Omitted Variables: Multiple Regression

- True model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$, but we omit x_k.
- Write the regression of x_k on the other independent variables:
  $x_k = \delta_0 + \delta_1 x_1 + \ldots + \delta_{k-1} x_{k-1} + v$.
- Substitute:

  $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_2(\delta_0 + \delta_1 x_1 + \ldots + \delta_{k\text{-}1} x_{k\text{-}1} + v) + u$$
  $$= (\beta_0 + \beta_k \delta_0) + (\beta_1 + \beta_k \delta_1) x_1 + \ldots + (\beta_{k\text{-}1} + \beta_k \delta_{k\text{-}1}) x_{k\text{-}1} + (\beta_k v + u)$$

- Determine the signs of β_k and δ_1 to estimate the direction of bias on x_1.
  - It's easy to guess sign of β_k.
  - It's trickier for δ_1 because it comes from a multiple regression.
    - It depends on how x_1 is correlated with other *x* values.

Example

- *wage* = β_0 + β_1 *educ* + β_2 *exper* + β_3 *abil* + *u*
- We omit *abil* because it can't be measured.
- Woolridge suggests that *exper* is approximately uncorrelated with *educ* and *abil*.
- This assumption makes it easier to reason about the regression of *abil* on *educ* and *exper*.
  - All variation in *educ* is unique.
  - Coefficient on *educ* should be same as from simple regression.
  - Assume it is positive.
- If we further assume that *abil* is positively correlated with *wage*, we know that the omitted variable bias is positive on *educ*.