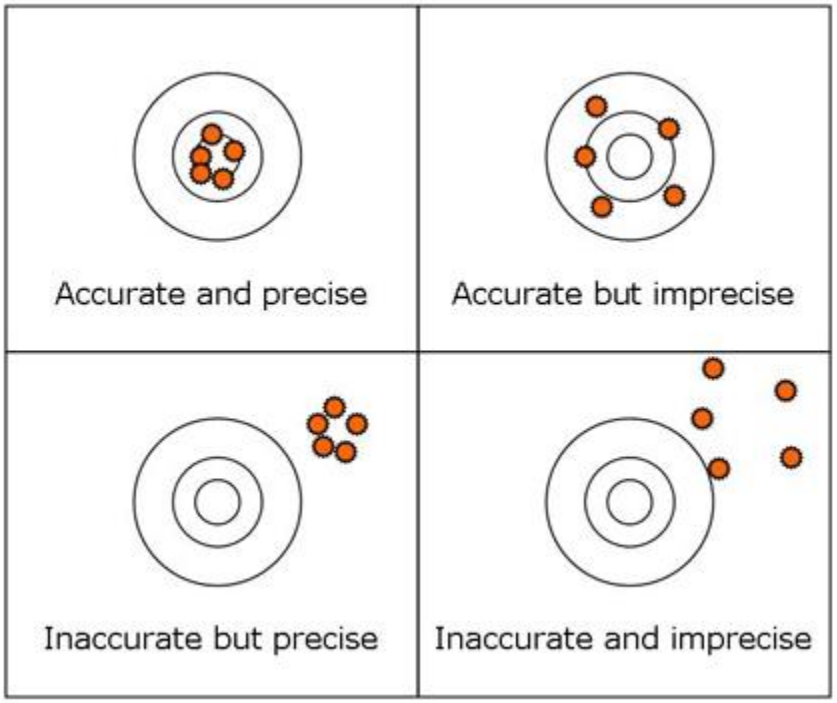
How do we evaluate the quality of a regression model?



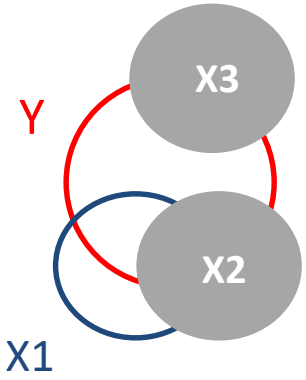
Our estimate of **program impact** should be:

- **Accurate** ("unbiased")
- **Precise** ("efficient")

unbiased *= no omitted variable bias*

efficient *= small standard errors*

Taxonomy of control variables



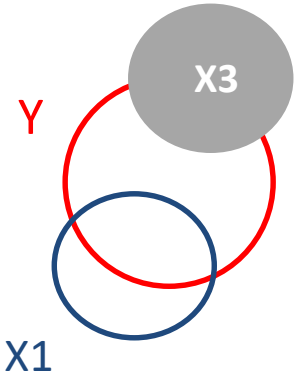
B1 = program impact

Full Model: $Y = B0 + B1 \cdot X1 + B2 \cdot X2 + B3 \cdot X3 + e$

policy variable

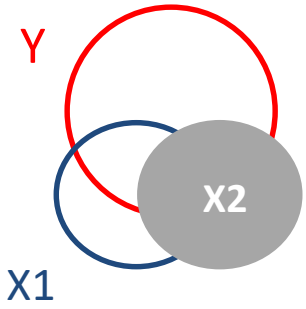
controls

Type A: Control is uncorrelated with X1



Explains extra Y
Smaller standard errors
More precise estimates

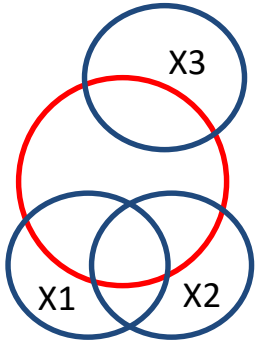
Type B: Control is correlated with X1



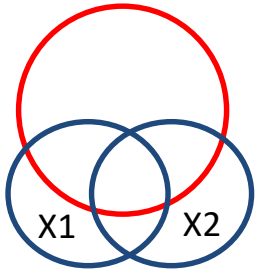
Removes bias from B1
More accurate estimates

How well will each model perform?

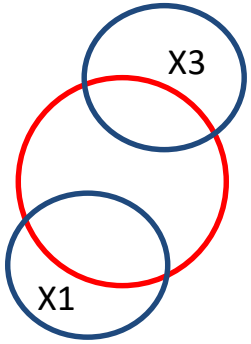
$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$



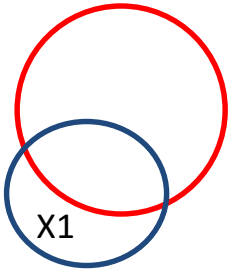
$$y = b_0 + b_1x_1 + b_2x_2$$

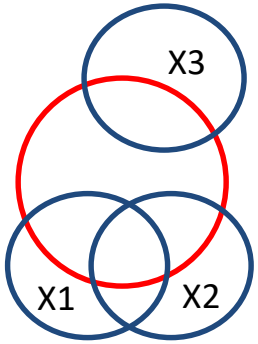


$$y = b_0 + b_1x_1 + b_3x_3$$

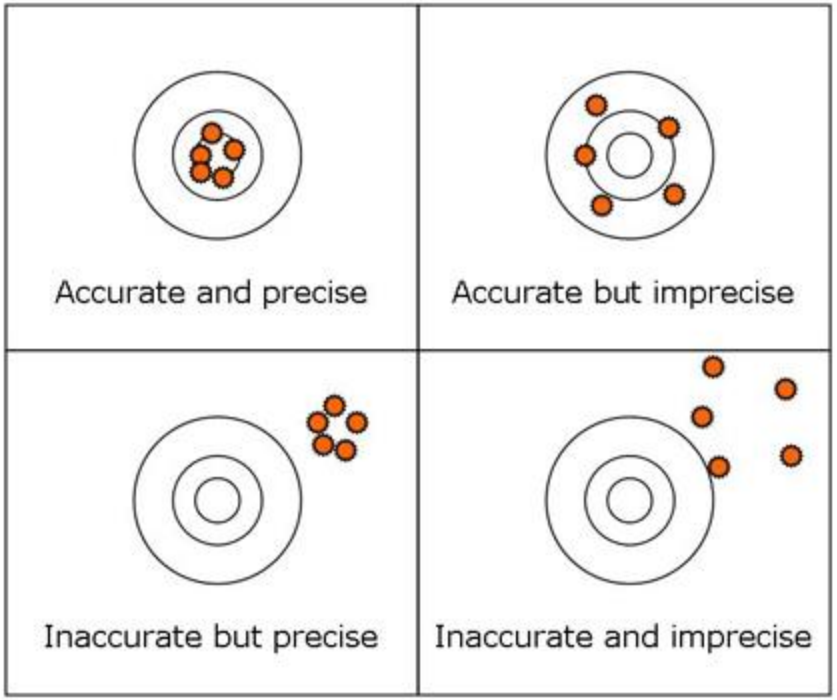


$$y = b_0 + b_1x_1$$

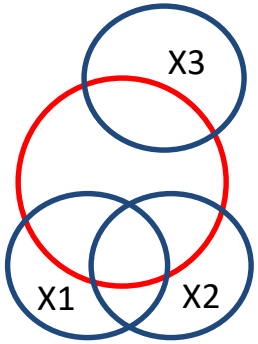




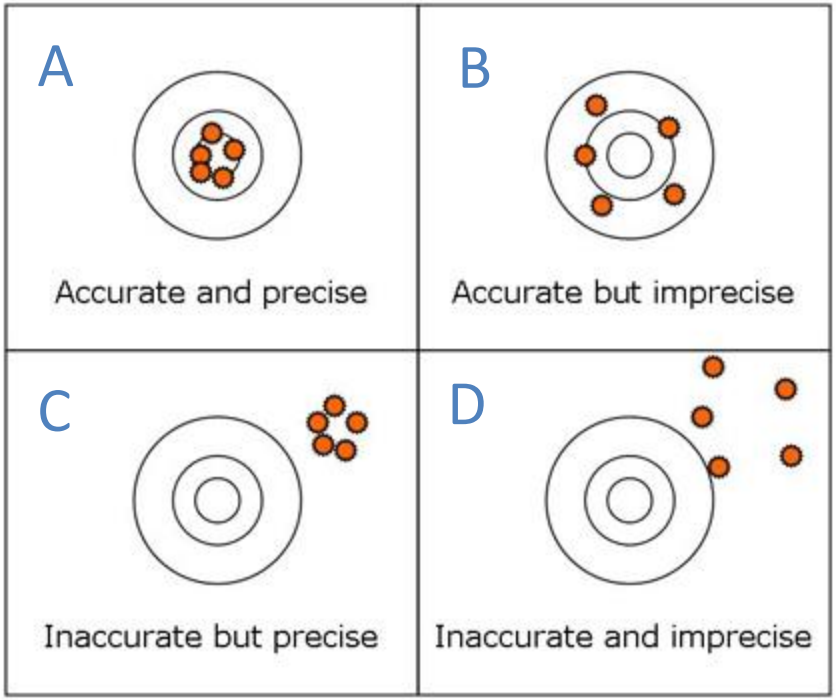
Full Model: $Y = B0 + B1 \cdot \underline{X1} + B2 \cdot X2 + B3 \cdot X3 + e$



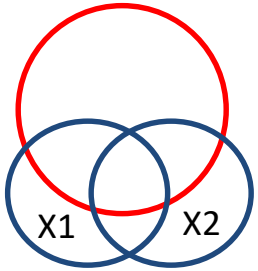
How do controls impact our model?



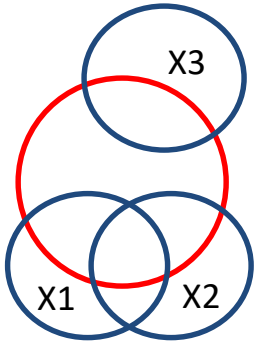
Full Model: $Y = B0 + B1 \cdot \underline{X1} + B2 \cdot X2 + B3 \cdot X3 + e$



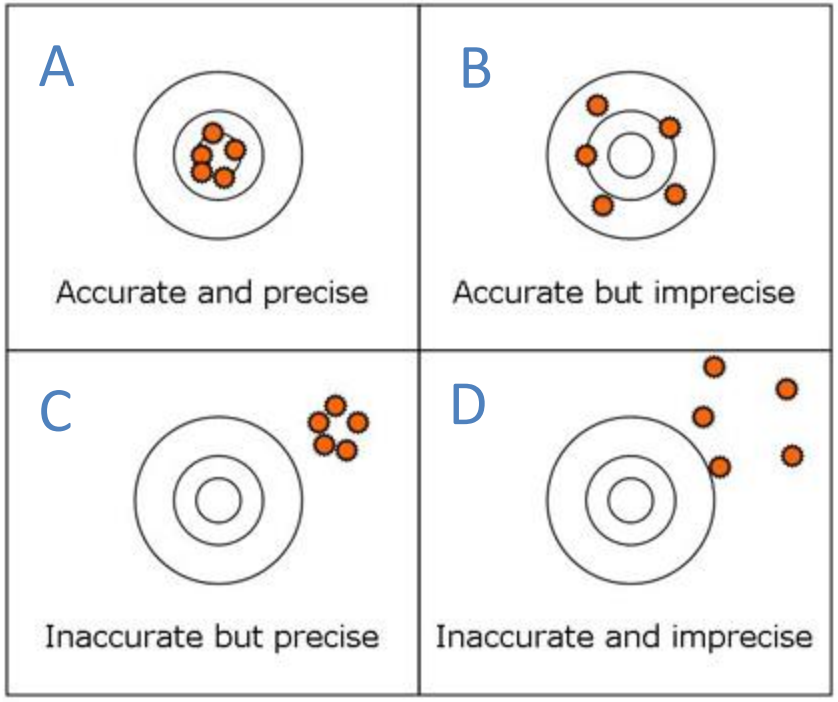
$$y = b_0 + b_1 x_1 + b_2 x_2$$



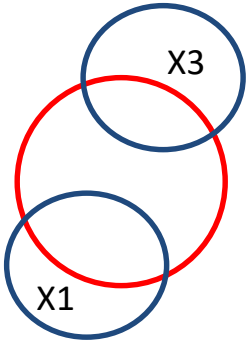
B: The competing hypotheses X2 helps isolate the independent contributions that X1, but the lack of the uncorrelated control variable X3 results in large standard errors.



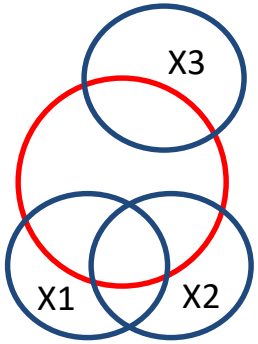
Full Model: $Y = B0 + B1 \cdot \underline{X1} + B2 \cdot X2 + B3 \cdot X3 + e$



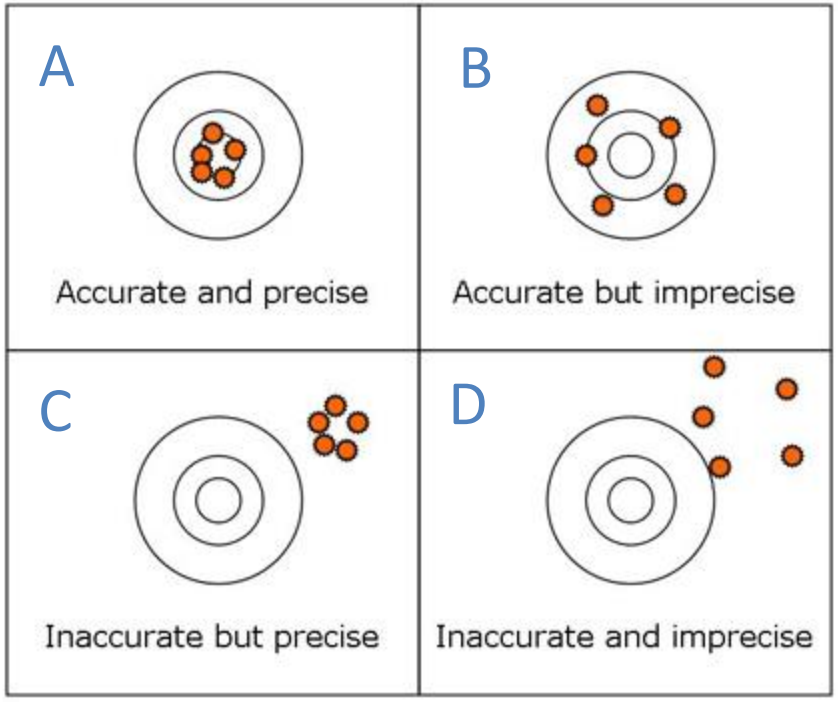
$$y = b_0 + b_1 x_1 + b_3 x_3$$



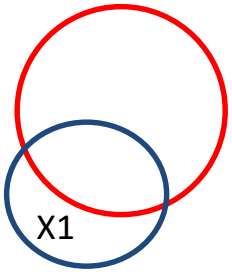
C: This is a complicated case because the uncorrelated control results in small standard error, which can give false confidence when the absence of X2 results in omitted variable bias.



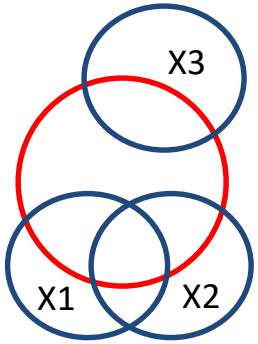
Full Model: $Y = B0 + B1 \cdot \underline{X1} + B2 \cdot X2 + B3 \cdot X3 + e$



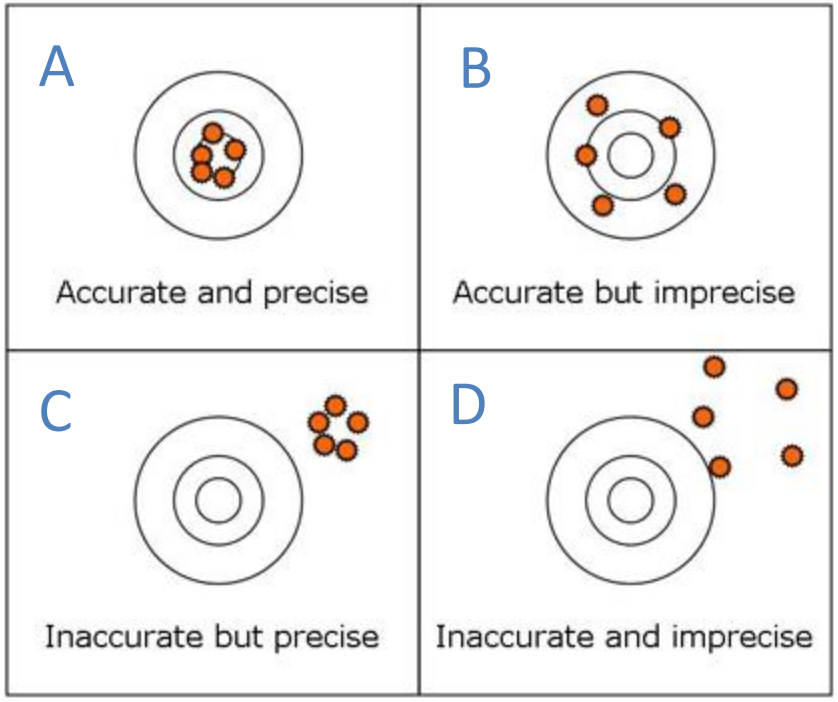
$$y = b_0 + b_1 x_1$$



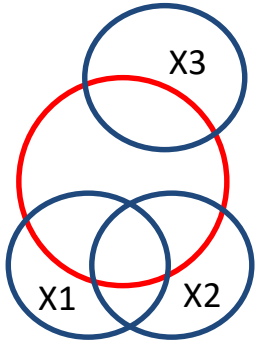
D: The absence of an important competing hypotheses (omitted variable X2) results in biased slopes, and the lack of beneficial controls leads to large standard errors.



Full Model: $Y = B0 + B1 \cdot \underline{X1} + B2 \cdot X2 + B3 \cdot X3 + e$

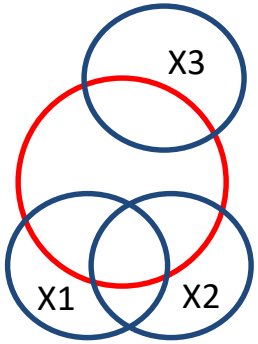


$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

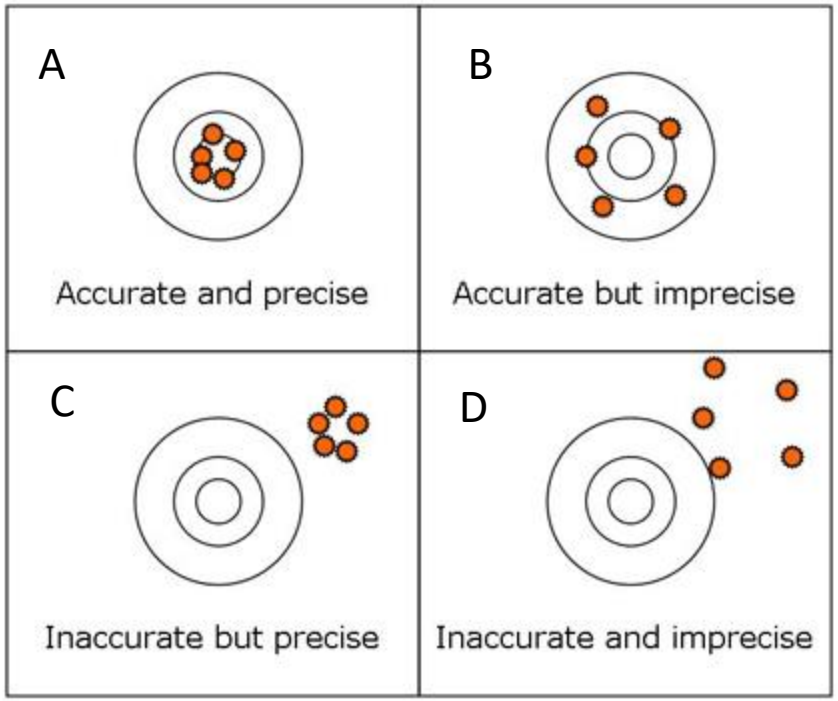


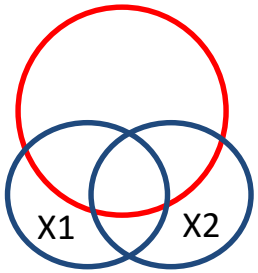
A: The competing hypotheses X2 helps isolate the independent contributions that X1 makes to the outcome, and the unrelated variable X3 reduces standard errors.

Exam question: match the cases

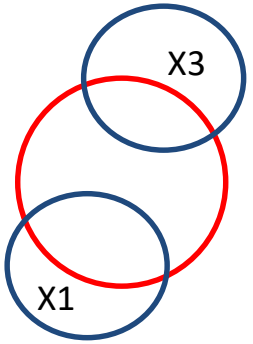


Full Model: $Y = B0 + B1 \cdot \underline{X1} + B2 \cdot X2 + B3 \cdot X3 + e$

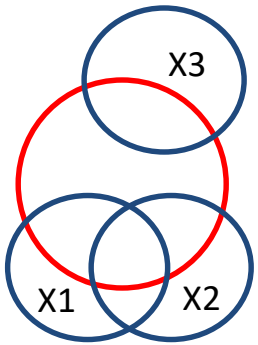




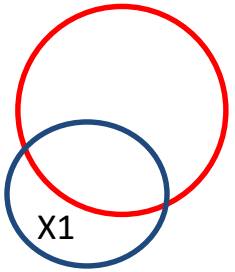
B



C

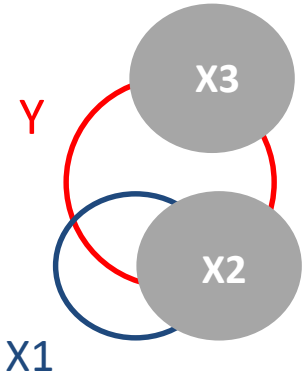


A



D

Taxonomy of control variables



B1 = program impact

controls

Full Model: $Y = B0 + \mathbf{B1}\cdot\underline{X1} + B2\cdot X2 + B3\cdot X3 + e$

policy variable

Quality of B1:

Omit X2 \ Omit X3	NO	YES
NO	$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$ **Unbiased & Precise**	$y = b_0 + b_1x_1 + b_2x_2$ **Unbiased & Imprecise**
YES	$y = b_0 + b_1x_1 + b_3x_3$ **Biased & Precise**	$y = b_0 + b_1x_1$ **Biased & Imprecise**