Instrumental Variables, Part I

PP346

Harris School of Public Policy
University of Chicago

Topic 3

Instrumental Variables Estimation

Homogeneous treatment effects

Suppose you have the following model

$$Y_i = \delta D_i + X_i\beta + \varepsilon_i$$

where Y_i is the outcome, D_i is the variable of interest, X_i is a vector of covariates including a constant, β is a vector of nuisance parameters, and ε_i is a regression error with mean zero and $cov(X_i, \varepsilon_i) = 0$.

- If $cov(D_i, \varepsilon_i) \neq 0$, then D_i is endogenous and OLS is inconsistent.
- Consider an example where the outcome is earnings, and D_i is a dummy equal to one if i has a college degree. Other covariates in X_i include experience and region of residence. Why might might college graduation be endogenous, that is, correlated with the unobservable component of earnings?

Think of the variable D_i as being composed of two parts

$$D = b\varepsilon + c$$

where $cov(c, \varepsilon) \equiv 0$. Using this decomposition, we can write

$$Y = \delta c + X\beta + (1 - \delta b)\varepsilon$$

If we observed the different components of D_i, you could regress the outcome on c, which is the exogenous part of D_i, and get a consistent estimate of δ.

In reality, we do not observe the components of D_i, so the best thing we can do is to use an IV.

The idea behind IV is to find a variable Z which is correlated with c, the exogenous part of D_i, and is uncorrelated with ε

Z is an instrument for D when the following two conditions are met:

- Exclusion restriction: $cov(Y, Z|X, D) = 0$. This implies that Z is exogenous, or $cov(Z, \varepsilon) = 0$
- Instrument condition: $Cov(Z, D) \neq 0$. It must be correlated with D

Then we can use two-stage least squares (2SLS) and recover an unbiased estimate of δ

Two Stage Least Squares (2SLS)

In the first stage, we regress the endogenous variable D on all the exogenous variables including the instrument

$$D_i = a_0 + a_1 Z_i + a_2 X_i + v$$

We recover the predicted values of D, $\hat{D}$, from the first stage

In the second stage, we regress the outcome variable on the predicted $\hat{D}$ and the other Xs

$$y_i = \alpha + \delta \hat{D}_i + X_i \beta + \varepsilon_i$$

$\hat{\delta}$ in the second stage is our IV estimate of the effect of treatment.

Instrumental Variables Estimation

Heterogeneous treatment effects

Homogeneous treatment effects were implicitly assumed in all empirical work until the mid 1990s.

With homogeneous treatment effects, IV was the solution to endogeneity issues. A valid instrument identified the treatment effects, which was the same for everyone.

If we relax the homogeneity assumption, we have to answer a key question:

Whose treatment effect is uncovered by IV?

Recall the decomposition of D

$$D = b\varepsilon + c$$

An instrument must be correlated with the component c and not the component $b\varepsilon$.

But in general, no instrument with be perfectly correlated with c

The portion of c that varies with one instrument may have a different average treatment effect than the portion of c that varies with another instrument.

Different instruments may induce unique sources of variation.

Consider the following model where Z is binary

$$y_i = \alpha + \delta D_i + \varepsilon_i$$
$$D_i = a_0 + a_1 Z_i + v_i$$

and the following four assumptions:

1. SUTVA (Stable Unit Treatment Value Assumption: outcomes of the ith individual are independent of others' outcomes
2. Exclusion restriction: $E(Y_i|z = 1, D) = E(Y_i|z = 0, D)$ for $i = 0, 1$
3. Instrument assumption: $E(D|z = 1) \neq E(D|z = 0)$
4. Monotonicity assumption $D(z = 1) - D(z = 0) \geq 0 \forall i$

Angrist, Imbens, and Rubin (1996) showed that under these assumptions, IV yields the following:

$$\begin{aligned}\delta &= \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} \\ &= E(Y_1 - Y_0|C)\end{aligned}$$

In words, the IV estimator (which in this case is also the Wald estimator) identifies the effect of treatment on the compliers, that is, on those people who are induced to switch treatment status as a result of the instrument

What does Assumption 4 really mean?

	$D(Z=0)=0$	$D(Z=0)=1$
$D(Z=1)=0$	never takers	Defiers
$D(Z=1)=1$	Compliers	always takers

The instrument assumption requires that Z be correlated with D. This means that *some* (not necessarily all) people must switch their treatment status in response to the value of Z.

The monotonicity assumption builds on the instrument condition. It requires that all people who react to the instrument change their treatment status the same way. The monotonicity assumption therefore rules out defiers.

With no defiers, the instrument traces out the impact of D for the compliers. If the monotonicity assumption does not hold and you allow for defiers, the resulting IV estimate may not give us a meaningful answer.

To see why, consider that under the monotonicity assumption,

$$p_A + p_N + p_C = 1$$

where p_A is the proportion of always takers, p_N is the proportion of never takers, and p_C is the proportion of compliers

We can express each of the terms in the numerator of the Wald estimator as follows

$$E(Y|Z{=}1){=}p_A E(Y_1|Z{=}1,A){+}p_N E(Y_0|Z{=}1,N){+}p_C E(Y_1|Z{=}1,C)$$
$$E(Y|Z{=}0){=}p_A E(Y_1|Z{=}0,A){+}p_N E(Y_0|Z{=}0,N){+}p_C E(Y_0|Z{=}0,C)$$

The numerator of the Wald estimator is

$$\begin{aligned} E(Y|Z=1) - E(Y|Z=0) &= p_C \left(E(Y_1|Z=1,C) - E(Y_0|Z=0,C)\right) \\ &= p_C \left(E(Y_1|C) - E(Y_0|C)\right) \end{aligned}$$

The denominator of the Wald estimator is

$$E(D|Z=1) - E(D|Z=0) = p_C$$

Thus, the IV estimator (which in this case is the Wald estimator) gives us the treatment impact for the compliers

$$
\begin{aligned}
\delta &= \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} \\
&= \frac{p_C \left[E(Y_1|C) - E(Y_0|C)\right]}{p_C} \\
&= E(Y_1 - Y_0|C) \\
&= \Delta^{compliers}
\end{aligned}
$$

What happens when there are defiers?

If the monotonicity assumption was violated and we had defiers, we would have

$$p_A + p_N + p_C + p_D = 1$$

$$E(Y|Z{=}1){=}p_A E(Y_1|Z{=}1,A){+}p_N E(Y_0|Z{=}1,N){+}p_C E(Y_1|Z{=}1,C){+}p_D E(Y_0|Z{=}1,D)$$

$$E(Y|Z{=}0){=}p_A E(Y_1|Z{=}0,A){+}p_N E(Y_0|Z{=}0,N){+}p_C E(Y_0|Z{=}0,C){+}p_D E(Y_1|Z{=}0,D)$$

The numerator of the Wald estimator becomes

$$E(Y|Z=1) - E(Y|Z=0) = p_C E(Y_1 - Y_0|C) - p_D E(Y_1 - Y_0|D)$$

The denominator of the Wald estimator becomes

$$E(D|Z=1) - E(D|Z=0) = p_C - p_D$$

This gives us the following Wald estimator

$$\begin{aligned} \delta^W &= \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} \\ &= \frac{p_C E(Y_1 - Y_0|C) - p_D E(Y_1 - Y_0|D)}{p_C - p_D} \end{aligned}$$

With the presence of defiers, the IV estimator is a weighted average of the treatment effect for the compliers and the defiers. The IV estimate is not equal to the treatment impact for either group and there is no way to separate out the effect for each group.

It is possible to get a negative IV estimate when in fact neither $\Delta^{Compliers}$ nor $\Delta^{Defiers}$ is negative

The IV estimator with defiers is meaningless unless we have a constant treatment effect

$$\begin{aligned}
\delta^W &= \frac{p_C E(Y_1 - Y_0|C) - p_D E(Y_1 - Y_0|D)}{p_C - p_D} \\
&= \frac{p_C \Delta - p_D \Delta}{p_C - p_D} \\
&= \Delta
\end{aligned}$$

Angrist, Imbens and Rubin showed that under their four assumptions, the IV estimate is gives us a consistent **LATE** (Local Average Treatment Effect) estimate. The IV estimate represents the impact of D for a special/local group of people, the compliers. It is the impact of D for those affected by the instrument

If different instruments affect different groups of people, the LATE obtained using different instruments may differ.

IV Implementation Problems

Problem 1: Weak Instruments

Weak instruments are instruments that are weakly correlated with treatment status. Any instrument has to satisfy the instrument condition, so any Z you will use must have a nonzero correlation with D. Weak instruments are poorly correlated with D.

When instruments are weak, they cause several problems.

Background

Asymptotic vs. Finite-Sample Properties of IV

When the exogeneity assumption and instrument conditions are satisfied, IV is consistent and asymptotically normal. This means that:

- $\hat{\delta}_{IV}$ approaches δ as n becomes large
- The sampling distribution for $\hat{\delta}_{IV}$ approaches the normal distribution as n becomes large

At the same time, $\hat{\delta}_{IV}$ is biased in finite samples, meaning $E(\hat{\delta}_{IV}) \neq \delta$

The problem

So long as $cov(Z, D)$ is large enough, finite sample bias is small and asymptotic normality holds

When $cov(Z, D)$ is small, things go bad:

- Finite sample bias is pronounced, and $\hat{\delta}_{IV}$ is biased toward $\hat{\delta}_{OLS}$
- Asymptotic normal approximation becomes inaccurate. Conventional test statistics do not have conventional distributions, so inference is flawed.

How do you know if $cov(Z, D)$ is large enough?

Estimate the first-stage regression.

If the F-statistic for the instrument (i.e. t^2) is less than 10, the instrument is weak.

The reason is that the bias from weak instruments is proportional to 1/F.

Solutions to the weak instruments problem

There are techniques for conducting inference with weak instruments, but they are complicated and fragile in their own right.

More to the point, how much would you believe results that are based on a bunch of restrictive conditions?

Problem 2: Large Standard Errors

IV estimates are much less precise (large standard errors) than OLS estimates. The tradeoff between consistency and efficiency is apparent .

To understand why IV estimates are always less efficient, think about the variation you are using to identify the effects.

We discussed how IV methods use only a portion of the total variation in the endogenous variable (in our case, D). Whenever we throw away some information/variation, standard errors get larger. We decided to use only a portion of the variation in the treatment status variable because we think this portion is uncorrelated with ϵ, so we are choosing consistency over efficiency.

In the case of one instrument, it can be shown that

$$Var\left(\hat{\delta}^{IV}\right) = \frac{Var\left(\hat{\delta}^{OLS}\right)}{\hat{\rho}^2_{Z,D}}$$

where $\hat{\rho}^2_{Z,D}$ is the square of the correlation coefficient of Z and D.

Weaker instruments mean smaller correlations between Z and D which leads to an even lower proportion of the variation in D being used. Again, this leads to larger standard errors.

Unfortunately, there is nothing one can do to get around the weak instrument and large standard errors issues. The only solution is to find a better instrument (or a bigger sample).