Instrumental Variables

Department of Economics
University of Wisconsin-Madison

September 30, 2018

Lets think about estimation of the model

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

α measures the causal effect of T_i on Y_i.

Our goal is to estimate α

A classic approach to this problem is Instrumental Variables-I know you have seen this before but it is an important enough concept that I want to cover it again (with perhaps more emphasis about how to think about it when doing applied work)

I want to think about three completely different approaches for estimating α

The first is the GMM approach and the second two will come from a Simultaneous Equations framework

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

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

and we continue to assume that

$$E(X_i u_i) = 0$$

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

Define

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

Then we can write this as a moment equation

$$E\left[Z_i^*\left(Y_i - X_i^{*\prime}B\right)\right] = 0$$

The sample analogue is

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

which we can solve as

$$\widehat{B}_{IV} \equiv \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^* X_i^{*\prime}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N} Z_i^* Y_i$$

And that is IV.

Consistency of IV

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

since

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

So (assuming iid sampling) this only took two assumptions. The moment conditions and the fact that you can invert

$$E\left(Z_i^* X_i^{*'}\right)$$

As we will discuss this assumption is typically a bigger deal than in OLS

Asymptotic Variance

Multiply by $\sqrt{N}$ then

$$\sqrt{N}\left(\widehat{B}_{IV}-B\right)=\left(\frac{1}{N}\sum_i Z_i^* X_i^{*\prime}\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_i Z_i^* u_i\right]$$

The CLT on term in brackets says

$$\frac{1}{\sqrt{N}}\sum_i Z_i^* u_i \sim N\left(0, E\left[Z_i^* Z_i^{*\prime} u_i^2\right]\right)$$

so

$$\sqrt{N}\left(\widehat{B}_{IV}-B\right)\sim N\left(0,\left(E\left[Z_i^* X_i^{*\prime}\right]\right)^{-1} E\left[Z_i^* Z_i^{*\prime} u_i^2\right]\left(E\left[X_i^* Z_i^{*\prime}\right]\right)^{-1}\right)$$

We approximate this by

$$E\left(Z_i^* X_i^{*'}\right) \approx \frac{1}{N} \sum_{i=1}^{N} Z_i^* X_i^{*'}$$

$$E\left(u_i^2 Z_i^* Z_i^{*'}\right) \approx \sum_{i=1}^{N} \widehat{u}_i^2 Z_i^* Z_i^{*'}$$

I want to think about the asymptotic bias of IV

Before that I want to take a detour and think about partitioned regression which will turn out to be really useful for this calculation

Partitioned Regression

Think about the standard regression model (in large matrix notation)

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

We will define

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

First notice that M_2 is symmetric.

Two more facts about M_2

Fact 1: M_2 is idempotent

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

Fact 2: $M_2 Y$ is the 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 the regression coefficients be be $\widehat{g}$ and generically let $\widetilde{Y}$ be residuals from a regression of Y on X_2 so that

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

An important special case of this is that if I regress something on itself, the residuals are all zero

That is

$$\begin{aligned} M_2 X_2 &= X_2 - X_2 (X_2' X_2)^{-1} X_2' X_2 \\ &= 0 \end{aligned}$$

as should be the case

If I think of the normal 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 $\widetilde{X}_1$

Since I derived this from the OLS normal equations, 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 the sample analogue of the moment equation as:

$$0 = Z'\left(Y - T\widehat{\alpha}_{IV} - X\widehat{\beta}_{IV}\right)$$
$$0 = X'\left(Y - T\widehat{\alpha}_{IV} - X\widehat{\beta}_{IV}\right)$$

The second can be solved as

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

Now plug this into the first

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

so

$$
\begin{aligned}
\widehat{\alpha}_{IV} &= \left(Z'M_X T\right)^{-1} Z'M_X Y \\
&= \frac{\widetilde{Z}'\widetilde{Y}}{\widetilde{Z}'\widetilde{T}} \\
&\approx \frac{cov(\widetilde{Z}_i, \widetilde{Y}_i)}{cov(\widetilde{Z}_i, \widetilde{T}_i)}
\end{aligned}
$$

(This last expression assumes that there is an intercept in the model. If not it would be expected values rather than covariances, but covariances make things easier to interpret-at least to me)

To see consistency from this perspective 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

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

This formula is helpful. In order for the model to be consistent you need

- $cov(\widetilde{Z}_i, \widetilde{u}_i) = 0$
- $cov(\widetilde{Z}_i, \widetilde{T}_i) \neq 0$

But more generally for the asymptotic 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

This means that in practice 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 asymptotic bias.

As a concrete example lets compare IV to OLS.

OLS is really just a special case of IV with $Z_i = T_i$

Then we get

$$\widehat{\alpha}_{IV} \approx \alpha + \frac{cov(\widetilde{Z}_i, \widetilde{u}_i)}{cov(\widetilde{Z}_i, \widetilde{T}_i)}$$
$$\widehat{\alpha}_{OLS} \approx \alpha + \frac{cov(\widetilde{T}_i, \widetilde{u}_i)}{cov(\widetilde{T}_i, \widetilde{T}_i)}$$

If cov$(\widetilde{Z}_i, \widetilde{u}_i) = 0$ and cov$(\widetilde{T}_i, \widetilde{u}_i) \neq 0$ then IV is consistent and OLS is not

However, cov$(\widetilde{Z}_i, \widetilde{u}_i) <$ cov$(\widetilde{T}_i, \widetilde{u}_i)$ does not guarantee less bias because it also depends on cov$(\widetilde{Z}_i, \widetilde{T}_i) = 0$ and cov$(\widetilde{T}_i, \widetilde{T}_i) \neq 0$

Simultaneous equations

The second and third way to see IV comes from the simultaneous equations framework

$$Y_i = \alpha T_i + X_i'\beta + u_i$$
$$T_i = \rho Y_i + X_i'\gamma + Z_i\delta + \nu_i$$

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 (we are still allowing this as a possibility as some of the γ coefficients could be zero)

The model with $\rho = 0$ simplifies things, but lets focus on what happens when it isn’t

We assume that

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

but notice that if $\rho \neq 0$, then 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

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

Note that $E(\nu_i^* \mid X_i, Z_i) = 0$, so one can obtain a consistent estimate of β_2^* and δ_2^* by regressing T_i on X_i and Z_i.

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

To me this is the classic definition of reduced form (you need to have a structural model)

We can obtain a consistent estimate of α as long as we have an exclusion restriction

That is we need some Z_i that affects T_i but not Y_i directly

I want to show this in two different ways

Method 1

We can also solve for the reduced form for Y_i

$$
\begin{aligned}
Y_i &= \alpha T_i + X_i'\beta + u_i \\
&= X_i \frac{\alpha\gamma + \beta}{1 - \alpha\rho} + Z_i \frac{\alpha\delta}{1 - \alpha\rho} + \frac{\alpha\nu_i + u_i}{1 - \alpha\rho} \\
&= X_i\beta_1^* + Z_i\delta_1^* + u_i^*
\end{aligned}
$$

with

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

Like the other reduced form, we can get a consistent estimate of β_1^* and δ_1^* by regressing Y_i on X_i and Z_i.

Notice then that

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

So we can get a consistent estimate of α simply by taking the ratio of the reduced form coefficients

In the exactly identified case (i.e. one Z_i), this is numerically identical to IV.

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

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

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

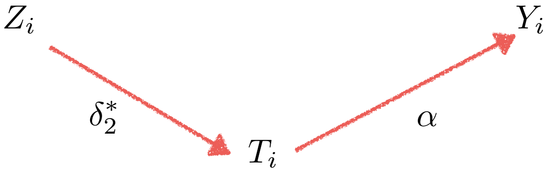
so

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

This is just math-it does not require that the “Structural equation" determining T_i be correct

It also gives another interpretation of IV:

- δ_2^* is the causal effect of Z_i on T_i
- δ_1^* is the causal effect of Z_i on Y_i-it only operates through T_i

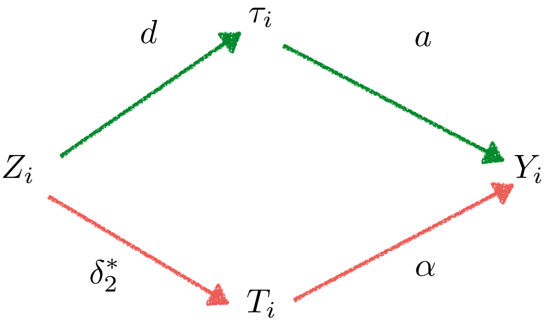


If we increase Z_i by one unit this leads T_i to increase by δ_2^* units which causes Y_i to increase by $\delta_2^* \alpha$ units

Thus the causal effect of Z_i on Y_i is $\delta_2^* \alpha$ units

This illustrates another important way to think about an instrument: the key assumption is that T_i is the only channel through which Z_i influences Y_i

Suppose there was another



Then the causal effect of Z_i on Y_i would be $\delta_2^* \alpha + da$ and IV would be

$$\frac{\delta_2^* \alpha + da}{\delta_2^*}$$

Method 2

Define

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

and suppose that T_i^f were known to the econometrician

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 the fitted values of a reduced form regression call this $\widehat{T}_i$ (it is crucial that the reduced form gives us consistent estimates of β_2^* and δ_2^*)

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 2SLS is also is numerically identical to IV (with 1 instrument)

Note that

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

so

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

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(using notation from earlier) that

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

Then (in the exactly identified case) we can write

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

3 Interpretations

Thus with 1 instrument we have 3 equivalent ways to derive IV:

1. GMM estimator or $(Z'T)^{-1}Z'Y$
2. Ratio of reduced form estimates-rescaling the reduced form
3. 2SLS-direct effect of fitted model

With more than one instrument only one of these procedures works-we'll worry about that later

Examples

There are three main reasons people use IV

1. Simultaneity bias: $\rho \neq 0$
2. Omitted Variable bias : There are unobservables that contribute to u_i that are correlated with T_i
3. Measurement Error: We do not observe T_i perfectly

While the first is the original reason for IV, in practice omitted variable bias is typically the biggest concern

A classic (perhaps the classic) example is the returns to schooling.

Returns to Schooling

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

Lets assume that

$$\log(W_i) = \alpha S_i + X_i'\beta + \varepsilon_i$$

where W_i is wages, S_i is schooling, and X_i is other stuff

The biggest problem is unobserved ability

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 unobserved 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	0.085 (0.001)	0.110 (0.024)
		Models that include parental education and earnings	0.083 (0.001)	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	0.518 (0.015)	0.947 (0.343)
		German Men	0.289 (0.031)	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:	0.070 (0.002)	0.164 (0.053)
		1981 Census:	0.062 (0.001)	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)	0.028 (0.007)	0.036 (0.021)
		IS Data (Dummy for 1–2 years of college relative to minimum schooling)	0.222 (0.020)	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

Heterogeneous Treatment Effects/Discount Rate Bias

We have been thinking about IV so far in a regression type framework, lets go back to our treatment effect framework

We can write

$$\begin{aligned} Y_i &\equiv T_i Y_{1i} + (1 - T_i)\, Y_{0i} \\ &= T_i\,(Y_{1i} - Y_{0i}) + Y_{0i} \\ &\equiv \beta_0 + \Delta_i T_i + \varepsilon_i \end{aligned}$$

(where $\beta_0 = E(Y_{0i})$ and $\varepsilon_i = Y_{0i} - \beta_0$)

Assume that we have an instrument Z_i that is correlated with T_i but not with Δ_i or ε_i (or equivalently Y_{0i} or Y_{1i})

Does IV estimate the ATE?

Lets abstract from other regressors

IV yields

$$\begin{aligned}
\widehat{\alpha} &\approx \frac{Cov(Z_i, Y_i)}{Cov(Z_i, T_i)} \\
&= \frac{Cov(Z_i, \varepsilon_i + \Delta_i T_i)}{Cov(Z_i, T_i)} \\
&= \frac{Cov(Z_i, \varepsilon_i)}{Cov(Z_i, T_i)} + \frac{Cov(Z_i, \Delta_i T_i)}{Cov(Z_i, T_i)} \\
&= \frac{Cov(Z_i, \Delta_i T_i)}{Cov(Z_i, T_i)}.
\end{aligned}$$

In the case in which treatment effects are constant so that $\Delta_i = \alpha$ for everyone

$$\frac{Cov(Z_i, \alpha T_i)}{Cov(Z_i, T_i)} = \alpha$$

However, more generally IV does not converge to the Average treatment effect

Imbens and Angrist (1994) consider the case in which there are not constant treatment effects

They consider a simple version of the model in which Z_i takes on 2 values, call them 0 and 1 for simplicity and without loss of generality assume that

$$Pr(T_i = 1 \mid Z_i = 1) > Pr(T_i = 1 \mid Z_i = 0)$$

There are 4 different types of people those for whom $T_i = 1$ when:

$\clubsuit$: $Z_i = 1, Z_i = 0$

$\diamondsuit$: Never

$\heartsuit$: $Z_i = 1$ only

$\spadesuit$: $Z_i = 0$ only

Imbens and Angrist's monotonicity rules out 4 as a possibility

Let $\mu_{\clubsuit}, \mu_{\diamondsuit}$, and $\mu_{\heartsuit}$ represent the sample proportions of the three groups and S_i an indicator of the type

We showed above that

$$\begin{aligned}\widehat{\alpha} \approx & \frac{Cov(Z_i, \Delta_i T_i)}{Cov(Z_i, T_i)} \\ = & \frac{E(\Delta_i T_i Z_i) - E\left(\Delta_i T_i\right) E\left(Z_i\right)}{E(T_i Z_i) - E\left(T_i\right) E\left(Z_i\right)}\end{aligned}$$

Let P_z denote the probability that $Z_i = 1$.

Lets look at the pieces

First the numerator

$$
\begin{aligned}
& E(\Delta_i T_i Z_i) - E\left(\Delta_i T_i\right) E\left(Z_i\right) \\
=& \left[P_z E(\Delta_i T_i \mid Z_i = 1) + (1 - P_z)\, 0\right] - E\left(\Delta_i T_i\right) P_z \\
=& P_z E(\Delta_i T_i \mid Z_i = 1) \\
& - \left[P_z E(\Delta_i T_i \mid Z_i = 1) + (1 - P_z)\, E(\Delta_i T_i \mid Z_i = 0)\right] P_z \\
=& P_z (1 - P_z) \left[E(\Delta_i T_i \mid Z_i = 1) - E(\Delta_i T_i \mid Z_i = 0)\right] \\
=& P_z (1 - P_z) \left[\mu_{\clubsuit} E(\Delta_i \mid S_i = \clubsuit) + \mu_{\diamondsuit} 0 + \mu_{\heartsuit} E(\Delta_i \mid S_i = \heartsuit) \right. \\
& \qquad\qquad \left. - \mu_{\clubsuit} E(\Delta_i \mid S_i = \clubsuit) - \mu_{\diamondsuit} 0 - \mu_{\heartsuit} 0\right] \\
=& P_z (1 - P_z) \mu_{\heartsuit} E(\Delta_i \mid S_i = \heartsuit)
\end{aligned}
$$

Next consider the denominator (which is really a special case of above with $\Delta_i = 1$)

$$
\begin{aligned}
& E(T_i Z_i) - E(T_i) E(Z_i) \\
=& P_z E(T_i \mid Z_i = 1) - E(T_i) P_z \\
=& P_z E(T_i \mid Z_i = 1) \\
& - \left[P_z E(T_i \mid Z_i = 1) + (1 - P_z) E(T_i \mid Z_i = 0)\right] P_z \\
=& P_z (1 - P_z) \left[E(T_i \mid Z_i = 1) - E(T_i \mid Z_i = 0)\right] \\
=& P_z (1 - P_z) \left[\mu_{\clubsuit} + \mu_{\heartsuit} - \mu_{\clubsuit}\right] \\
=& P_z (1 - P_z) \mu_{\heartsuit}
\end{aligned}
$$

Thus

$$\begin{aligned}\widehat{\alpha} \approx & \frac{P_z(1-P_z)\mu_{\heartsuit} E(\Delta_i \mid S_i = \heartsuit)}{P_z(1-P_z)\mu_{\heartsuit}} \\ = & E(\Delta_i \mid S_i = \heartsuit)\end{aligned}$$

They call this the local average treatment effect (LATE)

The Colonial Origins of Comparative Development: An Empirical Investigation

by Acemoglu, Johnson, and Robinson

Time for a non-education example-a very important paper on an extremely important issue

Their goal is to estimate the effects of “institutions” on Economic Development

The idea is that countries with better institutions should develop more

Specifically they want to estimate their equation:

$$\log(y_i) = \mu + \alpha R_i + X_i' \gamma + \varepsilon_i$$

where

- y_i is GDP per capita
- R_i is “protection against expropriation"
- X_i is other stuff

Lets take a preliminary look

TABLE 2—OLS REGRESSIONS

	Whole world (1)	Base sample (2)	Whole world (3)	Whole world (4)	Base sample (5)	Base sample (6)	Whole world (7)	Base sample (8)
	Dependent variable is log GDP per capita in 1995						Dependent variable is log output per worker in 1988	
Average protection against expropriation risk, 1985–1995	0.54 (0.04)	0.52 (0.06)	0.47 (0.06)	0.43 (0.05)	0.47 (0.06)	0.41 (0.06)	0.45 (0.04)	0.46 (0.06)
Latitude			0.89 (0.49)	0.37 (0.51)	1.60 (0.70)	0.92 (0.63)		
Asia dummy				−0.62 (0.19)		−0.60 (0.23)		
Africa dummy				−1.00 (0.15)		−0.90 (0.17)		
"Other" continent dummy				−0.25 (0.20)		−0.04 (0.32)		
R^2	0.62	0.54	0.63	0.73	0.56	0.69	0.55	0.49
Number of observations	110	64	110	110	64	64	108	61

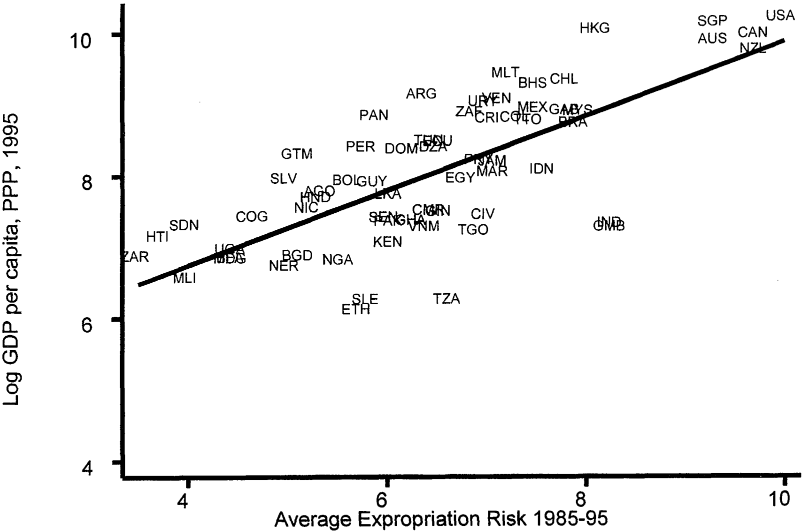


FIGURE 2. OLS RELATIONSHIP BETWEEN EXPROPRIATION RISK AND INCOME

The problem is that there is no reason to believe institutions are set at random

They suggest using mortalities of settlers as an instrument with the following argument

(potential) settler mortality $\Rightarrow$ settlements

$\Rightarrow$ early institutions $\Rightarrow$ current institutions

$\Rightarrow$ current performance.

Lets see what they find

TABLE 1—DESCRIPTIVE STATISTICS

	Whole world	Base sample	By quartiles of mortality (1)	(2)	(3)	(4)
Log GDP per capita (PPP) in 1995	8.3 (1.1)	8.05 (1.1)	8.9	8.4	7.73	7.2
Log output per worker in 1988 (with level of United States normalized to 1)	−1.70 (1.1)	−1.93 (1.0)	−1.03	−1.46	−2.20	−3.03
Average protection against expropriation risk, 1985–1995	7 (1.8)	6.5 (1.5)	7.9	6.5	6	5.9
Constraint on executive in 1990	3.6 (2.3)	4 (2.3)	5.3	5.1	3.3	2.3
Constraint on executive in 1900	1.9 (1.8)	2.3 (2.1)	3.7	3.4	1.1	1
Constraint on executive in first year of independence	3.6 (2.4)	3.3 (2.4)	4.8	2.4	3.1	3.4
Democracy in 1900	1.1 (2.6)	1.6 (3.0)	3.9	2.8	0.19	0
European settlements in 1900	0.31 (0.4)	0.16 (0.3)	0.32	0.26	0.08	0.005
Log European settler mortality	n.a.	4.7 (1.1)	3.0	4.3	4.9	6.3
Number of observations	163	64	14	18	17	15

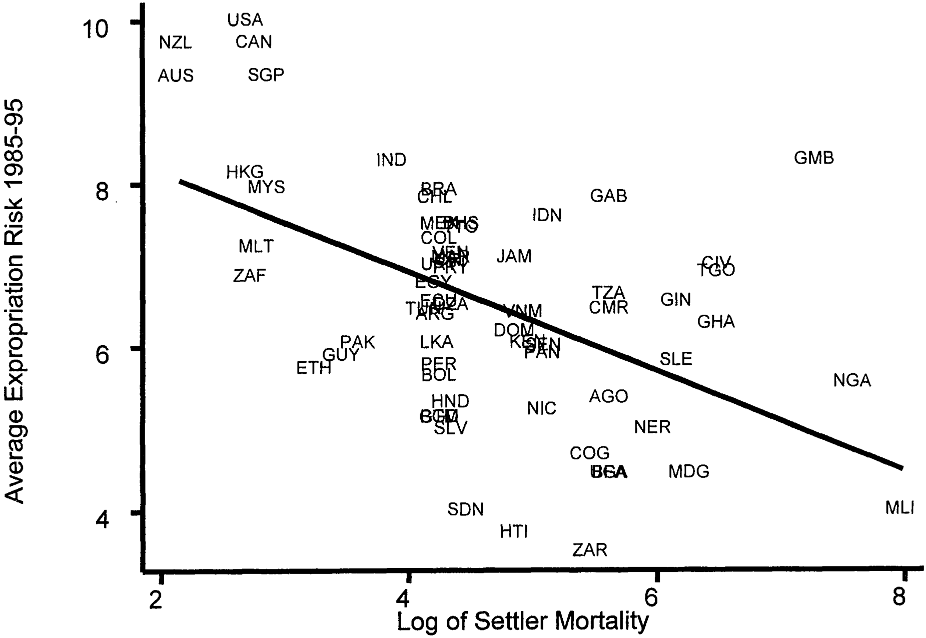
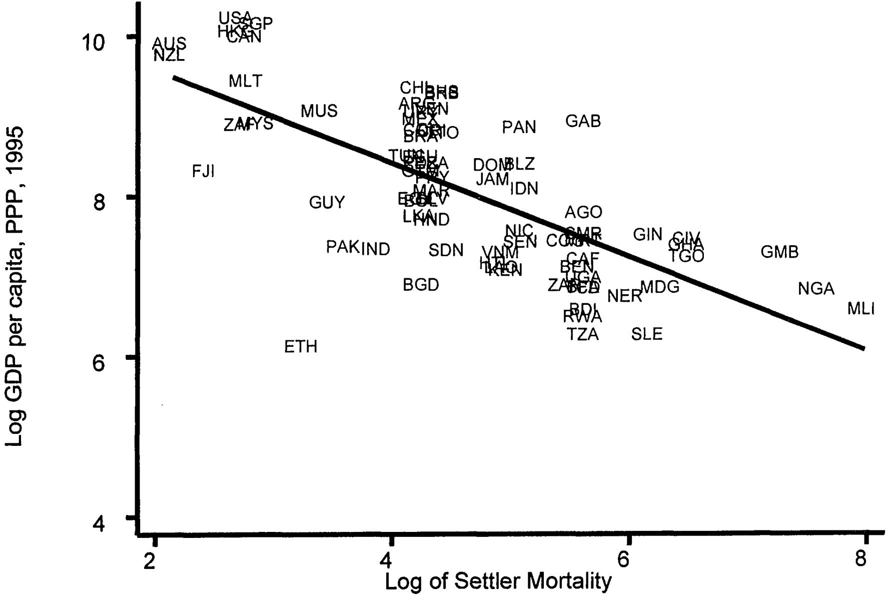

Average Expropriation Risk 1985-95
10
8
6
4
USA
NZL
CAN
AUS
SGP
HKG
MYS
IND
GMB
BRA
CHL
GAB
IDN
COL
MLT
JAM
ZAF
CIV
TGO
EGY
TZA
CMR
GIN
ARG
VNM
GHA
DOM
PAK
LKA
GUY
PAN
SLE
ETH
PER
BOL
NGA
HND
NIC
AGO
SLV
NER
COG
MDG
SDN
MLI
HTI
ZAR
2
4
6
8
Log of Settler Mortality

TABLE 3—DETERMINANTS OF INSTITUTIONS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A	Dependent Variable Is Average Protection Against Expropriation Risk in 1985–1995									
Constraint on executive in 1900	0.32	0.26								
	(0.08)	(0.09)								
Democracy in 1900			0.24	0.21						
			(0.06)	(0.07)						
Constraint on executive in first year of independence					0.25	0.22				
					(0.08)	(0.08)				
European settlements in 1900							3.20	3.00		
							(0.61)	(0.78)		
Log European settler mortality									−0.61	−0.51
									(0.13)	(0.14)
Latitude		2.20		1.60		2.70		0.58		2.00
		(1.40)		(1.50)		(1.40)		(1.51)		(1.34)
R^2	0.2	0.23	0.24	0.25	0.19	0.24	0.3	0.3	0.27	0.3
Number of observations	63	63	62	62	63	63	66	66	64	64

Panel B	Dependent Variable Is Constraint on Executive in 1900				Dependent Variable Is Democracy in 1900				Dependent Variable Is European Settlements in 1900	
European settlements in 1900	5.50	5.40			8.60	8.10				
	(0.73)	(0.93)			(0.90)	(1.20)				
Log European settler mortality			−0.82	−0.65			−1.22	−0.88	−0.11	−0.07
			(0.17)	(0.18)			(0.24)	(0.25)	(0.02)	(0.02)
Latitude		0.33		3.60		1.60		7.60		0.87
		(1.80)		(1.70)		(2.30)		(2.40)		(0.19)
R^2	0.46	0.46	0.25	0.29	0.57	0.57	0.28	0.37	0.31	0.47
Number of observations	70	70	75	75	67	67	68	68	73	73



Log GDP per capita, PPP, 1995
Log of Settler Mortality
10
8
6
4
2
4
6
8
USA
SGP
CAN
AUS
NZL
MLT
MUS
MYS
FJI
GUY
PAK
IND
ETH
BGD
SDN
CHL
BRA
PAN
GAB
JAM
IDN
AGO
NIC
SEN
GIN
TGO
GMB
NGA
MLI
CAF
NER
MDG
SLE
TZA
HND

TABLE 4—IV REGRESSIONS OF LOG GDP PER CAPITA

	Base sample (1)	Base sample (2)	Base sample without Neo-Europes (3)	Base sample without Neo-Europes (4)	Base sample without Africa (5)	Base sample without Africa (6)	Base sample with continent dummies (7)	Base sample with continent dummies (8)	Base sample, dependen variable i log outpu per worke (9)
				Panel A: Two-Stage Least Squares					
Average protection against expropriation risk 1985–1995	0.94	1.00	1.28	1.21	0.58	0.58	0.98	1.10	0.98
	(0.16)	(0.22)	(0.36)	(0.35)	(0.10)	(0.12)	(0.30)	(0.46)	(0.17)
Latitude		−0.65		0.94		0.04		−1.20	
		(1.34)		(1.46)		(0.84)		(1.8)	
Asia dummy							−0.92	−1.10	
							(0.40)	(0.52)	
Africa dummy							−0.46	−0.44	
							(0.36)	(0.42)	
"Other" continent dummy							−0.94	−0.99	
							(0.85)	(1.0)	
				Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995					
Log European settler mortality	−0.61	−0.51	−0.39	−0.39	−1.20	−1.10	−0.43	−0.34	−0.63
	(0.13)	(0.14)	(0.13)	(0.14)	(0.22)	(0.24)	(0.17)	(0.18)	(0.13)
Latitude		2.00		−0.11		0.99		2.00	
		(1.34)		(1.50)		(1.43)		(1.40)	
Asia dummy							0.33	0.47	
							(0.49)	(0.50)	
Africa dummy							−0.27	−0.26	
							(0.41)	(0.41)	
"Other" continent dummy							1.24	1.1	
							(0.84)	(0.84)	
R^2	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33	0.28

Measurement Error

Another reason people use instruments is for measurement error

In the classic model lets 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 other issues so assume that $cov(T_i, u_i) = 0.$

The complication 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?

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

Lets rewrite the model as

$$\begin{aligned} Y_i &= \beta_0 + \alpha T_i + u_i \\ &= \beta_0 + \alpha T_i + \alpha \xi_i + u_i - \alpha \xi_i \\ &= \beta_0 + \alpha \tau_{1i} + (u_i - \alpha \xi_i). \end{aligned}$$

You can see the problem with OLS: $\tau_{1i} = T_i + \xi_i$ is correlated with $(u_i - \alpha \xi_i)$

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.

Using this as an instrument gives us a solution.

τ_{2i} is correlated with τ_{1i} (through T_i), but uncorrelated with $(u_i - \alpha\xi_i)$ so we can use τ_{2i} as an instrument for τ_{1i}.

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

$$\log(w_{if}) = \theta_f + \alpha S_{if} + u_{if}$$

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

We can solve by differencing

$$\Delta \log(w_f) = \alpha \Delta S_f + \Delta u_f$$

if ΔS_f is uncorrelated with Δu_f, then we can use this to get consistent estimates of α

The problem here is that a little measurement error can screw up things quite a bit because the variance of ΔS_f 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 ΔS_{if}.

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 X_i^*

Now we have more equations then parameters so we can no longer solve for $\widehat{B}$ using

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

because this gives us K_Z equations in K_X unknowns.

We think of this as a GMM estimator and weight the moments by some $K_Z \times K_Z$ weighting matrix W and then minimize

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

which gives

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

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

We can solve directly for our estimator

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

Two staged least squares is a special case of this:

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

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

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

From GMM results we know that the efficient weighting matrix is

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

in which case the Covariance matrix simplifies to

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

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

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

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
- with random coefficients even if both instruments are valid you might get different parameters because they estimate different LATEs
- 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