Simultaneous Equations

The Chicken and the Egg

Jake Anderson

March 3, 2026

"My belt holds my pants up, but the belt loops hold my belt up.

I don't really know what's happening down there.

Who is the real hero?"

— Mitch Hedberg

Outline

The Problem: Jointly Determined Variables

In most regressions, we think of x as causing y. But sometimes causation runs **both ways simultaneously**.

Supply and demand:

- **Price** depends on how many people want boba (demand) and how many cups the shop can make (supply)
- **Quantity sold** also depends on the price
- P determines Q, and Q determines P: they settle together in **equilibrium**

When we observe (P_t, Q_t) data from a market, each data point is an equilibrium: the intersection of supply and demand.

$\implies$ Regressing Q on P with OLS does not recover the demand curve or the supply curve. It produces a confused mixture of both.

Bruin Boba: The Setup

Track weekly **price per cup** (P_t) and **cups sold** (Q_t) at Bruin Boba:

Week	**Shock**	P^*	Q^*
1	Normal	\$4.80	104
2	Hot week (demand ↑)	\$6.00	140
3	Pearl shortage (supply ↓)	\$6.00	80
4	Instagram viral (demand ↑)	\$5.60	128
5	Milk spike (supply ↓)	\$6.40	72

Notice: Weeks 2 and 3 have the **same price** (\$6.00) but very different quantities (140 vs 80). One is a demand shift, the other is a supply shift.

$\implies$ A single OLS line through (P_t, Q_t) cannot recover demand or supply.

Bruin Boba: The True Curves

The true (linear) curves:

$$\textbf{Demand:} \quad Q = 200 - 20P + u_D$$

$$\textbf{Supply:} \quad Q = -40 + 30P + v_S$$

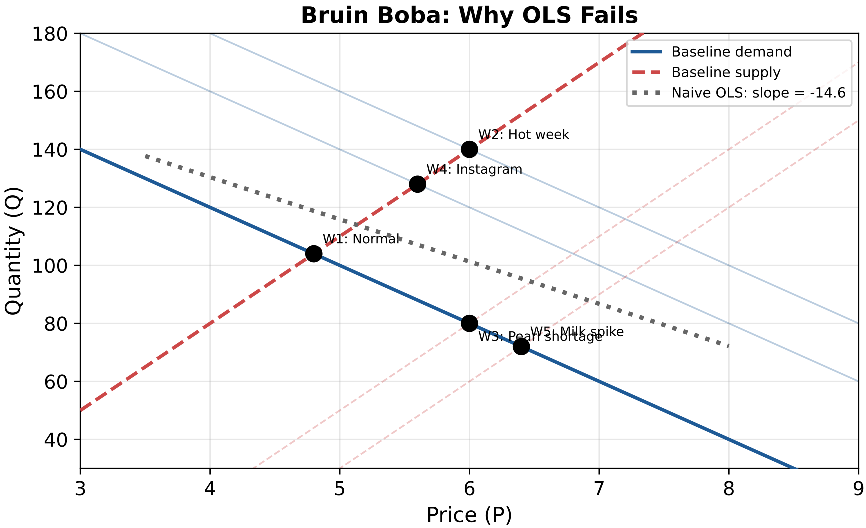
where u_D = demand shifter (weather, virality) and v_S = supply shifter (input costs).

Equilibrium: Set demand = supply:

$$200 - 20P + u_D = -40 + 30P + v_S$$

$$P^* = \frac{240 + u_D - v_S}{50}, \qquad Q^* = 200 - 20P^* + u_D$$

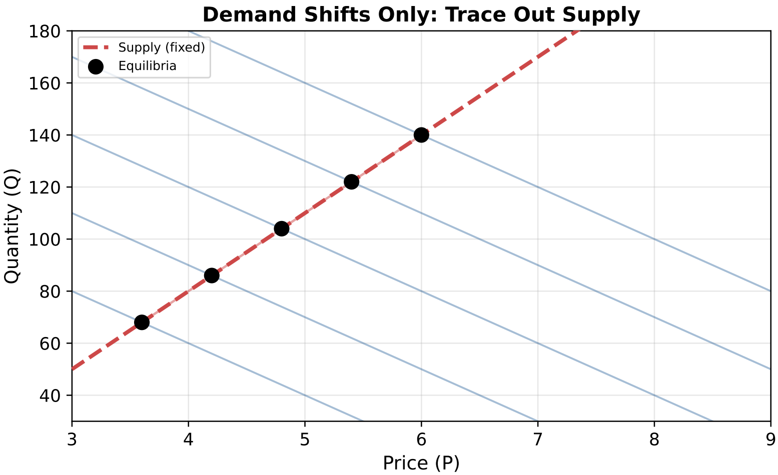
Visualizing the Problem



Outline

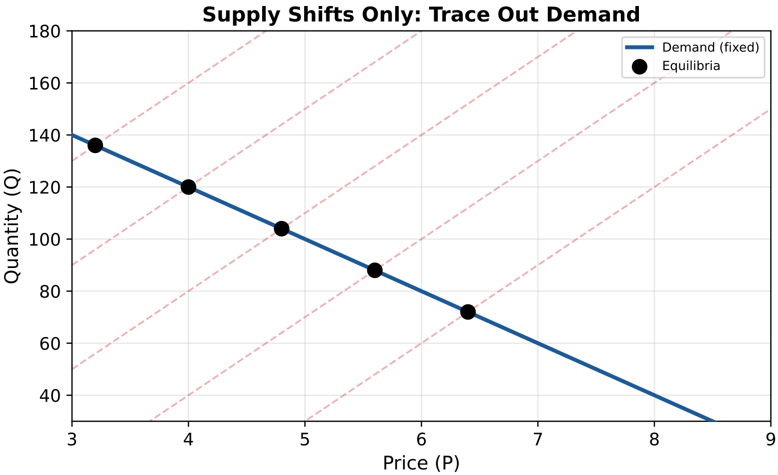
When Only Demand Shifts

If demand shifts while supply stays fixed, equilibria move **along the supply curve**.



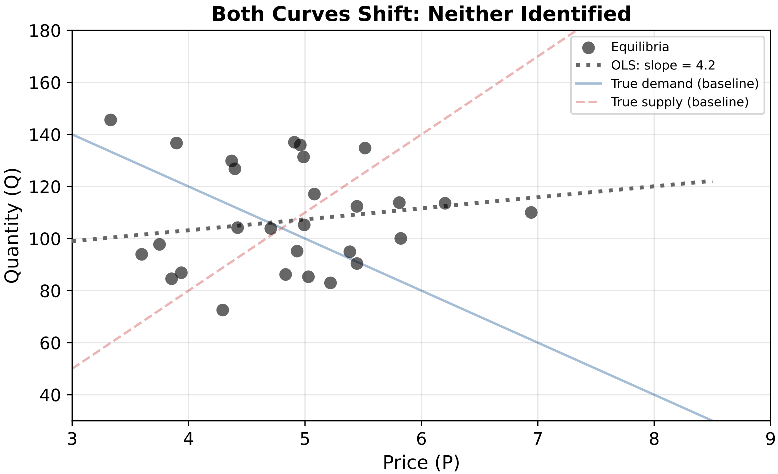
When Only Supply Shifts

If supply shifts while demand stays fixed, equilibria move **along the demand curve**.



When Both Shift: No Identification

If both curves shift simultaneously and we have no way to separate the shifts, the equilibria scatter in both directions.



Outline

The Structural Form

The **structural form** describes behavioral relationships from economic theory.

Truffle market:

$$\text{Demand:} \quad Q = \alpha_1 + \alpha_2 P + \alpha_3 \text{PS} + \alpha_4 \text{DI} + e^d$$

$$\text{Supply:} \quad Q = \beta_1 + \beta_2 P + \beta_3 \text{PF} + e^s$$

where:

- $P, Q =$ endogenous (determined *within* the system)
- $\text{PS} =$ price of substitute, $\text{DI} =$ disposable income (exogenous, demand side)
- $\text{PF} =$ price of a production factor (exogenous, supply side)

$\Longrightarrow$ In the demand equation, P is correlated with e^d because supply shocks affect P through equilibrium. OLS on structural equations is **biased and inconsistent**.

The Reduced Form

The **reduced form** expresses each endogenous variable as a function of *only* exogenous variables.

Derivation: Set demand = supply and solve for P:

$$P = \underbrace{\frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2}}_{\pi_{10}} + \underbrace{\frac{-\alpha_3}{\alpha_2 - \beta_2}}_{\pi_{11}} \text{PS} + \underbrace{\frac{-\alpha_4}{\alpha_2 - \beta_2}}_{\pi_{12}} \text{DI} + \underbrace{\frac{\beta_3}{\alpha_2 - \beta_2}}_{\pi_{13}} \text{PF} + v_1$$

Written compactly:

$$P = \pi_{10} + \pi_{11}\text{PS} + \pi_{12}\text{DI} + \pi_{13}\text{PF} + v_1$$

$$Q = \pi_{20} + \pi_{21}\text{PS} + \pi_{22}\text{DI} + \pi_{23}\text{PF} + v_2$$

$\implies$ OLS works on reduced-form equations because the RHS contains only exogenous variables. These are exactly the **first-stage regressions** in 2SLS.

Structural vs Reduced Form: Summary

Structural Form

Demand: $Q = \alpha_1 + \alpha_2 P + \alpha_3 PS + \alpha_4 DI + e^d$

Supply: $Q = \beta_1 + \beta_2 P + \beta_3 PF + e^s$

Solve for equilibrium →

Reduced Form

$P = \pi_{10} + \pi_{11} PS + \pi_{12} DI + \pi_{13} PF + v_1$

$Q = \pi_{20} + \pi_{21} PS + \pi_{22} DI + \pi_{23} PF + v_2$

RHS of reduced form = only exogenous variables ⇒ OLS is consistent

The reduced-form π coefficients are combinations of structural α's and β's. Recovering the structural parameters requires identification.

Outline

The Order Condition

In a system of M simultaneous equations, an equation is identified if it **excludes at least** $M - 1$ variables that appear elsewhere in the system.

Intuition: To trace out the demand curve, we need something that shifts *supply* but not demand. The excluded variables provide those shifts.

Exclusions	vs $M-1$	Status	Meaning
$< M - 1$	too few	Not identified	Cannot estimate
$= M - 1$	exact	Just identified	Exactly enough instruments
$> M - 1$	more than enough	Overidentified	Surplus instruments (testable)

Checking Identification: Truffle Market

With $M = 2$ equations, we need at least $M - 1 = 1$ exclusion per equation.

Demand equation:

- Contains: P, Q, PS, DI
- Excludes: **PF** (appears in supply, not demand)
- 1 exclusion $\geq 1 \implies$ **just identified**

Supply equation:

- Contains: P, Q, PF
- Excludes: **PS** and **DI** (appear in demand, not supply)
- 2 exclusions $\geq 1 \implies$ **overidentified**

$\implies$ Both equations are identified. The supply equation has a surplus instrument we can test.

Identification: Visual Summary

Identified: Truffle Market

Demand equation

Contains: P, Q, PS, DI
Excludes: PF

1 exclusion
$\geq$ M-1=1 ✓

Supply equation

Contains: P, Q, PF
Excludes: PS, DI

2 exclusions
$\geq$ M-1=1 ✓

Just identified / Overidentified

Not Identified: Same Variables

Demand equation

Contains: P, Q, DI
Excludes: nothing

0 exclusions
$<$ M-1=1 ✗

Supply equation

Contains: P, Q, DI
Excludes: nothing

0 exclusions
$<$ M-1=1 ✗

Neither equation identified

If both equations contain the same variables, there are no excluded instruments and neither equation is identified.

Outline

2SLS: The Procedure

Once an equation is identified, estimate it with **two-stage least squares**.

Estimating the demand equation ($Q = \alpha_1 + \alpha_2 P + \alpha_3 \text{PS} + \alpha_4 \text{DI} + e^d$):

Stage 1: Regress P on **all exogenous variables** (PS, DI, PF):

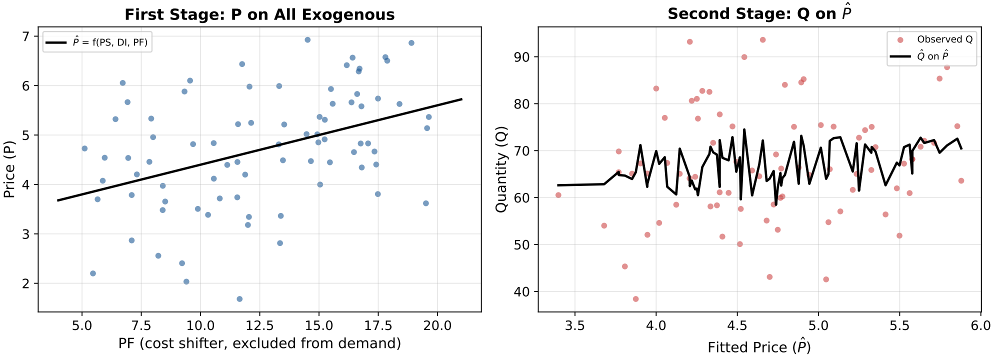
$$P = \pi_{10} + \pi_{11}\text{PS} + \pi_{12}\text{DI} + \pi_{13}\text{PF} + v_1 \quad \longrightarrow \quad \hat{P}$$

Stage 2: Replace P with $\hat{P}$ in the demand equation and run OLS:

$$Q = \alpha_1 + \alpha_2 \hat{P} + \alpha_3 \text{PS} + \alpha_4 \text{DI} + \text{residual}$$

$\Longrightarrow$ The first stage purges P of its correlation with e^d. The instrument for P in the demand equation is PF (the variable excluded from demand).

2SLS: The Two Stages Visually



Left: First stage projects P onto exogenous variables. **Right:** Second stage uses $\hat{P}$ (purged of endogeneity) to estimate the structural equation.

2SLS for the Supply Equation

Estimating supply ($Q = \beta_1 + \beta_2 P + \beta_3 \text{PF} + e^s$):

Stage 1: Same regression as before (P on all exogenous):

$$P = \pi_{10} + \pi_{11}\text{PS} + \pi_{12}\text{DI} + \pi_{13}\text{PF} + v_1 \quad \longrightarrow \quad \hat{P}$$

Stage 2: Replace P with $\hat{P}$ in the supply equation:

$$Q = \beta_1 + \beta_2 \hat{P} + \beta_3 \text{PF} + \text{residual}$$

The instruments for P in the supply equation are PS and DI (the demand shifters excluded from supply).

Same first stage, different second stage. Each structural equation uses different excluded variables as instruments.

Instrument Strength

Same rule as single-equation IV: check that the instruments are **strong**.

- The F-statistic from the first-stage regression should exceed **10** (Staiger–Stock rule of thumb)
- Weak instruments lead to biased and imprecise 2SLS estimates

In a simultaneous system:

- The first stage regresses the endogenous variable on **all** exogenous variables in the entire system
- Excluded instruments provide the identifying variation
- If the excluded instruments are weak predictors of P, identification breaks down even though the order condition is satisfied

$\Longrightarrow$ The order condition is *necessary* but not *sufficient*. Instrument strength is equally important.

Outline

Testing Surplus Instruments: Sargan Test

When an equation is **overidentified** ($L > B$, where $L =$ number of instruments, $B =$ number of endogenous regressors), we have **surplus instruments** we can test.

Sargan test procedure:

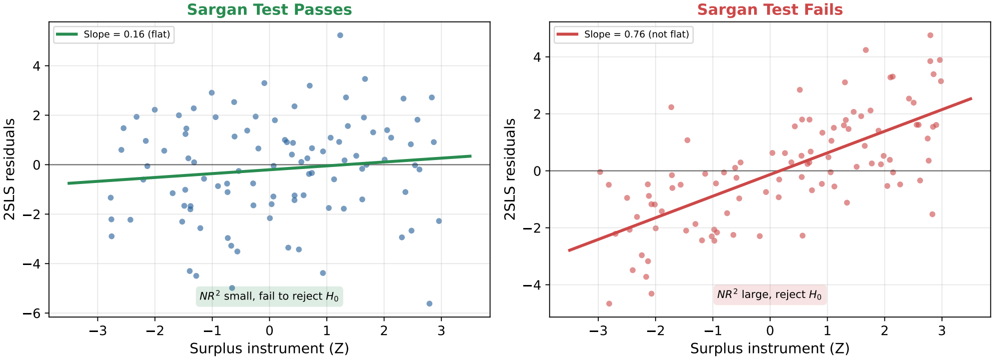
1. Estimate the structural equation by 2SLS. Obtain residuals $\hat{e}$.
2. Regress $\hat{e}$ on all exogenous variables (the instruments). Get R^2.
3. Test statistic: $NR^2 \sim \chi^2_{L-B}$ under H_0.

H_0: All surplus instruments are valid (uncorrelated with the structural error).

H_1: At least one surplus instrument is invalid.

$\implies$ If $NR^2 > \chi^2_{L-B,\,\alpha}$, reject. The surplus instruments appear to be correlated with the error, suggesting misspecification.

Sargan Test: Visual Intuition



Left: Residuals show no pattern with the surplus instrument $\implies$ valid. **Right:** Clear correlation $\implies$ instrument is endogenous.

Sargan Test: Truffle Supply Equation

The supply equation has $L = 2$ instruments (PS, DI) for $B = 1$ endogenous variable (P).

$\implies L - B = 1$ surplus instrument. Degrees of freedom $= 1$.

Example result: $NR^2 = 0.43$, critical value $\chi^2_{1,0.05} = 3.84$.

Since $0.43 < 3.84$, we **fail to reject** H_0.

$\implies$ No evidence that the surplus instrument is invalid. The 2SLS estimates for the supply equation appear reliable.

Limitation: The Sargan test assumes at least one instrument is valid. If *all* instruments are bad, the test has no power to detect it.

Outline

Truffle Market: Setup

Structural model:

$$\text{Demand:} \quad Q = \alpha_1 + \alpha_2 P + \alpha_3 \text{PS} + \alpha_4 \text{DI} + e^d$$

$$\text{Supply:} \quad Q = \beta_1 + \beta_2 P + \beta_3 \text{PF} + e^s$$

(a) Endogenous vs exogenous:

- **Endogenous:** P and Q (jointly determined by the intersection)
- **Exogenous:** PS (substitute price), DI (disposable income), PF (factor cost)

(b) Identification check:

- Demand: excludes PF (1 exclusion $\geq M - 1 = 1$) $\implies$ just identified
- Supply: excludes PS, DI (2 exclusions $\geq M - 1 = 1$) $\implies$ overidentified

Truffle Market: OLS vs 2SLS

(c) A researcher naively runs OLS on the demand equation:

$$\hat{\alpha}_2^{\text{OLS}} = -0.37$$

The 2SLS estimate (using PF as instrument for P):

$$\hat{\alpha}_2^{\text{2SLS}} = -0.53$$

Why do they differ?

- OLS is biased because $\text{Cov}(P, e^d) \neq 0$
- The OLS estimate (-0.37) **understates** the true price sensitivity: simultaneity bias pushes the estimate toward zero

$\implies$ Trust the 2SLS estimate (-0.53). Since PF shifts supply but not demand, it isolates movement along the demand curve.

Truffle Market: Overidentification Test

(d) For the supply equation, the Sargan overidentification test gives:

$$NR^2 = 0.43, \qquad \chi^2_{1,\,0.05} = 3.84$$

H_0: All surplus instruments are valid (uncorrelated with e^s).

Since $0.43 < 3.84$: **fail to reject** H_0.

The supply equation has $L = 2$ instruments (PS, DI) for $B = 1$ endogenous variable (P), so $L - B = 1$ surplus instrument can be tested.

$\Longrightarrow$ The instruments appear valid. No evidence of misspecification in the supply equation.

Summary

1. **Simultaneity bias:** When P and Q are jointly determined, OLS on structural equations is biased and inconsistent
2. **Identification:** To estimate a structural equation, we need variables excluded from that equation but present elsewhere (the order condition: $\geq M - 1$ exclusions)
3. **Reduced form:** Expresses endogenous variables as functions of only exogenous variables. OLS is consistent on reduced-form equations.
4. **2SLS:** First stage regresses the endogenous variable on all exogenous variables. Second stage replaces the endogenous variable with its fitted values.
5. **Sargan test:** When overidentified, regress 2SLS residuals on all exogenous variables. $NR^2 \sim \chi^2_{L-B}$ tests instrument validity.

Thank you!

jakeanderson@g.ucla.edu