Lec 3: Consistency of GMM

Eric Hsienchen Chu*

Spring, 2024

(⊛) Suggested readings: Newey and McFadden (1994), Ch2.5.

Overview. In Lec 2 we introduce the Consistency of MLE. We now turn to discuss **Consistency of GMM**. Recall that we use GMM when having overidentification– We don't need the "quadratic" form in *Just–ID* case since GMM collapses to MLE.

1 GMM

Assume $\mathbb{E}[g(Z;\theta)] = 0$, where $g(\zeta;\theta) = \begin{bmatrix} g_1(\zeta;\theta) \\ \vdots \\ g_r(\zeta;\theta) \\ \vdots \\ g_k(\zeta;\theta) \end{bmatrix}$, $\theta \in \Theta \subseteq \mathbb{R}^r,\ k > r.$

We have k equations with r unknowns $(\theta) \implies$ Overidentification

Definition 1.1 (GMM)**.** A GMM estimator is defined as the maximizer of $\hat{\mathbb{Q}}_n(\theta)$:

$$\begin{aligned} \hat{\theta}^{GMM} &= \arg\max_{\theta} \hat{\mathbb{Q}}_n(\theta) && (1.1) \\ &= \arg\max_{\theta} - \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i;\theta)\right)' \hat{\mathbf{W}} \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i;\theta)\right), && (1.2) \end{aligned}$$

where $\hat{\mathbf{W}}$ is symmetric p.d. weighting matrix.

$$\implies \mathbb{Q}_0(\theta) = -\mathbb{E}[g(Z;\theta)]'\mathbf{W}\mathbb{E}[g(Z;\theta)],\ \text{where } \mathbf{W} \text{ is } \text{——— matrix} \quad (1.3)$$

*Department of Economics, University of Wisconsin-Madison. hchu38@wisc.edu. This is lecture notes from the second half of ECON710: Economic Statistics and Econometrics II. Instructor: Prof. Harold Chiang. Materials and sources: Harold's handwritten notes.

2 Consistency of GMM

Theorem 2.1 (Consistency; GMM)**.** Suppose $(Z_i)_{i=1}^n \overset{iid}{\sim} f(\zeta;\theta)$ and $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ and:

① $\mathbf{W}$ is symmetric & p.d., and $\underline{\mathbf{W}\mathbb{E}[g(Z;\theta)] = 0 \textit{ only if } \theta = \theta_0}$ (★)

② $\theta_0 \in \Theta$, where Θ is compact

③ $\theta \mapsto g(Z;\theta)$ is continuous (a.e.) $\forall \theta \in \Theta, Z$ w.p.1.

④ $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|g(Z;\theta)\right\|\right] < \infty$

Then, $\hat{\theta}^{GMM} \xrightarrow{p} \theta_0$.

Remark. We require more equations in GMM to gain "efficiency" (smaller variance). So, we choose data-specific $\hat{\mathbf{W}}$ to ensure this efficiency.

Proof. Consider $\begin{cases} g_0(\theta) = \mathbb{E}[g(Z;\theta)] \\ \hat{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^n g(Z_i;\theta) \end{cases}$, then we define $\begin{cases} \mathbb{Q}_0(\theta) := -g_0(\theta)'\mathbf{W}g_0(\theta) \\ \hat{\mathbb{Q}}_n(\theta) := -\hat{g}_n(\theta)'\hat{\mathbf{W}}\hat{g}_n(\theta) \end{cases}$.

To apply **Consistency Theorem** (Theorem (1.1) in Lec 2), we need to verify its (i)–(iv).

- (i) θ_0 unique maximizer : we first note that, by ①, $0 \neq \mathbf{W}g_0(\theta)\ \forall \theta \neq \theta_0$ (♠).
  $\Longrightarrow$ By symmetric & p.d. of $\mathbf{W}$ (full rank),

$$\exists \mathbf{R} \text{ s.t. } \mathbf{W} = \mathbf{R}'\mathbf{R} \text{ and } \mathbf{R}^{-1} \text{ exists (non-singular)} \tag{2.1}$$

  We can thus rewrite (♠) by:

$$0 \neq \mathbf{W}g_0(\theta) = \mathbf{R}'\mathbf{R}g_0(\theta) \tag{2.2}$$
$$\Longrightarrow 0 \neq (\mathbf{R}')^{-1}\mathbf{R}'\mathbf{R}g_0(\theta) = \mathbf{R}g_0(\theta) \longleftarrow \text{since } \mathbf{R}^{-1} \text{ exists} \tag{2.3}$$

  Finally, we rewrite $\mathbb{Q}_0(\theta)$ and find that $\mathbb{Q}_0(\theta) < \mathbb{Q}_0(\theta_0)$, i.e., θ_0 unique maximizer:

$$\begin{aligned} \mathbb{Q}_0(\theta) &= -g_0(\theta)'\mathbf{W}g_0(\theta) && (2.4)\\ &= -g_0(\theta)'\mathbf{R}'\mathbf{R}g_0(\theta) && (2.5)\\ &= -\underbrace{\left(\mathbf{R}g_0(\theta)\right)'\left(\mathbf{R}g_0(\theta)\right)}_{\text{quadratic form: } > 0} && (2.6)\\ &< 0 && (2.7)\\ &= -g_0(\theta_0)'\underbrace{\mathbf{W}g_0(\theta_0)}_{= 0 \text{ by ①}} = \mathbb{Q}_0(\theta_0)\ (\checkmark) && (2.8) \end{aligned}$$

- (ii) Θ compact : by assumption ② (✓)

- (iii) continuity & (iv) uniform consistency: we invoke **ULLN** (see Lec 2).
  By ULLN with ②–④, we have

$$\begin{cases} g_0(\theta) \text{ is continuous } \forall \theta \in \Theta, Z \text{ w.p.1} \\ \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} g_j(Z_i;\theta) - \mathbb{E}[g_j(Z;\theta)] \right| \xrightarrow{p} 0 \text{ for } j = 1, \cdots, k \implies \sup_{\theta \in \Theta} \left\| \hat{g}_n(\theta) - g_0(\theta) \right\| \xrightarrow{p} 0 \end{cases}$$

  For (iii): continuity of $Q_0(\cdot)$, it is satisfied since $g_0(\cdot)$ is continuous by **ULLN** $\implies \mathbb{Q}_0(\cdot) = -g_0(\cdot)'\mathbf{W}g_0(\cdot)$ is continuous (✓)

  For (iv): $\hat{\mathbb{Q}}_n$ uniformly consistent for $\mathbb{Q}_0$, we **WTS** $\sup_{\theta \in \Theta} \left| \hat{\mathbb{Q}}_n(\theta) - \mathbb{Q}_0(\theta) \right| \xrightarrow{p} 0$
  By a clever way of "adding and subtracting 0", we obtain:

$$\begin{aligned} \left| \hat{\mathbb{Q}}_n - \mathbb{Q}_0 \right| &= \left| \hat{g}_n' \hat{\mathbf{W}} \hat{g}_n - g_0' \mathbf{W} g_0 \right| && (2.9) \\ &= \left| \Big((\hat{g}_n - g_0) + g_0 \Big)' \hat{\mathbf{W}} \Big((\hat{g}_n - g_0) + g_0 \Big) - g_0' \mathbf{W} g_0 \right| && (2.10) \\ &= \left| (\hat{g}_n - g_0)' \hat{\mathbf{W}} (\hat{g}_n - g_0) + 2 g_0' \hat{\mathbf{W}} (\hat{g}_n - g_0) + g_0' (\hat{\mathbf{W}} - \mathbf{W}) g_0 \right| && (2.11) \\ &\leq \left| (\hat{g}_n - g_0)' \hat{\mathbf{W}} (\hat{g}_n - g_0) \right| + 2 \left| g_0' \hat{\mathbf{W}} (\hat{g}_n - g_0) \right| + \left| g_0' (\hat{\mathbf{W}} - \mathbf{W}) g_0 \right| && (2.12) \\ &\leq \underbrace{\|\hat{g}_n - g_0\|^2}_{\text{ULLN: } \xrightarrow{p} 0} \cdot \left\| \hat{\mathbf{W}} \right\| + 2\|g_0\| \cdot \underbrace{\|\hat{g}_n - g_0\|}_{\text{ULLN: } \xrightarrow{p} 0} \cdot \left\| \hat{\mathbf{W}} \right\| + \|g_0\|^2 \cdot \underbrace{\left\| \hat{\mathbf{W}} - \mathbf{W} \right\|}_{\xrightarrow{p} 0} && (2.13) \\ &\xrightarrow{p} 0 && (2.14) \end{aligned}$$

  Equation (2.12) holds by $\triangle$-*ineq*. Equation (2.13) holds by *spectral norm* & $\|g_0\| = O(1)$ (*bounded*) as it's continuous on a compact set Θ. (✓)

Since we have checked (i)–(iv), by **Consistency Theorem** we conclude GMM is consistent.

□

Remark. Newey and McFadden (1994): "[T]he conditions of this result are quite weak, allowing for *discontinuity* in the moment functions. This Theorem remains true if the *i.i.d* assumption is replaced with the condition that $(Z_i)_{i=1}^n$ is stationary and ergodic."

References

Newey, W. K., & McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. Elsevier. https://doi.org/10.1016/S1573-4412(05)80005-4