Suppose the true data generating process in the examples below are of the form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 o_i + u_i,$$

where x_i and o_i are exogenous such that $\mathbb{E}[u_i|o, x] = 0$.
However, o is a variable that cannot be observed and so instead, we run the model:
$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ where $\varepsilon_i = \beta_2 o_i + u_i$
Recall that we can sign the bias by:

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \beta_2 \frac{Cov(X_i, Z_i)}{Var(X_i)}$$

Causal Effect of Migration on Earnings

1. Suppose we took survey data that included people's earnings and whether they have recently moved to a new city or not, and we estimated this model:

$$Earnings_i = \beta_0 + \beta_1 Migration_i + \varepsilon_i$$

Consider the unobservable variable absorbed in ε_i which is $o_i = ambition_i$.

(a) How does $Ambition$ correlate with $Earnings$? That is, would you expect β_2 to be positive, negative or 0?

(b) How does $Ambition$ correlate with $Migration$? That is, would you expect $Cov(migration, ambition)$ to be positive, negative or 0?

(c) We have to omit $ambition$ from the model because it is not observable. Considering your answers to parts (a) and (b), we should expect that $\hat{\beta}_1$ will be biased (circle one): up / down / not at all.

Causal Effect of Friends of the Opposite Sex on a High School Student's GPA

2. Suppose we took survey data that included high school student's GPAs and the number of friends of the opposite sex they have, and we estimated the following model:

$$GPA_i = \beta_0 + \beta_1 \text{Opposite Sex Friends}_i + \varepsilon_i$$

Consider the unobservable variable absorbed in ε_i which is $o_i = \text{strict parents}_i$.

(a) How does *Strict Parents* correlate with GPA? That is, would you expect β_2 to be positive, negative or 0?

(b) How does *Strict Parents* correlate with *Opposite Sex Friends*? That is, would you expect $Cov(\textit{Opposite Sex Friends}, \textit{Strict Parents})$ to be positive, negative, or 0?

(c) We have to omit *Strict Parents* from the model because it is not observable. Considering your answers to parts (a) and (b), we should expect that $\hat{\beta}_1$ will be biased (circle one): up / down / not at all.

3. For the questions below, suppose you estimate the following model:

$$weight = 0 - 140 \cdot female + 30 \cdot height + 20 \cdot height \times female + u$$

   (a) How much would you predict a 6 foot male weighs?

   (b) How much would you predict a 6 foot female weighs?

   (c) What is the estimated impact of a one-unit (one foot) increase in height for a male individual?

   (d) What is the estimated impact of a one-unit (one foot) increase in height for a female person?

   (e) Which group has the steeper slope when it comes to how height (x) impacts weight (y)?