How to use Instrumental Variables

Questions that can be answered with experiments:

1) If I give people aspirin, what happens to their body temperature?

2) If I use fertilizer on tomato plants, do I get more tomatoes?

Other conditions could affect outcome (hot room for body temperature, rainfall for plants)

If I have the power to hold other conditions fixed, I do so.
If I can't control other conditions,
I *randomize* treatment so that treatment is uncorrelated with other changes in conditions.

Questions that can't be answered with experiments:

1) What would happen to quantity demanded if I raise price,
leave other determinants of demand the same? (Price elasticity of demand)
Fool with price, see effect on quantity demanded. $P \rightarrow Q$

2) What would happen to real GDP if I raise real interest rate,
leave other determinants of real GDP the same? (Slope of IS curve)
Fool with real interest rate, see effect on real GDP. $r \rightarrow Y$

How you *can't* answer nonexperimental questions:

collect data where variables of interest wriggle

across time (time-series)

units (cross-section)

time & units (panel)

and observe co-movements between variables.

because other determinants/relevant conditions will be correlated with "treatment"

1) Anything that changes one economic variable changes others too -general equilibrium!

2) Reverse causality.

In quantity demanded, treatment is ΔP

1) other variables that affect Q^D, perhaps correlated with P: I, other P's

2) supply $Q^S(P)$

In IS curve, treatment is Δr

1) other variables that affect Y, perhaps correlated with r: G, Y^* *(exports)*, Y^e

2) Taylor rule $r = \bar{r} + \alpha(Y - \bar{Y}) + \beta(\pi - \pi^*)$

How to approach nonexperimental questions

1) Ignore reverse causality and correlation with other determinants.

- old papers
- econometrics field, where goal is to demonstrate technique not answer questions

2) Answer question for a model, not real world

In a model, you can exogenously wriggle variable of interest.
To quantify parameters, assume model is true and fit reduced forms to data

3) Natural experiments

Find some wriggles in treatment variable that are uncorrelated with other determinants

Often involves "instrumental variable:"

- something measureable that causes or indicates wriggles in treatment variable ("relevant")
- uncorrelated with other determinants ("valid")

e.g. in quantity demanded, variable that affects supply not demand

in IS curve, exogenous variations in central bank setting of r

In terms of regressions

$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (\varepsilon = \beta_3 X_3 + ...)$

X_1 "Troublesome variable" X_2 "Control variable" $X_3 \ldots$ Unmeasurable determinants

General equilibrium problems: $X_3 = \alpha_{31} X_1 + ...$ $X_1 = \alpha_{13} X_3 + ...$

$$X_1 = \alpha_{14} X_4 + ...$$
$$X_3 = \alpha_{34} X_4 + ...$$

Reverse causality:

$$X_1 = \alpha_{Y1} Y + \alpha_{15} X_5 + \ldots \quad \rightarrow \quad X_1 = \alpha_{Y1}\left(\beta_1 X_1 + \beta_2 X_2 + \varepsilon\right) + \alpha_{15} X_5 + \ldots$$

correlation with ε ↗

Instrumental variables *Z:*

1) Relevant: *Z* is correlated with X_1

$$X_1 = \gamma_{Z1} Z_1 + \gamma_{Z2} Z_2 + ...$$

2) Valid: *Z* is *uncorrelated* with ε (not cause or effect or common cause or...)

$$X_3 = 0 \times Z_1 + 0 \times Z_2 + ... \text{ "Exclusion restriction"}$$

Methods

1) 2SLS Good for small samples and/or probably-homoskedastic ε's

1) "First stage" Regress X_1 on Z.

2) Use first-stage coefficients to get predicted values $\hat{X}_1$
Note there's lots of variation in X_1 not captured by $\hat{X}_1$

3) "Second stage" Regress Y on X_1

Note that it wouldn't be right to use SE's from second stage to judge significance -
this wouldn't account for uncertainty about coefficients of first-stage regression.
2SLS SE's account for this.

2) GMM Good for large samples, heteroskedastic ε's
Can allow for correlations across observations' residuals.
E.g. "clustering" (common in panels):
ε's correlated across some observations, uncorrelated across others

Problem! if instruments are **"weak"** (relevant but only weakly correlated with X_1)
SE's *too small* and estimated coefficients *biassed* toward OLS values
(though *consistent* as $n \Rightarrow \infty$)
even in *very large* samples

How do you find instruments?
How do you know if possible instruments are *relevant* and *valid*?

1) Think about why X_1 varies. List all possible reasons.

2) Think about other determinants of Y. List all possible X_3 's.

3) Which reasons for X_1-variation are unrelated to X_3 and *measurable*?

Of course, you must think about what causes Z to vary, and on and on...

This is not a statistical issue. It's knowledge about the world (including economics).

but there are..

Statistical methods to test whether candidate *Z*'s are relevant, strong & valid

1) Look at "first-stage" regression.

Check sign, significance, magnitude (plausible?) of coefficient on *Z*.
Because of "weak instrument" problem, *t*-stats (or *F*-stats for multiple *Z*'s) need to be *lots bigger than 2*. "Rule of thumb": at least 10.

2) Look at "reduced form" regression of *Y* on *Z*

$$Y = \beta_1\left(\gamma_{Z1} Z_1 + \gamma_{Z2} Z_2 + \dots\dots\right) + \beta_2 X_2 + \varepsilon$$

Sign, magnitudes of coefficients should be consistent with story.

3) If you have more than one *Z* (X_1 is "overidentified") you can test one *Z* against others, see if results are consistent across *Z*'s.

Validity/exclusion restriction means all *Z's* *uncorrelated* with $(Y - \beta_1 X_1 - \beta_2 X_2)$

a) Get estimates $\hat{\beta}_1$, $\hat{\beta}_2$
b) Calculate $(Y - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2)$
c) Are *Z's* is uncorrelated with $(Y - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2)$?

Sargan test (for 2SLS), Hansen J-test (for GMM)

Z's are OK if you "fail to reject" hypothesis.

Are lagged *Y*'s and/or *X*'s valid instruments?

"One of the most mechanical and naive, yet common, approaches to the choice of instruments uses atheoretical and hard-to-assess assumptions about dynamic relationships to construct instruments from lagged variables in time series or panel data. The use of lagged endogenous variables..is problematic if the equation error or omitted variables are serially correlated"

(Angrist & Krueger, 2001)

E.g. in the Arellano-Bond panel estimator, "identification results from lack of serial correlation in the errors" (p. 293)

Are lagged *Y*'s and/or *X*'s valid instruments? (cont.)

Example: supply and demand

You want to estimate demand elasticity.

$$Q^D = -bP_t + \varepsilon_t \quad \textit{where} \quad \varepsilon_t = \rho_D \varepsilon_{t-1} + \hat{\varepsilon}_t \qquad \rho_D \geq 0$$

$$Q^S = cP_t + e_t \quad \textit{where} \quad e_t = \rho_S e_{t-1} + \hat{e}_t \qquad \rho_S \geq 0$$

$$(\beta_1 = -b \ , \ \alpha_{Y1} = 1/c)$$

so $P_t = \frac{1}{b+c}(\varepsilon_t - e_t)$ $\qquad Q_t = \frac{b}{b+c}(\varepsilon_t + e_t)$

1) For P_{t-1} or Q_{t-1} to be *relevant* instruments for P_t,
ρ_D and/or ρ_S must be greater than zero.

2) but if $\rho_D \neq 0$, they are not *valid*
because they are correlated with ε_t

Bottom line for you

In every empirical anything,

1) Try to tell me why X_1 varies across observations within your sample

2) Choose one from below

a) do not talk about "effects" (as in demand elasticity, IS curve slope, ...)

b) argue that most variation in X_1 is uncorrelated with unmeasurable determinants of Y variable

c) use BS (fitted model)

d) use instruments

3) If you choose d),

i) Tell me what causes Z's to vary, why they are correlated with X_1, why they are valid

ii) Show first-stage and reduced-form regressions

iii) Prove Z's are not *weak* with rule-of-thumb F-tests (or R^2's) or other tests

iv) Unless you can make very strong *a priori* case for validity, show Sargan or J-test