Instrumental Variables Regression (Stock and Watson, Ch. 12)

Giovanni Millo[1]

[1]DEAMS, University of Trieste

Outline

- IV regression: why and what; two-stage least squares
- The general IV regression model
- Checking instrument validity: weak/strong instruments and exogeneity
- Application: demand for cigarettes
- Examples: where do valid instruments come from?

IV Regression: Why?

- Three important threats to internal validity are:
- Omitted-variable bias from an unobserved variable correlated with X and inadequately controlled for.
- Simultaneous-causality bias (X causes Y and Y causes X).
- Errors-in-variables bias (X is measured with error).
- All three problems imply $E(u \mid X) \neq 0$.
- IV regression can eliminate this bias using an instrumental variable Z.

The IV Estimator with One Regressor and One Instrument

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- IV regression breaks X into a part that may be correlated with u and a part that is not.
- Isolating the uncorrelated part allows us to estimate β_1.
- Use an instrument Z_i correlated with X_i but uncorrelated with u_i.

Terminology: Endogeneity and Exogeneity

- An endogenous variable is correlated with u.
- An exogenous variable is uncorrelated with u.
- IV regression focuses on an endogenous X and an exogenous instrument Z.
- "Endogenous" literally means determined within the system; the broader IV definition also covers omitted-variable and measurement-error problems.

Two Conditions for a Valid Instrument

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

1. Instrument relevance: $\text{corr}(Z_i, X_i) \neq 0$
2. Instrument exogeneity: $\text{corr}(Z_i, u_i) = 0$

Suppose such a Z_i is available. How can it be used to estimate β_1?

The IV Estimator with One X and One Z

Explanation 1: Two-Stage Least Squares (TSLS)

1. Regress X on Z by OLS:

$$X_i = \pi_0 + \pi_1 Z_i + v_i.$$

Because Z_i is uncorrelated with u_i, the fitted component is uncorrelated with u_i.
Compute $\hat{X}_i$.

Two-Stage Least Squares (continued)

2. Replace X_i by $\hat{X}_i$ in the equation of interest and run OLS:

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i.$$

- $\hat{X}_i$ is uncorrelated with u_i in large samples.
- The resulting coefficient is the TSLS estimator $\hat{\beta}_1^{TSLS}$.

Two-Stage Least Squares: Summary

If Z_i is valid:

1. Relevance: $\text{corr}(Z_i, X_i) \neq 0$.
2. Exogeneity: $\text{corr}(Z_i, u_i) = 0$.

- Stage 1: regress X_i on Z_i (with an intercept) and obtain $\hat{X}_i$.
- Stage 2: regress Y_i on $\hat{X}_i$ (with an intercept).
- The coefficient on $\hat{X}_i$ is $\hat{\beta}_1^{TSLS}$, a consistent estimator of β_1.

The IV Estimator: Direct Algebraic Derivation

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\begin{aligned} \operatorname{cov}(Y_i, Z_i) &= \operatorname{cov}(\beta_0 + \beta_1 X_i + u_i, Z_i) \\ &= \beta_1 \operatorname{cov}(X_i, Z_i), \end{aligned}$$

so, because $\operatorname{cov}(u_i, Z_i) = 0$,

$$\boxed{\beta_1 = \frac{\operatorname{cov}(Y_i, Z_i)}{\operatorname{cov}(X_i, Z_i)}}.$$

The IV Estimator: Sample Covariances

$$\beta_1 = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(X_i, Z_i)}$$

Replace population covariances with sample covariances:

$$\boxed{\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}}}.$$

Here s_{YZ} and s_{XZ} are sample covariances. This is the TSLS estimator, obtained by a different derivation.

The IV Estimator from the Reduced Form

The reduced form relates Y and X to Z:

$$X_i = \pi_0 + \pi_1 Z_i + v_i, \qquad Y_i = \gamma_0 + \gamma_1 Z_i + w_i.$$

Because Z is exogenous, it is uncorrelated with v_i and w_i.
A one-unit change in Z changes X by π_1 and Y by γ_1. Hence the effect of an exogenous change in X is

$$\boxed{\beta_1 = \gamma_1 / \pi_1}.$$

IV from the Reduced Form: Algebra

$$
\begin{aligned}
Z_i &= -\frac{\pi_0}{\pi_1} + \frac{1}{\pi_1} X_i - \frac{1}{\pi_1} v_i, \\
Y_i &= \gamma_0 + \gamma_1 Z_i + w_i \\
&= \left(\gamma_0 - \frac{\pi_0 \gamma_1}{\pi_1} \right) + \frac{\gamma_1}{\pi_1} X_i + \left(w_i - \frac{\gamma_1}{\pi_1} v_i \right).
\end{aligned}
$$

Thus $Y_i = \beta_0 + \beta_1 X_i + u_i$, where $\beta_0 = \gamma_0 - \pi_0 \gamma_1 / \pi_1$, $\beta_1 = \gamma_1 / \pi_1$, and $u_i = w_i - (\gamma_1 / \pi_1) v_i$.

IV from the Reduced Form: Interpretation

$$X_i = \pi_0 + \pi_1 Z_i + v_i, \qquad Y_i = \gamma_0 + \gamma_1 Z_i + w_i$$

so

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad \beta_1 = \frac{\gamma_1}{\pi_1}.$$

Interpretation: an exogenous change in X of π_1 units is associated with a change in Y of γ_1 units.

Example 1: Effect of Studying on Grades

- What is the effect on grades of studying for an additional hour per day?
- $Y =$ GPA; $X =$ study time (hours per day).
- Data: grades and study hours of college freshmen.
- Would the OLS estimator of β_1 be unbiased? Why or why not?

Studying on Grades

- Stinebrickner, Ralph and Stinebrickner, Todd R. (2008), "The Causal Effect of Studying on Academic Performance," The B.E. Journal of Economic Analysis & Policy, Vol. 8, Issue 1.
- $n = 210$ freshmen at Berea College (Kentucky), 2001.
- $Y =$ first-semester GPA.
- $X =$ average study hours per day.
- Roommates were randomly assigned.
- $Z = 1$ if the roommate brought a video game, and 0 otherwise.
- Is Z relevant? Is it exogenous?

Studying on Grades: Reduced Form

$$X = \pi_0 + \pi_1 Z + v, \qquad Y = \gamma_0 + \gamma_1 Z + w.$$

$Y =$ GPA (4-point scale); $X =$ study time; $Z = 1$ if roommate brought a video game.

$$\hat{\pi}_1 = -0.668, \qquad \hat{\gamma}_1 = -0.241,$$

$$\hat{\beta}_1^{IV} = \frac{-0.241}{-0.668} = 0.360.$$

The original regressions included additional regressors.

Consistency of the TSLS Estimator

$$\hat{\beta}_1^{TSLS} = \frac{s_{YZ}}{s_{XZ}}$$

Sample covariances are consistent:

$$s_{YZ} \xrightarrow{p} \text{cov}(Y, Z), \qquad s_{XZ} \xrightarrow{p} \text{cov}(X, Z).$$

Therefore $\hat{\beta}_1^{TSLS} \xrightarrow{p} \beta_1$. Instrument relevance, $\text{cov}(X, Z) \neq 0$, prevents division by zero.

Example 2: Supply and Demand for Butter

IV regression was first developed to estimate demand elasticities for agricultural goods such as butter:

$$\ln Q_i^{butter} = \beta_0 + \beta_1 \ln P_i^{butter} + u_i.$$

β_1 is the price elasticity: the percentage change in quantity for a 1% change in price.
OLS suffers from simultaneous-causality bias because price and quantity are jointly determined by supply and demand.

Simultaneous Causality in Butter Demand

SlidesCap12-1.pdf

Supply-Demand Example: Shifting Supply

SlidesCap12-1.pdf

TSLS in the Supply-Demand Example

$$\ln Q_i^{butter} = \beta_0 + \beta_1 \ln P_i^{butter} + u_i.$$

Let $Z_i =$ rainfall in dairy-producing regions.

1. Relevant? $\text{corr}(rain_i, \ln P_i^{butter}) \neq 0$. Plausibly: insufficient rainfall $\rightarrow$ less grazing $\rightarrow$ less butter $\rightarrow$ higher prices.
2. Exogenous? $\text{corr}(rain_i, u_i) = 0$. Plausibly, rainfall in dairy regions should not affect demand for butter.

TSLS in the Supply-Demand Example: Continued

$$Z_i = rain_i.$$

1. Stage 1: regress $\ln P_i^{butter}$ on rainfall. The fitted value isolates price changes arising from supply.
2. Stage 2: regress $\ln Q_i^{butter}$ on the fitted log price.

This is the regression counterpart of using supply-curve shifts to trace the demand curve.

Example 3: Test Scores and Class Size

- California test-score/class-size regressions may suffer from omitted-variable bias, e.g. parental involvement.
- In principle, IV/TSLS can eliminate this bias.
- A valid instrument must be relevant: $\text{corr}(Z_i, STR_i) \neq 0$.
- It must be exogenous: $\text{corr}(Z_i, u_i) = 0$.

Test Scores and Class Size: Hypothetical Instrument

- Suppose some districts, randomly hit by an earthquake, "double up" classrooms.
- $Z_i = Quake_i = 1$ if hit by the earthquake, 0 otherwise.
- The earthquake makes the situation resemble random assignment, so the resulting variation in class size is exogenous.
- The first stage regresses class size on $Quake$, isolating the "as-if randomly assigned" component.

Inference Using TSLS

- In large samples, the TSLS estimator is approximately normally distributed.
- Hypothesis tests and confidence intervals proceed in the usual way, e.g. estimate $\pm 1.96SE$.
- The large-sample argument uses the CLT applied to averages of mean-zero i.i.d. random variables.
- The following slides sketch the algebra.

TSLS Sampling Distribution: Algebra I

$$\hat{\beta}_1^{TSLS} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Z_i - \bar{Z})}{\sum_{i=1}^n (X_i - \bar{X})(Z_i - \bar{Z})} = \frac{\sum_{i=1}^n Y_i (Z_i - \bar{Z})}{\sum_{i=1}^n X_i (Z_i - \bar{Z})}.$$

Substitute $Y_i = \beta_0 + \beta_1 X_i + u_i$ and simplify.

TSLS Sampling Distribution: Algebra II

$$\hat{\beta}_1^{TSLS} = \beta_1 + \frac{\sum_{i=1}^{n} u_i(Z_i - \bar{Z})}{\sum_{i=1}^{n} X_i(Z_i - \bar{Z})}.$$

Thus

$$\hat{\beta}_1^{TSLS} - \beta_1 = \frac{\sum u_i(Z_i - \bar{Z})}{\sum X_i(Z_i - \bar{Z})}.$$

TSLS Sampling Distribution: Algebra III

$$n(\hat{\beta}_1^{TSLS} - \beta_1) = \frac{\frac{1}{n}\sum(Z_i - \bar{Z})u_i}{\frac{1}{n}\sum X_i(Z_i - \bar{Z})}.$$

By the LLN,

$$\frac{1}{n}\sum X_i(Z_i - \bar{Z}) \xrightarrow{p} \text{cov}(X, Z) \neq 0.$$

By the CLT, $n^{-1}\sum(Z_i - \bar{Z})u_i$ is asymptotically normal.

Inference Using TSLS: Continued

$$\hat{\beta}_1^{TSLS} \approx N(\beta_1, \sigma^2_{\hat{\beta}_1^{TSLS}}),$$

where

$$\sigma^2_{\hat{\beta}_1^{TSLS}} = \frac{1}{n}\frac{\text{var}[(Z_i - \mu_Z)u_i]}{[\text{cov}(Z_i, X_i)]^2}.$$

- Standard inference is justified by large samples and valid instruments.
- Second-stage OLS standard errors are incorrect because they ignore first-stage estimation.
- Use a dedicated TSLS command and, as usual, heteroskedasticity-robust standard errors.

Example 4: Demand for Cigarettes

$$\ln Q_i^{cigarettes} = \beta_0 + \beta_1 \ln P_i^{cigarettes} + u_i.$$

Why is OLS likely biased?

- Panel data: annual cigarette consumption and average prices, including taxes, for the 48 contiguous U.S. states, 1985–1995.
- Proposed instrument: $Z_i =$ general sales tax per pack in the state ($SalesTax_i$).
- Check relevance and exogeneity.

Cigarette Demand: 1995 Data

First stage:

$$\ln(\hat{P}_i^{cigarettes}) = 4.63 + 0.031 SalesTax_i, \qquad n = 48.$$

Second stage:

$$\ln(\hat{Q}_i^{cigarettes}) = 9.72 - 1.08 \ln(\hat{P}_i^{cigarettes}), \qquad n = 48.$$

Combined TSLS with correct heteroskedasticity-robust SEs:

$$\ln(\hat{Q}_i^{cigarettes}) = 9.72 - 1.08 \ln(\hat{P}_i^{cigarettes}), \quad SE = (1.53, 0.32).$$

STATA Example: Cigarette Demand, First Stage

```
. reg lravgprs rtaxso if year==1995, r

Number of obs = 48     F(1,46) = 40.39
Prob > F = 0.0000      R-squared = 0.4710
Root MSE = .09394

             Coef.   Std. Err.      t    P>|t|
rtaxso       .0307289 .0048354    6.35   0.000
_cons        4.616546 .0289177 159.64   0.000

. predict lravphat
```

The predicted values from the first stage are now available.

Second Stage

```
. reg lpackpc lravphat if year==1995, r

Number of obs = 48     F(1,46) = 10.54
Prob > F = 0.0022      R-squared = 0.1525
Root MSE = .22645

             Coef.   Std. Err.      t    P>|t|
lravphat    -1.083586 .3336949 -3.25   0.002
_cons         9.719875 1.597119  6.09   0.000
```

These coefficients are the TSLS estimates, but the standard errors are wrong because they ignore first-stage estimation.

One-Command TSLS in STATA

```
. ivregress 2sls lpackpc (lravgprs = rtaxso) if year==1995, vce(robust)

Number of obs = 48     Wald chi2(1) = 12.05
Prob > chi2 = 0.0005   R-squared = 0.4011
Root MSE = .18635

lravgprs  -1.083587    .3122035    -3.47    0.001
_cons      9.719876   1.496143     6.50    0.000

Instrumented: lravgprs
Instruments:  rtaxso
```

The endogenous regressor is $lravgprs$; $rtaxso$ is the instrument.

$$\ln Q_i^{cigarettes} = 9.72 - 1.08 \ln P_i^{cigarettes}, \qquad n = 48.$$

Summary: IV Regression with One X and One Z

- A valid instrument satisfies relevance, $\text{corr}(Z_i, X_i) \neq 0$, and exogeneity, $\text{corr}(Z_i, u_i) = 0$.
- TSLS first regresses X on Z to obtain $\hat{X}$, then regresses Y on $\hat{X}$.
- The first stage isolates variation in X uncorrelated with u.
- With a valid instrument, the large-sample TSLS distribution is normal and usual inference applies.

The General IV Regression Model

- Extend the single-X, single-Z case to multiple endogenous regressors $X_1, \ldots, X_k$.
- Include exogenous variables/control variables $W_1, \ldots, W_r$.
- Allow multiple instruments $Z_1, \ldots, Z_m$; more relevant instruments can reduce TSLS variance.
- New terminology: identification and overidentification.

Identification

- A parameter is identified if different parameter values imply different data distributions.
- In IV regression, identification depends on the number of instruments m relative to the number of endogenous regressors k.
- If there are fewer instruments than endogenous regressors, $\beta_1, \ldots, \beta_k$ cannot all be estimated.

Identification: Exactly, Over-, and Underidentified

- Exactly identified if $m = k$: just enough instruments.
- Overidentified if $m > k$: more instruments than necessary; overidentifying restrictions can be tested.
- Underidentified if $m < k$: too few instruments; additional instruments are needed.

General IV Model: Terminology

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \cdots + \beta_{k+r} W_{ri} + u_i.$$

- Y_i: dependent variable.
- $X_{1i}, \ldots, X_{ki}$: endogenous regressors.
- $W_{1i}, \ldots, W_{ri}$: included exogenous regressors or controls.
- $Z_{1i}, \ldots, Z_{mi}$: excluded exogenous instruments.
- Overidentified if $m > k$, exactly identified if $m = k$, underidentified if $m < k$.

TSLS with One Endogenous Regressor

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i.$$

- Stage 1: regress X_1 on all exogenous variables $W_1, \ldots, W_r, Z_1, \ldots, Z_m$ and an intercept; obtain $\hat{X}_1$.
- Stage 2: regress Y on $\hat{X}_1$, $W_1, \ldots, W_r$, and an intercept.
- Second-stage coefficients are TSLS estimates, but second-stage OLS SEs are wrong.
- Use the software's single-command IV/TSLS estimator for correct SEs.

Example: Cigarette Demand with Income

Assume income is exogenous and estimate income elasticity as well:

$$\ln Q_i^{cigarettes} = \beta_0 + \beta_1 \ln P_i^{cigarettes} + \beta_2 \ln Income_i + u_i.$$

- Endogenous variable: $\ln P_i^{cigarettes}$.
- Included exogenous variable: $\ln Income_i$.
- Instruments: general sales tax and cigarette-specific tax.
- With one endogenous regressor and two excluded instruments, the model is overidentified.

Cigarette Demand: One Instrument

```
. ivreg lpackpc lperinc (lravgprs = rtaxso) if year==1995, r

lravgprs   -1.143375   .3723025   -3.07   0.004
lperinc     .214515   .3117467    0.69   0.495
_cons       9.430658  1.259392    7.49   0.000

Instrumented: lravgprs
Instruments:  lperinc rtaxso
```

A single-command IV regression produces correct robust SEs.

Cigarette Demand: Two Instruments

```
. ivreg lpackpc lperinc (lravgprs = rtaxso rtax) if year==1995, r

lravgprs   -1.277424   .2496099   -5.12   0.000
lperinc     .2804045   .2538894    1.10   0.275
_cons       9.894955   .9592169   10.32   0.000

Instrumented: lravgprs
Instruments:  lperinc rtaxso rtax
```

With two instruments the SE on the price elasticity is smaller.

TSLS Estimates: One vs. Two Instruments

With sales tax only ($m = 1$):

$$\ln Q = 9.43 - 1.14 \ln P + 0.21 \ln Income,$$

with SEs $(1.26, 0.37, 0.31)$.
With sales tax and cigarette tax ($m = 2$):

$$\ln Q = 9.89 - 1.28 \ln P + 0.28 \ln Income,$$

with SEs $(0.96, 0.25, 0.25)$.

- Two instruments provide more information and more "as-if random" variation.
- Income elasticity is small and not statistically different from zero.
- Price elasticity is surprisingly large.

General Assumptions for Instrument Validity

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \beta_{k+1} W_{1i} + \cdots + \beta_{k+r} W_{ri} + u_i.$$

1. Exogeneity: $\text{corr}(Z_{ji}, u_i) = 0$ for every instrument j.
2. Relevance: in the general multiple-X case, the population first stage must provide enough variation to avoid perfect collinearity in the infeasible second stage.

For one X, this means at least one instrument enters the first stage and the W variables are not perfectly collinear.

IV Regression Assumptions

1. $E(u_i \mid W_{1i}, \ldots, W_{ri}) = 0$: the included exogenous regressors are exogenous.
2. $(Y_i, X_{1i}, \ldots, X_{ki}, W_{1i}, \ldots, W_{ri}, Z_{1i}, \ldots, Z_{mi})$ are i.i.d.
3. X, W, Z, Y have nonzero finite fourth moments.
4. The instruments are valid.

Under these assumptions, TSLS and its t statistic are asymptotically normal. The critical requirement is instrument validity.

Control Variables W

- Often W variables are included to control for omitted factors so that, conditional on W, Z is uncorrelated with u.
- In this case the W variables need not themselves be exogenous; they need to be effective controls.
- Technically, effective controls satisfy $E(u_i \mid W_i, Z_i) = E(u_i \mid W_i)$.

Control Variables W: Continued

An alternative to the first IV assumption is conditional mean independence:

$$E(u_i \mid W_i, Z_i) = E(u_i \mid W_i).$$

This is the IV analogue of the conditional-mean-independence assumption used for controls elsewhere. In many applications, controls are needed to make Z plausibly exogenous.

Studying and Grades: Why Might Z Be Endogenous?

- $Y =$ first-semester GPA; $X =$ average study hours per day.
- $Z = 1$ if roommate brought a video game; roommates were randomly assigned.
- Even random assignment may not guarantee Z is uncorrelated with u if another characteristic is related to both treatment assignment and grades.
- What other determinants of grades are contained in u?

Studying and Grades: Gender as a Control

- Suppose women obtain higher grades than men, holding study time fixed.
- Suppose men are more likely to bring a video game.
- Then $\mathrm{corr}(Z_i, u_i) < 0$.
- This is another form of omitted-variable bias; control for gender.

Studying and Grades: IV with Controls

Include gender as a control variable:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i.$$

The reported TSLS estimate includes gender and other controls, such as the student's major.

Checking Instrument Validity

- Recall the two requirements:
- 1. Relevance: at least one instrument must enter the population first-stage equation (special case of one X).
- 2. Exogeneity: every instrument must be uncorrelated with the error term.
- What happens if a requirement fails? How can we check? Which instruments should be used if there are several?

Checking Assumption 1: Instrument Relevance

For one included endogenous regressor,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i.$$

First stage:

$$X_i = \pi_0 + \pi_1 Z_{1i} + \cdots + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + \cdots + \pi_{m+k} W_{ki} + v_i.$$

- Instruments are relevant if at least one $\pi_j \neq 0$.
- They are weak if all π_j are zero or close to zero.

Consequences of Weak Instruments

- With weak instruments, the sampling distribution of TSLS and its t statistic is not normal even when n is large.
- For one X, $\hat{\beta}_1^{TSLS} = s_{YZ}/s_{XZ}$.
- If $\mathrm{cov}(X, Z)$ is near zero, the denominator is near zero and the estimator becomes unstable.
- Usual large-sample normal inference can be highly unreliable.

Sampling Distribution with Weak Instruments

SlidesCap12-1.pdf

Why the Normal Approximation Fails

SlidesCap12-1.pdf

Measuring Instrument Strength: First-Stage F

- With one endogenous X, the first stage regresses X on $Z_1, \ldots, Z_m, W_1, \ldots, W_k$.
- Completely irrelevant instruments correspond to all coefficients on the Z variables being zero.
- The first-stage F statistic tests whether the instruments jointly enter the first stage.
- Weak instruments imply a low first-stage F.

Weak Instruments with One X

- Compute the first-stage F statistic.
- Rule of thumb: if $F < 10$, the instrument set is weak.
- Then TSLS can be biased and standard errors, hypothesis tests, and confidence intervals can be misleading.

Why Compare the First-Stage F with 10?

- Rejecting the null that instrument coefficients are zero is not enough; the instruments need meaningful predictive power for the normal approximation.
- Comparing F with 10 corresponds to whether the TSLS bias relative to OLS is below roughly 10%.
- If $F < 10$, the relative bias may exceed 10%.

What to Do with Weak Instruments

- Find better instruments (easier said than done).
- If there are many instruments, discard the weakest; removing irrelevant instruments can raise the first-stage F.
- If all instruments are weak, use IV methods beyond TSLS.
- Separate estimation of β_1 from construction of valid confidence intervals.

Confidence Intervals with Weak Instruments

- TSLS confidence intervals are invalid with weak instruments.
- Two valid large-sample approaches are:
- 1. Anderson–Rubin confidence intervals.
- 2. Moreira conditional likelihood-ratio confidence intervals.

Weak-Instrument Confidence Intervals: Moreira

- The Anderson–Rubin interval inverts the Anderson–Rubin test of $\beta_1 = \beta_{1,0}$.
- The Moreira conditional likelihood-ratio interval inverts the corresponding conditional likelihood-ratio test.
- The latter is often narrower, especially with many instruments.
- Specialized software is required for these procedures.

Estimation with Weak Instruments

- There are no unbiased estimators when instruments are weak or irrelevant.
- Some estimators nevertheless have sampling distributions more centered on β_1 than TSLS.
- The key lesson is that weak instruments require inference methods designed for the weak-instrument problem.

Checking Assumption 2: Instrument Exogeneity

Instrument exogeneity requires all instruments to be uncorrelated with the error:

$$\text{corr}(Z_{1i}, u_i) = 0, \ldots, \text{corr}(Z_{mi}, u_i) = 0.$$

Unlike relevance, exogeneity generally cannot be verified from the data alone.

Testing Overidentifying Restrictions

Consider the simple model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

with two valid instruments Z_{1i}, Z_{2i}. More generally, suppose $m > k$ so the model is overidentified. Overidentification gives additional restrictions that can be tested.

The J Test of Overidentifying Restrictions

For $m > k$:

1. Estimate the structural equation by TSLS using all m instruments and obtain TSLS residuals $\hat{u}_i$.
2. Regress $\hat{u}_i$ on all instruments and included exogenous variables.
3. Compute the F statistic testing that the coefficients on all instruments are jointly zero.
4. Define $J = mF$.

The J Test: Distribution

Under the null that all instruments are exogenous,

$$J \sim \chi^2_{m-k}.$$

If $m = k$, then $J = 0$: there are no overidentifying restrictions. If some instruments are endogenous, J tends to be large and the null is rejected.

Checking Instrument Validity: Summary

- Relevance: at least one instrument must enter the population first stage.
- If instruments are weak, TSLS is biased and its t statistic is non-normal.
- For one endogenous regressor, check the first-stage F: $F > 10$ suggests strong instruments; $F < 10$ suggests weak instruments.
- Exogeneity: all instruments must be uncorrelated with u.
- If $m > 1$, the J test can partially test exogeneity; rejection means at least some instruments are endogenous.

2. Exogeneity

- All instruments must be uncorrelated with the error term.
- If $m > 1$, the J test evaluates the null that all instruments are exogenous against the alternative that some are endogenous.
- If the J test rejects, some instruments must be discarded or reconsidered.

Application: Demand for Cigarettes

- Why estimate cigarette-demand elasticity?
- Optimal taxation: the optimal tax rate depends inversely on price elasticity.
- Smoking creates externalities, including health effects of second-hand smoke and monetary externalities.

Cigarette Demand: Panel Data

- Annual cigarette consumption, average consumer prices including tax, personal income, and tax rates.
- 48 contiguous U.S. states, 1985–1995.
- IV is needed to handle simultaneous causality from supply-demand interaction.
- State indicators are control variables for time-invariant unobserved state characteristics.

Fixed-Effects Model of Cigarette Demand

$$\ln Q_{it}^{cigarettes} = \alpha_i + \beta_1 \ln P_{it}^{cigarettes} + \beta_2 \ln Income_{it} + u_{it}.$$

- $i = 1, \ldots, 48$; $t = 1985, \ldots, 1995$.
- Price is plausibly correlated with the error because of supply-demand interactions.
- α_i captures time-invariant state characteristics such as attitudes toward smoking.
- Use panel methods to eliminate α_i and TSLS to handle simultaneity.
- Use 1985–1995 changes to estimate a long-run response.

The "Changes" Method with $T = 2$

Model long-run effects using ten-year changes:

$$\Delta_{10} \ln Q_i = \beta_1 \Delta_{10} \ln P_i + \beta_2 \Delta_{10} \ln Income_i + \Delta_{10} u_i.$$

For example,

$$\Delta_{10} \ln P_i = \ln P_{i,1995} - \ln P_{i,1985}.$$

This is equivalent to using the original data with state fixed-effect indicators.

STATA: Cigarette Demand, Ten-Year Changes

```
. gen dlpackpc = log(packpc/packpc[_n-10])
. gen dlavgprs = log(avgprs/avgprs[_n-10])
. gen dlperinc = log(perinc/perinc[_n-10])
. gen drtaxs   = rtaxs-rtaxs[_n-10]
. gen drtax    = rtax-rtax[_n-10]
. gen drtaxso  = rtaxso-rtaxso[_n-10]
```

Here "_n-10" is the ten-year lagged observation.

TSLS with Ten-Year Changes

```
. ivregress 2sls dlpackpc dlperinc (dlavgprs = drtaxso), r

Number of obs = 48      F(2,45) = 12.31
Prob > F = 0.0001       R-squared = 0.5499

 dlavgprs  -.9380143   .2075022  -4.52  0.000
 dlperinc   .5259693   .3394942   1.55  0.128
 _cons      .2085492   .1302294   1.60  0.116
```

Estimated elasticity $= -0.94$ (SE $= 0.21$); income elasticity is small and statistically insignificant. All variables are ten-year changes.

Checking Instrument Relevance: First-Stage F

```
. reg dlavgprs drtaxso dlperinc

Number of obs = 48     F(2,45) = 23.86
Prob > F = 0.0000      R-squared = 0.5146

drtaxso   .0254611   .0037374    6.81    0.000
dlperinc -.2241037   .2119405   -1.06    0.296
_cons      .5321948   .031249    17.03    0.000

. test drtaxso
F(1,45) = 46.41    Prob > F = 0.0000
```

With one instrument, $F = t^2 = 6.81^2 = 46.41$. Since $46.5 > 10$, the instrument is not weak. Exogeneity cannot be tested because $m = k$.

Cigarette Demand, Ten-Year Changes: Two IVs

```
. ivregress 2sls dlpackpc dlperinc (dlavgprs = drtaxso drtax), vce(r)

Number of obs = 48     Wald chi2(2) = 45.44
Prob > chi2 = 0.0000   R-squared = 0.5466

dlavgprs  -1.202403  .1906896  -6.31  0.000
dlperinc   .4620299  .2995177   1.54  0.123
_cons      .3665388  .1180414   3.11  0.002
```

$drtaxso$ is the general sales-tax change; $drtax$ is the cigarette-specific tax change. Estimated elasticity is -1.2.

First-Stage F: Both Instruments

```
. reg dlavgprs drtaxso drtax dlperinc

Number of obs = 48     F(3,44) = 51.36
Prob > F = 0.0000      R-squared = 0.7779

drtaxso   .013457   .0030498    4.41   0.000
drtax     .0075734  .0010488   7.22   0.000
dlperinc -.0289943  .1474923  -0.20  0.845
_cons      .4919733  .0220923   22.27  0.000

. test drtaxso drtax
F(2,44)=75.65    Prob > F=0.0000
```

$75.65 > 10$, so the instruments are not weak. Because $m > k$, overidentifying restrictions can be tested.

Testing the Overidentifying Restrictions

```
. predict e, resid
. reg e drtaxso drtax dlperinc

. test drtaxso drtax
F(2,44)=2.47
```

The initial calculation gives $J = mF = 2(2.47) = 4.93$, but the displayed F uses the wrong degrees of freedom. The correct degrees of freedom for J are $m - k$.

Correct Degrees of Freedom for the J Statistic

$$J = mF.$$

Under the null of exogeneity,

$$J \sim \chi^2_{m-k}.$$

Here $J = 4.93$ and $m - k = 1$. The 5% critical value is 3.84, so the null is rejected at the 5% level.
STATA reports $J = 4.9319853$ and $p = 0.02636401$.

Interpreting Rejection of the J Test

- The J test rejects the null that both instruments are exogenous.
- Either $rtaxso$ is endogenous, or $rtax$ is endogenous, or both.
- The J test does not tell us which instrument is problematic; judgment is required.
- Cigarette-specific taxes may respond to political pressure associated with smoking prevalence.
- This reasoning does not apply in the same way to the general sales tax, so use the general sales tax alone.

Demand for Cigarettes: Summary of Empirical Results

SlidesCap12-1.pdf

Assessing the Validity of the Study

- Remaining internal-validity threats: omitted variables, functional-form misspecification, simultaneous causality, measurement error, and selection.
- Fixed effects control for unobserved factors that vary across states but not over time.
- If the general sales tax is a valid instrument after fixed effects, remaining simultaneous-causality bias is addressed.
- Instrument relevance and exogeneity are additional IV-specific threats.

Assessing Validity: External Validity

- The estimate is a long-run elasticity; it may not generalize to a short-run elasticity.
- Smoking levels today are lower than in 1985–1995.
- Cultural attitudes toward smoking have shifted against smoking.
- These changes may threaten generalization from the historical sample to current policy.

Where Do Valid Instruments Come From?

- The hard part of IV analysis is finding valid instruments.
- Method 1: variables in another equation, such as supply shifters that do not affect demand.
- Method 2: exogenous variation that is "as if" randomly assigned, does not directly affect Y, but affects X.
- Examples: rainfall shifts butter supply but not demand; sales taxes shift cigarette supply but not demand.

Example: Cardiac Catheterization

- McClellan, McNeil, and Newhouse (1994): Does more intensive treatment of acute myocardial infarction reduce mortality?
- $Y_i =$ survival time (days) of a heart-attack patient.
- $X_i = 1$ if the patient receives cardiac catheterization, 0 otherwise.
- Clinical trials show an effect, but is the treatment effective in the field?

Cardiac Catheterization: Endogeneity

$$SurvivalDays_i = \beta_0 + \beta_1 CardCath_i + u_i.$$

The treatment decision is endogenous because EMT personnel choose treatment using information related to unobserved patient health in u_i. If healthier patients are more likely to be catheterized, OLS overstates the treatment effect. Proposed instrument: differential distance to the nearest cardiac-catheterization hospital versus the nearest regular hospital.

Cardiac Catheterization: Instrument

- $Z =$ differential distance to a cardiac-catheterization hospital.
- Relevant? Greater distance reduces the probability of being taken to a CC hospital and receiving CC.
- Exogenous? If distance affects survival only through treatment, then $\mathrm{corr}(Z, u) = 0$.
- If patient location is effectively random, differential distance is "as if" randomly assigned.
- First stage: a linear probability model for treatment.
- OLS estimates a significant, large effect; TSLS estimates a small, often insignificant effect.

Example: Crowding Out of Private Charitable Spending

- Gruber and Hungerman (2005): "Faith-Based Charity and Crowd Out During the Great Depression."
- Question: Does government social-service spending crowd out private charitable spending?
- $Y =$ private charitable spending; $X =$ government spending.
- Proposed instrument: strength of the Congressional delegation.

Private Charitable Spending: Data

- Panel data by state, annually, 1929–1939, U.S.
- $Y =$ benevolent spending by six church denominations.
- $X =$ federal relief spending under New Deal programs.
- $Z =$ tenure of state representatives on House and Senate Appropriations Committees, in months.
- $W =$ many fixed effects.

Private Charitable Spending: Validity Assessment

- Instrument validity: relevance? exogeneity?
- Other internal-validity threats: omitted variables, functional form, measurement error, selection, simultaneous causality.
- External validity: generalization to the U.S. today or to aid in developing countries?

Example: School Competition

- Hoxby (2000), "Does Competition Among Public Schools Benefit Students and Taxpayers?"
- $Y =$ 12th-grade test scores.
- $X =$ measure of school-district choice, a function of the number of districts in the metropolitan area.
- Motivation for IV: school competition may be endogenous.
- Proposed instrument: number of small streams in the metropolitan area.

School Competition: Data

SlidesCap12-1.pdf

School Competition: Validity Assessment

- Instrument validity: relevance? exogeneity?
- Other internal-validity threats: omitted variables, functional form, measurement error, selection, simultaneous causality.
- External validity: relevance to the U.S. today and to developing-country contexts?

Conclusion

- A valid instrument isolates a part of X uncorrelated with u; that part identifies the effect of a change in X on Y.
- IV regression requires valid instruments.
- Relevance: check the first-stage F statistic.
- Exogeneity: with overidentification, test overidentifying restrictions using the J statistic.
- A valid instrument isolates variation in X that is "as if" randomly assigned.
- The fundamental validity requirement cannot be tested mechanically; it requires substantive judgment.

IV Frequently Asked Questions: When to Use IV?

- Use IV when X is correlated with u and a valid instrument is available.
- Possible sources of correlation:
- Omitted variables, e.g. ability bias in returns to education.
- Measurement error, e.g. years of education measured with error.
- Selection bias, e.g. patients select treatment.
- Simultaneous causality, e.g. supply and demand for butter or cigarettes.

IV FAQs: Threats to Internal Validity

- The main threat is failure of the valid-instrument assumption.
- Given controls W, instruments must be relevant and exogenous.
- Relevance: assess whether instruments are weak or strong; a first-stage $F > 10$ is the common rule of thumb.
- Exogeneity: the J statistic can test overidentifying restrictions when there are more instruments than endogenous regressors.
- In general, instrument exogeneity must also be assessed using expert knowledge of the application.