Instrumental Variables

Ethan Kaplan

1 Instrumental Variables: Intro.

- Bias in OLS:

  – Consider a linear model:

$$Y = X\beta + \epsilon$$

  – Suppose that

$$cov\left(X, \epsilon\right) = \rho$$

  – then OLS yields:

$$\begin{aligned} \hat{\beta}_{OLS} &= \left(X'X\right)^{-1} X'Y = \\ & \left(X'X\right)^{-1} X'\left(X\beta + \epsilon\right) \\ \Longrightarrow & \quad E\hat{\beta}_{OLS} = \beta + \left(X'X\right)\rho \end{aligned}$$

- Two Stage Least Squares
  - One solution to the problem of bias in OLS is to find variables correlated with Y only through their correlation with X and use only the variation in X correlated with these other variables (called instruments) Z. First run:

$$X = Z\gamma + \delta$$

  - From this we get an estimate of γ which we call $\hat{\gamma}$ and a predicted X:

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

  - then run:

$$\left(\begin{array}{c} \left[X'Z\left(Z'Z\right)^{-1}Z'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1} \\ X'Z\left(Z'Z\right)^{-1}Z'Y \end{array}\right)$$

- Three cases:

  1. Under-identified: number of regressors in Z $<$ number of regressors in X

  2. Just Identified: number of regressors in Z $=$ number of regressors in X

  3. Over identified: number of regressors in Z $>$ number of regressors in X

  - In the under-identified case, the model can not be estimated

  - In the just identified case the dimension of $X\prime Z$ is the dimension of $Z\prime Z$ in which case:

$$\begin{pmatrix} \left[(Z'X)^{-1}(Z'Z)(Z'Z)^{-1}(Z'Z)(X'Z)^{-1}\right] \\ X'Z(Z'Z)^{-1}Z'Y \end{pmatrix}$$

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

- Weak Instruments Problem
  - One problem is that if $Z'X \approx 0 = Z'Y$, then the distribution $(Z'X)^{-1} Z'Y$ is the ratio of two normals with mean zero and is approximated well even in very large samples by a Cauchy Distribution, whose mean and variance do not exist. This can be very problematic.
  - What is?

$$E\left[\left(Z'X\right)^{-1} Z'Y - \beta\right]$$

  - In general, we dont know. However,

$$\begin{aligned}
& p\lim\left[\left(Z'X\right)^{-1} Z'Y - \beta\right] \\
= \ & p\lim\left[\left(Z'X\right)^{-1} Z'\left(X\beta + \epsilon\right) - \beta\right] \\
= \ & p\lim\left[\beta + \left(Z'X\right)^{-1} Z'\epsilon - \beta\right] \\
= \ & p\lim\left[\beta - \beta\right] = 0
\end{aligned}$$

- Bias in OLS vs. 2SLS: $\frac{(Z'X)^{-1}Z'\epsilon}{X'\epsilon}$

  - Re-expressed:

$$\frac{\sigma_{Z\epsilon}}{\sigma_{XZ}\sigma_{X\epsilon}}$$

  - Another way to write the 2SLS estimator is:

$$\left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'\epsilon = \frac{\sigma_{\hat{X}\epsilon}}{\sigma^2_{\hat{X}}}$$

  - as opposed to the OLS bias:

$$\left(X'X\right)^{-1}X'\epsilon = \frac{\sigma_{X\epsilon}}{\sigma^2_{X}}$$

  - So the small bias of the 2SLS is in the direction of the OLS estimator.

- Wald Estimator:
  - One special example is the so-called Wald Estimator:

$$\begin{aligned} Y_i &= \alpha_1 + \beta X_i + \epsilon \\ X_i &= \alpha_2 + \gamma D_i + \delta \end{aligned}$$

  - where D_i is a dummy variable taking on the values of $\{0, 1\}$. Then:

$$\hat{\beta}_{WALD} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$$

  - where $\bar{Y}_1, \bar{Y}_0$ are the average Y when $Z = 1, 0$ respectively and $\bar{X}_1, \bar{X}_0$ are the average X when $Z = 1, 0$ respectively.

- Small Sample Bias of 2SLS:
  - Approximate Bias Formula for small samples (derived using power series approximations):

$$\begin{aligned} & \frac{\sigma_{Z,\delta}}{\gamma' Z' Z \gamma}(K-2) \\ = \; & \frac{\sigma_{Z,\delta}}{\sigma_\delta^2} \frac{\sigma_\delta^2}{\gamma' Z' Z \gamma}(K-2) \end{aligned}$$

  - where K is the number of excluded instruments.

$$\tau^2 = \frac{\sigma_\delta^2}{\gamma' Z' Z \gamma}$$

  - τ^2 is called a concentration parameter and is equal to $\frac{1}{R^2}$ from the first stage regression.
  - Commonly thought that bias is proportional to K. In fact, this is only true in the case where $\sigma_{XZ} = 0$ (or $\sigma_{XZ} \approx 0$). Otherwise $\gamma' Z' Z \gamma$ and thus τ^2 depend upon K.
  - So is adding more instruments a good thing? Depends if they are correlated with LHS variables

conditional on the other instruments. Similar to out of sample prediction... not always a good idea.

– Can you test if instruments are too weak? You can run a joint F-test on the first stage (essentially the concentration parameter). Usually you want at least F-Statistic of 4 or 5 in the literature. Some will want at least 10.

- Limited Information Maximum Likelihood

  – Can also estimate with Limited Information Maximum Likelihood. It turns out that though the asymptotic distribution of the 2SLS estimator and the LIML estimator are the same, the small sample distributions can be quite different in overidentified models. In particular, with LIML, the parameter being estimated is close to its population median rather than mean. The formula for LIML is:

$$L(\beta, \pi, \Omega) = \sum_{n=1}^{N} \begin{pmatrix} -\frac{1}{2}|\Omega| - \frac{1}{2}\binom{Y_i - \beta\gamma Z_i}{X_i - \gamma Z_i}' \\ \Omega^{-1}\binom{Y_i - \beta\gamma Z_i}{X_i - \gamma Z_i} \end{pmatrix}$$

  – Example with real and random instruments:

	Single Instr.	500 Instr. 2SLS
Real	0.089 (0.011)	0.073 (0.008)***
Random	−1.958 (18.116)	0.059 (0.085)

	500 Instr. LIML
Real	0.095 (0.017)***
Random	−0.330 (0.1001)***

- 2SLS Inference:
  - Suppose you run 2SLS in two stages. Then you compute SEs as:

$$\left(\hat{X}'\hat{X}\right)^{-1}\hat{\sigma}^2$$

  - Instead you should take the assymptotic variance of:

$$\left(\begin{array}{c} \left[(Z'X)^{-1}(Z'Z)(Z'Z)^{-1}(Z'Z)(X'Z)^{-1}\right] \\ X'Z(Z'Z)^{-1}Z'Y \end{array}\right)$$

  - In which case you get:

$$\left(X'Z\right)\left(Z'Z\right)^{-1}\left(Z'X\right)^{-1}\hat{\sigma}^2$$

  - You can show that the true SEs are large than the second stage OLS because they include the variation from the first stage which the second stage OLS standard errors do not.

2 Average Treatment Effects

- Setup: Binary Instrument and Binary Endogenous RHS Variable.
  - Note, according to Angrist (Journal of Econometrics, 1991): Grouped-data estimation and testing in simple labor-supply models, continuous IV models can be reduced to binary IV models.

- Define Four Types of Reactions to Instrument:

	$D_i(0) = 0$	$D_i(0) = 1$
$D_i(1) = 0$	Never-Taker	Defier
$D_i(1) = 1$	Complier	Always-Taker

- Then if we see the following combinations of instrument and RHS variable, we know that:

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	Complier/Never-Taker	Never-Taker/Defier
$D_i = 1$	Always-Taker/Defier	Complier/Always-Taker

- Assuming monotonicity $\left(D_i(1) \geq D_i(0)\ \ \forall i\right)$, we can eliminate defiers. Then from combinations of instrument and RHS variable, we can figure out:

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	Complier/Never-Taker	Never-Taker
$D_i = 1$	Always-Taker	Complier/Always-Taker

- So we define fraction complier $= \alpha_C$, fraction Never-Taker $= \alpha_N$ and fraction Always-Taker $= \alpha_A$
  - Then $\alpha_C + \alpha_N + \alpha_A = 1$
  - Morever we get that $P\left(D_i = 1 | Z_i = 0\right) = \alpha_A$ and $P\left(D_i = 1 | Z_i = 0\right) = \alpha_N$ and finally $\alpha_C = 1 - \alpha_N - \alpha_A$
  - So, under the assumption that there are no defiers, we can recover, α_C, α_N, and α_A

- With one regressor:

$$\begin{aligned}
& \left(Z'D\right)^{-1} Z'Y \\
= \;& \left(Z'D\right)^{-1} \left(Z'Z\right) \left(Z'Z\right)^{-1} \left(Z'Y\right) \\
= \;& \left(\left(Z'Z\right)^{-1}\right)^{-1} \left(Z'D\right)^{-1} \left(Z'Z\right)^{-1} \left(Z'Y\right) \\
= \;& \frac{(Z'Z)^{-1} (Z'Y)}{(Z'Z)^{-1} (Z'D)}
\end{aligned}$$

  In other words, we can interpret the IV coefficient as the ratio of the regression coefficient of the outcome variable on the instrument to the regression coefficient of the endogenous explanatory variable on the instrument.

- Now look at the numerator of this formula:

$$E\left(Y|Z=1\right) - E\left(Y|Z=0\right)$$

  - We can break it up into the expectation conditional on $Z = 0$ and the expectation conditional on $Z = 1$. Starting with $Z = 0$:

$$
\begin{aligned}
E(Y|Z=0) &= \\
& E(Y|Z=0,C)\,P(C|Z=0)+ \\
& E(Y|Z=0,N)\,P(N|Z=0)+ \\
& E(Y|Z=0,A)\,P(A|Z=0)
\end{aligned}
$$

– And now turning to $Z = 1$:

$$
\begin{aligned}
E(Y|Z=1) &= \\
& E(Y|Z=1,C)\,P(C|Z=1)+ \\
& E(Y|Z=1,N)\,P(N|Z=1)+ \\
& E(Y|Z=1,A)\,P(A|Z=1)
\end{aligned}
$$

- Note that Always-Takers and Never-Takers are not affected by the instrument:

$$E(Y|Z=1,N) = E(Y|Z=0,N)$$

$$E(Y|Z=1,A) = E(Y|Z=0,A)$$

- Also since Z is randomized, probabilities of getting assigned the instrument are independent of type:

$$
\begin{aligned}
P(A|Z=1) &= P(A|Z=0) \\
P(N|Z=1) &= P(N|Z=1) \\
P(C|Z=1) &= P(C|Z=0)
\end{aligned}
$$

- Now we can compute the numerator conditioning on type:

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

$$
\begin{aligned}
&E(Y|Z=1,C)\,P(C|Z=1) + \\
&E(Y|Z=1,N)\,P(N|Z=1) + \\
&E(Y|Z=1,A)\,P(A|Z=1) - \\
&E(Y|Z=0,C)\,P(C|Z=0) - \\
&E(Y|Z=0,N)\,P(N|Z=0) - \\
&E(Y|Z=0,A)\,P(A|Z=0)
\end{aligned}
$$

- But the conditional expectations and probabilities for the Never-Takers and Always-Takers second two terms are the same (the always and never takers are not affected by the instrument) and they thus cancel out, leaving:

$$
\begin{aligned}
& \left(Z'Z\right)^{-1}\left(Z'Y\right) \\
= & \left[E\left(Y|Z=1,C\right)-E\left(Y|Z=0,C\right)\right]\alpha_C
\end{aligned}
$$

- Similarly (without showing computations) for the denominator:

$$
\begin{aligned}
\left(Z'Z\right)^{-1}\left(Z'D\right) &= \\
\begin{bmatrix} P\left(C\right)1+P\left(A\right)1+P\left(N\right)0 \\ -P\left(A\right)1-P\left(C\right)0-P\left(N\right)0 \end{bmatrix} &= \\
\alpha_C+\alpha_A-\alpha_A &= \alpha_C
\end{aligned}
$$

- Finally, we get our expression:

$$
\begin{aligned}
\beta_{IV} &= \left(Z'D\right)^{-1}Z'Y = \\
& \frac{\left[E\left(Y|Z=1,C\right)-E\left(Y|Z=0,C\right)\right]\alpha_C}{\alpha_C} \\
&= E\left(Y|Z=1,C\right)-E\left(Y|Z=0,C\right)
\end{aligned}
$$

or in other words, the IV instrument gives the local average treatment effect for the compliers to the instrument (and thus since different instruments will have different sets of compliers, different instruments may yield different IV estimates).

3 Control Function Approach

- Equivalence of controlling for first stage residuals and standard 2SLS approach of putting in fitted values from first stage. Assume model:

$$
\begin{aligned}
Y &= X\beta + \gamma W + \epsilon \\
W &= X\beta + \mu Z + \nu
\end{aligned}
$$

  – We are interested in γ; X is a set of controls, W is an endogenous variable, Z is a valid instrument

$$
\begin{aligned}
cov\left(W, \epsilon\right) &\neq 0 \\
cov\left(Z, \epsilon\right) &\neq 0 \\
cov\left(Z, \nu\right) &= 0
\end{aligned}
$$

  – 1st stage: regress

$$W = X\beta + \mu Z + \nu$$

  – Obtain first stage residuals $\hat{\nu}$

– 2nd stage: plug in residuals into first equation. Regress:

$$Y = X\beta + \gamma W + \phi\hat{\nu} + \epsilon$$

– Then

$$\hat{\gamma} = \gamma_{2SLS}$$

– The coefficient matrix $[\gamma|\beta]$ can be obtained using the Frisch-Waugh-Lovell Theorem:

$$[\gamma|\beta] = \left[V'(I-P)V\right]^{-1} V'[I-P]Y$$

where

– Note that the coefficient on γ using this method is not just assymptotically equivalent to γ_{2SLS}, it is identical. Therefore:

$$V = [W|X]$$

$$Q = [Z|X]$$

$$P = \left[I - Q\left(Q'Q\right)^{-1}Q'\right]Q$$

* The standard errors on γ will be identical to the γ_{2SLS} standard errors and
* The second stage OLS standard errors will not be equal to the true standard errors.

3.1 Random Coefficients

- Now we relax that coefficients on the impact of W are the same for the entire population. First we assume that

$$\begin{aligned} Y &= X\beta + \gamma W + \epsilon \\ W &= X\beta + \mu Z + \nu \end{aligned}$$

$$\gamma = \bar{\gamma} + \delta$$

$$cov\,(\delta, W) = 0$$

In this case, the 2SLS estimator consistently estimates the average effect of W:

$$\text{plim}\,(\hat{\gamma}_{2SLS}) = \bar{\gamma}$$

- However, often times the impact of W may be different for different values of W :

$$cov\,(\delta, W) \neq 0$$

  - In this case, $\hat{\gamma}_{2SLS}$ does not estimate an average treatment effect but rather a weighted average of treatment effects (weighted by W).

  - In this case, we can still estimate (with a linearity assumption) a control function:

$$\begin{aligned} Y &= X\beta + \gamma W + \rho\hat{\nu} + \eta\hat{\nu}W + \theta \\ W &= X\beta + \mu Z + \nu \end{aligned}$$

- ρ captures endogeneity bias

- η captures selectivity (a positive η means that those likely to select into higher W are more likely to have higher residual Y; a negative η means that those likely to select into lower W are more likely to have higher residual Y)

- Note that only in the case of $cov\,(\delta, W) = 0$ is δ (as the population average of consistently estimated with normal IV.

- Also, note that this is a more general model (assymptotically). Anytime that $\hat{\gamma}_{2SLS}$ consistently estimates the true $\bar{\gamma}$, then so does $\hat{\gamma}_{CF}$ (the control function $\hat{\gamma}$). However, if $cov\,(\delta, W) \neq 0$, $\hat{\gamma}_{CF}$ still consistently estimates $\bar{\gamma}$ but $\hat{\gamma}_{2SLS}$ does not.