ECON 6200 : Section 6

2/28/25

Agenda

- Review of GMM
- Efficient GMM
  - ↳ under homoscedasticity
- TSGMM Procedure + Intuition
- Overidentification Test
- 2013 Midterm Q1
- 2023 Midterm Q2

GMM Estimator

The GMM estimator is defined as

$$\hat{\theta}(W) = \underset{\theta}{\operatorname{argmin}}\, J_n(\theta)$$

where $J_n(\theta) = n\,\bar{g}_n(\theta)'\,W\,\bar{g}_n(\theta)$, and $\bar{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\cdot\,;\theta)$

In this context of this class, we focus on the linear case where:

$$\bar{g}_n(\beta) = \frac{1}{n}\sum g_i(\beta) = \frac{1}{n}\sum_i^n Z_i(Y_i - X_i'\beta) = \underbrace{\frac{1}{n}(Z'(Y-X\beta))}_{\text{cannot set to 0 bc model overident}}$$

sample moment condition

⇓ get as close as possible to 0

$$g_i(\beta) = Z_i(Y_i - X_i'\beta) = 0$$

population moment condition

in theory

GMM Assumptions

1. We observe i.i.d. realizations (Y_i, X_i, Z_i), $i = 1, \ldots, n$.
2. $\mathbb{E}(Z(Y - X'\beta)) = 0$.
3. $\mathbb{E}(|Y^4|) < \infty$,
4. $\mathbb{E}(\|X\|^4) < \infty$,
5. $\mathbb{E}(\|Z\|^4) < \infty$,
6. $\boldsymbol{Q} \equiv \mathbb{E}(ZX')$ has full rank k,
7. $\boldsymbol{W}$ is positive definite,
8. $\Omega \equiv \mathbb{E}(ZZ'\varepsilon^2)$ is positive definite.

Efficient GMM

(Hansen pg 429)

We say that an estimator is "efficient" if it has the smallest asymptotic variance in the class of GMM estimators with this set of moment conditions.

↳ weak concept of optimality

↳ only consider alternative weight matrices $\hat{W}$

BUT, turns out GMM estimator is semiparametrically efficient
(ie: attains best possible asymptotic variance)

Goal: If $\hat{\beta}_{GMM}$ = efficient GMM estimator

$\tilde{\beta}_{GMM}$ = some other GMM estimator

Show that

$$\text{avar}[\hat{\beta}_{GMM}] \leq \text{avar}[\tilde{\beta}_{GMM}]$$

Last week, we showed the asymptotic distribution of the GMM estimator to be:

$$\sqrt{n}(\hat{\beta}_{GMM}(\boldsymbol{W}) - \beta) \xrightarrow{d} N\Big(0, \underbrace{(\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{W}\Omega\boldsymbol{W}\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q})^{-1}}\Big).$$

Let

$$V = \underbrace{(Q'WQ)^{-1}Q'W}_{A'}\,\Omega\,\underbrace{WQ(Q'WQ)^{-1}}_{A}$$

$$= A'\Omega A \quad \text{where } A = WQ(Q'WQ)^{-1}$$

$$V^{*} = \underbrace{(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}}_{B'}\,\Omega\,\underbrace{\Omega^{-1}Q(Q'\Omega^{-1}Q)^{-1}}_{B}$$

$$= B'\Omega B \quad \text{where } B = \Omega^{-1}Q(Q'\Omega^{-1}Q)^{-1}$$

Note: If I were to simplify $B'\Omega B$

$$B'\Omega B = (Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}\cancel{\Omega}\,\cancel{\Omega^{-1}}Q(Q'\Omega^{-1}Q)^{-1}$$

$$= (Q'\Omega^{-1}Q)^{-1}\cancel{Q'}\,\cancel{\Omega^{-1}}\,\cancel{Q}\,\cancel{Q^{-1}}\,\cancel{\Omega}\,\cancel{(Q')^{-1}}$$

$$= (Q'\Omega^{-1}Q)^{-1}$$

Then

$$B'\Omega A = \underbrace{(Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}}_{B}\,\Omega\,\underbrace{WQ(Q'WQ)^{-1}}_{A}$$

$$= Q^{-1}\Omega\,\cancel{(Q')^{-1}}\,\cancel{Q'}\,\cancel{\Omega^{-1}}\,\cancel{\Omega}\,\cancel{W}\,\cancel{Q}\,\cancel{Q^{-1}}\,\cancel{W^{-1}}(Q')^{-1}$$

correct

$$= (Q'\Omega^{-1}Q)^{-1} = B'\Omega B$$

Thus,

$$B'\Omega(A-B) = B'\Omega A - B'\Omega B = 0 \qquad (*)$$

Now, to prove that $V \geq V^*$

$$V = A'\Omega A$$

$$= (B+(A-B))'\Omega(B+(A-B))$$

$$= \underbrace{B'\Omega B}_{V^*} + \underbrace{(A-B)'\Omega B}_{=0 \text{ by } (*)} + \underbrace{B'\Omega(A-B)}_{=0 \text{ by } (*)} + \underbrace{(A-B)'\Omega(A-B)}_{\text{p.s.d}}$$

since Ω p.s.d $\Rightarrow$ quadratic form will be psd

$$\geq V^*$$

Simplification Under Homoskedasticity

Assume homoskedasticity, $E[\varepsilon^2 | Z] = \sigma^2$

Then $\Omega = E[ZZ'\varepsilon^2] = \sigma^2 E[ZZ']$

oracle $W = \Omega^{-1}$

$$
\begin{aligned}
\hat{\beta}_{GMM}(\Omega^{-1}) &= (X'ZWZ'X)^{-1}(X'ZWZ'Y) \\
&= (X'Z\,\sigma^{-2}E[ZZ']^{-1}Z'X)^{-1}(X'Z\,\sigma^{-2}E[ZZ']^{-1}Z'Y) \\
&= (X'Z\,E[ZZ']^{-1}Z'X)^{-1}(X'Z\,E[ZZ']^{-1}Z'Y) \\
&= \beta_{TSLS}
\end{aligned}
$$

We can estimate $E[ZZ']$ by $E_n[ZZ']$ by $E_n[ZZ'] = \frac{1}{n}Z'Z$

$$
\hat{\beta}_{GMM}(\hat{\Omega}^{-1}) = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) = \hat{\beta}_{TSLS}
$$

Note: TSLS estimator is efficient under homoscedasticity

Two Stage Efficient GMM

Goal: Find a $\hat{W}$ s.t. $\hat{W} \xrightarrow{p} W^* = \Omega^{-1}$

(Assuming no misspecification)

Issue: Don't know Ω^{-1}

Recall $\Omega^{-1} = var(\frac{1}{n}Z'\varepsilon)$
$= E[g_i(\beta) g_i(\beta)']^{-1}$
var-cov of moment conditions

Recall our goal of finding $\hat{W}$ is to find:

Issue: $\hat{W}$ $\hat{\beta}_{GMM}$ start?

$$\hat{\beta}_{GMM} = \arg\min \bar{g}_n(\beta) \hat{W}' \bar{g}_n(\beta)$$

Stage 1 Perform TSLS (as $\hat{\beta}_{TSLS} \xrightarrow{p} \beta$), so my $\hat{W}_1 = (Z'Z)^{-1}$.

$\Rightarrow$ $\hat{\beta}_{TSLS}$ is a first stage estimate of β_{GMM}

$\Rightarrow$ Get residuals $\tilde{e}_i$

$$Y = X\hat{\beta}_{TSLS} - \tilde{e}$$

$\Rightarrow$ Calculate my updated weighting matrix

$$\hat{W} = E_n[ZZ'\tilde{e}^2]^{-1}$$

Note: First stage avar (in this example avar($\hat{\beta}_{TSLS}$)) is not efficient because it does not account for heteroscedasticity and correlations in the moment conditions

Stage 2 Use estimated weights to get efficient GMM estimator

$$\hat{\beta}_{TSGMM} = \hat{\beta}(\hat{W})$$
$$= (X'Z\hat{W}Z'X)^{-1}(X'Z\hat{W}Z'Y)$$

Comment: We know $\hat{W} \xrightarrow{P} W$ bc

1) $\hat{\beta}_{TSLS} \xrightarrow{P} \beta$

2) $\bar{g}_n(\beta) \xrightarrow{P} g(\beta)$

3) $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\beta})\, g_i(\hat{\beta})' \xrightarrow{P} \Omega = E[g_i(\beta)\, g_i(\beta)']$

4) $\hat{W} = \mathbb{E}_n(ZZ'\tilde{e}^2)^{-1} \xrightarrow{P} W = E(ZZ'e^2)$

which we know that $W = \Omega^{-1}$ achieves the lowest possible asymptotic variance.

Why do we perform TSGMM?

- don't know $W = \Omega^{-1}$, so we perform TSGMM to get $\hat{W} = \hat{\Omega}^{-1}$ in practice

Comment: We can also define our 1st stage β using the weighting matrix $\hat{W}_1 = I$

Possible reasons may be that we suspect heteroscedasticity or correlation in the errors

Overidentification Test

Issue: In the overidentified model ($\ell \geq k$), we may not be able to find a β such that $E[Z\varepsilon] = 0$

↳ overidentifying restrictions are testable

Hence, we can test joint validity of moments.

(ie: $H_0 : E[Z\varepsilon] = 0$)

Note: Be careful with interpretation

Overidentification test is a partial check.

↳ If pass ⟹ conditionally on enough of my instruments are valid, there is no strong evidence that the remaining instruments are invalid

Thm

Under the assumptions above,

$$J_n \equiv J(\hat{\beta}_{TSGMM}) \xrightarrow{d} \chi^2_{\ell - k}$$

2013 Q1 Midterm

1 Consider the following assumptions about a model characterized by outcome Y, (potentially endogenous) regressors X and instruments Z:

1. i.i.d realizations of $(Y_i, X_i, Z_i), i = 1, \ldots, n$
2. $Y = X'\beta + \varepsilon$
3. $E[Z(Y - X'\beta)] = 0$
4. $E[|Y^4|] < \infty$
5. $E[||X^4||] < \infty$
6. $E[||Z^4||] < \infty$
7. $Q \equiv E[ZX']$ has maximal rank
8. $\Omega \equiv E[ZZ'\varepsilon^2]$ is positive definite

Consider the estimator defined by

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (Z_i (Y_i - X_i'\beta))' \left(\sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \sum_{i=1}^{n} (Z_i (Y_i - X_i'\beta)).$$

1.1 Solve for the estimator in closed form.

1.2 Derive the estimator's asymptotic distribution.

1.3 Is the estimator the best you can do under the conditions? If not, provide conditions that make this estimator best.

1.4 Under the above assumptions, the object

$$\sum_{i=1}^{n} (Z_i (Y_i - X_i'\beta))' \left(\sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \sum_{i=1}^{n} (Z_i (Y_i - X_i'\beta))$$

is of which stochastic order?

1.1 & 1.2) Refer to last weeks notes

1.3) Note that this problem defines W for $\hat{\beta}(W)$ as $W = (Z'Z)^{-1}$.
If you remember the tree I make you draw every week, you'll remember that $W = (Z'Z)^{-1}$ is <u>only</u> the efficient weighting matrix if we have conditional homoscedasticity (ie: TSLS).

1.4) Recall $J_n(\beta) = \sum (Z_i(Y_i - X'\beta))' \left(\sum Z_i Z_i'\right)^{-1} \sum (Z_i(Y_i - X_i'\beta))$

which is also the test statistic to the overidentification test.

$\Rightarrow$ Over ID. test

$$J_n(\beta) \to \underbrace{\chi^2_{\ell-k}}_{\text{obviously a distribution}} = O_p(1)$$

2023 Midterm Q2

2 This question is about linear GMM. Suppose that there are $L = 8$ moments and $K = 5$ parameters. Throughout this question, suppose that data are i.i.d., that moments exist as needed, and furthermore that:

$$\boldsymbol{Y} = \boldsymbol{X}\theta + \varepsilon$$
$$\mathbb{E}\left(\boldsymbol{Z}'(\boldsymbol{Y} - \boldsymbol{X}\theta)\right) = \mathbf{0}$$
$$\boldsymbol{Q} \equiv \mathbb{E}(\boldsymbol{Z}'\boldsymbol{X}) \text{ is of full column rank.}$$
$$\Omega \equiv \mathbb{E}\left(\boldsymbol{Z}'\boldsymbol{Z}\varepsilon^2\right) \text{ is nonsingular.}$$

You may assume the following theorem: Let $\hat{\theta}(\hat{\boldsymbol{W}})$ be the GMM estimator and let $\hat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$, where $\boldsymbol{W}$ is symmetric with full rank. Then (under above assumptions) $\hat{\theta}(\hat{\boldsymbol{W}}) \xrightarrow{p} \theta$ and $\sqrt{n}(\hat{\theta}(\hat{\boldsymbol{W}}) - \theta) \xrightarrow{d} \mathcal{N}(0, \text{avar}(\hat{\theta}(\hat{\boldsymbol{W}})))$.

A researcher wants to estimate θ by GMM with weighting matrix

$$\tilde{\mathbf{W}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_5 & \underset{5\times3}{\mathbf{0}} \\ \underset{3\times5}{\mathbf{0}} & \underset{3\times3}{\mathbf{0}} \end{bmatrix}.$$

2.1 Write down the optimization problem that defines $\hat{\theta}(\tilde{\mathbf{W}})$ in data matrix notation. Show in closed form that $\hat{\theta}(\tilde{\mathbf{W}}) = \left(\boldsymbol{X}'\boldsymbol{Z}\tilde{\boldsymbol{W}}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}\tilde{\boldsymbol{W}}\boldsymbol{Z}'\boldsymbol{Y}$.

2.2 The above theorem does not immediately apply to $\hat{\theta}(\tilde{\mathbf{W}})$! Why not?

2.3 However, $\hat{\theta}(\tilde{\boldsymbol{W}})$ is consistent and asymptotically normal anyway. Why? (Hint: Is the closed form from 2.1 the most succinct expression of $\hat{\theta}(\tilde{\boldsymbol{W}})$?)

2.4 Is it possible that $\hat{\theta}(\tilde{\mathbf{W}})$ is efficient in the usual sense of attaining the lower bound on asymptotic variance for *all* data generating processes that fulfil our assumptions? Is it possible that $\hat{\theta}(\tilde{\mathbf{W}})$ is efficient for *some* such data generating processes? (Just give general answers, do not try to construct examples.)

2.5 Is the estimator

$$\tilde{\theta} = \hat{\theta}(\hat{\Omega}^{-1})$$
$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}(Z_i Z_i'(Y_i - X_i'\hat{\theta}(\tilde{\boldsymbol{W}})^2)) \qquad \left[= \frac{1}{n}\boldsymbol{Z}'\hat{\varepsilon}\hat{\varepsilon}'\boldsymbol{Z}\right]$$

efficient? Explain why or why not.

2.1) See last week's discussion

2.2) $\tilde{W}$ not full rank!

2.3) Originally we have

$$(L = 8 \text{ moments}) \geq (K = 5 \text{ params})$$

However, the last 3 columns of my $\tilde{W}$ are all 0s

$\Rightarrow$ Essentially giving 0 weight to last 3 moments (ie: dropping last 3 moments)

If I drop the last 3 moments

$$(L' = 5 \text{ moments}) = (K = 5 \text{ params})$$

$\Rightarrow$ just-identified!

$\Rightarrow$ basically just an IV problem

$$\Rightarrow \hat{\theta}(\tilde{W}) = (\underset{n \times 5}{Z}' X)^{-1} (\underset{n \times 5}{Z}' Y)$$

2.4) No, in general $\hat{\theta}(\tilde{W})$ is only efficient if $\tilde{W} = \Omega^{-1}$.

This weighting matrix is only efficient when we know the likelihood & the first 5 conditions are the score equations.

2.5) Yes! This is the TSGMM (assuming we choose an initial $\hat{\beta} \xrightarrow{p} \beta$)