Exogeneity

Lecture for Economics 240A

Douglas G. Steigerwald

UC Santa Barbara

September 2011
Initial Question

Does additional education lead to higher wages?

- Stochastic Model - $i$ indexes workers
  - $W_i$ - (log) wage
  - $S_i$ - years of schooling (education)
  - $J_i$ - years at current job (tenure)
  - $E_i$ - years of workforce experience
  - $U_i$ - other factors that determine the wage

$$W_i = \beta_0 + \beta_1 S_i + \beta_2 J_i + \beta_3 E_i + U_i$$

- How would you determine if additional education increases a worker’s wage?
Stochastic Model Interpretation

\[ W_i = \beta_0 + \beta_1 S_i + \beta_2 J_i + \beta_3 E_i + U_i \]

- \[ \beta_1 = \frac{\partial W_i}{\partial S_i} = \frac{\partial \ln \text{wage}_i}{\partial S} \]

\[ \beta_1 = \frac{1}{\text{wage}_i} \cdot \frac{\partial \text{wage}_i}{\partial S_i} \]

- *percentage* change in wage from an additional year of education, holding tenure and experience constant

- \[ \beta_2 \] - percentage change in wage from an additional year of tenure, holding education and experience constant

- \[ \beta_3 \] - percentage change in wage from an additional year of experience, holding education and tenure constant

- Education increases wages if \( \beta_1 > 0 \)
A latent variable is a variable that cannot be measured

- $W_i$, $S_i$, $J_i$, $E_i$
  - all are measured, not latent
- $U_i$ - other factors that determine the wage
  - one factor, *latent* ability
  - higher ability workers receive higher wages
- $U_i$ contains latent factors that determine the wage
Coefficient Interpretation

- $\beta_1$ - percentage change in the wage from an additional year of schooling, *holding all else constant*
- tenure and experience
  - included, held constant by construction
- ability
  - not included, held constant by assumption
  - identification assumption: $\mathbb{E}(S_i U_i) = 0$

1. if ability moves in a systematic way with education, then ability cannot be held constant
2. if ability cannot be held constant, then:
   1. the identification assumption is violated
   2. $\beta_1$ captures the impact on the wage that results from both a change in schooling and ability
   3. hence, $\beta_1$ no longer measures the wage change from an additional year of schooling
Framework: Endogenous and Exogenous Variables

- An endogenous variable is impacted by the (latent) factors in the error
- An exogenous variable is not impacted by the (latent) factors in the error
- $W_i$ - endogenous by construction
- $X_i^T = (1 \ S_i \ J_i \ E_i)$ - must determine if exogenous
- $\mathbb{E} (S_i U_i) = 0$
  - ability (and all other factors in $U_i$) can be held constant as education varies
  - $S_i$ is exogenous
- $\mathbb{E} (S_i U_i) \neq 0$
  - ability (and all other factors in $U_i$) cannot be held constant as education varies
  - $S_i$ is endogenous
Identification Assumption Implication
Mean of Error is Zero

\[ X_i^T = (1 \ S_i \ J_i \ E_i) \]

- Identification Assumption: \( \mathbb{E} (X_i U_i) = 0 \)

\[
\mathbb{E} \begin{pmatrix}
U_i \\
S_i U_i \\
J_i U_i \\
E_i U_i
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
0 \\
0
\end{pmatrix}
\]

implies \( \mathbb{E} (U_i) = 0 \)
For each regressor (take $S_i$ for example)

- $\mathbb{E}(S_i U_i) = 0$
- $\mathbb{E}(U_i) = 0$

Covariance

$$\text{Cov} (S_i, U_i) \equiv \mathbb{E} [(S_i - \mathbb{E}S_i) (U_i - \mathbb{E}U_i)]$$

$$= \mathbb{E} (S_i U_i) - \mathbb{E} (\mathbb{E} (S_i) \cdot U_i)$$

$$= \mathbb{E} (S_i U_i) - \mathbb{E} (S_i) \cdot \mathbb{E} (U_i)$$

$$= \mathbb{E} (S_i U_i)$$

- $\mathbb{E} (S_i U_i) = 0$ implies $\text{Cov} (S_i, U_i) = 0$
Covariance: Interpretation

\[ \text{Cov} (S_i, U_i) \equiv \mathbb{E} \left[ (S_i - \mathbb{E}S_i)(U_i - \mathbb{E}U_i) \right] \]

- **Interpret positive covariance**
  - when \( S_i \) is above \( \mathbb{E}S_i \), \( U_i \) tends to be above \( \mathbb{E}U_i \)

- **Interpret negative covariance**
  - when \( S_i \) is above \( \mathbb{E}S_i \), \( U_i \) tends to be below \( \mathbb{E}U_i \)

- **Interpret zero covariance**
  - when \( S_i \) is above \( \mathbb{E}S_i \), we have no information about \( U_i \) relative to \( \mathbb{E}U_i \)
Covariance Interpretation: Wage Regression

- \( W_i = \beta_0 + \beta_1 S_i + \beta_2 J_i + \beta_3 E_i + U_i \)
  - \( W_i \) - (log) wage
  - \( S_i \) - years of schooling (education)
  - \( J_i \) - years at current job (tenure)
  - \( E_i \) - years of workforce experience
  - \( U_i \) - other factors that determine the wage

- \( U_i \) includes latent ability, \( A_i \)
  - high ability workers \( U_i > \mathbb{E} U_i \)
  - high ability workers likely have more schooling \( S_i > \mathbb{E} S_i \)

- implies \( \text{Cov} (S_i, U_i) > 0 \)
  - identification assumption fails - schooling is endogenous
  - we cannot measure the effect of schooling on wages, holding all else constant, because ability varies with schooling
  - estimates of \( \beta_1 \) will tend to be too large (confounding the effects of schooling and ability)
Framework: Correlation

- **Covariance**
  - sign is easy to interpret
  - magnitude is difficult to interpret

- example: $\text{Cov} (S_i, U_i)$ units are years * "ability units"

- **Correlation** - sign and units easy to interpret

\[
\text{Corr} (S_i, U_i) = \frac{\text{Cov} (S_i, U_i)}{\sqrt{\text{Var} (S_i) \cdot \text{Var} (U_i)}}
\]

- $-1 \leq \text{Corr} (S_i, U_i) \leq 1$

- $\text{Cov} (S_i, U_i) = 0 \Rightarrow \text{Corr} (S_i, U_i) = 0$

- Identification assumption: regressors and error are uncorrelated
samples of workers in many studies

\[ b_1 = 0.08 \]

**Interpretation - consider 2 workers**
- identical characteristics
  - tenure & experience (by construction)
  - other factors - ability (by assumption)
- different characteristic
  - worker 2 has 1 more year of schooling

you predict the wage of worker 2 to be 8 percent higher than that of worker 1

**Difficulty** - cannot logically establish that schooling and ability are uncorrelated

combined effect of ability and schooling increases wages by 8 percent - we do not know the precise effect of schooling on wage
How should firms pay workers to elicit maximum effort?

- Model in vector form

\[ E_i = X_i^T \beta + U_i \quad X_i = \begin{pmatrix} 1 \\ W_i \\ C_i \end{pmatrix} \]

- What is the identification assumption?
  - \( \mathbb{E} (X_i U_i) = 0 \)
  - wage and co-worker wage are uncorrelated with other factors that determine effort

- In a firm, wage and effort are likely jointly determined
  - to estimate this effect, we'll need experimental evidence, where wages are set exogenously