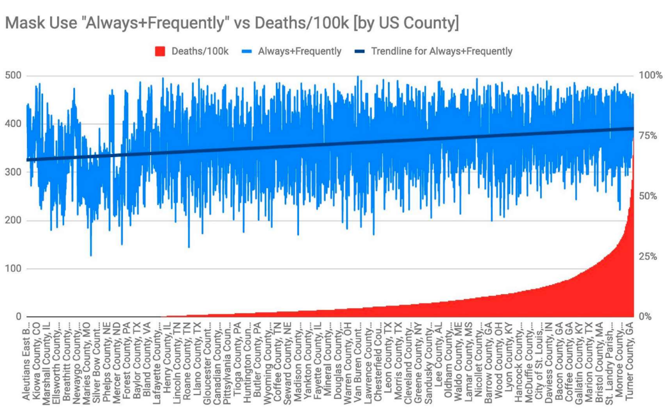
Instrumental Variables

Monday, November 13, 2023 1:58 PM

From Khoa Vu, University of Minnesota:

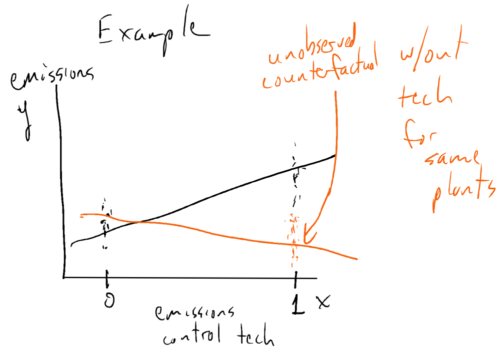
From some guy on Twitter:



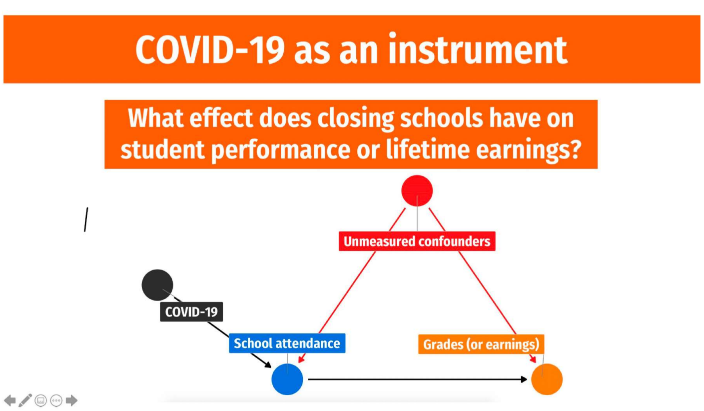
$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$E(x_t \cdot \varepsilon_t) = 0$$

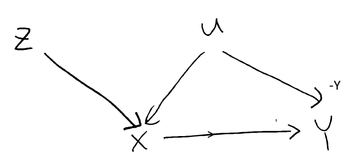
Example



From Andrew Heiss, Georgia State University:



unobserved confounders



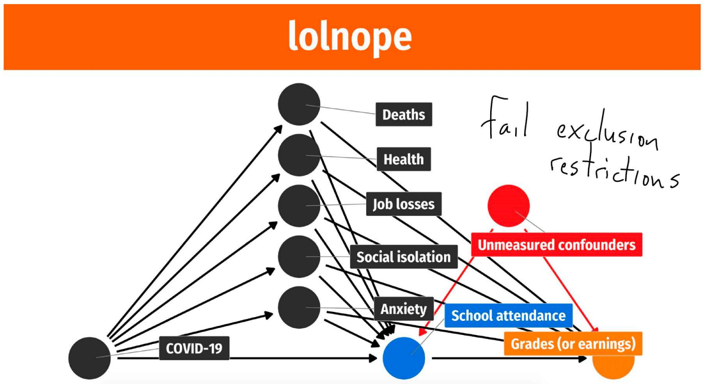
Z affects X, not U

and only affects Y through X

Z an Instrumental Variable "IV"

① Relevance: $Cov(Z, X) \neq 0$

(2) Exclusion Restriction : $Cov(z, u) = 0$



fail exclusion restrictions

In commodity markets → want to know shape of supply + demand curves

$$q_{st} = \alpha + \beta P_t + u_t$$

$$q_{dt} = c - d P_t + v_t$$

In practice we observe equilibrium outcomes

$$\alpha + \beta P_t + u_t = c - d P_t + v_t$$

solve for $P_t = \frac{c - \alpha}{\beta + d} + \frac{v_t + u_t}{\beta + d}$

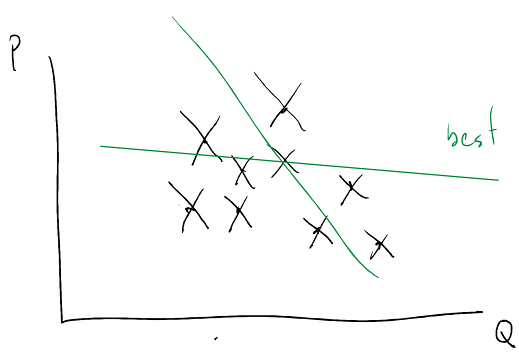
Naive Regression :

$$q_t = \phi_0 + \phi_1 P_t + \varepsilon_t$$

contains $v_t + u_t$

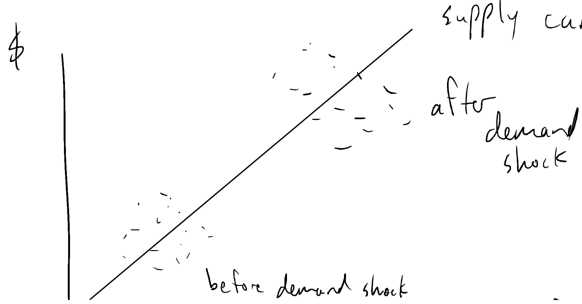
contains $v_t + u_t$

correlated → both contain unobserved demand + supply shocks



best fit line not demand or supply.

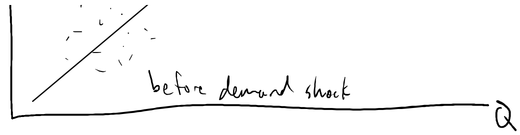
Maybe I can find some demand shifters to trace out supply curve



supply curve identified by exogenous demand shifters

↓

"Instrumental Variables"



Variables"

Basic Idea:

Endogenous Variable X, Exogenous (Instrumental Variable) Z

1) regress X on Z, predict $\hat{X}$ from that regression

$$x_t = \underbrace{\phi_0 + \phi_1 z_t}_{\text{exogenous part of } x} + \underbrace{v_t}_{\text{endogenous part of } x}$$

$$\hat{x}_t = \hat{\phi}_0 + \hat{\phi}_1 z_t$$

→ contains variation only from z → shared variation between x + z

2) regress Y on $\hat{X}$

$$y_t = \beta_0 + \beta_1 \hat{x}_t + e_t$$

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_t + \hat{e}_t \quad \text{fitted regression}$$

Back to supply + demand

$$q_{st} = \beta_0 + \beta_1 p_t + \beta_2 z_{st} + u_t$$

z_{st} = input prices, technology shocks war, weather shocks

$$q_{dt} = \alpha_0 - \alpha_1 p_t + \alpha_2 z_{dt} + v_t$$

z_{dt} = consumer income

suppose only P_t is endogenous in each regression

$E(z_{st} u_t) = 0 \qquad E(z_{dt} v_t) = 0$

instruments for demand: z_{dt} for itself, z_{st} for P_t

" " supply: z_{st} for itself, z_{dt} for P_t

Focus on demand for now:

$$q_{dt} = \underline{x}_t^T \cdot \underline{\alpha} + v_t = \begin{bmatrix} 1 & p_t & z_{dt} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ -\alpha_1 \\ \alpha_2 \end{bmatrix} + v_t$$

every row

$$\underset{T\times 1}{\underline{q}_d} = \underset{T\times 3}{\underset{\sim}{X}} \; \underset{3\times 1}{\underline{\alpha}} + \underset{T\times 1}{\underline{v}}$$

$\hat{\ } \quad (\ \cdot^T \ )^{-1} \cdot^T$ — matrix VARIAN

$\underline{Cov(q_d, x)}$

$$\begin{array}{ccc} \overset{v}{T\times 1} & T\times 3 \quad 3\times 1 & T\times 1 \end{array}$$

From OLS $\quad \hat{\underline{\alpha}} = (\underset{3\times T}{X^T}\underset{T\times 3}{X})^{-1}\underset{3\times T}{X^T}\underset{3\times 1}{q_d} \quad$ = matrix version of $\quad \dfrac{Cov(q_d, x)}{Var(x)}$

For IV (or 2-stage least squares 2SLS)

1) regress X on Z : $\underbrace{(Z^T Z)^{-1} Z^T X}$ coefficients in first stage

predict $\hat{X} = Z\left[(Z^T Z)^{-1} Z^T X\right]$

right hand side of stage 1

estimated coefficients in stage 1

2) regress q_d on $\hat{X}$

$$\hat{\underline{\alpha}}_{IV} = (\hat{X}^T \cdot \hat{X})^{-1} \hat{X}^T \cdot q_d$$

cancel

$$= \left[X^T Z \cancel{(Z^T Z)^{-1} Z^T Z} (Z^T Z)^{-1} Z^T X\right]^{-1} X^T Z (Z^T Z)^{-1} Z^T q_d$$

$$= \left[X^T Z (Z^T Z)^{-1} Z^T X\right]^{-1} X^T Z (Z^T Z)^{-1} Z^T q_d$$

general case w/ more IV's (supply shifters) than endogenous variables (price)

In special case w/ 1 IV for each endogenous variable, formula simplification:

$$\hat{\underline{\alpha}}_{IV} = (Z^T X)^{-1} Z^T q_d \quad \text{matrix version of} \quad \frac{Cov(z, y)}{Cov(z, x)}$$

q_d

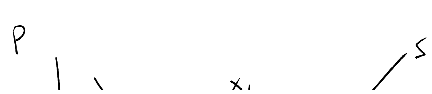
use these formulas to get standard errors:

$$\hat{\underline{\alpha}}_{IV} = (Z^T X)^{-1} Z^T (X \underline{\alpha} + \underline{v})$$

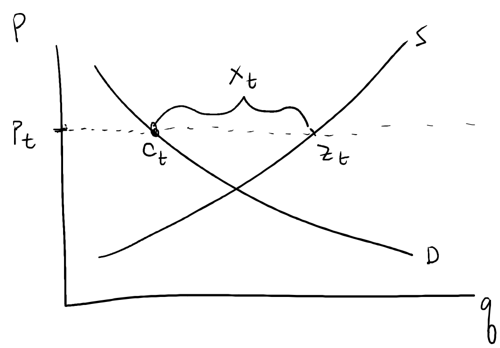
Examples from commodity markets:

storable commodity

- Roberts + Schlenker American Economic Review 2013
  - → ethanol
- Hausman + Kellogg Brookings Economic Papers 2015
  - → shale gas



x_t = amount stored in t carried into $t+1$ as inventory



x_t = amount stored in t carried into $t+1$ as inventory

$$Z_{t+1} = x_t + \underbrace{f(\lambda_t, \omega_{t+1})}_{\text{production function}}$$

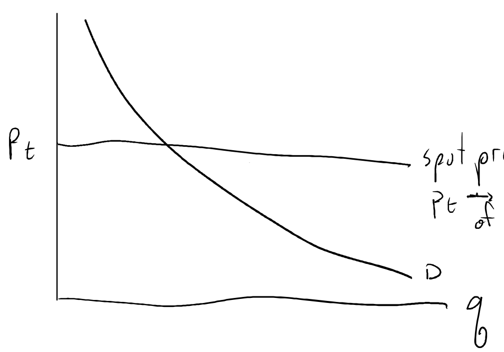
x_t: storage decisions previously

λ_t: inputs decisions chosen previously

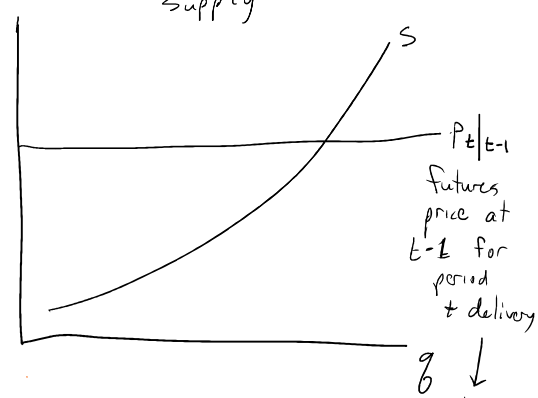
ω_{t+1}: productivity shock unforecastable

Insight: if we can capture incentives for storage in earlier periods, we can trace out supply and demand.

· Demand



Supply



New production at time t = $\log(\lambda_{t-1} \cdot \omega_t) = q_{st}$

as a supply curve:

2nd stage

$$q_{st} = \alpha_s + \beta_s P_{t|t-1} + \gamma_s \omega_{st} + \underbrace{f_{sz}(t)}_{\text{monthly dummies, trend}} + u_t$$

can't be correlated

first stage of IV for $P_{t|t-1}$

$$P_{t|t-1} = \delta_s + \mu_{s0} \cdot \omega_{st} + \mu_{s1} \omega_{st-1} + \mu_d \omega_{dt-1} + \underbrace{f_{s1}(t)}_{\text{monthly dummies, trends}} + \varepsilon_t$$

correlated

good IV has two requirements:

① IV's strongly correlated w/ endogenous X's

② NOT correlated with u_t → the main equation residual

Demand side:

spot price

$$q_{dt} = \alpha_d + \beta_d P_t + \omega_{dt} + f_{d2}(t) + v_t$$

first stage

$$P_t = \delta_1 + \mu_{1n} \cdot \omega_{dt} + f_{d1}(t) + \mu_{d1} \omega_{dt-1} + \mu_{\_} \omega_{st-1} + e_t$$

first stage

$$p_t = \delta_d + \mu_{d0} \cdot \omega_{dt} + f_{d1}(t) + \underbrace{\mu_{d1}\omega_{dt-1} + \mu_s \omega_{st-1}}_{\text{Instruments}} + e_t$$

What to use ? for ω_{dt-1}, ω_{st-1}

Roberts + Schlenker (ethanol) unexpected corn yield shocks
→ regress yield on time trends, seasonal dummies, take residuals → ω_{st}

Hausman + Kellogg (natural gas)
→ neighboring states demand shocks (weather in neighboring states)
→ neighbor's demand shock is a supply shock for me.

→ lagged cumulative weather shock in neighboring states over a year
→ want shocks to inventory!

Cournot Econometrics (Vitamins Paper)

Cournot Problem : small # firms, each influences market price

Competitive market each firm "i"
$$\max_{q_i} \pi(q_i) = p \cdot q_i - c_i(q_i)$$

$$\text{FOC} : \; p - c_i'(q_i) = 0$$

Oligopoly w/ quantity competition

$$\max_{q_i} \pi_i(q_i) = \underbrace{P(q_1 + \dots + q_i + \dots + q_N)}_{\text{inverse demand function}} \cdot q_i - c_i(q_i)$$

$$= P(Q) \cdot q_i - c_i(q_i) \qquad Q = \sum_{i=1}^{N} q_i$$

$$\text{FOC}: \; \underbrace{P(Q) + \frac{dP}{dQ} \cdot q_i}_{\text{marginal revenue}} - \underbrace{c_i'(q_i)}_{\text{marginal cost}} = 0$$

In Vitamins Paper : marginal costs are constant

$$P_t + \frac{dP}{dQ} \cdot q_{it} = c_{mt} + \gamma_i + \eta_{it}$$

Each firm! data

regression residual

parameters to

marginal cost of

firm i's marginal cost

residual unexpected cost shocks

data

parameters to estimate

marginal cost of market leader

firm i's marginal cost increment

residual unexpected cost shocks

approach: use c_{mt} as IV / supply shifter

① in demand function to get $\frac{dP}{dQ}$ demand slope.

② use demand shifters as IV's ~~to~~ to estimate the γ_i (the marginal cost functions)

→ disposable income in high income countries