Today: GMM

1] Cf 2SLS as minimum distance

2] Cf White's efficient IV estimator

- nonlinear case - can improve upon White-Newey-West

Function $g_i(\beta) = g(z_i, \underset{p\times 1}{\beta})$ $m\times 1$ vector $g: \mathbb{R}^p \to \mathbb{R}^m$

- can think of g as the FOCs.
- $E[g_i(\beta)] = 0$

Let $\hat{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta)$

$\hat{A}$ pd with $\hat{A} \to A$

Then $\hat{\beta}_{GMM} = \underset{\beta}{\operatorname{argmin}}\ \hat{g}(\beta)'\hat{A}\hat{g}(\beta)$

- this is referred to as the generalized method of moments.
- cf $\hat{\beta}_{2SLS} = (y_1 - X_1\delta_1)' P_Z (y_1 - X_1\delta_1)$

$$= \underbrace{(y_1 - X_1\delta_1)'Z}_{\hat{g}'} \underbrace{(Z'Z)^{-1}}_{\hat{A}} \underbrace{Z'(y_1 - X_1\delta_1)}_{\hat{g}}$$

$$\hat{\beta}_{OIV} = (y_1 - X_1\delta_1)'Z\,(Z'\Omega Z)^{-1} Z'(y_1 - X_1\delta_1)$$

If your system is just identified, it doesn't matter what your weight matrix is.
Typically, we have overidentification: $m > p$.

$$y_i = X_i'\beta + \varepsilon_i \qquad E[z_i \varepsilon_i] = 0 \quad i = 1, \dots, N$$

Moment functions: $g_i(\beta) = z_i(y_i - X_i'\beta)$

$$\Rightarrow \hat{g}(\beta) = \frac{1}{n} Z'(y - X\beta)$$

Thus, $\hat{A} = \left[\frac{1}{n}(Z'Z)\right]^{-1}$, which gives us

$$\hat{\beta}_{GMM} = \arg\min \hat{g}(\beta)' \hat{A} \hat{g}(\beta)$$
$$= \arg\min (y - X\beta)' Z (Z'Z)^{-1} Z'(y - X\beta)$$

Want to find asymptotic variance of $\sqrt{n}\,\hat{g}(\beta_0)$

$$Var(\sqrt{n}\,\hat{g}(\beta_0)) \to \Omega \equiv \underbrace{\Lambda_0}_{\text{White matrix}} + \sum_{l=1}^{\infty} (\Lambda_l + \Lambda_l'), \text{ where}$$

$$\Lambda_l = E[g_i(\beta_0)\, g_{i+l}(\beta_0)']$$

Optimal GMM has $\hat{A} \xrightarrow{p} \Omega^{-1}$

Examples: 2SLS with heterskedasticity but no serial correlation. (White's efficient IV estimator)

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} Z_i Z_i' \underbrace{\hat{\varepsilon}_i^2}_{\text{2SLS residuals.}}$$

If you have auto correlation,

$$\hat{\Lambda}_\ell = \frac{1}{n}\sum_{i=1}^{n-\ell} Z_i Z_{i+\ell}' \hat{\varepsilon}_i \hat{\varepsilon}_{i+\ell}$$

Thus, $\hat{\beta}^{GMM} = (X'Z\hat{\Omega}^{-1}Z'X)^{-1}X'Z\hat{\Omega}^{-1}Z'y$

GMM is, in finite samples,

1] biased, which gets worse as $m >> p$, and
- even more biased than 2SLS.

2] test of overid (J-test) is "biased" toward rejection.

3] se's are biased downwards.

Can deal with 1] by using continuously updated estimator "CUE." (CUE is similar to LIML).

$$\hat{\beta} = \min_{\beta} \hat{g}(\beta)[\hat{\Omega}(\beta)]^{-1}\hat{g}(\beta)$$

- still has a bias originating from $\hat{\Omega}$ term.
- median unbiased
- no moments (has not been proven for nonlinear case
- very difficult to estimate
  - objective fcn is "flat" in β.

Want to show that optimal $A = \Omega^{-1}$

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$$

where $V = (G'AG)^{-1} G'A\Omega A'G (G'AG)^{-1}$

$$G = E\left[\frac{\partial g_i(\beta_0)}{\partial \beta}\right]$$

setting $A = \Omega^{-1}$, we get $V = (G'\Omega^{-1}G)^{-1}$

For any other $\tilde{A}$, get $\tilde{V} = (G'\tilde{A}G)^{-1} G'\tilde{A}\Omega\tilde{A}'G(G'\tilde{A}G)^{-1}$

$$\Rightarrow \tilde{V} - V = F\underbrace{(I - H(H'H)^{-1}H')}_{Q_H}F' \geq 0$$

J-test (or Sargan test):

an efficient GMM estimator.

$$T = n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) \xrightarrow{d} \chi(m-p)$$

- can use the objective function as a test statistic.

If $g_i(\beta) = Z_i(y - X_i'\beta)$, we have

$$T = \frac{1}{n}\hat{\varepsilon}' Z\hat{\Omega}^{-1}Z'\hat{\varepsilon}$$

Can do Wald test, sum of squared residuals test (cf LR), and LM test.

Two examples:

Intertemporal CAPM

- C_i is consumption at time i
- R_i is asset return
- $u(c, \gamma_0)$, Z_i

$\to$ Cf $MRS_{i,i+1} = \frac{1}{\alpha R_i}$ = price ratio

FOCs: $$y_i(\beta) = Z_i \Big[\underbrace{\frac{R_i \cdot \overbrace{\alpha}^{\text{discounted}} \cdot u_c(c_{i+1}, \gamma)}{u_c(c_i, \gamma)} - 1}_{\text{can be seen as a residual}} \Big]$$

use lags of c_{i+1} and R_i

Dynamic Panel Data

$$E^*[y_{it} \mid y_{it-1}, \ldots, y_{i,0}, \alpha_i] = \overbrace{\beta_0}^{\text{very near 1}} y_{i,t-1} + \overbrace{\alpha_i}^{\text{fixed effect}}$$

Let $\eta_{it} = y_{it} - E^*[\cdot]$. By orthogonality conditions,

$$E[y_{i,t-j}\,\eta_{it}] = 0 \quad (1 \le j \le t) \quad t = 1, \ldots, T \ (T=5)$$

Let $\Delta y_{it} = y_{it} - y_{i,t-1}$

Orthogonality: $$E[y_{i,t-j} \underbrace{(\Delta y_{it} - \beta_0 \Delta y_{i,t-1})}_{\text{first differences of regression model}}] = 0 \quad 2 \le j \le t$$

min bias:
- \# instruments $\simeq 22$
- optimal uses 1 instrument
- best to use is long differences

since we are local to unit circle,

If want to minimize MSE: • want to use 3-4 instruments

Lectures on weak instruments available on web.

$Y_1 = Y_2\beta + Z_1\gamma + u$

$Y_2 = Z_1\pi_1 + Z_2\pi_2 + V$

$Q_{Z_1} = I - Z_1'(Z_1'Z_1)^{-1}Z_1'$

$$\left.\begin{array}{l} y_1 = \beta y_2 + \varepsilon \\ y_2 = Z\pi_2 + v_2 \end{array}\right\} \text{ after partialing out the } Z_1\text{'s.}$$

$\begin{pmatrix} \varepsilon_{1i} \\ v_{2i} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_{vv} \end{bmatrix}\right)$ assume $\sigma_{\varepsilon\varepsilon} = \sigma_{vv} = 1$ and $\sigma_{\varepsilon v} = \rho$

• $Var(y_2) = \frac{1}{1-R^2}$

assume $\pi_2 \neq 0 \Rightarrow$ identification

• $E\left[\frac{Z'\varepsilon}{n}\right]$

Then

$$E[\hat{\beta}_{2SLS} - \beta] \approx \frac{K\sigma_{\varepsilon v}}{n\Theta} \approx \frac{K\sigma_{\varepsilon v}}{nR^2\underbrace{var(y_2)}_{=\frac{1}{1-R^2}}} = \frac{K\rho(1-R^2)}{nR^2}$$

If $K \to \infty$ as $n \to \infty$, and $\frac{K}{n} \to \alpha$, 2SLS will be inconsistent. (Bekker asymptotics)

Here, $\Theta = \frac{\pi'Z'Z\pi}{n}$

$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \rightarrow N\left(\frac{\sigma_{\varepsilon v}\,\mu}{\Theta}, V_{2SLS}\right)$, $V_{2SLS} = \frac{\sigma_{\varepsilon\varepsilon}}{\Theta}$

concentration parameter is $\frac{nR^2}{1-R^2}$

We will look at the 'reverse' regression

$y_2 = \delta y_1 + \tilde{\varepsilon}$

$y_1 = Z\pi_1 + v$

$\delta = \frac{1}{\beta}$, $\tilde{\varepsilon} = \frac{1}{\beta}$

Nagar: variance is $\frac{1-R^2}{R^2} + \frac{K(1+\rho^2)}{n}\left(\frac{1-R^2}{R^2}\right)^2$
JN

Fuller, LIML variance is $\frac{1-R^2}{R^2} + \frac{K(1-\rho^2)}{n}\left(\frac{1-R^2}{R^2}\right)^2$

- Fuller/LIML do better than Nagar/JN
- Fuller estimator has moments.