Lecture 20: GMM

- Key: Set sample moments equal to theoretical moments and solve parameters.

- Generalized moments: Expectations of functions

Eg.

$$\begin{aligned} E(y-\mu) &= 0 \\ \text{set } \frac{1}{n}\sum(y_i-\hat{\mu}) &= 0 \Leftrightarrow \hat{\mu}=\bar{y} \end{aligned}$$

Regression

$E(y - X\beta) = 0$

set $\frac{1}{n}\sum(y_i - X_i\beta) = 0$?

- Too many solutions

- Suppose we group into K groups

- Solve simultaneously

- Illustrates arbitrariness of choice of moment conditions

- A better moment condition:

$$E(X'(y - X\beta)) = 0$$

solve

$$X'(y - X\hat{\beta}) = 0 \text{ for } \hat{\beta} = (X'X)^{-1}X'y$$

a GMM estimator!

Often we have overidentifying restrictions

$$\begin{aligned} E(W'(y - X\beta)) &= 0 \\ W &: n \times p, \, p > k \end{aligned}$$

Then $W'(y - X\hat{\beta}) = 0$ is p linear equations in K unknonws.

Write $W'y = W'X\beta + \varepsilon$ and do GLS.

If

$$V(y) = \sigma^2 I,\ V(\varepsilon) = \sigma^2 W'W.$$

Then

$$\hat{\beta}_{GLS} = \hat{\beta}_{GMM} = (X'W(W''W)^{-1}W'X)^{-1}X'W(W'W)^{-1}$$

Which is the solution to

$$\min(y - X\hat{\beta})'W(W'W)^{-1}W'(y - X\hat{\beta}).$$

With $V(y) = V$:

$$\hat{\beta} = (X'W(W'VW)^{-1}W'X)^{-1}X'W(W'VW)^{-1}W'y.$$

Solution to

$$\min(y - x\hat{\beta})'W(W'VW)^{-1}W'(y - X\hat{\beta})$$

Distribution Theory

- Assume $\frac{W'\varepsilon}{\sqrt{n}} \rightarrow N(0, W'VW)$ plausible?

Then

$$\hat{\beta} \approx \beta + (X'W(W'VW)^{-1}X'W)^{-1}X'W(W'VW)^{-1}W'\varepsilon$$

so

$$\sqrt{n}(\hat{\beta} - \beta) \tilde{} N(0, (X'W(W'VW)^{-1}W'X)^{-1})$$

The trick is choosing the moments (or instruments)

- More cannot hurt.

Asymptotically we have:

$$E(W'(y - X\beta)) = 0$$

Instead use:

$$\begin{aligned} E(A'W'(y - X\beta)) &= 0 \\ A \quad : \quad & m \times p, \; m < p \end{aligned}$$

(Fewer conditions)

Then

$$\begin{aligned} V(\beta_A) &= [X'WA(A'W'VWA)^{-1}A'W'X]^{-1} \\ V(\hat{\beta})^{-1} - V(\beta_A)^{-1} &= X'W[(W'VW)^{-1} - A(A'W'VWA \end{aligned}$$

Letting

$$\begin{aligned} CC' &= (W'VW)^{-1} \\ V(\hat{\beta})^{-1} - V(\beta_A)^{-1} &= X'WC[I - C^{-1}A(A'C'^{-1}C^{-1}A)^{-} \end{aligned}$$

p.s.d., so

$$V(\beta_A) \geq V(\hat{\beta})$$

- Note that more may not help if conditions are chosen right.

$$(W' = X' \text{ for OLS})$$

- Look more closely at the case $V(y) = I$: If we let $W = [X \;\; Z]$

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

(after a little work)

- First factor is $(X'X)^{-1}$

- Note $W(W'W)^{-1}W'X$ is the matrix of predicted values from the regression of X on W (namely, X itself)

Nonlinear Models

$E(W'f(\theta)) = 0$, W: $n \times p$, $f(\theta)$: $n \times 1$, θ: $k \times 1$

Same principle as linear:

Let $F = f_\theta$: $n \times k$ and $E(ff') = V$

Solving by nonlinear GLS minimizes

$$f'W(W'VW)^{-1}W'f$$

and

$$\sqrt{n}(\hat{\theta} - \theta) \tilde{} N(0, n(F'W(W'VW)^{-1}W'F)^{-1})$$

choice of instruments?

Try $W = V^{-1}F$

Then $V(\hat{\theta}) = (F'V^{-1}F(F'V^{-1}VV^{-1}F)^{-1}F'V^{-1}F)^{-1} = (F'V^{-1}F)^{-1}$

- Smallest in this class (why?) (note there are only k of these)

Generalization

- We have only considered covariances so far.

Consider G: $N \times p$ where we let each column represent different functions of data. We have N observations on each function, sothe moment conditions become

$$E(f_{ti}(y_t, \theta)) = 0, \;\; i = 1, ..., k \;\; t = 1, ..., N.$$

Where earlier we had specified

$$f_{it}(y, \theta) = w_{it} f_{it}$$

- Now GLS can be applied:

Resulting in $\min 1' G A G' 1$

That is, the moment conditions are $1'G = 0$ and A is a p.d. weighting matrix.

For efficiency, A should be equal to (well, proportional to) $V(1'G)^{-1}$ (this is $E(FF')$ in earlier notation).

To develop intuition for this, consider the linear case where G has elements

$$f_{ti} = w_{ti}(y_t - x_t\beta)$$

And the $1'$ just sums over observations, so $V(1'G)$ is just $(W'W)$.

Note $1'G$ is taking the place of $(y - X\beta)'W$.

Asymptotic Distribution of GMM

Write $Q = 1'GAG'1$, the function to be minimized to calculate the GMM estimator.

Taking the Taylor expansion of the first order condition $Q_\theta = 0$

$$0 = Q_\theta(\theta) + Q_{\theta\theta}(\theta^* - \theta)$$

and just as in the ML case, we solve for the vector of estimation errors as

$$(\theta^* - \theta) = -(Q_{\theta\theta})^{-1}Q_\theta$$

and apply a LLN to the second derivative matrix and a CLT to the "scores."

The notation can get cumbersome here, but let $g_{ij} =$ the derivative of the ith column of $1'G$ with respect to θ_j, and let $g = \{g_{ij}\}$ be the associated matrix. Then

$$V(n^{1/2}(\theta^* - \theta)) = (g'Ag)^{-1}g'AEFF'Ag(g'Ag)^{-1}.$$

This comes from evaluating the derivatives, bringing in the scaling factors in n and generally simplifying. The result should look familiar.

Note that when $A = EFF'$, the formula simplifies to

$$V(n^{1/2}(\theta^* - \theta)) = (g'Ag)^{-1}.$$

As in the usual GLS case!

Questions: How many moments to use? Note the FOC only use k.... Nice analogy with 2SLS - with lots of IV, still with 2SLS, we reduce to the "just identified" case by using the optimal linear combination of instruments.

Discussion?

In fact k moments are sufficient for efficiency if there are k parameters. What are the k moments?

The real use for GMM is when the LF is too complicated or unknown (better, not plausibly known).

Conditional GMM

Basically generates more moment conditions. When we take $X'(y - X\beta) = 0$, we are imposing that the error is uncorrelated with X.

But the property may be stronger, e.g. that $E((y - X\beta)|X) = 0$. This implies that $EX'(y - X\beta) = 0$ but also that any other function of X is uncorrelated with $(y - X\beta)$. This comes up a lot in RE modeling and provides a source of lots of moment conditions.

Note tradeoff between introduction of noise and gains from using more moments.

Conclusion

- GMM requires fewer assumptions than ML
- Can be somewhat arbitrary
- Can be very inefficient relative to ML