14 2SLS. Control Function. Endogeneity and overidentification tests.

14.1 Under, just and overidentification

Consider again the linear regression model, with $\vec{x}_{1i}$ exogenous and $\vec{x}_{2i}$ endogenous.

$$y_i = \beta_0 + x_{1i}'\beta_1 + x_{2i}'\beta_2 + u_i$$

Then take instrument:

$$w_i = \begin{pmatrix} x_{1i} \\ z_i \end{pmatrix}$$

with x_{1i} instrumenting for themselves (included exogenous variables) and z_i instrumenting for x_{2i} (excluded exogenous variables).

If w_i l-dimensional and x_i k-dimensional:

$$\underbrace{E[w_i y_i]}_{l\times 1} = \underbrace{E[w_i x_i']}_{l\times k}\underbrace{\beta}_{k\times 1}$$

- If $l < k$, then we have **underidentification**
- If $l = k$, then we have **just identification**
- If $l > k$, then we have **overidentification**

The relevance condition, $E[w_i x_i']$ full column rank, <u>rules out</u> underidentification. This is because now l rows will be fewer than k columns, and since column rank = row rank, we must have deficient column rank.
If $l < k$ we have more equations than unknowns and $E[w_i x_i']$ is no longer invertible. We could throw away extra variables but better instead to use 2SLS, since we want to extract as much exogenous variation from our endogenous variables as possible.

14.2 2SLS

For now assume $E[\varepsilon_i|w_i] = 0$. Then:

$$0 = E[\varepsilon_i|w_i] = E[y_i - x_i'\beta|w_i] = E[y_i|w_i] - E[x_i'|w_i]\beta$$

$$\Rightarrow E[y_i|w_i] = E[x_i'|w_i]\beta$$

Suppose we also know

$$E[x_i'|w_i] = w_i'\pi$$

Then we have:

$$E[y_i|w_i] = (w_i'\pi)\beta$$

This suggest the following procedure:

Definition 14.2.1

2SLS

Stage 1:

- Regress $X_{n\times k}$ on $W_{n\times l}$ to get $\hat{\pi} = (W'W)^{-1}W'X$
- Use the results to form $\hat{X} = W\hat{\pi}$

Note: $\hat{X} = W\hat{\pi} = W(W'W)^{-1}W'X = P_W X$
For the exogenous variables columns in $\hat{X}$ this will correspond exactly to the original values, but for the endogenous variables columns, they will be formed as a linear combination of both the relevant instruments and exogenous variables.

Stage 2:

Regress $Y_{n\times 1}$ on $\hat{X}_{n\times k}$ to find:

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y = (X'P_W'P_WX)^{-1}X'P_W'Y$$
$$= (X'P_WX)^{-1}X'P_WY$$

Consider the following **IV assumptions** for the model $y_i = x_i'\beta + \varepsilon_i$:

- (IV0) y_i, x_i, w_i is an i.i.d sequence
- (IV1) $E[w_iw_i'] < \infty$ non-singular; $E[w_ix_i']$ has full column rank (relevance)
- (IV2) $E[\varepsilon_i|w_i] = 0$ ($\Rightarrow$) (IV2') $E(w_i\varepsilon_i) = 0$ (exogeneity)
- (IV3) $E[\varepsilon_i^2|w_i] = \sigma^2$ (homoskedasticity) or (IV3') $V = Var(w_i\varepsilon_i)$ is finite non singular (Under IV(3): $V = E[w_iw_i'\varepsilon_i^2] - 0 = E[E[w_iw_i'\varepsilon_i^2|w_i]] = \sigma^2 E[w_iw_i']$)

Theorem 14.2.1. 2SLS consistency

Under IV(0) IV(1) IV(2')
$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta$$

Proof.
$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}(\hat{X}Y) = (X'P_WX)^{-1}X'P_WY$$
$$= \beta + (X'P_WX)^{-1}X'P_W\varepsilon$$
$$\hat{\beta}_{2SLS} - \beta = [X'W(W'W)^{-1}W'X]^{-1}X'W(W'W)^{-1}W'\varepsilon$$
$$= \left[\frac{1}{n}\sum x_iw_i'(\frac{1}{n}\sum w_iw_i')^{-1}\frac{1}{n}\sum w_ix_i'\right]^{-1}\frac{1}{n}\sum x_iw_i'(\frac{1}{n}\sum w_iw_i')^{-1}(\frac{1}{n}\sum w_i\varepsilon_i)$$
$$\xrightarrow{p} [E(x_iw_i')E(w_iw_i')^{-1}E(w_ix_i')]^{-1}E(x_iw_i')E(w_iw_i')^{-1}E(w_i\varepsilon_i)$$

By IV(2'), $E(w_i\varepsilon_i) = 0$ and by IV(1) $E(w_iw_i')$ is non-singular to a finite constant matrix (also assume $E(x_iw_i') < \infty$). Thus
$$\hat{\beta}_{2SLS} - \beta \xrightarrow{p} 0$$

□

In general $dimW \neq dimX$. In the case where they do: $\hat{\beta}_{2SLS} \equiv \hat{\beta}_{IV}$, since $W'X$ now invertible. The 2SLS procedure ensures that $dim\hat{X} = dimX$, so that $\hat{\beta}_{2SLS} \equiv \hat{\beta}_{IV}$, using $\hat{X}$ as an instrument. Explicitly: $(X'P_W X)^{-1}X'P_W Y = (X'P_W' X)^{-1}X'P_W' Y = (\hat{X}'X)^{-1}(\hat{X}'Y) = \hat{\beta}_{IV}$

Theorem 14.2.2. 2SLS asymptotic distribution

Under IV0-1-2'-3':

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, (D'C^{-1}D)^{-1}D'C^{-1}VC^{-1}D(D'C^{-1}D)^{-1})$$

where $V = Var(w_i\varepsilon_i)$, $C = E[w_i w_i']$ and $D = E[w_i x_i']$.

Proof.

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) =$$

$$\left[\frac{1}{n}\sum x_i w_i' (\frac{1}{n}\sum w_i w_i')^{-1}\frac{1}{n}\sum w_i x_i'\right]^{-1}\frac{1}{n}\sum x_i w_i'(\frac{1}{n}\sum w_i w_i')^{-1}(\frac{1}{\sqrt{n}}\sum w_i\varepsilon_i)$$

By Lindeberg-Levy CLT:

$$\frac{1}{\sqrt{n}}\sum w_i\varepsilon_i \xrightarrow{d} N(0, V)$$

By Slutsky's theorem:

$$\xrightarrow{d} [D'C^{-1}D]^{-1}D'C^{-1}N(0,V)$$
$$= N(0, (D'C^{-1}D)^{-1}D'C^{-1}VC^{-1}D(D'C^{-1}D)^{-1})$$

Under (IV3) (homoskedasticity):

$$V = Var(w_i\varepsilon_i) = E[w_i w_i' \varepsilon_i^2] - 0 = \sigma^2 E[w_i w_i'] = \sigma^2 C$$

Thus much of the asymptotic variance cancels, leaving

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, \sigma^2(D'C^{-1}D)^{-1})$$

□

Note:-

In general for two full column rank conformable matrices A, B:
We have AB full column rank.

Proof: Suppose AB not full column rank.
Then $\exists\ x \neq 0$ such that $ABx = 0$ (by the rank-nullity theorem).
$\Rightarrow Bx \neq 0$ as B full rank implies its null space is only $\{0\}$.
$\Rightarrow A(Bx) \neq 0$ as A also full rank with only trivial null space.
Contradiction

We apply this proof to argue $D'C^{-1}D$ is full column rank, and hence invertible.

14.2.1 Linear Hypothesis Testing with β_{2SLS}

We can estimate the asymptotic variance of $\sqrt{n}(\hat{\beta}_{2SLS} - \beta)$ by:

$$\hat{V} = \hat{\sigma}^2(\tfrac{1}{n}\hat{X}'\hat{X})^{-1}$$

where $\hat{\sigma}^2 = \frac{1}{n}\hat{\varepsilon}'\hat{\varepsilon}$ and $\hat{\varepsilon} = Y - \hat{X}'\hat{\beta}_{2SLS}$

Homoskedasticity or robust variance estimates of $\hat{\beta}_{2SLS}$ can be used to form F-statistics for testing linear hypotheses in the usual way. Asymptotically, such F-statistics would be distributed as $\chi^2(p)/p$, where p is the number of restrictions. However, finite sample distribution of the F-statistics would not be $F(p, n-k)$ even if ε_i is normally distributed.

Asymptotically the Wald statistic for testing $H_0 : R\beta = r$ is:

$$W = (R\hat{\beta}_{2SLS} - r)'[R\hat{V}_{2SLS}R']^{-1}(R\hat{\beta}_{2SLS} - r) \xrightarrow{d} \chi^2(p)$$

where p is the number of restrictions.

Note:-

Under (IV3') (hetereoskedasticity) we can use White's estimate as in earlier discussions:

$$\hat{V}_{het} = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 w_i w_i'$$

14.3 Control function approach

This is an alternative approach to 2SLS, which is useful when we have multiple endogenous variables.

Consider again the model:

$$y_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + \varepsilon_i$$

where x_{1i} is exogenous and x_{2i} is endogenous.

Instead of extracting the *exogenous* part $w_i'\pi$ of x_i to use in the second stage, we could instead extract the **endogenous part** of x_i (the control function) and add it to the regression as an additional regressor.

Theorem 14.3.1. The two approaches are equivalent in a linear model (but not in non-linear) $\hat{\beta}_{CF} \equiv \hat{\beta}_{2SLS}$

Proof. The exogenous part $w_i'\pi$ of x_i is simply the best linear predictor of x_i given w_i.

The first stage regression:

$$x_i' = w_i'\pi + u_i', \text{ where } \pi \text{ is } l \times k$$

is called a *reduced form* regression, because it does not have any structural interpretation. We just want to predict x_i by a linear function of w_i in the best possible way (thus exogeneity is not required). Recall w_i contains components of both included exogenous variables x_{1i} and excluded exogenous variables z_i.

Thus we partition the reduced form equations into:

$$x_{1i}' = x_{1i}'\pi_{11} + z_i'\pi_{12} + u_{1i}'$$

$$x_{2i}' = x_{1i}'\pi_{21} + z_i'\pi_{22} + u_{2i}'$$

where π_{ij} is a $k_j \times k_i$ matrix.

Of course the BLP of x_{1i} given x_{1i} and z_{1i} is just x_{1i} so the first of the above equations is trivial $x_{1i}' = x_{1i}'$ For the second equation we drop the first subscript and rewrite as:

$$x_{2i}' = x_{1i}'\pi_1 + z_i'\pi_2 + u_i'$$

In 2SLS this regression would be estimated, obtain $\hat{x}'_{2i}$, form $\hat{x}_i$ by combining x_{1i}, with $\hat{x}_{2i}$ and proceeding to second stage.

But note x_{2i} can only be endogenous if $E(u_i\varepsilon_i) \neq 0$, that is, the error of the first stage u_i is correlated with the structural error ε_i. Alternatively, note $x_2 i$ can only be endogenous if $E(\vec{u}_i\varepsilon_i) \neq \vec{0}$*

That is, the error of the first stage regression, u_i is correlated with the structural error ε_i. The error u_i has *soaked up* the endogeneity in x_{2i} thus adding it to the structural equation would control for the endogeneity and so get consistent estimates for the other structural parameters.

Consdier the BLP of ε_i given u_i:

$$\varepsilon_i = u_i'\alpha + e_i$$

By definition the error of the BLP is uncorrelated to the dependent ε_i, else it would have been taken into account in the regression.

Substituting this into the strucutural equation, we obtain

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + u_i'\alpha + e_i$$

where:

$$E(u_ie_i) = 0$$

$$E(x_{1i}e_i) = E(x_{1i}(\varepsilon_i - u_i'\alpha)) = 0$$

$$E(x_{2i}e_i) = E((\pi_1'x_{1i} + \pi_2'z_i + u_i)e_i) = E(\pi_2'z_ie_i) = \pi_2E(z_i(\varepsilon - u_i'\alpha)) = 0$$

Thus OLS2' satisfied and the OLS estimates of β_1, β_2, and α should be consistent. But we do not observe u_i so it must first be estimated from the first stage regression before insertion.

Let $\hat{U}$ be the matrix with rows $\hat{u}_i'$. Then by the partitioned regression formula (FW - theorem):

$$\hat{\beta}_{CF} \equiv (X'M_{\hat{U}}X)^{-1}X'M_{\hat{U}}X$$

But $\hat{U} = M_WX_2$ so that:

$$M_{\hat{U}} = I - \hat{U}(\hat{U}'\hat{U})^{-1}\hat{U}' = I - M_WX_2(X_2'M_WX_2)^{-1}X_2'M_W$$

Since X_1 is a part of W, $M_WX_1 = 0$, and

$$M_{\hat{U}}X_1 = X_1 = P_WX_1$$

Further

$$M_{\hat{U}}X_2 = X_2 - M_WX_2(X_2'M_WX_2)^{-1}X_2'M_WX_2 = P_WX_2$$

Therefore

$$M_{\hat{U}}X = P_WX$$

and so

$$\hat{\beta}_{CF} \equiv (X'M_{\hat{U}}X)^{-1}X'M_{\hat{U}}Y = (X'P_WX)^{-1}X'P_WY = \hat{\beta}_{2SLS}$$

$*E(x_{2i}\varepsilon_i) \neq \vec{0} \Rightarrow E[(w_i'\pi + u_i')'\varepsilon] \neq 0 \Rightarrow \pi\cancel{E[w_i\varepsilon]} + E[u_i\varepsilon_i] \neq 0$

$\square$

14.4 Endogeneity and Overidentification test

Endogeneity test: If x_{2i} is not endogenous, then OLS is efficient (BLUE) and 2SLS is not.

Test

$$H_0 : E(x_{2i}\varepsilon_i) = 0 \text{ against } H_1 : E(x_{2i}\varepsilon_i) \neq 0$$

Recall the CF regression:

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + u'_i\alpha + e_i$$

where

$$\alpha = E(u_iu'_i)^{-1}E(u_i\varepsilon_i)(\text{ the coefficient of BLP} for \varepsilon_i given u_i)$$

We have $E(x_{2i}\varepsilon_i \neq 0)$ if and only if $E(u_i\varepsilon_i) \neq 0$. Therefore hypothesis test equivalent to:

$$H_0 : \alpha = 0 \text{ against } H_1 : \alpha \neq 0$$

Therefore a natural test would be the Wald statistic for testing linear restrictions $\alpha = 0$ in the control function regression, with u_i replaced with $\hat{u}_i$. It turns out this replacement des not affect the asymptotic distribution of the test stastistic under the null, and remains $\chi^2(k_2)$ where k_2 is the $dim(\alpha) = dim(x_{2i})$ This follows from a general result on the asymptotic distribution of the OLS estimates of regression coefficients with 'generated' regressions (i.e. the hats consistently estimating the true) H(12-26,12-27) In stata this occurs after estat endoggy WUFF WUFF after ivregress. Het robust s.e. then reported as 'robust regression F' otherwise if default daniel homoskedasticity then reported as 'Wu-Hausman F'

Overidentification test: With $l > k$ (instruments ¿ endoggy regressors) we can test the hypothesis that instruments are exogenous, that is

$$H_0 : E(w_i\varepsilon_i) = 0$$

Let us assume the homoskedasticity, so that $E(\varepsilon_i^2|w_i) = \sigma^2$. Then consider a reduced form regression:

$$\varepsilon_i = w'_i\alpha + e_i$$

, where

$$\alpha = (E(w_iw'_i))^{-1}E(w_i\varepsilon_i)$$

We see that $E(w_i\varepsilon_i) \neq 0$ if and only if $\alpha \neq 0$. We cannot regress ε_i on w_i because we do not observe ε_i. But we can try to replace ε_i with $\hat{\varepsilon}_i$, (the residuals from the 2SLS esimtate of β NOTE this is not the same as the second stage residuals).

Sargan proposed to use nR^2 from this regression as the test stat for H_0 vs H_1:

$$S = nR^2 = n\frac{SSE}{SST} = n\frac{\hat{\varepsilon}'W(W'W)^{-1}W'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}$$

Asymptotic Distribution of S: Note S is invariant wrt transformations $W \to W \times A$ where A is any invertible matrix. Therefore wlog we assume W rotated and scaled so that $W(w_iw'_i) = I_l$ As $n \to \infty$:

$$\frac{1}{\sqrt{n}}W'\varepsilon = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\varepsilon_i \xrightarrow{d} N(0, Var(w_i\varepsilon_i)) = N(0, \sigma^2 I_l) = \sigma N(0, I_l)$$

$$\frac{1}{n}W'W \xrightarrow{p} E(w_iw'_i)^{-1} = I_l$$

and $\frac{1}{n}W'X \xrightarrow{p} E(w_ix'_i) = Q$ where Q is some full column rank matrix. On the other hand:

$$\frac{1}{\sqrt{n}}W'\hat{\varepsilon} = \frac{1}{\sqrt{n}}W'(Y - X\hat{\beta}_{2SLS}) = \frac{1}{\sqrt{n}}W'(Y - X(X'P_WX)^{-1}X'P_WY)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}} W'(\varepsilon + X(X'P_W X)^{-1} X' P_W \varepsilon) \\
&= (I - W'X(X'P_W X)^{-1} X' P_W) \frac{1}{\sqrt{n}} W' \varepsilon \\
&\xrightarrow{d} (I - Q(Q'Q)^{-1} Q') \sigma N(0, I_l)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\hat{\varepsilon}' W (W'W)^{-1} W' \hat{\varepsilon} &= \frac{1}{\sqrt{n}} \hat{\varepsilon}' W (\frac{1}{n} W'W)^{-1} \frac{1}{\sqrt{n}} W' \hat{\varepsilon} \\
&\xrightarrow{d} \sigma^2 N' (I - Q(Q'Q)^{-1} Q') N
\end{aligned}
$$

Lemma 14.4.1. $N'(I - Q(Q'Q)^{-1}Q')N \sim \chi^2(l-k)$

Proof. We have $Q'Q = I_k$ and $Q : l \times k$ where $l > k$ We define Q_c as the $l \times (l-k)$ orthonormal complement matrix such that $[Q\, Q_c]$ together form an $l \times l$ complete orthogonal matrix. Thus $[Q\, Q_c][Q\, Q_c]' = I_l$

$$
\begin{aligned}
&\Rightarrow QQ' + Q_c Q_c' = I_l \\
&\Rightarrow Q_c Q_c' = I_l - QQ'
\end{aligned}
$$

Thus

$$
\begin{aligned}
N'(I - Q(Q'Q)^{-1}Q')N &= N' Q_c Q_c' N \\
&= (Q_c' N)'(Q_c' N)
\end{aligned}
$$

But $Q_c' N \sim N(0, Q_c' I_l Q_c) = N(0, I_{l-k})$ Thus

$$
(Q_c'N)'(Q_c'N) = \sum_{i=1}^{l-k} (z_i)^2 \sim \chi^2(l-k)
$$

□

Thus:

$$
\hat{\varepsilon}' W (W'W)^{-1} W' \hat{\varepsilon} \xrightarrow{d} \sigma^2 \chi^2(l-k)
$$

Finally, $\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} \xrightarrow{p} \sigma^2$ (sim to lec 8 proof) Therefore:

$$
S = n \frac{\hat{\varepsilon}' W (W'W)^{-1} W' \hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}} \xrightarrow{d} \chi^2(l-k)
$$

We reject the null of the instrument exogeneity when s is larger than a critical value of $\chi^2(l-k)$

Note:-

The test cannot be performed in the just-identified situation ($l = k$). Then $W'X$ has full rank and so is thus invertible.

$$
\begin{aligned}
\frac{1}{\sqrt{n}} W' \hat{\varepsilon} &= (I - W'X(X'P_W X)^{-1} X' W (W'W)^{-1}) \frac{1}{\sqrt{n}} W' \varepsilon \\
&= (I - W'X(X'W(W'W)^{-1}W'X)^{-1} X'W(W'W)^{-1}) \frac{1}{\sqrt{n}} W' \varepsilon \\
&= (I - W'X(W'X)^{-1} W'W (X'W)^{-1} X'W(W'W)^{-1}) \frac{1}{\sqrt{n}} W' \varepsilon \\
&(I - I) \frac{1}{\sqrt{n}} W' \varepsilon = 0
\end{aligned}
$$

14.5 Appendix

14.5.1 Chi-squared asymptotic result

Lemma 14.5.1. For $\vec{z} \sim N(0, V)$ We have

$$z'V^{-1}z \xrightarrow{d} \chi^2(p)$$

where p is the number of elements in z.

Proof. As V symmetric we can write its spectral decomposition:

$$V = Q\Lambda Q' = Q\Lambda^{1/2}\Lambda^{1/2}Q'$$

where Q orthogonal and Λ diagonal with eigenvalues $\lambda_1, ..., \lambda_p$.

$$\therefore z'V^{-1}z = z'(Q\Lambda^{1/2}\Lambda^{1/2}Q')^{-1}z$$

$$= ((\Lambda^{1/2}Q)^{-1}z)'((\Lambda^{1/2}Q)^{-1}z)$$

But

$$(\Lambda^{1/2}Q)^{-1}z \sim N(0, (\Lambda^{1/2}Q)^{-1}V(Q'\Lambda^{1/2})^{-1})$$
$$= N(0, (\Lambda^{1/2}Q)^{-1}Q\Lambda^{1/2}\Lambda^{1/2}Q'(Q'\Lambda^{1/2})^{-1})$$
$$= N(0, I_p)$$

Therefore $(\Lambda^{1/2}Q)^{-1}z$ is a vector of p independent standard normals.
Therefore $((\Lambda^{1/2}Q)^{-1}z)'((\Lambda^{1/2}Q)^{-1}z)$ is the sum of p independent standard normals squared, which is $\chi^2(p)$. □

14.5.2 Limited Info Maximum Likelihood

- no finite sample moments the same as 2SLS (so will have outliers)
- but better than 2sls with weak instruments

Recall the same linear regression model:

$$y_i = x_i'\beta + \varepsilon_i$$
$$x_i' = w_i'\pi + u_i'$$
$$\Rightarrow y_i = w_i'\pi\beta + u_i'\beta + \varepsilon_i$$

Let $(y_i, x_i) = Y_i'$

$$\Rightarrow Y_i' = w_i'(\pi\beta, \pi) + (u_i'\beta + \varepsilon_i, u_i')$$

Transposing

$$Y_i = \underbrace{\begin{pmatrix} \beta'\pi' \\ \pi' \end{pmatrix}}_{\Gamma(\beta,\pi)} w_i + \underbrace{\begin{pmatrix} \beta' u_i + \varepsilon_i \\ u_i \end{pmatrix}}_{e_i}$$

$$\Rightarrow Y_i = \Gamma(\beta, \pi)w_i + e_i$$

Assume:

$$e_i | w_i \sim N(0, \Omega)$$

We can then write likelihood function, and maximise wrt parameters to find $\hat{\beta}_{ML} = \hat{\beta}_{LIML}, \hat{\pi}_{ML}$ and $\hat{\Omega}_{ML}$