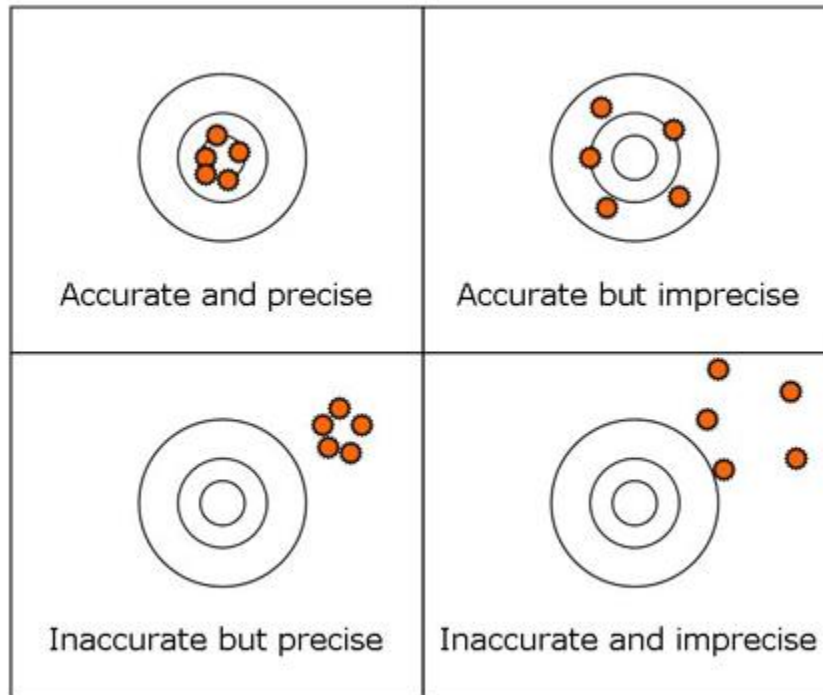


How do we evaluate the quality of a regression model?



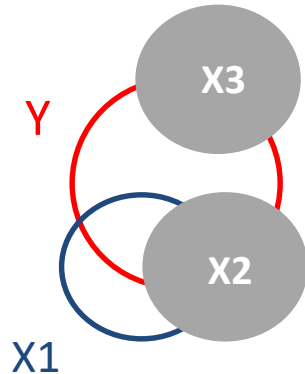
Our estimate of **program impact** should be:

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

unbiased = no omitted variable bias

efficient = small standard errors

Taxonomy of control variables



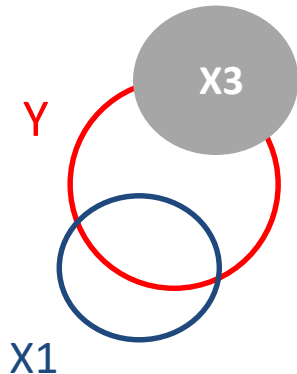
B1 = program impact

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

policy variable

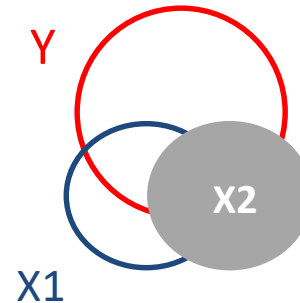
controls

Type A: Control is uncorrelated with X1



Explains extra Y
Smaller standard errors
More precise estimates

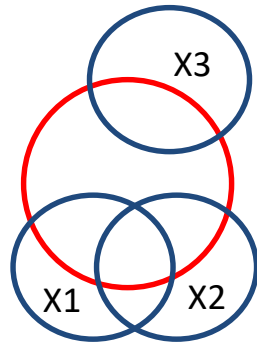
Type B: Control is correlated with X1



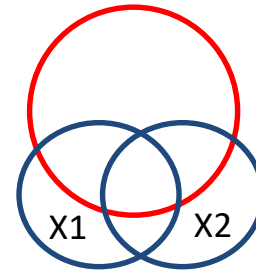
Removes bias from B1
More accurate estimates

How well will each model perform?

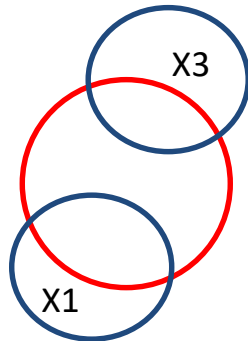
$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$



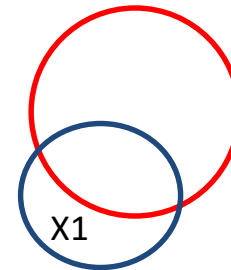
$$y = b_0 + b_1x_1 + b_2x_2$$

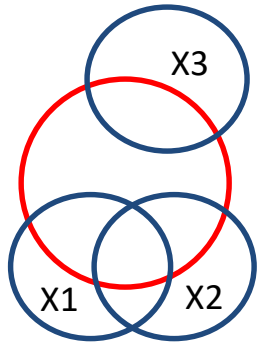


$$y = b_0 + b_1x_1 + b_3x_3$$



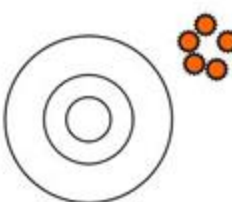
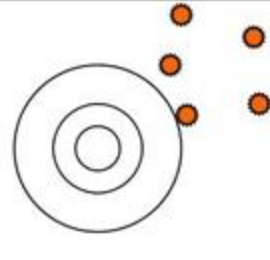


$$y = b_0 + b_1x_1$$

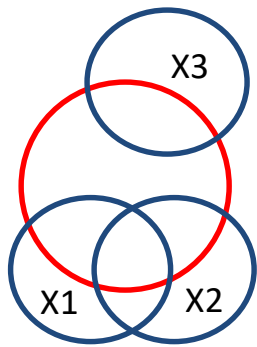




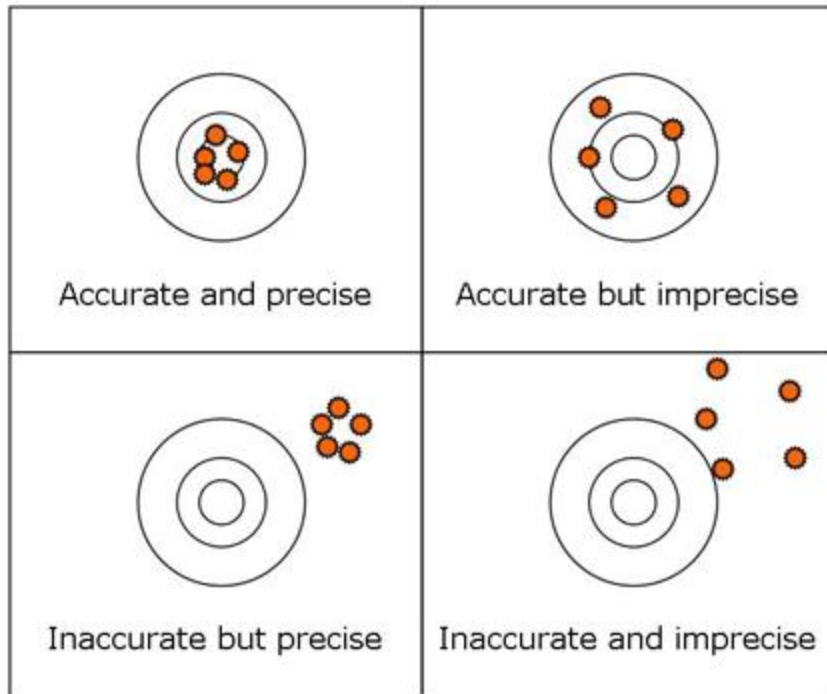
Full Model: $Y = B_0 + B_1 \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$

 <p>Accurate and precise</p>	 <p>Accurate but imprecise</p>
 <p>Inaccurate but precise</p>	 <p>Inaccurate and imprecise</p>

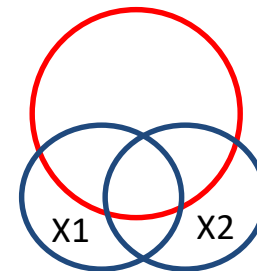
How do
controls
impact our
model?

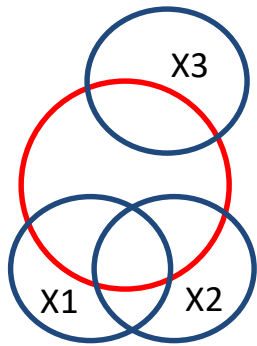


$$\text{Full Model: } Y = B_0 + B_1 \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$$



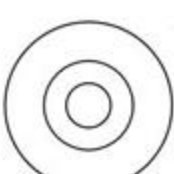



$$y = b_0 + b_1 x_1 + b_2 x_2$$

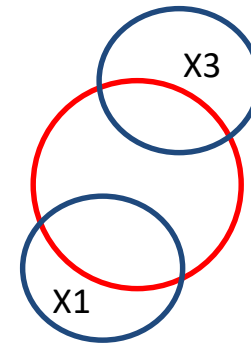


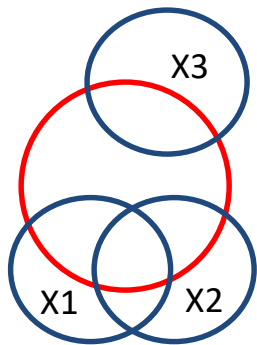


$$\text{Full Model: } Y = B_0 + B_1 \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$$



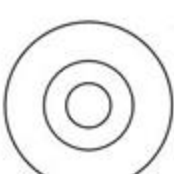

 <p>Accurate and precise</p>	 <p>Accurate but imprecise</p>
 <p>Inaccurate but precise</p>	 <p>Inaccurate and imprecise</p>

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

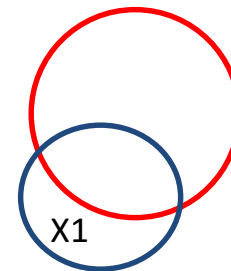


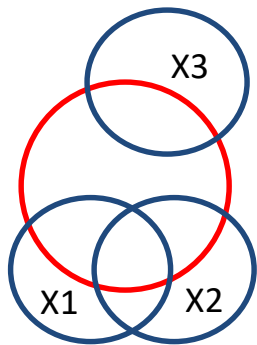


Full Model: $Y = B_0 + B_1 \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$



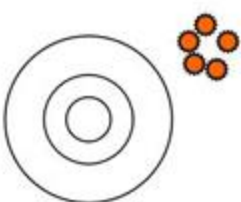
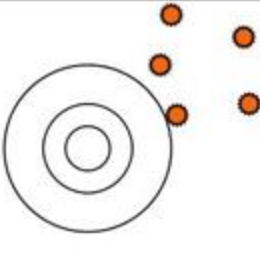
 <p>Accurate and precise</p>	 <p>Accurate but imprecise</p>
 <p>Inaccurate but precise</p>	 <p>Inaccurate and imprecise</p>

$$y = b_0 + b_1 x_1$$

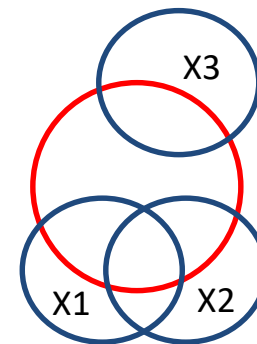




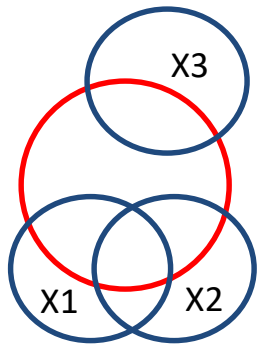
Full Model: $Y = B_0 + B_1 \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$

 <p>Accurate and precise</p>	 <p>Accurate but imprecise</p>
 <p>Inaccurate but precise</p>	 <p>Inaccurate and imprecise</p>



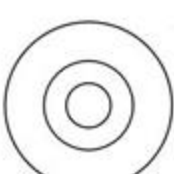

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

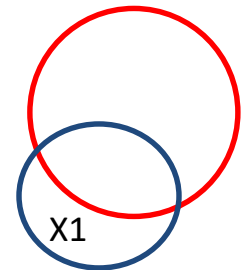
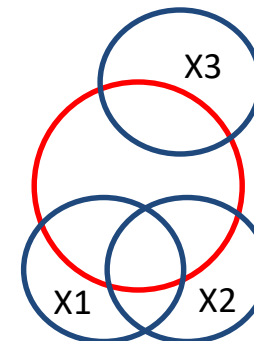
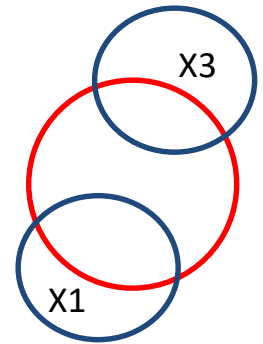
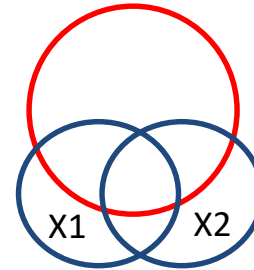


Exam question: match the cases

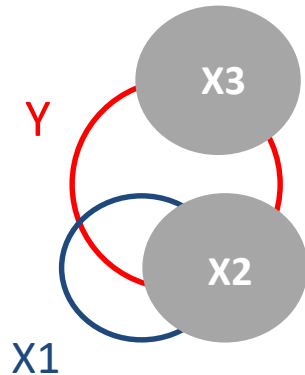


$$\text{Full Model: } Y = B_0 + B_1 \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$$

<p>A</p>  <p>Accurate and precise</p>	<p>B</p>  <p>Accurate but imprecise</p>
<p>C</p>  <p>Inaccurate but precise</p>	<p>D</p>  <p>Inaccurate and imprecise</p>



Taxonomy of control variables



Full Model: $Y = B_0 + \text{B1} \cdot \underline{X_1} + B_2 \cdot X_2 + B_3 \cdot X_3 + e$

Annotations:
 - **B1 = program impact** (points to B1)
 - **policy variable** (points to $\underline{X_1}$)
 - **controls** (points to X_2 and X_3)

Quality of B1:

Omit X3

NO

YES

Omit X2

		Quality of B1:	
		NO	YES
Omit X2	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