# Chapter 12
# Instrumental Variables: A Solution to the Endogeneity Problem

# Learning Objectives

- State the two necessary properties of a good instrumental variable
- In real-world settings, articulate the two properties of a good instrument and critique the instruments used by researchers.
- Apply the instrumental variables, or two-stage least squares, estimator to solve the endogeneity problem

# Three Sources of "Endogeneity"

1. Measurement error in *X*.

2. *X* and *Y* determined jointly.

3. Omitted *X* variable.

Some economists use the word "endogeneity" only for jointly determined variables, but modern econometrics uses the word for any of the three settings.

# Consumption Example

- Household consumption in a particular year for 3,254 households in the Dominican Republic

$$C_i = \beta_0 + \beta_1 X_i + \beta_2 SIZE_i + \varepsilon_i$$

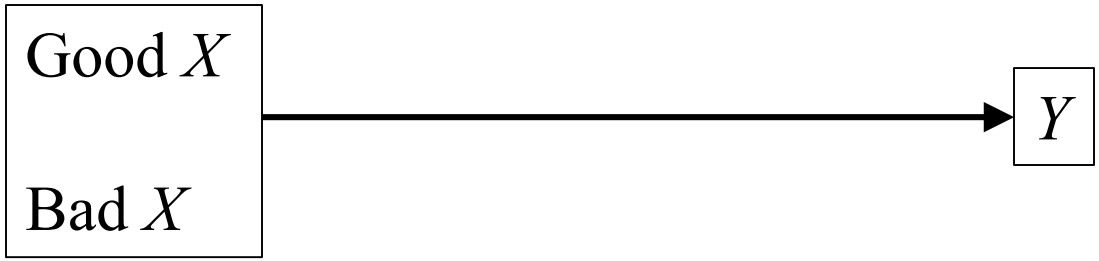
- **Problem:** consumption depends on *permanent* income not *annual* income
- **Solution:** Schooling and experience are **instruments** for permanent income

$$X_i = \alpha_0 + \alpha_1 S_i + \alpha_2 EX_i + \alpha_3 EX_i^2 + \alpha_4 SIZE_i + \varepsilon_{Xi}$$

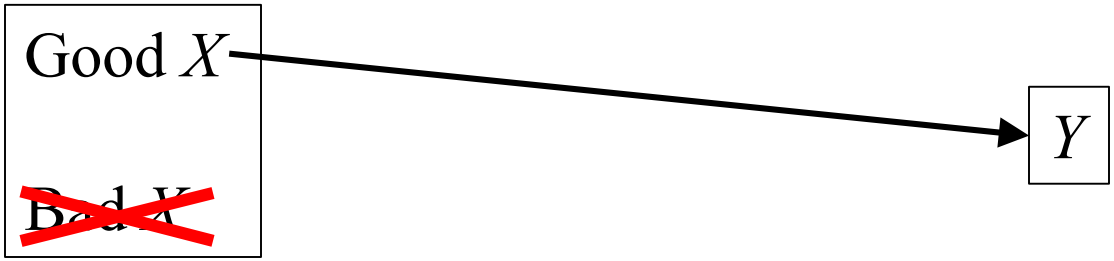
$$C_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 SIZE_i + \varepsilon_i$$

# Endogeneity Problem

OLS regression

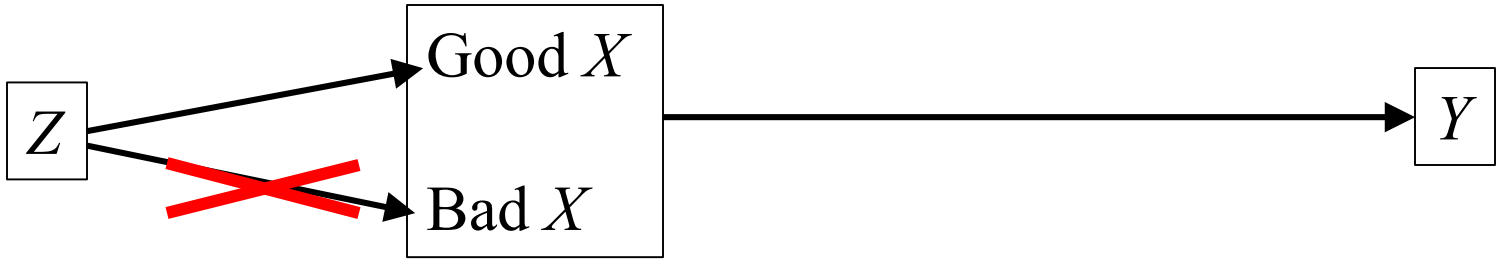



What we want

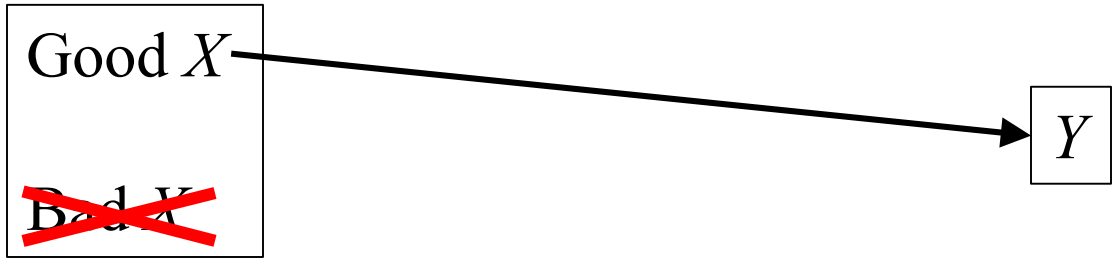


# Instrumental Variables

IV regression




What we want



# Characteristics of Good Instruments

- ***IV1*****:** an instrument must be strongly correlated with the endogenous right-hand-side variable.

- ***IV2*****:** an instrument must not be correlated with the error term in the main equation of interest.

# Consumption Example

- Household consumption in a particular year for 3,254 households in the Dominican Republic

$$C_i = \beta_0 + \beta_1 X_i + \beta_2 SIZE_i + \varepsilon_i$$

- **Problem:** consumption depends on *permanent* income not *annual* income
- **Solution:** Schooling and experience are **instruments** for permanent income

$$X_i = \alpha_0 + \alpha_1 S_i + \alpha_2 EX_i + \alpha_3 EX_i^2 + \alpha_4 SIZE_i + \varepsilon_{Xi}$$

$$C_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 SIZE_i + \varepsilon_i$$

# First Stage Regression

| Variable | Panel A: Without Family Size | | | Panel B: With Family Size | | |
|---|---|---|---|---|---|---|
| | Estimated Coefficient | Robust Standard Error | t-Statistic | Estimated Coefficient | Robust Standard Error | t-Statistic |
| **Schooling** | 0.21 | 0.02 | 9.49 | 0.20 | 0.02 | 9.23 |
| **Experience** | 0.039 | 0.005 | 7.16 | 0.028 | 0.005 | 5.31 |
| **Experience-squared** | -0.00050 | 0.00067 | -7.43 | -0.00036 | 0.00064 | -5.69 |
| **Family size** | | | | 0.11 | 0.01 | 12.00 |
| **Constant** | 10.18 | 0.12 | 85.87 | 9.96 | 0.12 | 84.02 |
| **Sample size** | | 3,252 | | | 3,252 | |
| **$R^2$** | | 0.057 | | | 0.098 | |
| **F-statistic for $H_0$: $\alpha_1=\alpha_2=\alpha_3=0$** | | 50.18 | | | 43.24 | |

# Second Stage Regression

| Variable | Panel A: OLS | | | Panel B: IV Method | | |
|---|---|---|---|---|---|---|
| | Estimated Coefficient | Standard Error | t-Statistic | Estimated Coefficient | Standard Error | t-Statistic |
| **Income** | 0.35 | 0.01 | 24.44 | 0.75 | 0.06 | 12.14 |
| **Family size** | 0.14 | 0.01 | 20.02 | 0.09 | 0.01 | 9.06 |
| **Constant** | 7.27 | 0.16 | 46.11 | 2.93 | 0.67 | 4.38 |
| **Sample size** | | 3,252 | | | 3,252 | |
| **$R^2$** | | 0.34 | | | | |

# Example: Institutions and Economic Performance

## The Colonial Origins of Comparative Development: An Empirical Investigation

*By* DARON ACEMOGLU, SIMON JOHNSON, AND JAMES A. ROBINSON*

*We exploit differences in European mortality rates to estimate the effect of institutions on economic performance. Europeans adopted very different colonization policies in different colonies, with different associated institutions. In places where Europeans faced high mortality rates, they could not settle and were more likely to set up extractive institutions. These institutions persisted to the present. Exploiting differences in European mortality rates as an instrument for current institutions, we estimate large effects of institutions on income per capita. Once the effect of institutions is controlled for, countries in Africa or those closer to the equator do not have lower incomes.* (*JEL* O11, P16, P51)

# Economic Development

- Do rich countries have good institutions because they are rich?
- Or do good institutions make countries rich?
- **Institutions: secure property rights, economic freedom, etc.**

| | Base sample (1) | Base sample (2) |
|---|---|---|
| Average protection against expropriation risk 1985–1995 | 0.94 (0.16) | 1.00 (0.22) |
| Latitude | | −0.65 (1.34) |
| Asia dummy | | |
| Africa dummy | | |
| "Other" continent dummy | | |
| Panel B: First Stage for A | | |
| Log European settler mortality | −0.61 (0.13) | −0.51 (0.14) |
| Latitude | | 2.00 (1.34) |
| Asia dummy | | |
| Africa dummy | | |
| "Other" continent dummy | | |
| $R^2$ | 0.27 | 0.30 |
| Average protection against expropriation risk 1985–1995 | 0.52 (0.06) | 0.47 (0.06) |
| Number of observations | 64 | 64 |

(potential) settler mortality $\Rightarrow$ settlements

$\Rightarrow$ early institutions $\Rightarrow$ current institutions

$\Rightarrow$ current performance.

$$GDP_i = \beta_0 + \beta_1 INST_i + \varepsilon_i^{\beta}$$

$$INST_i = \alpha_0 + \alpha_1 MORT_i + \varepsilon_i^{\alpha}$$

# Example: Schooling and Wages

## Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment

*By* ESTHER DUFLO*

*Between 1973 and 1978, the Indonesian government engaged in one of the largest school construction programs on record. Combining differences across regions in the number of schools constructed with differences across cohorts induced by the timing of the program suggests that each primary school constructed per 1,000 children led to an average increase of 0.12 to 0.19 years of education, as well as a 1.5 to 2.7 percent increase in wages. This implies estimates of economic returns to education ranging from 6.8 to 10.6 percent.* (*JEL* I2, J31, O15, O22)

# Does schooling increase wages?

$$\ln(W_{ijk}) = \beta_0 + \beta_1 S_{ijk} + \beta_2 X_{ijk} + \varepsilon_{ijk}$$

- Individual *i* in region *j* in year *k*
- Schooling may be endogenous – if only we could randomize schooling …..
- Use school building program as an instrument

# Setting for the Study

- Between 1973 and 1978, 61,000 elementary schools were constructed in Indonesia
- Enrollment rates increased from 69% to 83%
- In 1995, Indonesian GDP per capita was $1/30^{th}$ the size of the US

# Data

- Collected from a 1995 survey of men born between 1950 and 1972
- 152,989 individuals in sample
- 60,633 work for a wage in 1995

TABLE 1—DESCRIPTIVE STATISTICS

| | Mean |
|---|---|
| *Panel A: Individual Level Means* | |
| Education (whole sample $N = 152{,}989$) | 7.98 |
| Education (sample with valid wage data $N = 60{,}663$) | 9.00 |
| INPRES schools built per 1,000 children | 1.98 |
| INPRES schools built per 1,000 children (sample with valid wage data) | 1.89 |
| INPRES schools built per 1,000 children (High program regions) | 2.44 |
| INPRES schools built per 1,000 children (Low program regions) | 1.54 |
| Log(hourly wage) | 6.87 |
| Monthly earnings (SUPAS 1995), thousands Rupiah | 13 |
| Monthly earnings (SUSENAS 1993) of wage earners, thousands Rupiah | 205 |
| Monthly earnings (SUSENAS 1993) of self-employed individuals, thousands Rupiah | 152 |

# Did building more schools increase schooling and wages?

$$\hat{S}_{ijk} = a_0 + a_1 P_j T_i + a_2 X_{ijk}$$

$$\ln(W_{ijk}) = \beta_0 + \beta_1 \hat{S}_{ijk} + \beta_2 X_{ijk} + \varepsilon_{ijk}^{\beta}$$

- $P_j$ is schools built per 1000 students in region *j*
- $T_i$ is a dummy variable for whether individual is young in 1974

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

| | Years of education | | | Log(wages) | | |
|---|---|---|---|---|---|---|
| | Level of program in region of birth | | | Level of program in region of birth | | |
| | High (1) | Low (2) | Difference (3) | High (4) | Low (5) | Difference (6) |
| *Panel A: Experiment of Interest* | | | | | | |
| Aged 2 to 6 in 1974 | 8.49 | 9.76 | −1.27 | 6.61 | 6.73 | −0.12 |
| | (0.043) | (0.037) | (0.057) | (0.0078) | (0.0064) | (0.010) |
| Aged 12 to 17 in 1974 | 8.02 | 9.40 | −1.39 | 6.87 | 7.02 | −0.15 |
| | (0.053) | (0.042) | (0.067) | (0.0085) | (0.0069) | (0.011) |
| Difference | 0.47 | 0.36 | 0.12 | −0.26 | 0.29 | 0.026 |
| | (0.070) | (0.038) | (0.089) | (0.011) | (0.0096) | (0.015) |
| *Panel B: Control Experiment* | | | | | | |
| Aged 12 to 17 in 1974 | 8.02 | 9.40 | −1.39 | 6.87 | 7.02 | −0.15 |
| | (0.053) | (0.042) | (0.067) | (0.0085) | (0.0069) | (0.011) |
| Aged 18 to 24 in 1974 | 7.70 | 9.12 | −1.42 | 6.92 | 7.08 | −0.16 |
| | (0.059) | (0.044) | (0.072) | (0.0097) | (0.0076) | (0.012) |
| Difference | 0.32 | 0.28 | 0.034 | 0.056 | 0.063 | 0.0070 |
| | (0.080) | (0.061) | (0.098) | (0.013) | (0.010) | (0.016) |

*Notes*: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

TABLE 4—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN COHORT DUMMIES AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN THE REGION OF BIRTH

| | | Dependent variable | | | | | |
|---|---|---|---|---|---|---|---|
| | | Years of education | | | Log(hourly wage) | | |
| | Observations | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A: Experiment of Interest: Individuals Aged 2 to 6 or 12 to 17 in 1974 (Youngest cohort: Individuals ages 2 to 6 in 1974)* | | | | | | | |
| Whole sample | 78,470 | 0.124 | 0.15 | 0.188 | | | |
| | | (0.0250) | (0.0260) | (0.0289) | | | |
| Sample of wage earners | 31,061 | 0.196 | 0.199 | 0.259 | 0.0147 | 0.0172 | 0.0270 |
| | | (0.0424) | (0.0429) | (0.0499) | (0.00729) | (0.00737) | (0.00850) |
| *Panel B: Control Experiment: Individuals Aged 12 to 24 in 1974 (Youngest cohort: Individuals ages 12 to 17 in 1974)* | | | | | | | |
| Whole sample | 78,488 | 0.0093 | 0.0176 | 0.0075 | | | |
| | | (0.0260) | (0.0271) | (0.0297) | | | |
| Sample of wage earners | 30,225 | 0.012 | 0.024 | 0.079 | 0.0031 | 0.00399 | 0.0144 |
| | | (0.0474) | (0.0481) | (0.0555) | (0.00798) | (0.00809) | (0.00915) |
| *Control variables:* | | | | | | | |
| Year of birth*enrollment rate in 1971 | | No | Yes | Yes | No | Yes | Yes |
| Year of birth*water and sanitation program | | No | No | Yes | No | No | Yes |

# Does Schooling Affect Hourly Wages?

TABLE 7—EFFECT OF EDUCATION ON LABOR MARKET OUTCOMES: OLS AND 2SLS ESTIMATES

| Method | Instrument | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| *Panel A: Sample of Wage Earners* | | | | | |
| *Panel A1: Dependent variable: log(hourly wage)* | | | | | |
| OLS | | 0.0776 | 0.0777 | 0.0767 | |
| | | (0.000620) | (0.000621) | (0.000646) | |
| 2SLS | Year of birth dummies*program | 0.0675 | 0.0809 | 0.106 | 0.0908 |
| | intensity in region of birth | (0.0280) | (0.0272) | (0.0222) | (0.0541) |
| | | [0.96] | [0.9] | [0.93] | [0.9] |
| 2SLS | (Aged 2–6 in 1974)*program | 0.0752 | 0.0862 | 0.104 | |
| | intensity in region of birth | (0.0338) | (0.0336) | (0.0304) | |
| | | (0.0338) | (0.0336) | (0.0304) | |
| Control variables: | | | | | |
| Year of birth*enrollment rate in 1971 | | No | Yes | Yes | Yes |
| Year of birth*water and sanitation program | | No | No | Yes | No |
| Propensity score, propensity score squared | | No | No | No | Yes |

# Does Schooling Affect Hourly Wages?

TABLE 7—EFFECT OF EDUCATION ON LABOR MARKET OUTCOMES: OLS AND 2SLS ESTIMATES

| Method | Instrument | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| *Panel B: Whole Sample* | | | | | |
| *Panel B1: Dependent variable: participation in the wage sector* | | | | | |
| OLS | | 0.0328 | 0.0327 | 0.0337 | |
| | | (0.00311) | (0.000311) | (0.000319) | |
| 2SLS | Year of birth dummies*program intensity in region of birth | 0.101 | 0.118 | 0.0892 | |
| | | (0.0210) | (0.0197) | (0.0162) | |
| | | [0.66] | [0.93] | [1.12] | |
| Control variables: | | | | | |
| Year of birth*enrollment rate in 1971 | | No | Yes | Yes | Yes |
| Year of birth*water and sanitation program | | No | No | Yes | No |
| Propensity score, propensity score squared | | No | No | No | Yes |

# Conclusions

- Additional 0.12 - 0.19 years of education for each new school built per 1000 children.

- Additional 1.5% - 2.7% in wages for each new school built per 1000 children.

- IV: An extra year of schooling implies 6.8% - 10.6% higher earnings.

# What We Learned

- Good instruments must satisfy two conditions in order to convincingly solve the endogeneity problem.

  IV1: an instrument must be strongly correlated with the endogenous right-hand-side variable.

  IV2: an instrument must not be correlated with the error term in the main equation of interest.

- Finding a good instrument is really hard