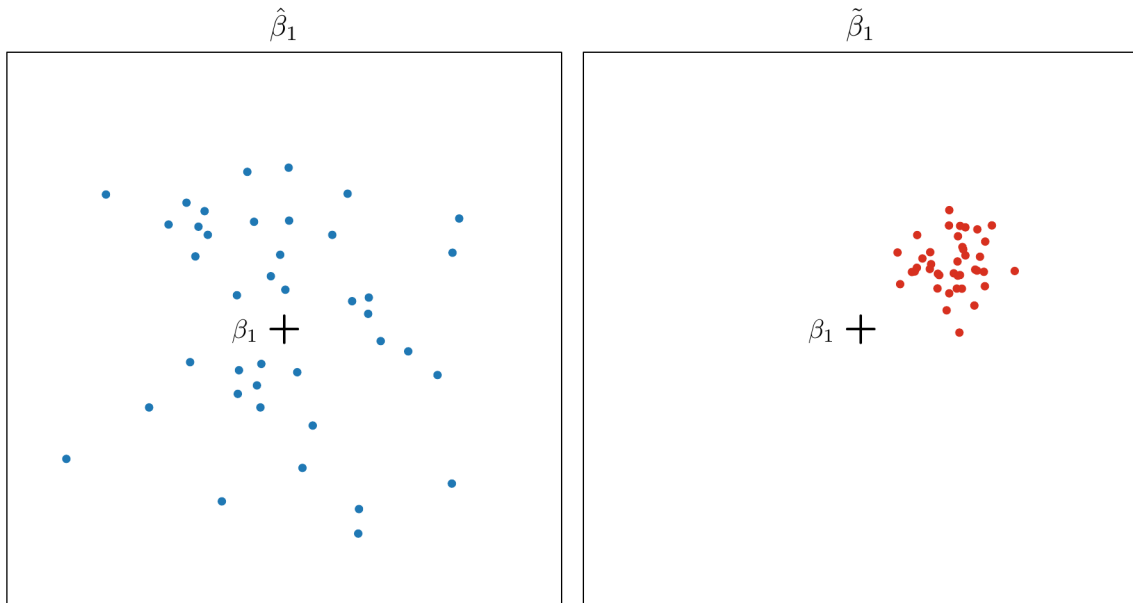
Bias against variance

- If $\beta_2 \neq 0$:

$\hat{\beta}_1$: unbiased, larger variance

$\tilde{\beta}_1$: biased in general, smaller variance

- Estimates from repeated samples, $k_1 = 2$; cross at the true β_1 :



Mean squared error (MSE)

- $\hat{\theta}_{1 \times 1}$: estimator of $\theta_{0, 1 \times 1}$.
- Expected loss, add and subtract $E[\hat{\theta}]$:

$$\begin{aligned} E[(\hat{\theta} - \theta_0)^2] &= E[(\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta_0)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta_0)^2 \\ &\quad + 2E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta_0)] \end{aligned}$$

- Cross term, $E[\hat{\theta}] - \theta_0$ not random:

$$\begin{aligned} E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta_0)] &= (E[\hat{\theta}] - \theta_0) \underbrace{E[\hat{\theta} - E[\hat{\theta}]]}_{=0} \\ &= 0 \end{aligned}$$

- Therefore:

$$\boxed{E[(\hat{\theta} - \theta_0)^2] = \underbrace{E[(\hat{\theta} - \theta_0)^2]}_{\text{MSE}} = \underbrace{E[(\hat{\theta} - E[\hat{\theta}])^2]}_{\text{Var}(\hat{\theta})} + \underbrace{(E[\hat{\theta}] - \theta_0)^2}_{\text{Bias}^2}}$$

- A biased estimator can have a smaller MSE.

Estimated variances

Two estimated variances

- Long regression, $\hat{U} = Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2$:

$$\widehat{\text{Var}}(\hat{\beta}_1 | X_1, X_2) = s^2(X_1^\top M_2 X_1)^{-1}, \quad s^2 = \frac{\hat{U}^\top \hat{U}}{n - k}$$

- Short regression, $\tilde{U} = Y - X_1\tilde{\beta}_1$:

$$\widehat{\text{Var}}(\tilde{\beta}_1 | X_1, X_2) = \tilde{s}^2(X_1^\top X_1)^{-1}, \quad \tilde{s}^2 = \frac{\tilde{U}^\top \tilde{U}}{n - k_1}$$

- Standard errors, j -th entries:

$$\begin{aligned} \text{std. err}(\hat{\beta}_{1,j}) &= \sqrt{s^2[(X_1^\top M_2 X_1)^{-1}]_{jj}} \\ \text{std. err}(\tilde{\beta}_{1,j}) &= \sqrt{\tilde{s}^2[(X_1^\top X_1)^{-1}]_{jj}} \end{aligned}$$

- Third lecture: $E[s^2 | X_1, X_2] = \sigma^2$.
- Sums of squared residuals, in every sample:

$$\begin{aligned} \hat{U}^\top \hat{U} &= \|Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2\|^2 \\ &\leq \|Y - X_1\tilde{\beta}_1\|^2 \\ &= \tilde{U}^\top \tilde{U} \end{aligned}$$

- In terms of s^2 and $\tilde{s}^2$:

$$s^2 = \frac{\hat{U}^\top \hat{U}}{n - k}$$

$$\tilde{s}^2 = \frac{\tilde{U}^\top \tilde{U}}{n - k_1}$$

- Two estimated variances, factor by factor:

$$\widehat{\text{Var}}(\hat{\beta}_1 \mid X_1, X_2) = \frac{s^2}{\mathbb{E}[s^2 \mid X_1, X_2] = \sigma^2} \underbrace{(X_1^\top M_2 X_1)^{-1}}_{\text{bigger}}$$

$$\widehat{\text{Var}}(\tilde{\beta}_1 \mid X_1, X_2) = \frac{\tilde{s}^2}{\mathbb{E}[\tilde{s}^2 \mid X_1, X_2] \geq \sigma^2} \underbrace{(X_1^\top X_1)^{-1}}_{\text{smaller}}$$

Proof that $\mathbb{E}[\tilde{s}^2 \mid X_1, X_2] \geq \sigma^2$

- M_1 , third lecture:

$$M_1 = I_n - X_1(X_1^\top X_1)^{-1}X_1^\top$$

- M_1 : symmetric, idempotent, p.s.d., $M_1 X_1 = 0$, $\text{tr}(M_1) = n - k_1$.
- $\tilde{U}$ under the true model:

$$\begin{aligned} \tilde{U} &= Y - X_1 \tilde{\beta}_1 \\ &= M_1 Y \\ &= M_1 (X_1 \beta_1 + X_2 \beta_2 + U) \\ &= \underbrace{M_1 X_1}_{=0} \beta_1 + M_1 X_2 \beta_2 + M_1 U \\ &= M_1 X_2 \beta_2 + M_1 U \end{aligned}$$

- We will use, for $a, b \in \mathbb{R}^n$:

$$\begin{aligned} (a + b)^\top (a + b) &= a^\top a + b^\top b + \underbrace{b^\top a + a^\top b}_{=2a^\top b} \\ &= a^\top a + b^\top b + 2a^\top b \end{aligned}$$

- Sum of squared residuals:

$$\begin{aligned} \tilde{U}^\top \tilde{U} &= \left(\underbrace{M_1 X_2 \beta_2}_a + \underbrace{M_1 U}_b \right)^\top (M_1 X_2 \beta_2 + M_1 U) \\ &= \beta_2^\top X_2^\top M_1 X_2 \beta_2 + U^\top M_1 U + 2\beta_2^\top X_2^\top M_1 U \quad (M_1^\top M_1 = M_1) \end{aligned}$$

- Conditional mean of $\tilde{U}^\top \tilde{U}$ term by term, trace argument (third lecture):

$$\begin{aligned} \mathbb{E}[\tilde{U}^\top \tilde{U} \mid X_1, X_2] &= \beta_2^\top X_2^\top M_1 X_2 \beta_2 + \underbrace{\mathbb{E}[U^\top M_1 U \mid X_1, X_2]}_{=\sigma^2 \text{tr}(M_1) = \sigma^2(n-k_1)} \\ &\quad + 2\beta_2^\top X_2^\top M_1 \underbrace{\mathbb{E}[U \mid X_1, X_2]}_{=0} \\ &= \beta_2^\top X_2^\top M_1 X_2 \beta_2 + \sigma^2(n - k_1) \end{aligned}$$

- Divide by $n - k_1$:

$$\begin{aligned} \mathbb{E}[\tilde{s}^2 \mid X_1, X_2] &= \frac{\mathbb{E}[\tilde{U}^\top \tilde{U} \mid X_1, X_2]}{n - k_1} \\ &= \boxed{\sigma^2 + \frac{\beta_2^\top X_2^\top M_1 X_2 \beta_2}{n - k_1}} \\ &\geq \sigma^2 = \mathbb{E}[s^2 \mid X_1, X_2] \end{aligned}$$

- $\tilde{s}^2$: biased upward in general.

Including an irrelevant regressor

Irrelevant regressor: both estimators unbiased

- True model, estimated by the long regression $Y \mid X_1, X_2$:

$$\begin{aligned} Y &= X_1\beta_1 + X_2\beta_2 + U, \quad \beta_2 = 0 \\ \implies Y &= X_1\beta_1 + U \end{aligned}$$

- Short regression:

$$\begin{aligned} \tilde{\beta}_1 &= (X_1^\top X_1)^{-1} X_1^\top (X_1\beta_1 + U) \\ &= \beta_1 + (X_1^\top X_1)^{-1} X_1^\top U \end{aligned}$$

- Conditional mean of $\tilde{\beta}_1$, by Assumption 2:

$$E[\tilde{\beta}_1 \mid X_1, X_2] = \beta_1$$

- Long regression:

$$\begin{aligned} \hat{\beta}_1 &= (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 (X_1\beta_1 + U) \\ &= \beta_1 + (X_1^\top M_2 X_1)^{-1} X_1^\top M_2 U \end{aligned}$$

- Conditional mean of $\hat{\beta}_1$, by Assumption 2:

$$E[\hat{\beta}_1 \mid X_1, X_2] = \beta_1$$

Irrelevant regressor: variances

- Same conditional variances as for $\beta_2 \neq 0$, by Assumption 3:

$$\begin{aligned} \text{Var}(\hat{\beta}_1 \mid X_1, X_2) &= \sigma^2 (X_1^\top M_2 X_1)^{-1} \\ \text{Var}(\tilde{\beta}_1 \mid X_1, X_2) &= \sigma^2 (X_1^\top X_1)^{-1} \end{aligned}$$

- By the Claim on the larger variance of the long regression:

$$\text{Var}(\hat{\beta}_1 \mid X_1, X_2) - \text{Var}(\tilde{\beta}_1 \mid X_1, X_2) \geq 0$$

- $\tilde{s}^2$ and s^2 both unbiased, trace argument, $\text{tr}(M_1) = n - k_1$, $\text{tr}(M_X) = n - k$:

$$E[\tilde{s}^2 \mid X_1, X_2] = E[s^2 \mid X_1, X_2] = \sigma^2$$

Two cases: $\beta_2 \neq 0$ and $\beta_2 = 0$

- Suppose $X_1^\top X_2 \neq 0$:

- $\beta_2 \neq 0 \implies \tilde{\beta}_1$: biased in general
- $\beta_2 = 0 \implies \hat{\beta}_1$: larger variance than $\tilde{\beta}_1$

Orthogonal regressors

Orthogonal blocks: $M_2 X_1 = X_1$

- Suppose $X_1^\top X_2 = 0$:

$$X_1^\top X_2 = 0 \iff \text{columns of } X_1 \in S^\perp(X_2)$$

- Then:

$$\begin{aligned} M_2 X_1 &= X_1 - P_2 X_1 \\ &= X_1 - X_2 (X_2^\top X_2)^{-1} \underbrace{X_2^\top X_1}_{=0} \\ &= X_1 \end{aligned}$$

- Exactly orthogonal dummies, high school (HS) and college (C), $k_1 = k_2 = 1$:

$$\begin{aligned} X_{i,1} &= \mathbf{1}(\text{Education}_i = \text{HS}) \\ X_{i,2} &= \mathbf{1}(\text{Education}_i = \text{C}) \end{aligned}$$

- Never both equal to 1:

$$\begin{aligned} X_{i,1} X_{i,2} &= 0 \quad \text{for every } i \\ \implies X_1^\top X_2 &= \sum_{i=1}^n X_{i,1} X_{i,2} = 0 \end{aligned}$$

Same estimate, same variance

- Estimators coincide, by $M_2^\top = M_2$:

$$\begin{aligned} \hat{\beta}_1 &= (X_1^\top \underbrace{M_2 X_1}_{=X_1})^{-1} \underbrace{X_1^\top M_2}_{=(M_2 X_1)^\top = X_1^\top} Y \\ &= (X_1^\top X_1)^{-1} X_1^\top Y \\ &= \tilde{\beta}_1 \end{aligned}$$

- Both unbiased:

$$E[\hat{\beta}_1 | X_1, X_2] = E[\tilde{\beta}_1 | X_1, X_2] = \beta_1$$

- Equal conditional variances:

$$\begin{aligned} \text{Var}(\hat{\beta}_1 | X_1, X_2) &= \sigma^2 (X_1^\top M_2 X_1)^{-1} \\ &= \sigma^2 (X_1^\top X_1)^{-1} \\ &= \text{Var}(\tilde{\beta}_1 | X_1, X_2) \end{aligned}$$

- $X_1^\top M_2 X_1$ against $X_1^\top X_1$:

$$\begin{aligned} X_1^\top X_1 - X_1^\top M_2 X_1 &\geq 0 \\ X_1^\top M_2 X_1 &= X_1^\top X_1 \quad \text{if } X_1^\top X_2 = 0 \end{aligned}$$

- Adding X_2 with $X_1^\top X_2 = 0 \implies$ variance unchanged.

Which regression to run

- $X_1^\top X_2 = 0$; two options: long $Y \mid X_1, X_2$, short $Y \mid X_1$.
- Short regression's error:

$$Y = X_1\beta_1 + \underbrace{X_2\beta_2 + U}_V$$

- $M_1 X_2 = X_2$, by the same chain as $M_2 X_1 = X_1$.
- Short regression's residuals keep $X_2\beta_2$:

$$\begin{aligned}\tilde{U} &= M_1 X_2 \beta_2 + M_1 U \\ &= X_2 \beta_2 + M_1 U\end{aligned}$$

- $\tilde{U}$ in terms of the long regression's fit, $\hat{\beta}_1 = \tilde{\beta}_1$:

$$\begin{aligned}\tilde{U} &= Y - X_1 \tilde{\beta}_1 \\ &= Y - X_1 \hat{\beta}_1 \\ &= X_2 \hat{\beta}_2 + \hat{U}\end{aligned}$$

- $X_2 \hat{\beta}_2$: not removed from $\tilde{U}$.

Controlling for orthogonal X_2 : smaller estimated variance in general

- $E[\tilde{s}^2 \mid X_1, X_2]$ with $M_1 X_2 = X_2$:

$$\begin{aligned}E[\tilde{s}^2 \mid X_1, X_2] &= \sigma^2 + \frac{\beta_2^\top X_2^\top M_1 X_2 \beta_2}{n - k_1} \\ &= \sigma^2 + \frac{\beta_2^\top X_2^\top X_2 \beta_2}{n - k_1} \\ &= \sigma^2 + \frac{\|X_2 \beta_2\|^2}{n - k_1} \\ &> \sigma^2 = E[s^2 \mid X_1, X_2] \quad \text{if } X_2 \beta_2 \neq 0\end{aligned}$$

- Estimated variances, same estimate $\hat{\beta}_1 = \tilde{\beta}_1$:

$$\begin{aligned}\widehat{\text{Var}}(\hat{\beta}_1 \mid X_1, X_2) &= s^2 (X_1^\top X_1)^{-1} \\ \widehat{\text{Var}}(\tilde{\beta}_1 \mid X_1, X_2) &= \tilde{s}^2 (X_1^\top X_1)^{-1}\end{aligned}$$

- Short regression's estimated variance: biased upward if $X_2 \beta_2 \neq 0$.
- $\implies$ Use $\hat{\beta}_1$, the long regression.
- Fourth and fifth lectures: std. err $\downarrow \implies$ confidence interval length $\downarrow$, power $\uparrow$.

Summary

Omitted and irrelevant regressors

- In every sample: $\tilde{\beta}_1 = \hat{\beta}_1 + (X_1^\top X_1)^{-1} X_1^\top X_2 \hat{\beta}_2$; equal if $X_1^\top X_2 = 0$.
- $E[\tilde{\beta}_1 \mid X_1, X_2] = \beta_1 + (X_1^\top X_1)^{-1} X_1^\top X_2 \beta_2$, bias = 0 if $\beta_2 = 0$ or $X_1^\top X_2 = 0$; $E[\hat{\beta}_1 \mid X_1, X_2] = \beta_1$.
- $\text{Var}(\hat{\beta}_1 \mid X_1, X_2) = \sigma^2 (X_1^\top M_2 X_1)^{-1} = \sigma^2 (\tilde{X}_1^\top \tilde{X}_1)^{-1}$, $\text{Var}(\tilde{\beta}_1 \mid X_1, X_2) = \sigma^2 (X_1^\top X_1)^{-1}$.

- $X_1^\top X_1 - X_1^\top M_2 X_1 = X_1^\top P_2 X_1 \geq 0 \implies \text{Var}(\hat{\beta}_1 | X_1, X_2) - \text{Var}(\tilde{\beta}_1 | X_1, X_2) \geq 0$.
- $\beta_2 \neq 0 \implies \hat{\beta}_1$ unbiased, larger variance; $\tilde{\beta}_1$ biased in general, smaller variance.
- $\text{MSE} = \text{Var}(\hat{\theta}) + \text{Bias}^2$.
- $\beta_2 = 0 \implies$ both unbiased, $\hat{\beta}_1$ larger variance; $\text{E}[s^2 | X_1, X_2] = \text{E}[\tilde{s}^2 | X_1, X_2] = \sigma^2$.

Estimated variances and orthogonal regressors

- $\widehat{\text{Var}}(\hat{\beta}_1 | X_1, X_2) = s^2(X_1^\top M_2 X_1)^{-1}$, $s^2 = \hat{U}^\top \hat{U} / (n - k)$; $\widehat{\text{Var}}(\tilde{\beta}_1 | X_1, X_2) = \tilde{s}^2(X_1^\top X_1)^{-1}$, $\tilde{s}^2 = \tilde{U}^\top \tilde{U} / (n - k_1)$; $\hat{U}^\top \hat{U} \leq \tilde{U}^\top \tilde{U}$.
- $\tilde{U} = M_1 X_2 \beta_2 + M_1 U$; $\text{E}[\tilde{s}^2 | X_1, X_2] = \sigma^2 + \beta_2^\top X_2^\top M_1 X_2 \beta_2 / (n - k_1) \geq \sigma^2 = \text{E}[s^2 | X_1, X_2]$.
- $X_1^\top X_2 = 0 \implies M_2 X_1 = X_1 \implies \hat{\beta}_1 = \tilde{\beta}_1$ and $\text{Var}(\hat{\beta}_1 | X_1, X_2) = \text{Var}(\tilde{\beta}_1 | X_1, X_2) = \sigma^2(X_1^\top X_1)^{-1}$.
- $X_1^\top X_2 = 0$, $X_2 \beta_2 \neq 0 \implies \tilde{U} = X_2 \beta_2 + M_1 U$, $\text{E}[\tilde{s}^2 | X_1, X_2] = \sigma^2 + \beta_2^\top X_2^\top X_2 \beta_2 / (n - k_1) > \sigma^2 \implies$ use $\hat{\beta}_1$, the long regression.