Instrumental Variables

Department of Economics
University of Wisconsin-Madison

January 23, 2012

Treatment Effects

Throughout my part of the course we will focus on the "Treatment Effect Model"

For now take that to be

$$Y_i = \alpha T_i + X_i'\beta + u_i$$

α measures the causal effect of T_i on Y_i.

We don't want to do OLS because we are worried that T_i is not randomly assigned, that is that T_i and u_i are correlated.

There are a number of different reasons that might be true-I think the main thing that we are worried about in the treatment effects literature is omitted variables.

In this sub-course we are going to think about a lot of different ways of dealing with this potential problem and estimating α.

Instrumental Variables

Lets start with the instrumental variables way of thinking of things

I basically want to think about three completely different approaches for identifying α

The first is the GMM approach (the second two will come from simultaneous equations)

To justify OLS we would need

$$E\left(T_i u_i\right) = 0$$
$$E(X_i u_i) = 0$$

The focus of IV is to try to relax the first assumption

(There is much less concern about the second)

Lets suppose that we have an instrument Z_i for which

$$E\left(Z_i u_i\right) = 0$$

We also will stick with the exactly identified case (1 dimensional Z_i)

Define

$$Z_i^* = \begin{bmatrix} Z_i \\ X_i \end{bmatrix}, W_i = \begin{bmatrix} T_i \\ X_i \end{bmatrix}, B = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Then we know

$$E\left[Z_i^* \left(Y_i - W_i' B\right)\right] = 0$$

Turning this into a GMM problem we get

$$\begin{aligned} 0 =& \frac{1}{N}\sum_{i=1}^{N} Z_i^* \left(Y_i - W_i'\widehat{B}\right) \\ =& \frac{1}{N}\sum_{i=1}^{N} Z_i^* Y_i - \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)\widehat{B} \end{aligned}$$

which we can solve as

$$\widehat{B} = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} Z_i^* Y_i$$

And that is IV.

To see it is consistent note that

$$
\begin{aligned}
\widehat{B} &= \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} Z_i^* Y_i \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} Z_i^* \left(W_i' B + u_i\right) \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right) B \\
&\quad + \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} Z_i^* u_i \\
&= B + \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} Z_i^* u_i \\
&\approx B
\end{aligned}
$$

since

$$\frac{1}{N}\sum_{i=1}^{N} Z_i^* u_i \approx 0$$

I take the notation $\approx$ to mean that the difference between the two objects converges in probability to zero.

Furthermore we get standard errors from

$$\sqrt{N}\left(\widehat{B}-B\right) \approx \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* W_i'\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i^* u_i\right]$$

Since under standard conditions:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Z_i^* u_i \approx N(0, E(u_i^2 Z_i^* Z_i^{*'})$$

so

$$\sqrt{N}\left(\widehat{B}-B\right) \approx N\left(0, E(Z_i^* W_i')^{-1} E(u_i^2 Z_i^* Z_i^{*'}) E(W_i Z_i^{*'})^{-1}\right)$$

And that is the formal econometrics, but lets get under the hood

Partitioned Regression

This is a really useful trick that I want to review-we will use it quite a few times in the next few weeks

Think about the model (in large matrix notation)

$$Y = X_1\beta_1 + X_2\beta_2 + u$$

We will define

$$M_2 \equiv I - X_2\left(X_2'X_2\right)^{-1}X_2'.$$

Two facts about M_2

M_2 is idempotent

$$
\begin{aligned}
M_2 M_2 &= \left(I - X_2 \left(X_2' X_2\right)^{-1} X_2'\right)\left(I - X_2 \left(X_2' X_2\right)^{-1} X_2'\right) \\
&= I - 2X_2 \left(X_2' X_2\right)^{-1} X_2' + X_2 \left(X_2' X_2\right)^{-1} X_2' X_2 \left(X_2' X_2\right)^{-1} X_2' \\
&= I - X_2 \left(X_2' X_2\right)^{-1} X_2' \\
&= M_2
\end{aligned}
$$

Residuals from Regression

For any potential dependent variable (say Y), $M_2 Y$ is the residuals I would get if I regressed Y on X_2

To see that let those regressors be $\widehat{g}$ and generically let $\widetilde{W}$ be those residuals so that

$$
\begin{aligned}
\widetilde{W} \equiv & W - X_2 \widehat{g} \\
= & W - X_2 \left(X_2' X_2 \right)^{-1} X_2' W \\
= & \left[I - X_2 \left(X_2' X_2 \right)^{-1} X_2' \right] W \\
= & M_2 W.
\end{aligned}
$$

If I think of the GMM moment equations for least squares I get the "two equations"

$$0 = X_1' \left(Y - X_1 \widehat{\beta}_1 - X_2 \widehat{\beta}_2 \right)$$
$$0 = X_2' \left(Y - X_1 \widehat{\beta}_1 - X_2 \widehat{\beta}_2 \right)$$

The second can be solved as

$$\widehat{\beta}_2 = \left(X_2' X_2 \right)^{-1} X_2' \left(Y - X_1 \widehat{\beta}_1 \right)$$

Now plug this into the first

$$\begin{aligned} 0 &= X_1' \left(Y - X_1 \widehat{\beta}_1 - X_2 \widehat{\beta}_2 \right) \\ &= X_1' \left(Y - X_1 \widehat{\beta}_1 - X_2 \left(X_2' X_2 \right)^{-1} X_2' \left(Y - X_1 \widehat{\beta}_1 \right) \right) \\ &= X_1' M_2 Y - X_1' M_2 X_1 \widehat{\beta}_1 \end{aligned}$$

Or

$$\begin{aligned}\widehat{\beta}_1 &= \left(X_1' M_2 X_1\right)^{-1} X_1' M_2 Y \\ &= \left(\widetilde{X}_1' \widetilde{X}_1\right)^{-1} \widetilde{X}' \widetilde{Y}\end{aligned}$$

That is if I

1. Run a regression of X_1 on X_2 and form its residuals $\widetilde{X}_1$
2. Run a regression of Y on X_2 and form its residuals $\widetilde{Y}$
3. Run a regression of $\widetilde{Y}$ on $\widetilde{X}_1$

This gives me exactly the same result as if I had run the full regression of Y on X_1 and X_2

It turns out the same idea works for IV.

Put everything we had before into large Matrix notation and we can write GMM as:

$$
\begin{aligned}
0 &= Z'\left(Y - T\widehat{\alpha} - X\widehat{\beta}\right) \\
0 &= X'\left(Y - T\widehat{\alpha} - X\widehat{\beta}\right)
\end{aligned}
$$

The second can be solved as

$$
\widehat{\beta} = \left(X'X\right)^{-1} X'\left(Y - T\widehat{\alpha}\right)
$$

Now plug this into the first

$$
\begin{aligned}
0 &= Z'\left(Y - T\widehat{\alpha} - X\widehat{\beta}\right) \\
&= Z'\left(Y - T\widehat{\alpha} - X\left(X'X\right)^{-1} X'\left(Y - T\widehat{\alpha}\right)\right) \\
&= Z'M_X Y - Z'M_X T\widehat{\alpha}
\end{aligned}
$$

so

$$\widehat{\alpha} = \frac{\widetilde{\widetilde{Z}}' \widetilde{Y}}{\widetilde{\widetilde{Z}}' \widetilde{\widetilde{T}}} \approx \frac{cov(\widetilde{\widetilde{Z}}_i, \widetilde{Y}_i)}{cov(\widetilde{\widetilde{Z}}_i, \widetilde{T}_i)}$$

Next note that

$$\begin{aligned}\widetilde{Y} =& M_x Y \\ =& \alpha M_x T + M_x X\beta + M_x u \\ =& \alpha \widetilde{T} + \widetilde{u}\end{aligned}$$

so

$$\begin{aligned}\widehat{\alpha} &\approx \frac{cov(\widetilde{Z}_i, \widetilde{Y}_i)}{cov(\widetilde{Z}_i, \widetilde{T}_i)} \\ &\approx \frac{cov(\widetilde{Z}_i, \alpha\widetilde{T}_i + \widetilde{u}_i)}{cov(\widetilde{Z}_i, \widetilde{T}_i)} \\ &= \alpha + \frac{cov(\widetilde{Z}_i, \widetilde{u}_i)}{cov(\widetilde{Z}_i, \widetilde{T}_i)}\end{aligned}$$

This formula is helpful. In order for the bias to be small you want

- $cov(\widetilde{Z}_i, \widetilde{u}_i)$ to be small
- $cov(\widetilde{Z}_i, \widetilde{T}_i)$ to be large

You can also see that there is some tradeoff between them.

If your instrument is not very powerful, a little bit of correlation in $cov(\widetilde{Z}_i, \widetilde{u}_i)$ could lead to a large bias.

Simultaneous equations

A second way to see IV is from the simultaneous equations framework

$$\begin{aligned} Y_i &= \alpha T_i + X_i'\beta + u_i \\ T_i &= \rho Y_i + X_i'\gamma + Z_i\delta + \nu_i \end{aligned}$$

Thee are called the “structural equations”

Note the difference between X_i and Z_i in that we restrict what can affect what.

We could also have stuff that affects Y_i but not not T_i but lets not worry about that (some of the γ coefficients could be zero)

We assume that

$$E(u_i \mid X_i, Z_i) = 0$$
$$E(v_i \mid X_i, Z_i) = 0$$

but notice that almost for sure T_i is correlated with u_i because u_i influences T_i through Y_i

It is useful to calculate the "reduced form" for T_i, namely

$$\begin{aligned} T_i &= \rho Y_i + X_i'\gamma + Z_i\delta + \nu_i \\ &= \rho\left[\alpha T_i + X_i'\beta + u_i\right] + X_i'\gamma + Z_i\delta + \nu_i \\ &= \rho\alpha T_i + X_i'\left[\rho\beta + \gamma\right] + Z_i'\delta + (\rho u_i + \nu_i) \\ &= X_i'\frac{\rho\beta + \gamma}{1 - \rho\alpha} + Z_i'\frac{\delta}{1 - \rho\alpha} + \frac{\rho u_i + \nu_i}{1 - \rho\alpha} \\ &= X_i'\beta_2^* + Z_i'\delta_2^* + \nu_i^* \end{aligned}$$

where

$$\begin{aligned} \beta_2^* &= \frac{\rho\beta + \gamma}{1 - \rho\alpha} \\ \delta_2^* &= \frac{\delta}{1 - \rho\alpha} \end{aligned}$$

Note that $E(\nu_i \mid X_i, Z_i) = 0$, so one can identify β_2^* and δ_2^*.

This is called the “reduced form” equation for T_i

Note that the parameters here are not the fundamental structural parameters themselves, but the are a known function of these parameters

This model is identified if we have "exclusion restrictions"

That is we can identify α as long as we have some Z_i

I want to show this two different ways

Method 1

We can also solve for the reduced for for Y_i

$$\begin{aligned} Y_i &= \alpha T_i + X_i'\beta + u_i \\ &= \alpha\left(\rho Y_i + X_i'\gamma + Z_i\delta + \nu_i\right) + X_i'\beta + u_i \\ &= \alpha\rho Y_i + X_i\left(\alpha\gamma + \beta\right) + \alpha\delta + \alpha\nu_i + u_i \\ &= X_i\frac{\alpha\gamma + \beta}{1 - \alpha\rho} + Z_i\frac{\alpha\delta}{1 - \alpha\rho} + \frac{\alpha\nu_i + u_i}{1 - \alpha\rho} \\ &= X_i\beta_1^* + Z_i\delta_2^* + u_i^* \end{aligned}$$

with

$$\begin{aligned} \beta_1^* &= \frac{\alpha\gamma + \beta}{1 - \alpha\rho} \\ \delta_1^* &= \frac{\alpha\delta}{1 - \alpha\rho} \end{aligned}$$

Notice then that

$$\frac{\delta_1^*}{\delta_2^*} = \alpha$$

So we can get the coefficient α simply by taking the ratio of the reduced form coefficients

In the exactly identified case, this is numerically identical to IV.

To see why this is identical note that in a regression of Y_i and T_i on (X_i, Z_i) yields

$$\widehat{\delta}_1 = \frac{\widetilde{Z}_i' \widetilde{Y}_i}{\widetilde{Z}_i' \widetilde{Z}_i}$$

$$\widehat{\delta}_2 = \frac{\widetilde{Z}_i' \widetilde{T}_i}{\widetilde{Z}_i' \widetilde{Z}_i}$$

so

$$\begin{aligned}\frac{\widehat{\delta}_1}{\widehat{\delta}_2} &= \frac{\widetilde{Z}_i' \widetilde{Y}_i}{\widetilde{Z}_i' \widetilde{T}_i} \\ &= \widehat{\alpha}_{IV}\end{aligned}$$

Method 2

Define

$$T_i^f \equiv X_i' \beta_2^* + W_i' \delta_2^*$$

we have shown that this is identified

Now notice that

$$\begin{aligned} Y_i &= \alpha T_i + X_i' \beta + u_i \\ &= \alpha \left[T_i^f + \nu_i^* \right] + X_i' \beta + u_i \\ &= \alpha T_i^f + X_i' \beta_2 + (\alpha \nu_i^* + u_i) \end{aligned}$$

One could get a consistent estimate of α by regressing Y_i on X_i and T_i^f.

Two Stage Least Squares

In practice we don't know T_i^f but can get a consistent estimate of it from a fitted regression call this $\widehat{T}_i$

That is:

1. Regress T_i on X_i and Z_i, form the predicted value $\widehat{T}_i$
2. Regress Y on X_i and $\widehat{T}_i$

To run the second regression one needs to be able to vary $\widehat{T}_i$ separately from X_i which can only be done if there is a Z_i

It turns out that the sample analogue of this also is numerically identical to IV

Note that

$$\widehat{T} = Z^* \left(Z^{*'} Z^* \right)^{-1} Z^{*'} T$$

Then we can write

$$\widehat{B}_{2SLS} = \left(\begin{bmatrix} \widehat{T} & X \end{bmatrix}' \begin{bmatrix} \widehat{T} & X \end{bmatrix} \right)^{-1} \begin{bmatrix} \widehat{T} & X \end{bmatrix}' Y$$

However, note that we can write

$$X = Z^* \left(Z^{*'} Z^* \right)^{-1} Z^{*'} X$$

That is projecting X on (X, Z) and using it to predict X will be a perfect fit.

That means that

$$\left[\begin{array}{cc} \widehat{T} & X \end{array}\right] = Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}W$$

Then (in the exactly identified case) we can write

$$\begin{aligned}
\widehat{B}_{2SLS} =& \left(W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}W\right)^{-1} \times \\
& W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}Y \\
=& \left(W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}W\right)^{-1}W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}Y \\
=& \left(Z^{*'}W\right)^{-1}\left(Z^{*'}Z^{*}\right)\left(W'Z^{*}\right)^{-1}W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}Y \\
=& \left(Z^{*'}W\right)^{-1}\left(Z^{*'}Z^{*}\right)\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}Y \\
=& \left(Z^{*'}W\right)^{-1}Z^{*'}Y \\
=& \widehat{\beta}_{IV}
\end{aligned}$$

Experiments

Forget the X we just want to estimate the effect of some treatment on Y,

$$Y_i = \alpha_0 + \alpha T_i + u_i$$

Easiest way to solve the OLS problem is random assignment

So flip a coin, if heads then assign treatment (i.e. give the drug), if tails don't give treatment (i.e. give a placebo)

Then we can get consistent estimates of α by running a regression of Y_i on T_i which is equivalent to taking the difference in means between the treatment and control

The Problem

People are not like plants, just because we assign the treatment doesn't mean they take their medicine.

This is a bigger deal with social experiments like schools, job training programs, or the military

You can get the other problem as well-some controls might get the treatment.

This invalidates the experiment. As long as the decision not to take the treatment was not done at Random, we no longer have random assignment and T_i is correlated with u_i

For example, treatments who find jobs early are less likely to take job training

You can’t fix this problem easily because you don’t know which controls would have refused treatment had it been assigned

The Solution

Let R_i be a dummy variable for random assignment (the coin flip)

R_i is correlated with T_i by construction

R_i is uncorrelated with u_i by construction

Thus we can use R_i as an instrument for T_i

This is done routinely in randomized trials

To see why this works think about the ratio of reduced forms interpretation of IV

For simplicity assume no controls receive the treatment

$$
\begin{aligned}
E(T_i \mid R_i) &= Pr(T_i = 1 \mid R_i = 1)R_i \\
E(Y_i \mid R_i) &= E(\alpha_0 + \alpha T_i + u_i \mid R_i) \\
&= \alpha_0 + \alpha Pr(T_i = 1 \mid R_i = 1)R_i
\end{aligned}
$$

Thus the ratio of the coefficients is

$$
\frac{\alpha Pr(T_i = 1 \mid R_i = 1)}{Pr(T_i = 1 \mid R_i = 1)} = \alpha
$$

Returns to Schooling

This comes from the Card's chapter in the 1999 Handbook of Labor Economics

Lets assume that

$$\log(W_i) = a + bS_i + g(X_i) + \varepsilon_i$$

where W_i is wages, S_i is schooling, and X_i is experience.

We are worried about ability bias we want to use instrumental variables

A good instrument should have two qualities:

- It should be correlated with schooling
- It should be uncorrelated with ability (and other unobservables)

Many different things have been tried. Lets go through some of them

Family Background

If my parents earn quite a bit of money it should be easier for me to borrow for college

Also they might put more value on education

This should make me more likely to go

This has no direct effect on my income-Wisconsin did not ask how much education my Father had when they made my offer

But is family background likely to be uncorrelated with unobserved ability?

Closeness of College

If I have a college in my town it should be much easier to attend college

- I can live at home
- If I live on campus
  - I can travel to college easily
  - I can come home for meals and to get my clothes washed
- I can hang out with my friends from High school

But is this uncorrelated with unobserved ability?

Quarter of Birth

This is the most creative

Consider the following two aspects of the U.S. education system (this actually varies from state to state and across time but ignore that for now),

- People begin Kindergarten in the calendar year in which they turn 5
- You must stay in school until you are 16

Now consider kids who:

- Can't stand school and will leave as soon as possible
- Obey truancy law and school age starting law
- Are born on either December 31,1972 or January 1,1973

Those born on December 31 will

- turn 5 in the calendar year 1977 and will start school then (at age 4)
- will stop school on their 16th birthday which will be on Dec. 31, 1988
- thus they will stop school during the winter break of 11th grade

Those born on January 1 will

- turn 5 in the calendar year 1978 and will start school then (at age 5)
- will stop school on their 16th birthday which will be on Jan. 1, 1989
- thus they will stop school during the winter break of 10th grade

The instrument is a dummy variable for whether you are born on Dec. 31 or Jan 1

This is pretty cool:

- For reasons above it will be correlated with education
- No reason at all to believe that it is correlated with unobserved ability

The Fact that not everyone obeys perfectly is not problematic:

An instrument just needs to be correlated with schooling, it does not have to be perfectly correlated

In practice we can't just use the day as an instrument, use "quarter of birth" instead

Policy Changes

Another possibility is to use institutional features that affect schooling

Here often institutional features affect one group or one cohort rather than others

TABLE II

OLS AND IV ESTIMATES OF THE RETURN TO EDUCATION WITH INSTRUMENTS BASED ON FEATURES OF THE SCHOOL SYSTEM

Author	Sample and Instrument		Schooling Coefficients OLS	IV
1. Angrist and Krueger (1991)	1970 and 1980 Census Data, Men. Instruments are quarter of birth interacted with year of birth. Controls include quadratic in age and indicators for race, marital status, urban residence.	1920–29 cohort in 1970	0.070 (0.000)	0.101 (0.033)
		1930–39 cohort in 1980	0.063 (0.000)	0.060 (0.030)
		1940–49 cohort in 1980	0.052 (0.000)	0.078 (0.030)
2. Staiger and Stock (1997)	1980 Census, Men. Instruments are quarter of birth interacted with state and year of birth. Controls are same as in Angrist and Krueger, plus indicators for state of birth. LIML estimates.	1930–39 cohort in 1980	0.063 (0.000)	0.098 (0.015)
		1940–49 cohort in 1980	0.052 (0.000)	0.088 (0.018)
3. Kane and Rouse (1993)	NLS Class of 1972, Women. Instruments are tuition at 2 and 4-year state colleges and distance to nearest college. Controls include race, part-time status, experience. Note: Schooling measured in units of college credit equivalents.	Models without test score or parental education	0.080 (0.005)	0.091 (0.033)
		Models with test scores and parental education	0.063 (0.005)	0.094 (0.042)
4. Card (1995b)	NLS Young Men (1966 Cohort) Instrument is an indicator for a nearby 4-year college in 1966, or the interaction of this with parental education. Controls include race, experience (treated as endogenous), region, and parental education	Models that use college proximity as instrument (1976 earnings)	0.073 (0.004)	0.132 (0.049)
		Models that use college proximity × family background as instrument	—	0.097 (0.048)

5. Conneely and Uusitalo (1997)	Finnish men who served in the army in 1982, and were working full time in civilian jobs in 1994. Administrative earnings and education data. Instrument is living in university town in 1980. Controls include quadratic in experience and parental education and earnings.	Models that exclude parental education and earnings Models that include parental education and earnings	0.085 (0.001) 0.083 (0.001)	0.110 (0.024) 0.098 (0.035)
6. Harmon and Walker (1995)	British Family Expenditure Survey 1978–86 (men). Instruments are indicators for changes in the minimum school leaving age in 1947 and 1973. Controls include quadratic in age, survey year, and region.		0.061 (0.001)	0.153 (0.015)
7. Ichino and Winter-Ebmer (1998)	Austria: 1983 Census, men born before 1946. Germany: 1986 GSOEP for adult men. Instrument is indicator for 1930–35 cohort. (Second German IV also uses dummy for father's veteran status). Controls include age, unemployment rate at age 14, and father's education (Germany only). Education measure is dummy for high school or more.	Austrian Men German Men	0.518 (0.015) 0.289 (0.031)	0.947 (0.343) 0.590/0.708 (0.844) (0.279)
8. Lemieux and Card (1998)	Canadian Census, 1971 and 1981: French-speaking men in Quebec and English-speaking in Ontario. Instrument is dummy for Ontario men age 19–22 in 1946. Controls include full set of experience dummies and Quebec-specific cubic experience profile.	1971 Census: 1981 Census:	0.070 (0.002) 0.062 (0.001)	0.164 (0.053) 0.076 (0.022)
9. Meghir and Palme (1999)	Swedish Level of Living Survey (SLLS) data for men born 1945–55, with earnings in 1991, and Individual Statistics (IS) sample of men born in 1948 and 1953, with earnings in 1993. Instrument is dummy for attending "reformed" school system at age 13. Other controls include cohort, father's education, and county dummies. Models for IS data also include test scores at age 13.	SLLS Data (Years of education) IS Data (Dummy for 1–2 years of college relative to minimum schooling)	0.028 (0.007) 0.222 (0.020)	0.036 (0.021) 0.245 (0.082)

TABLE II—*Continued*

Author	Sample and Instrument		Schooling Coefficients OLS	IV
10. Maluccio (1997)	Bicol Multipurpose Survey (rural Philippines): male and female wage earners age 20–44 in 1994, whose families were interviewed in 1978. Instruments are distance to nearest high school and indicator for local private high school. Controls include quadratic in age and indicators for gender and residence in a rural community.	Models that do not control for selection of employment status or location	0.073 (0.011)	0.145 (0.041)
		Models with selection correction for location and employment status	0.063 (0.006)	0.113 (0.033)
11. Duflo (1999)	1995 Intercensal Survey of Indonesia: men born 1950–72. Instruments are interactions of birth year and targeted level of school building activity in region of birth. Other controls are dummies for year and region of birth and interactions of year of birth and child population in region of birth. Second IV adds controls for year of birth interacted with regional enrollment rate and presence of water and sanitation programs in region.	Model for hourly wage	0.078 (0.001)	0.064/0.091 (0.025) (0.023)
		Model for monthly wage with imputation for self-employed.	0.057 (0.003)	0.064/0.049 (0.017) (0.013)

Notes: See text for sources and more information on individual studies.

Consistently IV estimates are higher than OLS

Why?

- Bad Instruments
- Ability Bias
- Measurement Error
- Publication Bias
- Discount Rate Bias

Measurement Error

Another way people use instruments is for measurement error

In the classic model suppose we get rid of $X's$ so we want to measure the effect of T on Y.

$$Y_i = \beta_0 + \alpha T_i + u_i$$

and lets not worry about endogeneity so assume that $cov(T_i, u_i) = 0$.

The problem is that I don't get to observe T_i, I only get to observe a noisy version of it:

$$\tau_{1i} = T_i + \xi_i$$

where ξ_i is i.i.d measurement error with variance σ_ξ^2

What happens if I run the regression on τ_{1i} instead of T_i?

$$\begin{aligned}\widehat{\alpha} &\approx \frac{Cov\left(\tau_{1i}, Y_i\right)}{Var(\tau_{1i})} \\ &= \frac{Cov\left(T_i + \xi_i, \beta_0 + \alpha T_i + u_i\right)}{Var(T_i + \xi_i)} \\ &= \alpha \frac{Var\left(T_i\right)}{Var(T_i) + \sigma_\xi^2}\end{aligned}$$

Now suppose we have another measure of T_i,

$$\tau_{2i} = T_i + \eta_i$$

where η_i is uncorrelated with everything else in the model.

Note that we can write

$$Y_i = \beta_0 + \alpha\tau_{1i} + u_i - \alpha\xi_i.$$

You can see the problem with OLS:, τ_{1i} is correlated with ξ_i

However, IV gives us a solution.

τ_{2i} is correlated with τ_{1i} (through T_i),but uncorrelated with ξ_i(and u_i) so we can use one measure as an instrument for the other.

Twins

(Here we will think about both measurement error and fixed effect approaches)

$$\log(w_{if}) = \alpha_f + \beta S_{if} + u_{if}$$

The problem is that α_f is correlated with S_{if}

We can solve by differencing

$$E\left(\log(w_{if}) - \log(w_{jf})\right) = \beta E\left(S_{if} - S_{jf}\right)$$

Use this to get consistent estimates of β

The problem here is that a little measurement error can screw up things quite a bit because the variance of $S_{if} - S_{jf}$ is small.

A solution of this is to get two measures on schooling

- Ask me about my schooling
- Also ask my brother about my schooling
- do the same think for my brother's schooling

This gives us two different measure of $S_{if} - S_{jf}$.

Use one as an instrument for the other

Table 6
Cross-sectional and within-family differenced estimates of the return to education for twins[a]

Author	Sample and specification		Cross-sectional OLS	Differenced	
				OLS	IV
1. Ashenfelter and Rouse (1998)	1991–1993 Princeton Twins Survey. Identical male and female twins. Controls Basic controls include quadratic	Basic	0.110 (0.010)	0.070 (0.019)	0.088 (0.025)
	in age, gender and race. Added controls include tenure, marital status and union status.	Basic + added controls	0.113 (0.010)	0.078 (0.018)	0.100 (0.023)
2. Rouse (1997)	1991–1995 Princeton Twins Survey. Identical male and female twins. Basic controls as above.		0.105 (0.008)	0.075 (0.017)	0.110 (0.023)
3. Miller et al. (1995)	Australian Twins Register. Identical and fraternal twins. Controls include quadratic in	Identical twins	0.064 (0.002)	0.025 (0.005)	0.048 (0.010)
	age, gender, marital status. Incomes imputed from occupation	Fraternal twins	0.066 (0.002)	0.045 (0.005)	0.074 (0.008)
4. Behrman et al. (1994)	NAS-NRC white male twins born 1917–1927, plus male twins born 1936–1955 from Minnesota Twins	Identical twins	0.071 (0.002)	0.035 (0.005)	0.056 –
	Registry. Controls include quadratic in age[b]	Fraternal twins	0.073 (0.003)	0.057 (0.005)	0.071 –
5. Isacsson (1997)	Swedish same-sex twins with both administrative and survey measures of schooling.	Identical twins	0.049 (0.002)	0.023 (0.004)	0.024 (0.008)
	Controls include sex, marital status, quadratic in age, and residence in a large city[c]	Fraternal twins	0.051 (0.002)	0.040 (0.003)	0.054 (0.006)

Overidentification

What happens when we have more than one instrument?

Lets think about a general case in which Z_i is multidimensional

- Let K_Z be the dimension of Z_i^*
- Let K_X denote the dimension of W_i

Now we have more equations then parameters so we can no longer solve

$$0 = Z^{*'} \left(Y - W\widehat{B} \right)$$

This gives us K_Z equations in K_X unknowns.

A simple solution is follow GMM and weight the moments by some $K_Z \times K_Z$ weighting matrix Ω and then minimize

$$\left[Z^{*'}\left(Y - W\widehat{B}\right)\right]' \Omega \left[Z^{*'}\left(Y - W\widehat{B}\right)\right]$$

which gives

$$-2W'Z^{*}\Omega Z^{*'}\left(Y - W\widehat{B}\right) = 0$$

(notice that in the exactly identified case $W'Z^{*}\Omega$ drops out)

We can solve directly for our estimator

$$\widehat{B}_{GMM} = \left(W'Z^{*}\Omega Z^{*'}W\right)^{-1} W'Z^{*}\Omega Z^{*'}Y$$

Two staged least squares is a special case of this:

$$\widehat{B}_{2SLS} = \left(W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}W\right)^{-1}W'Z^{*}\left(Z^{*'}Z^{*}\right)^{-1}Z^{*'}Y$$

Notice that this is the same as $\widehat{B}_{GMM}$ when

$$\Omega = \left(Z^{*'}Z^{*}\right)^{-1}$$

Consistency

$$
\begin{aligned}
\widehat{B}_{GMM} =& \left(\frac{1}{N}W'Z^{*}\Omega\frac{1}{N}Z^{*'}W\right)^{-1}\frac{1}{N}W'Z^{*}\Omega\frac{1}{N}Z^{*'}\left(WB+U\right)\\
=& B+\left(\frac{1}{N}W'Z^{*}\Omega\frac{1}{N}Z^{*'}W\right)^{-1}\frac{1}{N}W'Z^{*}\Omega\frac{1}{N}Z^{*'}U\\
\approx& B+\left(E\left(W_iZ_i^{*'}\right)\Omega E\left(Z_i^{*}W_i^{'}\right)\right)^{-1}E\left(W_iZ_i^{*'}\right)\Omega E(Z_iu_i)\\
=& B
\end{aligned}
$$

Inference

$$\sqrt{N}\left(\widehat{B}-B\right)=\left(\frac{1}{N}W'Z^{*}\Omega\frac{1}{N}Z^{*'}W\right)^{-1}\frac{1}{N}W'Z^{*}\Omega\frac{1}{\sqrt{N}}Z^{*'}U$$

Using a standard central limit theorem with i.i.d. data

$$\begin{aligned}\frac{1}{\sqrt{N}}Z^{*'}U &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}Z_i^{*}u_i\\ &\approx N\left(0,E\left(u_i^2Z_i^{*}Z_i^{*'}\right)\right)\end{aligned}$$

Thus

$$\sqrt{N}\left(\widehat{B}-B\right)\approx N\left(0,A'VA\right)$$

with

$$\begin{aligned}V &= E\left(W_iZ_i^{*'}\right)\Omega E\left(u_i^2Z_i^{*}Z_i^{*'}\right)\Omega E\left(Z_i^{*}W_i'\right)\\ A &= \left(E\left(W_iZ_i^{*'}\right)\Omega E\left(Z_i^{*}W_i'\right)\right)^{-1}\end{aligned}$$

From GMM results we know that the efficient weighting matrix is

$$\Omega = E\left(u_i^2 Z_i^* Z_i^{*'}\right)^{-1}$$

in which case the Covariance matrix simplifies to

$$\left(E\left(W_i Z_i^{*'}\right) E\left(u_i^2 Z_i^* Z_i^{*'}\right)^{-1} E\left(Z_i^* W_i'\right)\right)^{-1}$$

This also means that under homoskedasticity two staged least squares is efficient.

Overidentification Tests

Lets think about testing in the following way.

Suppose we have two instruments so that we have three sets of moment conditions

$$0 = Z_1' \left(Y - T\widehat{\alpha} - X\widehat{\beta} \right)$$
$$0 = Z_2' \left(Y - T\widehat{\alpha} - X\widehat{\beta} \right)$$
$$0 = X' \left(Y - T\widehat{\alpha} - X\widehat{\beta} \right)$$

As before we can use partitioned regression to deal with the X's and then write the first two moment equations as

$$0 = \widetilde{Z}_1' \left(\widetilde{Y} - \widetilde{T}\widehat{\alpha} \right)$$
$$0 = \widetilde{Z}_2' \left(\widetilde{Y} - \widetilde{T}\widehat{\alpha} \right)$$

Thus the way I see the overidentification test is whether we can find an $\widehat{\alpha}$ that solves both equations.

That is let

$$\widehat{\alpha}_1 = \frac{\widetilde{Z}_1'\widetilde{Y}}{\widetilde{Z}_2'\widetilde{T}}$$

$$\widehat{\alpha}_2 = \frac{\widetilde{Z}_1'\widetilde{Y}}{\widetilde{Z}_2'\widetilde{T}}$$

If

$$\widehat{\alpha}_1 \approx \widehat{\alpha}_2$$

then the test will not reject the model, otherwise it will

For this reason I am not a big fan of overidentification tests:

- If you have two crappy instruments with roughly the same bias you will fail to reject
- Why not just estimate $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ and look at them? It seems to me that you learn much more from that than a simple F-statistic