# Instrumental Variables and Endogeneity: Structural vs Reduced Form

Dr. Yang Liang

## 1 Structural vs Reduced-Form Models

**Structural Model**

- Derived from **economic theory** — describes behavioral relationships.

- Each equation has a **causal interpretation** (e.g. demand, supply, production).

- Often **simultaneous**: variables on both sides are determined together.

  **Example: Supply and Demand**

$$Q_d = \alpha_0 - \alpha_1 P + u_d \qquad \text{(Demand)}$$
$$Q_s = \beta_0 + \beta_1 P + u_s \qquad \text{(Supply)}$$
$$Q_d = Q_s = Q \qquad \text{(Market equilibrium)}$$

Structural equations reflect economic mechanisms.

**Reduced-Form Model**

- Expresses endogenous variables (e.g. $Q, P$) as functions of exogenous variables and instruments.

- No behavioral meaning — just statistical relationships.

- Used to estimate the causal effect implied by the structural system.

  **Example:**
$$P = \pi_0 + \pi_1 Z + v_P, \quad Q = \gamma_0 + \gamma_1 Z + v_Q,$$

where $Z$ are exogenous variables (e.g. supply shifters).

| Feature | Structural Model | Reduced Form |
|---|---|---|
| Derived from | Theory | Data relationships |
| Interpretation | Behavioral / Causal | Statistical |
| Endogeneity | Possible | None (in reduced form) |
| Purpose | Explain mechanism | Estimate total effect |

# 2   Sources of Endogeneity (Bias in OLS)

OLS assumes $E[X'u] = 0$. When violated $\Rightarrow$ biased and inconsistent estimates.

## (a) Simultaneity Bias

**Definition:** When $X$ and $Y$ are determined jointly (e.g. in equilibrium), causing correlation between $X$ and $u$.
   **Example:** Demand and supply

$$Q_d = \alpha_0 - \alpha_1 P + u_d, \quad Q_s = \beta_0 + \beta_1 P + u_s, \quad Q_d = Q_s.$$

**Show:** Equilibrium price $P$ depends on both $u_d$ and $u_s$. Hence $\text{Cov}(P, u_d) \neq 0$.
   **Solution:** Use an instrumental variable (e.g. cost shifter that affects supply but not demand).

## (b) Reverse Causality

**Definition:** The dependent variable $Y$ affects the regressor $X$, rather than the other way around.
   **Example:** Health and income

$$\text{Health}_i = \beta_0 + \beta_1 \text{Income}_i + u_i,$$

but healthier individuals may earn higher income $\Rightarrow \text{Cov}(\text{Income}, u) \neq 0$.
   **Solution:** Find exogenous variation in income (e.g. randomized tax credit, policy change).

## (c) Omitted Variable Bias (OVB)

**Definition:** A relevant variable affecting $Y$ is missing and correlated with $X$.
   **Example:** Wage regression

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Education}_i + u_i.$$

Ability affects both education and wages $\Rightarrow$ omitted variable bias.
   **Solution:**

- Control for ability proxies (e.g. test scores), or

- Use IV: instrument for education (e.g. proximity to college, compulsory schooling laws).

## (d) Measurement Error

**Definition:** Observed regressor $X^*$ measured with noise:

$$X^* = X + v.$$

Classical measurement error $\Rightarrow$ attenuation bias.
   **Example:** Self-reported income or years of schooling.
   **Solution:** Use IV or validation data (e.g. administrative records).

| Source of Endogeneity | Example | Direction of Bias | Solution |
|---|---|---|---|
| Simultaneity | Supply & demand | Ambiguous | Supply/demand shifters as IVs |
| Reverse causality | Health $\leftrightarrow$ Income | Ambiguous | Lagged/exogenous shocks |
| Omitted variable | Education & wages | Typically positive | Add controls or IV |
| Measurement error | Misreported income | Toward zero | Use IV or better data |

# 3  Instrumental Variables (IV)

**Purpose**

Find a variable $Z$ that isolates exogenous variation in the endogenous regressor $X$.

**IV Conditions**

1. **Relevance:** $\text{Cov}(Z, X) \neq 0$ — instrument must be correlated with $X$.

2. **Exogeneity (Exclusion Restriction):** $\text{Cov}(Z, u) = 0$ — instrument affects $Y$ only through $X$.

**IV Estimator (Single Instrument, Single $X$)**

$$\hat{\beta}_{IV} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)}.$$

**IV in Structural Models**

- Goal: identify **structural parameters** (e.g. demand slope, supply elasticity).

- Instruments come from **economic theory**: variables that shift one equation but not the other.

**Example (Demand):**
$$Q = \alpha_0 - \alpha_1 P + u_d, \quad P = f(Z, u_d, u_s),$$

Use a supply shifter $Z$ (e.g. weather, cost shock) as an instrument for $P$.

**IV in Reduced-Form Models**

- Focus on causal effects, not structural interpretation.

- IV estimates the effect of $X$ on $Y$ for compliers — **LATE (Local Average Treatment Effect)**.

**Example (Education and wages):**

$$Y_i = \beta_0 + \beta_1 \text{Educ}_i + u_i, \quad Z_i = \text{Distance to college.}$$

$Z$ affects education decisions (first stage). Under exclusion, IV identifies the effect of education for those whose schooling responds to $Z$.

| Context | Purpose | Interpretation | |
|---|---|---|---|
| Structural IV | Identify economic parameters | Causal mechanisms (e.g. demand elasticity) | Supply & |
| Reduced-form IV | Identify treatment effects | LATE / policy-relevant effect | |

## Summary Equation

$$\text{Structural:} \quad Y = \beta_0 + \beta_1 X + u, \quad E[Xu] \neq 0,$$

$$\text{Instrument: } Z \text{ such that } E[Zu] = 0, \ E[ZX] \neq 0,$$

then

$$\hat{\beta}_{IV} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, X)}.$$