Economics 8379
Spring 2020
Ben Williams

IV estimators and "forbidden regressions"

Preliminary results

Consider the triangular model with first stage given by

$$x_{i2} = \gamma_1' X_{i1} + \gamma_2 Z_i + \nu_i$$

and second stage given by

$$y_i = \beta_1' X_{i1} + \beta_2 x_{i2} + u_i$$

To use matrix notation, let y denote the $n \times 1$ vector $y = (y_1, \ldots, y_n)'$, let X denote the $n \times k$ matrix with i^{th} row given by (X_{i1}', x_{i2}), and let $\tilde{Z}$ denote the $n \times q$ matrix with i^{th} row given by (X_{i1}', Z_i'). The two stage least squares (2SLS) estimator of $\beta = (\beta_1', \beta_2)'$ is given by

$$\hat{\beta}^{2SLS} = (\hat{X}'X)^{-1}\hat{X}'y$$

where $\hat{X} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X$. We can also define an indirect least squares estimator of β as

$$\hat{\beta}^{ILS} = (Z^{*\prime}X)^{-1}Z^{*\prime}y$$

for some matrix Z^*. The simplest example of an indirect least squares estimator is when $q = k$ so that $\tilde{Z}'X$ is a square matrix and we can calculate an indirect least squares estimator using $Z^* = \tilde{Z}$.

Result 1. $\hat{\beta}^{2SLS}$ **is a version of the ILS estimator with** $Z^* = \hat{X}$**. If** $q = k$ **then** $\hat{\beta}^{2SLS}$ **is also equivalent to an ILS estimator with** $Z^* = \tilde{Z}$**.**

Proof. First note that $\hat{X}'\hat{X} = X'P_{\tilde{Z}}P_{\tilde{Z}}'X$ where $P_{\tilde{Z}}$ is the projection matrix $P_{\tilde{Z}} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'$. Since $P_{\tilde{Z}}P_{\tilde{Z}}' = P_{\tilde{Z}}$, (that is, $P_{\tilde{Z}}$ is an idempotent matrix) , we can conclude that $\hat{X}'\hat{X} = \hat{X}'X$.

For the second claim, if $\tilde{Z}'X$ is full rank then it is also invertible. Therefore,

$$\begin{aligned}
\hat{\beta}^{2SLS} &= (\hat{X}'X)^{-1}\hat{X}'y \\
&= \left(X'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X\right)^{-1} X'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'y \\
&= (\tilde{Z}'X)^{-1}(\tilde{Z}'\tilde{Z})(X'\tilde{Z})^{-1}X'\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'y \\
&= (\tilde{Z}'X)^{-1}\tilde{Z}'y
\end{aligned}$$

□

Result 2. If $E(Z_i u_i) = 0$ and $E(X_{1i}u_i) = 0$ then, under standard regularity conditions, $\hat{\beta}^{2SLS} \to_p \beta$

Proof. Using result 1,

$$\begin{aligned}
\hat{\beta}^{2SLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\
&= (\hat{X}'X)^{-1}\hat{X}'y \\
&= (\hat{X}'X)^{-1}\hat{X}'(X\beta + u) \\
&= \beta + (\hat{X}'X)^{-1}\hat{X}'u
\end{aligned}$$

Under standard regularity conditions, $(\hat{X}'X)^{-1}\hat{X}'u$ converges in probability to 0 if $E(Z_i u_i) = 0$. □

Similarly, we can state the following conditions for consistency of an ILS estimator.

Result 3. If $n^{-1}Z^{*\prime}X \to_p M_{Z^*X}$, where M_{Z^*X} is a nonsingular matrix, and $n^{-1}Z^{*\prime}u \to_p 0$ then $\hat{\beta}^{ILS} \to_p \beta$

Forbidden regressions

In general, a two stage estimator only produces the 2SLS estimator if the first stage predicted values are calculated from an OLS regression of the equation

$$x_{i2} = \gamma_1' X_{i1} + \gamma_2 Z_i + \nu_i$$

If the full set of controls, X_{i1}, is not included, or if this is estimated with via another estimator (for example, weighted least squares or probit in the case where x_{i2} is binary)

Using the calculations in the previous section, we can show the following three results:

1. If the $\hat{X}$ in the 2SLS estimator, $(\hat{X}'\hat{X})^{-1}\hat{X}'y$, is replaced with a consistent estimator for $E(X_i \mid \tilde{Z}_i)$ then we obtain a consistent estimator of β. However, this is not the standard 2SLS estimator.

2. If the model used to estimate $\hat{X}$ is misspecified (so that $\hat{X}$ is not consistent for $E(X \mid \tilde{Z})$) then generally the modified 2SLS estimator will not be consistent. The standard 2SLS estimator is the exception to this rule – the OLS first stage generally is not a consistent estimator for $E(X \mid \tilde{Z})$.

3. An ILS estimator that uses a misspecified model to estimate $\hat{X}$ and then uses $Z^* = \hat{X}$ provides a consistent estimator.

Let $\hat{X}$ denote *some* first stage predicted value, not necessarily $\hat{X} = \tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X$. First I will consider a general class of formulas for $\hat{X}$. Suppose $\hat{X}$ is a consistent estimator of some function $\mu(\tilde{Z})$. So we can use the notation $\hat{\mu}(\tilde{Z}) := \hat{X}$. In some cases $\mu(\tilde{Z}) = E(X \mid \tilde{Z})$, but this is only if the first stage model is correctly specified.

I will start with the first result. Note that we can write $y = X'\beta + u = \hat{X}'\beta - (\hat{X} - X)\beta + u$. Therefore, for the modified 2SLS estimator is

$$\begin{aligned}\hat{\beta}^{2SLS,modified} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y \\ &= \beta - (\hat{X}'\hat{X})^{-1}\hat{X}'(\hat{X} - X)\beta + (\hat{X}'\hat{X})^{-1}\hat{X}'u\end{aligned}$$

Under standard regularity conditions, $(\hat{X}'\hat{X})^{-1}\hat{X}'u = (\hat{X}'\hat{X})^{-1}\mu(\tilde{Z})'u + o(1)$, so if $E(u_i \mid \tilde{Z}_i) = 0$ then we can ignore this term. Thus this modified 2SLS estimator is consistent if $(\hat{X}'\hat{X})^{-1}\hat{X}'(\hat{X} - X) \to_p 0$. Since $X_i = E(X_i \mid \tilde{Z}_i) + \eta_i$ where $E(\eta_i \mid \tilde{Z}_i) = 0$, we can write, in matrix form, $\hat{X} - X = \hat{\mu}(\tilde{Z}) - \mu(\tilde{Z}) + \mu(\tilde{Z}) - E(X \mid \tilde{Z}) + \eta$. Then we are essentially done. Under sufficient regularity conditions,

$$(\hat{X}'\hat{X})^{-1}\hat{X}'(\hat{\mu}(\tilde{Z}) - \mu(\tilde{Z})) \to_p 0$$

if $\hat{\mu}(\tilde{Z})$ is a (uniformly) consistent estimator of $\mu(\tilde{Z})$. Next,

$$(\hat{X}'\hat{X})^{-1}\hat{X}'\eta = (\hat{X}'\hat{X})^{-1}\mu(\tilde{Z})'\eta + o(1) \to 0$$

since $E(\eta \mid \tilde{Z}) = 0$ by definition. Therefore, we conclude that $\hat{\beta}^{2SLS,modified}$ is consistent if $E(\mu(\tilde{Z}_i)'(\mu(\tilde{Z}_i) - E(X_i \mid \tilde{Z}_i)) = 0$. This happens if the model used to estimate $\hat{X}$ is correctly specified, so that $\mu(\tilde{Z}_i) = E(X_i \mid \tilde{Z}_i)$, or if (a less likely scenario) the specification error, $\mu(\tilde{Z}_i) - E(X_i \mid \tilde{Z}_i)$ is mean independent of $\tilde{Z}_i$.

So why does the standard 2SLS work, even when $E(X \mid \tilde{Z})$ is not linear? Because in that case $\hat{X}'(\hat{X} - X) = X'P_{\tilde{Z}}'(P_{\tilde{Z}} - I)X = 0$ since $P_{\tilde{Z}}$ is idempotent. The OLS residuals are uncorrelated with the OLS predicted values by construction. But the residuals from a nonlinear model are not generally uncorrelated with the predicted values.

Lastly, consider the ILS estimator based on $\hat{X} = \hat{\mu}(\tilde{Z})$. We can write

$$\begin{aligned}
\hat{\beta}_{ILS} &= (\hat{X}'X)^{-1}\hat{X}'y \\
&= \beta + (\hat{X}'X)^{-1}\hat{X}'u \\
&= \beta + (\hat{X}'X)^{-1}\mu(\tilde{Z})'u + (\hat{X}'X)^{-1}(\hat{\mu}(\tilde{Z}) - \mu(\tilde{Z}))'u
\end{aligned}$$

The second term will generally converge to 0 provided that $E(u_i \mid \tilde{Z}_i) = 0$. The third term will converge to 0 because $\hat{\mu}(\tilde{Z})$ is a (uniformly) consistent estimator for $\mu(\tilde{Z})$. So this estimator will generally be consistent. Note that this is equivalent to the 2SLS estimator that regresses X on $\hat{\mu}(\tilde{Z})$.

Examples

The three examples considered in the slides are (1) using a subset of X_{i1} in the first stage, (2) using a probit or logit in the first stage when x_{i2} is binary, and (3) using nonlinear functions of predicted values from the first stage rather than taking the nonlinear transformation before estimating the first stage.

In case (1), suppose $X_i = (X_{i11}', X_{i12}', x_{i2})'$ where X_{i11} is included in the first stage but X_{i12} is not. Then the modified (incorrect) 2SLS estimator, that is the so-called forbidden regression, uses

$\tilde{Z} = (X'_{i11}, Z_i)$ to get $\hat{x}_{i2}$. If $E(x_{i2} \mid X'_{i11}, X'_{i12}, Z_i) = E(x_{i2} \mid X'_{i11}, Z_i)$, then this modified 2SLS estimator is consistent. Otherwise it is not. On the other hand, the ILS estimator that uses the "incorrect" first stage predicted values as an instrument will produce a consistent estimator. This ILS estimator will in general differ from the correct 2SLS estimator though. It can, in fact, be more efficient.

In case (2), the modified 2SLS estimator is generally ill-advised as it will only be consistent if the probit or logit used in the first stage is correctly specified. However, the ILS estimator using the probit/logit predicted values as an instrument will be consistent *and* can be more efficient than the standard 2SLS. See the last paragraph in 6.4.2 in Cameron and Trivedi, where they show that this is in fact the optimal estimator if there is not heteroskedasticity in the outcome equation.

In case (3), the "forbidden regression" will *never* be correct. Consider, for example, the model where $y_i = \beta_0 + \beta_1 x_{i2} + \beta_2 x_{i2}^2 + u_i$. If $\hat{x}_{i2}$ denotes the OLS predicted values then $\hat{x}_{i2}^2$ will never be a consistent estimator for $E(x_{i2}^2 \mid Z_i)$. Therefore, the correct procedure is to estimate two first stage equations with the same set of regressors (just Z_i, which in this case must be at least 2 variables) with x_{i2} and x_{i2}^2 as dependent variables. This produces the correct 2SLS estimator.

Final notes

When using any of these estimators, the estimation error in the first stage must be accounted for in the standard errors. The ivreg command in Stata does this for the standard 2sls estimator but when constructing the ILS estimator, care should be taken in getting the correct standard error.