GMM

1. OLS as a Method of Moment Estimator

Consider a simple cross-sectional case

$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$ $\quad$ i=1,..,N $\quad$ $\boldsymbol{\beta}$ true coeff

If $E(\mathbf{x_i}' u_i) = 0 \;\Rightarrow\; E[\mathbf{x_i}' (y_i - \mathbf{x}_i \boldsymbol{\beta})]=0$ $\quad$ [OLS assumption]

The MM estimator solves the sample moment condition:

$$\frac{1}{N}\sum_i \mathbf{x}_i'\left(y_i - \mathbf{x}_i'\hat{\beta}\right) = 0$$ giving you $\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{y}\right)$ [OLS]

2. Instrumental Variable Estimation

Now assume that some of the **x** variables are correlated with the error term. OLS estimator is inconsistent. We use instrumental variable estimation using say **z** as instruments. Assume number of instruments=L and L ≥ K.

The population moment conditions are: $E(\mathbf{z_i}'u_i)=0$

Then IV estimation solves:

$$\frac{1}{N}\sum_i \mathbf{z}_\mathbf{i}'\left(y_i - \mathbf{x}_i'\tilde{\beta}\right) = 0 \qquad (1)$$

The above involves L equations in K unknowns.

Note if L<K, we can't solve for our estimates.

If L=K, we have K equations and K unknowns and hence have a unique solution giving you

$$\tilde{\beta}=\left(Z'\mathbf{X}\right)^{-1}\left(Z'\mathbf{y}\right). \qquad \text{[Simple IVE]}$$

However, if L > K, then we have more equations than unknowns. One inefficient solution would be to just select K instruments out of the set of

L. But instead, it is better to do something that is more efficient. This is the GMM estimator.

GMM chooses $\tilde{\beta}$ to make (1) as small as possible using quadratic loss. i.e. GMM estimator $\tilde{\beta}$ minimises

$$Q_N(\beta)= \left[\frac{1}{N}\sum_i \mathbf{z}_i'(y_i - \mathbf{x}_i'\beta)\right]' \mathbf{W}_N \left[\frac{1}{N}\sum_i \mathbf{z}_i'(y_i - \mathbf{x}_i'\beta)\right]$$

$\mathbf{W}_\mathbf{N}$ is an LxL matrix of weights which is chosen ‘optimally’ [i.e. giving you the smallest variance GMM estimator].

The solution is

$$\hat{\boldsymbol{\beta}} = \left[\left(\sum_i \mathbf{X_i'Z_i}\right)\mathbf{W}\left(\sum_i \mathbf{Z_i'X_i}\right)\right]^{-1}\left[\left(\sum_i \mathbf{X_i'Z_i}\right)\mathbf{W}\left(\sum_i \mathbf{Z_i'y_i}\right)\right] \text{ (N cancels)}$$

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

This optimal weighting matrix should be a consistent estimate up to a multiplicative constant of the inverse of the variance of the orthogonality conditions:

$\mathbf{W} = \mathbf{\Lambda}^{-1}$ where $\mathbf{\Lambda} \equiv E(\mathbf{Z_i}' u_i u_i' \mathbf{Z_i}) = Var(\mathbf{Z_i}' u_i)$

Can show that the GMM estimator is consistent.

The asymptotic variance of the optimal GMM estimator is estimated using

$$\text{AVAR } \hat{\boldsymbol{\beta}} = \left[\left(\mathbf{X'Z}\right) \left(\sum_i \mathbf{Z_i}' \hat{u}_i \hat{u}_i' \mathbf{Z_i} \right)^{-1} \left(\mathbf{Z'X}\right) \right]^{-1} \quad \text{(N cancels)} \quad (2)$$

In order to generate the residuals, use GMM with $\mathbf{W} = \left[\frac{1}{N} \sum \mathbf{Z_i}'\mathbf{Z_i} \right]^{-1}$

This is the GIVE (2SLS) estimator:

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

This is the same as assuming that

$$\mathrm{Var}(\mathbf{Z_i}'u_i) = E(\mathbf{Z_i}'u_i u_i'\mathbf{Z_i}) = \sigma^2 (\mathbf{Z_i}'\mathbf{Z_i})$$

- Need to be able to invert the above matrices….rank condition!

- In order to generate the residuals, use GMM with $\mathbf{W} = \left[\frac{1}{N} \sum \mathbf{Z_i}'\mathbf{Z_i} \right]^{-1}$
- Can think of the equation (2) as giving you a covar matrix under heteroskedasticity and serial correlation of unknown form.
- When L=K and hence **X'Z** is a square, then W does not matter.
- Stop with the second step (not a lot of efficiency gains in continuing).
- Simulation studies show - very little efficiency gain in doing 2-step GMM even in the presence of considerable heteroskedasticity.

- Additionally, since 2-step GMM depends on 1st step coeff est, the std. error calculations tend to be too small…. Windmeijer provides a correction (now implemented in some software – PcGive).

Test for over identifying restrictions (Sargan/Hansen)

Moment condition: $E(\mathbf{z_i}'u_i)=0$ [has L eqns]

Min: $$Q=\left[\sum_i \mathbf{Z}_\mathbf{i}'\mathrm{u}_\mathrm{i}\right]'\mathbf{W}\left[\sum_i \mathbf{Z}_\mathbf{i}'\mathrm{u}_\mathrm{i}\right]$$

- When L=K, $Q(\hat{\boldsymbol{\beta}})=0$;
- When L>K, then $Q(\hat{\boldsymbol{\beta}})>0$ although $Q(\hat{\boldsymbol{\beta}})\to 0$ in probability.
- So use this to derive the test by comparing the value of the criterion function Q with its expected value under the null that the restrictions are valid.
- This is simple when the optimal weighting matrix is used:

N $Q(\hat{\boldsymbol{\beta}})$ dist asym as a $\chi^2(L-K)$ under H_0. (**Only valid under homosk**.)

- If you only suspect that L_1 are ok but L_2 are not (where $L=L_1 + L_2$) then can use $N(Q-Q_1)$ is asymp $\chi^2(L_2)$ where Q_1 is the minimand when the non-suspect instruments L_1 are used.