

- Suppose we want to estimate:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- But we know that  $x_i$  is *endogenous* (that is,  $Cov(x_i, u_i) \neq 0$ ) and we can't reasonably find control variables to remedy this problem. What can we do?
- One possibility is to look for an 'instrument' variable  $z_i$  that only affects our outcome  $y_i$  through it's effect on  $x_i$ . So that:

$z_i$  is a *relevant* instrument:  $Cov(z_i, x_i) \neq 0$  ( )

$z_i$  is a *valid* instrument (exogenous):  $Cov(z_i, u_i) = 0$