ECON 6200 : Section 5

2/21/25

Agenda

- GMM Intuition
- Review Moments
- GMM Definition
- Estimation
    - ↳ Derivation
- Relationship w/ TSLS, IV, OLS
- GMM Asymptotics

GMM Intuition

GMM generalizes the classical method of moments by allowing for $\underbrace{\text{more equations than unknown parameters}}_{l \geq k}$

Like method of moments, GMM starts by postulating moment conditions, which we assume to be true in the population
↳ ie: there exists population parameters that solve the system of moment equation

How is GMM more "general"?

1) Overidentification ($l \geq k$)
   ↳ # moment eq ≥ # of regressors
   ↳ cannot solve for a exact solution AND meeting our sample moment conditions

   Solution: Minimize "distance" of sample moment conditions to 0

2) Heteroscedasticity
   ↳ Previously, OLS, IV, and (implicitly) TSLS assume homoscedasticity

Review: Method of Moments

- Random vectors D
- Parameter of interest $\theta \in \mathbb{R}^k$, for some $k < \infty$
- Known function $g(\cdot)$

We have a moment condition that looks like

$$E[g(D;\theta)] = 0$$

The method of moments estimator $\hat{\theta}$ can be constructed by solving the sample analog of the moment condition

$$\frac{1}{n}\sum_{i=1}^{n} g(D;\hat{\theta}) = 0$$

GMM Formal Definition

Given random vectors $Y \in \mathbb{R}$, $X \in \mathbb{R}^k$, $Z \in \mathbb{R}^\ell$, and moment conditions

$$E[g(Y, X, Z; \theta)] = 0$$

where $\theta \in \mathbb{R}^k$ and $g(\cdot)$ is a known smooth function mapping into $\mathbb{R}^\ell$, $\ell \geq k$.

Moreover, for now, we assume $g(\cdot)$ is <u>linear</u>

Hence, in the context of this class / overidentified IV, we can write the moment condition as

$$E[Z(Y - X'\beta)] = 0$$

Estimation

The GMM estimator is defined as

weights

$$\hat{\theta}(\underset{\ell \times \ell}{W}) = \underset{\theta}{\arg\min}\, J_n(\theta)$$

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

scaling factor to derive convergence in distribution
↳ doesn't affect minimization problem

W: symmetric matrix

Derivation of $\hat{\beta}_{GMM}$

Under the linear case

$$J_n(\beta) = n(\underbrace{Z'(Y-X\beta)}_{\bar{g}_n(\beta)})' W (\underbrace{Z'(Y-X\beta)}_{\bar{g}_n(\beta)})$$

$$= n(Z'Y - Z'X\beta)' W (Z'Y - Z'X\beta)$$

Some algebra:

$$(Y'Z - \beta'X'Z) W (Z'Y - Z'X\beta)$$
$$= (Y'ZW - \beta'X'ZW)(Z'Y - Z'X\beta)$$
$$= Y'ZWZ'Y - Y'ZWZ'X\beta - \beta'X'ZWZ'Y + \beta'X'ZWZ'X\beta$$

Recall

- if a, b $k\times 1$ vectors $\dfrac{\partial a'b}{\partial b} = \dfrac{\partial b'a}{\partial b} = a$

FOC

$[\beta]$ $\quad \cancel{2}X'ZWZ'X\hat{\beta} - \cancel{2}X'ZWZ'Y = 0$

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

Special Cases

- OLS estimator

  If $X = Z$
  $$\hat{\beta} = (X'XWX'X)^{-1} X'XWX'Y$$
  $$= (X'X)^{-1}\cancel{W^{-1}}\cancel{(X'X)^{-1}}\,\cancel{X'X}\,\cancel{W}X'Y$$
  $$= (X'X)^{-1}(XY)$$

- IV estimator

  If $(X'Z)$ square and invertible (ie: $\ell = k$)
  $$\hat{\beta} = (Z'X)^{-1}\cancel{W^{-1}}\cancel{(X'Z)^{-1}}\,\cancel{X'Z}\,\cancel{W}Z'Y$$
  $$= (Z'X)^{-1}Z'Y$$

- TSLS estimator

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

  vs.

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

GMM Asymptotics

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, $\Rightarrow$ $l \geq k$
7. $\boldsymbol{W}$ is positive definite,
8. $\Omega \equiv \mathbb{E}(ZZ'\varepsilon^2)$ is positive definite.

(3–5:) need to calculate var

Asymptotic Distribution

The following algebra generalizes previous algebra for OLS:

$Y = X\beta + \varepsilon$

$$
\begin{aligned}
\hat{\beta}_{GMM}(\boldsymbol{W}) &= (\boldsymbol{X'ZWZ'X})^{-1}\boldsymbol{X'ZWZ'Y} \\
&= \beta + (\boldsymbol{X'ZWZ'X})^{-1}\boldsymbol{X'ZWZ'}\varepsilon \\
&= \beta + (\tfrac{1}{n}\boldsymbol{X'ZW}\tfrac{1}{n}\boldsymbol{Z'X})^{-1}\tfrac{1}{n}\boldsymbol{X'ZW}\tfrac{1}{n}\boldsymbol{Z'}\varepsilon \\
&= \beta + (\mathbb{E}(XZ')\boldsymbol{W}\mathbb{E}(ZX'))^{-1}\mathbb{E}(XZ')\boldsymbol{W}\tfrac{1}{n}\boldsymbol{Z'}\varepsilon + o_P(1) \\
&= \beta + (\boldsymbol{Q'WQ})^{-1}\boldsymbol{Q'W}\underbrace{\tfrac{1}{n}\boldsymbol{Z'}\varepsilon}_{\xrightarrow{p} 0} + o_P(1)
\end{aligned}
$$

$\pm E[XZ']$

conv in prob

(whole term) $\xrightarrow{p} 0$

By WLLN, $\frac{1}{n}Z'\varepsilon \xrightarrow{p} E[Z'\varepsilon] = \underbrace{E[Z'(Y - X\beta)] = 0}_{\text{true by assumptions}}$

1) Consistency

$$\hat{\beta}_{GMM}(\boldsymbol{W}) - \beta \quad \xrightarrow{p} \quad 0,$$

var(AX)
= A var(X) A'

2) Asymptotic Normality

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

By CLT, $\sqrt{n}\left(\frac{1}{n}Z'\varepsilon\right) \xrightarrow{d} N(0, \Omega)$

$$\begin{aligned}
\hat{\beta}_{GMM}(\boldsymbol{W}) &= (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{Y} \\
&= \beta + (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\varepsilon \\
&= \beta + (\tfrac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\tfrac{1}{n}\boldsymbol{Z}'\boldsymbol{X})^{-1}\tfrac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\tfrac{1}{n}\boldsymbol{Z}'\varepsilon \\
&= \beta + (\mathbb{E}(XZ')\boldsymbol{W}\mathbb{E}(ZX'))^{-1}\mathbb{E}(XZ')\boldsymbol{W}\tfrac{1}{n}\boldsymbol{Z}'\varepsilon + o_P(1) \\
&= \beta + (\boldsymbol{Q}'\boldsymbol{W}\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{W}\underline{\tfrac{1}{n}\boldsymbol{Z}'\varepsilon} + o_P(1)
\end{aligned}$$

Since $E[Z'\varepsilon] = 0$

$$\sqrt{n}\Big(\underbrace{\tfrac{1}{n}Z'\varepsilon}_{\text{sample avg}} - \underbrace{E[Z'\varepsilon]}_{\text{expection}}\Big) \xrightarrow{d} N(0, \Omega)$$