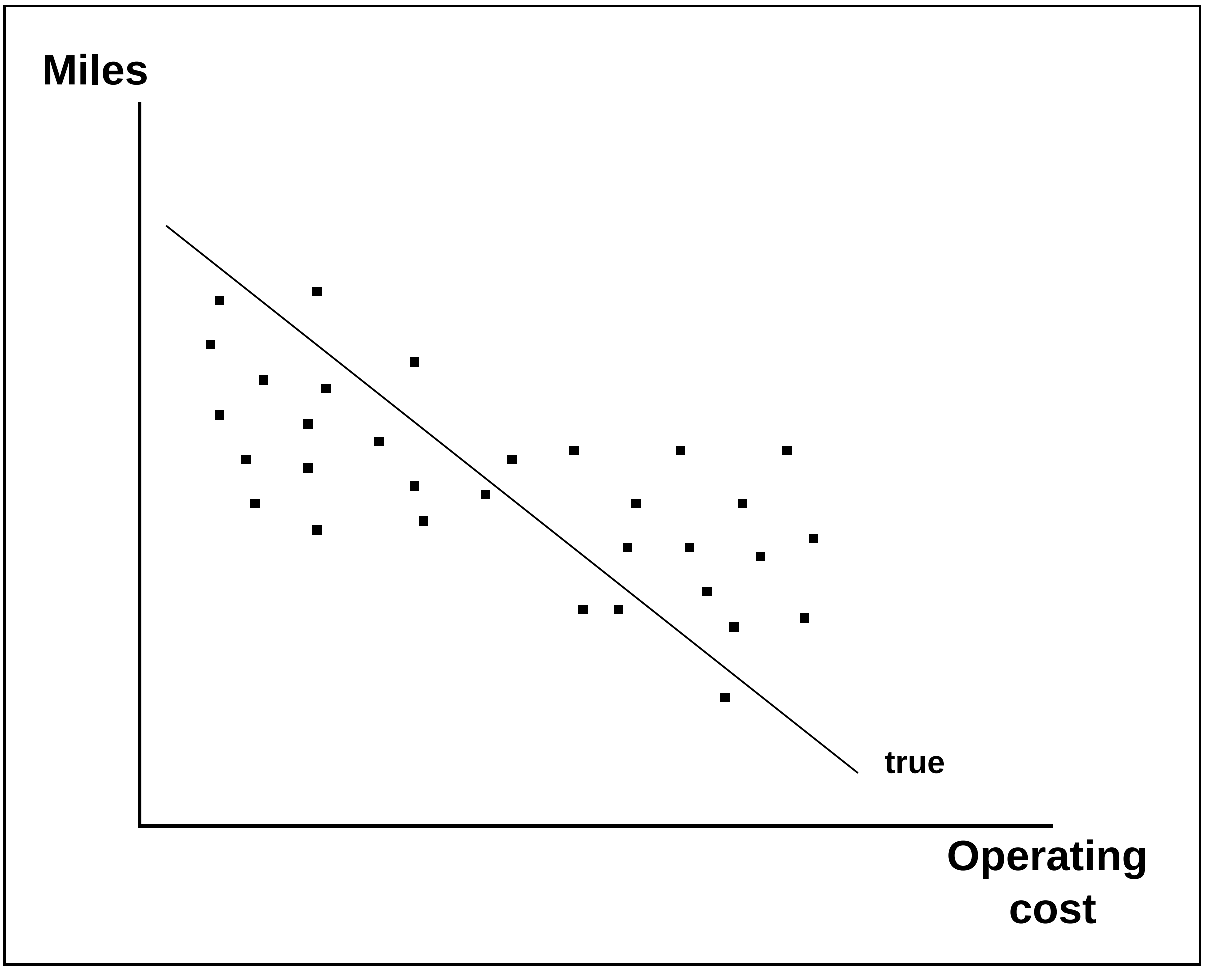
LECTURE / DISCUSSION

Instrumental Variables

Instrumental Variables Estimation

Problem: error is correlated with explanatory variable.



$$\text{Miles} = \alpha + \beta(\text{op. cost}) + \varepsilon$$

Solution #1: Include the variable that causes the correlation, if possible.

$$(\text{Miles}) = \alpha + \beta(\text{op. cost}) + \theta(\text{distance to work}) + \varepsilon$$

Solution #2: Instrumental variables estimation

$$Y_n = \alpha + \beta X_n + \varepsilon_n$$

$$\mathrm{Corr}(X,\varepsilon) \neq 0$$

Find a variable, Z, that is
- correlated with X
- not correlated with ε

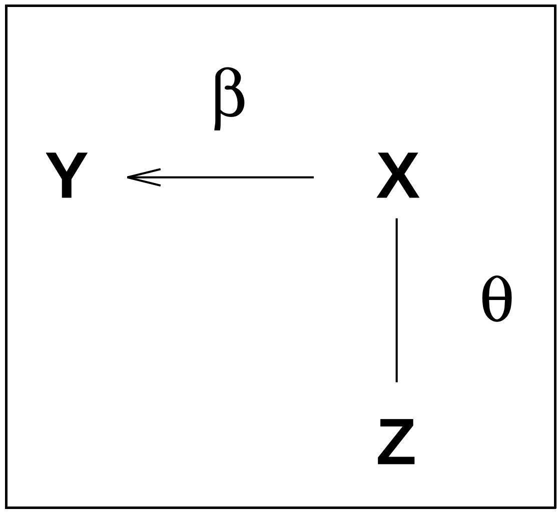
Z is called an **instrument.**

$$\hat{\beta}_{IV} = \frac{\sum z_n y_n}{\sum z_n x_n} = \frac{\mathrm{cov}(z,y)}{\mathrm{cov}(z,x)}$$

Example: Miles = α + β(op. cost) + ε
Instrument: gas price

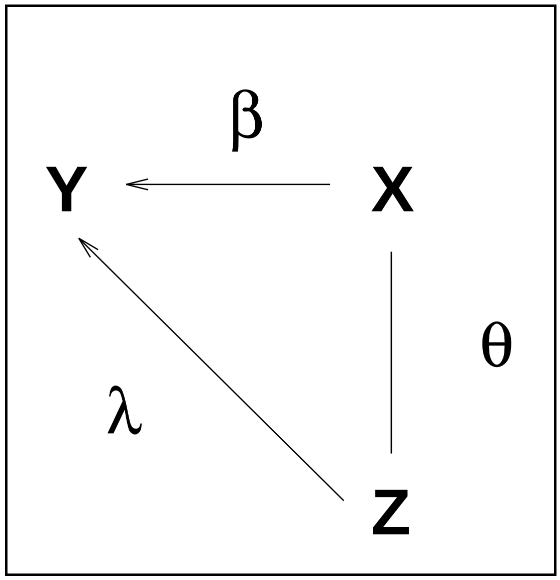
$$\hat{\beta}_{IV} = \frac{\mathrm{cov}(\text{gas price}, \text{miles})}{\mathrm{cov}(\text{gas price}, \text{op. cost})}$$

Why Does IV Work?



$$E(\hat{\beta}_{IV}) = \frac{\mathrm{cov}(zy)}{\mathrm{cov}(zx)} = \frac{\beta\theta}{\theta} = \beta$$

When IV is Biased



$$E(\hat{\beta}_{IV}) = \frac{\text{cov}(zy)}{\text{cov}(zx)} = \frac{\theta\beta + \lambda}{\theta} = \beta + \frac{\lambda}{\theta}$$

Why Is IV Unbiased?

$$y_n = \beta x_n + \varepsilon_n$$

$$\sum z_n y_n = \beta \sum z_n x_n + \sum z_n \varepsilon_n$$

$$\begin{aligned} E(\sum z_n y_n) &= \beta E(\sum z_n x_n) + E(\sum z_n \varepsilon_n) \\ &= \beta E(\sum z_n x_n) \end{aligned}$$

So that z_n isolates the covariation between y_n and x_n. This leads to

$$\hat{\beta}_{IV} = \frac{\sum z_n y_n}{\sum z_n x_n}$$

☞ Efficiency is greater with instruments that are more highly correlated with X, while still uncorrelated with the error terms.

Can Apply IV Through 2SLS

$$Y_n = \alpha + \beta X_n + \varepsilon_n$$

instrument Z_n

Step 1: Run OLS regression

$$X_n = \lambda + \theta Z_n + \mu_n$$

Get predicted $\hat{X}_n = \hat{\lambda} + \hat{\theta} Z_n$

Step 2: Run OLS regression using $\hat{X}$

$$Y_n = \alpha + \beta \hat{X}_n + \varepsilon_n^*$$

Example

$$\text{Miles} = \alpha + \beta(\text{op. cost}) + \varepsilon$$

instrument: gas price

Step 1: Regress

$$(\text{op. cost}) = \lambda + \theta(\text{gas price}) + \mu$$

Calculate

$$(\text{op. cost}) = \hat{\lambda} + \hat{\theta}(\text{gas price})$$

Step 2: Regress

$$\text{Miles} = \alpha + \beta(\text{op. cost}) + \varepsilon^*$$

2SLS Interpretation of IV

$$Y_n = \alpha + \beta X_n + \varepsilon_n$$

Step 1: X is decomposed into a part that is uncorrelated with ε and a part that is correlated with ε.

$$X_n = \underbrace{\lambda + \theta Z_n}_{\text{uncorrelated with } \varepsilon} + \underbrace{\mu_n}_{\text{correlated with } \varepsilon}$$

Step 2: Use the portion of X that is uncorrelated with ε in estimating the original equation.

$$Y_n = \alpha + \beta \hat{X} + \varepsilon_n^*$$

$$\hat{X}_n = \hat{\lambda} + \hat{\theta} z_n$$

$$\varepsilon_n^* = \varepsilon_n + \beta \hat{\mu}_n$$

The part of X that was correlated with ε gets moved into the error.

The part of X that was uncorrelated with ε stays as an explanatory variable.

2SLS method also shows that many instruments can be used in IV.

$$(\text{Miles}) = \alpha + \beta(\text{op. cost}) + \varepsilon$$

Instruments: gas price
relative prices for large and small cars

Not instruments: income
family size

Note: Income and family size **can** serve as instruments if they also enter the original regression as explanatory variables.

$$\text{Miles} = \alpha + \beta(\text{op. cost}) + \theta(\text{income}) + \lambda(\text{family size}) + \varepsilon$$

Instruments: gas price
relative prices for large and small cars
income
family size

Because income and family size enter directly, they are not part of ε and hence are not correlated with ε.

Equivalence of IV and 2SLS

2SLS

Step 1:

$$x_n = \theta z_n + \varepsilon_n$$

$$\hat{\theta}_{OLS} = \frac{cov(zx)}{var(z)}$$

$$\hat{x}_n = \hat{\theta}_{OLS} z_n$$

Step 2:

$$y_n = \beta \hat{x}_n + \varepsilon_n$$

$$\hat{\beta}_{2SLS} = \frac{cov(\hat{x}y)}{var(\hat{x})}$$

$$\hat{\beta}_{2SLS} = \frac{cov(\hat{\theta}_{OLS} zy)}{var(\hat{\theta}_{OLS} z)} = \frac{\hat{\theta}_{OLS} cov(zy)}{\hat{\theta}^2_{OLS} var(z)}$$

$$= \frac{cov(zy)}{\hat{\theta}_{OLS} var(z)} = \frac{cov(zy)}{\frac{cov(zx)}{var(z)} \cdot var(z)}$$

$$= \frac{cov(zy)}{cov(zx)} = \hat{\beta}_{IV}$$

TSP Commands

Three ways to estimate

$$Y = \alpha + \beta X + \varepsilon \quad \text{with instrument } Z$$

1. 2SLS explicitly as two OLS regressions

```
olsq x c,z;
genr xhat = @fit;
olsq y c,xhat;
```

2. 2SLS implicitly

```
2sls(inst = (c,z)) y c,x;
```

3. IV

```
inst y c,x invr c,z;
```