# Lec 5: Asymptotic Normality of GMM

Eric Hsienchen Chu*

Spring, 2024

(⊛) Suggested readings: Newey and McFadden (1994), Ch3.3.

**Motivation.** Recall, given moment equation $\mathbb{E}[g(Z;\theta_0)] = 0$, we can form GMM estimators by the "quadratic" structure and choose a symmetric weighting matrix $\hat{\mathbf{W}}_{\mathrm{r}\times\mathrm{r}}$:

$$\begin{cases} \hat{\mathbb{Q}}_n(\theta) = -\hat{g}_n(\theta)'\hat{\mathbf{W}}\hat{g}_n(\theta), \ \text{where } \hat{g}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(Z_i;\theta) \in \mathbb{R}^r, \ \theta_0 \in \mathbb{R}^k \\ \mathbb{Q}_0(\theta) = -g_0(\theta)'\mathbf{W}g_0(\theta), \ \text{where } g_0(\theta) = \mathbb{E}[g(Z;\theta)] \end{cases}$$

⊛ Note that at true $\theta_0$ we have $\mathbb{E}[g(Z;\theta_0)] = 0$ and that $k < r$: Over–ID case!

# 1 AN for GMM-type

**Theorem 1.1** (AN for GMM-type (★)). Suppose $\hat{\theta} = \arg\max_{\theta\in\Theta} \hat{\mathbb{Q}}_n(\theta)$, consistency: $\hat{\theta} \xrightarrow{p} \theta_0$ and $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$. And, in addition:

(A1) $\theta_0 \in int(\Theta)$;

(A2) $\hat{g}_n(\theta) \in \mathcal{C}^1(\mathcal{N})$ for open $\mathcal{N}$ s.t. $\theta_0 \in \mathcal{N} \subseteq \Theta$ (continuously differentiable);

(A3) $\sqrt{n}\hat{g}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma})$ for some $\mathbf{\Sigma} > 0$ (distribution of sample analogs; STRONG);

(A4) $\exists \mathbf{G}(\theta) \in \mathbb{R}^{r\times k}$ continuous at $\theta_0$ and $\sup_{\theta\in\mathcal{N}} \left\| \nabla_{\theta'}\hat{g}_n(\theta) - \mathbf{G}(\theta) \right\| \xrightarrow{p} 0$ (uniform consistency);

(A5) $\mathbf{G} := \mathbf{G}(\theta_0)$ s.t. $\mathbf{G}'\mathbf{W}\mathbf{G}$ nonsingular

Then,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \underline{(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}}\mathbf{G}'\mathbf{W}\mathbf{\Sigma}\mathbf{W}\mathbf{G}\underline{(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}}\right), \tag{1.1}$$

*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.

*Proof.* **WTS.** $\sqrt{n}(\hat{\theta}-\theta_0)$, where $\hat{\theta} = \arg\max_{\theta\in\Theta} \hat{\mathbb{Q}}_n(\theta) = -\hat{g}_n(\theta)\hat{\mathbf{W}}\hat{g}_n(\theta)$. Let's take FOC:

$$\begin{aligned}
FOC:\ 0 &= \cancel{-}2\nabla_\theta \hat{\mathbb{Q}}_n(\hat{\theta}) && (1.2)\\
&= \underbrace{\left[\frac{1}{n}\sum_{i=1}^n \nabla_{\theta'} g(Z_i;\hat{\theta})\right]'}_{\equiv \hat{\mathbf{G}}_n(\hat{\theta})'} \hat{\mathbf{W}} \underbrace{\left[\frac{1}{n}\sum_{i=1}^n g(Z_i;\hat{\theta})\right]}_{\equiv \hat{g}_n(\hat{\theta})} && (1.3)\\
&= \hat{\mathbf{G}}_n(\hat{\theta})'\mathbf{W} \underbrace{\left[\frac{1}{n}\sum_{i=1}^n g(Z_i;\boldsymbol{\theta_0}) + \frac{1}{n}\sum_{i=1}^n \nabla_{\theta'} g(Z_i;\bar{\boldsymbol{\theta}})(\hat{\theta}-\theta_0)\right]}_{\text{MV expansion for } \hat{g}_n(\hat{\theta}) \text{ s.t. properly centered at } \theta_0!} && (1.4)\\
&= \hat{\mathbf{G}}_n(\hat{\theta})'\mathbf{W}\left[\hat{g}_n(\boldsymbol{\theta_0}) + \hat{\mathbf{G}}_n(\bar{\boldsymbol{\theta}})(\hat{\theta}-\theta_0)\right]\ (\bigstar) && (1.5)
\end{aligned}$$

By (A5), denote $\mathbf{G} \equiv \mathbf{G}(\theta_0)$. We notice that:

$$\begin{aligned}
\left\|\hat{\mathbf{G}}_n(\hat{\theta}) - \mathbf{G}\right\| &= \left\|\hat{\mathbf{G}}_n(\hat{\theta}) - \mathbf{G}(\hat{\theta}) + \mathbf{G}(\hat{\theta}) - \mathbf{G}\right\| && (1.6)\\
&\leq \left\|\hat{\mathbf{G}}_n(\hat{\theta}) - \mathbf{G}(\hat{\theta})\right\| + \left\|\mathbf{G}(\hat{\theta}) - \mathbf{G}(\theta_0)\right\| \longleftarrow \text{by } \triangle\text{–ineq} && (1.7)\\
&\leq \underbrace{\sup_{\theta\in\mathcal{N}}\left\|\nabla_{\theta'}\hat{g}_n(\hat{\theta}) - \mathbf{G}(\hat{\theta})\right\|}_{\xrightarrow{p} 0 \text{ by (A4) U.C.}} + \underbrace{\left\|\mathbf{G}(\hat{\theta}) - \mathbf{G}(\theta_0)\right\|}_{\xrightarrow{p} 0 \text{ by } \mathbf{G} \text{ cont. \& } \hat{\theta} \xrightarrow{p} \theta_0} && (1.8)\\
&\xrightarrow{p} 0 \quad (\text{as } \hat{\theta} \in \mathcal{N} \text{ w.p. approaching 1}) && (1.9)
\end{aligned}$$

Same argument applies to $\bar{\theta}$. So, we now have $\hat{\mathbf{G}}_n(\hat{\theta}) = \mathbf{G} + o_p(1)$, $\hat{\mathbf{G}}_n(\bar{\theta}) = \mathbf{G} + o_p(1)$, and $\hat{\mathbf{W}} = \mathbf{W} + o_p(1)$ (since $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$). Jointly, the three stochastic order notations give us:

$$\begin{aligned}
\hat{\mathbf{G}}_n(\hat{\theta})'\hat{\mathbf{W}}\hat{\mathbf{G}}_n(\bar{\theta}) &= \mathbf{G}'\mathbf{W}\mathbf{G} + o_p(1) && (1.10)\\
(CMT):\ \left(\hat{\mathbf{G}}_n(\hat{\theta})'\hat{\mathbf{W}}\hat{\mathbf{G}}_n(\bar{\theta})\right)^{-1} &= \left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1} + o_p(1) && (1.11)
\end{aligned}$$

We can apply CMT to Eqn (1.11) since (A5): $\mathbf{G}'\mathbf{W}\mathbf{G}$ nonsingular ($> 0$).
Then, by Equation ($\bigstar$):

$$\begin{aligned}
\sqrt{n}(\hat{\theta}-\theta_0) &= -\underbrace{\left(\hat{\mathbf{G}}_n(\hat{\theta})'\hat{\mathbf{W}}\hat{\mathbf{G}}_n(\bar{\theta})\right)^{-1}}_{\xrightarrow{p}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}} \underbrace{\hat{\mathbf{G}}_n(\hat{\theta})'\hat{\mathbf{W}}}_{\xrightarrow{p}\mathbf{G}'\mathbf{W}} \underbrace{\left[\sqrt{n}\hat{g}_n(\theta_0)\right]}_{\mathcal{N}(0,\boldsymbol{\Sigma})} && (1.12)\\
&\xrightarrow{d} -\left[\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}\right]\mathcal{N}(0,\boldsymbol{\Sigma}) \longleftarrow \text{by CLT} && (1.13)\\
&= \mathcal{N}\left(0, \left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\mathbf{G}\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\right) && (1.14)
\end{aligned}$$

Eqn (1.14) holds by Slutsky's Theorem. We successfully show the AN for GMM-type. □

**Question.** What are $\mathbf{G}$ & $\mathbf{\Sigma}$?

**Answer.** By construction, we have:

- $\mathbf{G} = \mathbb{E}\left[\nabla_{\theta'} g(Z;\theta_0)\right]$ (derivative of moment equation, evaluated at $\theta_0$)
- $\mathbf{\Sigma} = \mathbb{E}[g(Z;\theta_0)g(Z;\theta_0)'] = \text{Var}\left(g(Z;\theta_0)\right)$ (since $\mathbb{E}\left[g(Z;\theta_0)\right] = 0$)

**Question.** How to choose $\mathbf{W}$ "optimally"?

**Answer.** We set $\mathbf{W} = \mathbf{\Sigma}^{-1}$, then

$$\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Sigma}\mathbf{W}\mathbf{G}\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1} = \left(\mathbf{G}'\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}, \tag{1.15}$$

which is more concise & smaller (& more **efficient** ⊛)

## 2 Variance Estimation

**Motivation.** Since we claim "efficient", we need to show the variance of GMM estimator at Eqn (1.15) (with optimal $\mathbf{W} = \mathbf{\Sigma}^{-1}$) is smaller.

**Claim 2.1** ("GMM is efficient"). $\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Sigma}\mathbf{W}\mathbf{G}\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1} - \left(\mathbf{G}'\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} \geq 0$

*Proof.* We rely on an algebraic trick with *idempotence*:

$$\begin{aligned}
\Longrightarrow\quad & \left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Sigma}\mathbf{W}\mathbf{G}\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1} - \left(\mathbf{G}'\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1} && (2.1)\\
= \quad & \underbrace{\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Sigma}^{\frac{1}{2}}}_{\equiv \mathcal{A}}\underbrace{\left[\mathbf{I} - \mathbf{\Sigma}^{\frac{-1}{2}}\mathbf{G}\left(\mathbf{G}'\mathbf{\Sigma}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{\Sigma}^{\frac{-1}{2}}\right]}_{\equiv \mathbf{I}-\mathcal{B}}\underbrace{\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{W}\mathbf{G}\left(\mathbf{G}'\mathbf{W}\mathbf{G}\right)^{-1}}_{\equiv \mathcal{A}} && (2.2)\\
= \quad & \mathcal{A}\left[\mathbf{I} - \mathcal{B}\right]\mathcal{A}' && (2.3)\\
= \quad & \mathcal{A}\left[\mathbf{I} - \mathcal{B}\right]\left[\mathbf{I} - \mathcal{B}\right]\mathcal{A}' \longleftarrow \text{since } [\mathbf{I} - \mathcal{B}] \text{ idempotent \& } \mathcal{B} \text{ symmetric} && (2.4)\\
\geq \quad & 0 && (2.5)
\end{aligned}$$

Eqn (2.5) holds since being a quadratic form. □

**Remark.** $\mathbf{W} = \mathbf{\Sigma}^{-1}$ is called **efficient weighting matrix**. But it is actually *not feasible* since $\mathbf{\Sigma} = \mathbb{E}\left[g(Z;\theta_0)g(Z;\theta_0)'\right]$ is unknown (precisely, we don't know $\theta_0$). In practice, we use **2-step GMM**.

**Definition 2.1** (2-step GMM)**.** We employ 2-step GMM to get away with the unknown $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma}^{-1}$) (the variance of moment equation evaluated at true $\theta_0$):

① Estimate $\theta$ by first choosing $\hat{\mathbf{W}} = \mathbf{I}_{\mathrm{r}} \implies$ get $\theta^{\text{1st}}$ (not efficient, but consistent)

② Estimate $\boldsymbol{\Sigma}$ by sample analog $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} g(Z_i; \theta^{\text{1st}})g(Z_i; \theta^{\text{1st}})'$

③ Estimate $\theta$ again by $\hat{\mathbf{W}} = \hat{\boldsymbol{\Sigma}} \implies$ get $\theta^{\text{2nd}}$ □

**Summary** (Comparison: Variance Estimation)**.** In general, we have "MLE–type" or "GMM-type" estimators and estimate each of their variance by:

⊛ **MLE–type**: $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{H}^{-1}\mathcal{J}\mathbf{H}^{-1}\right)$

$$\implies \text{Var estimation by } \hat{\mathbf{V}}(\hat{\theta}) = \hat{\mathbf{H}}^{-1}\hat{\mathcal{J}}\hat{\mathbf{H}}^{-1}, \text{ where } \begin{cases} \hat{\mathbf{H}} = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\theta\theta'} g(Z_i; \hat{\theta}) \\ \hat{\mathcal{J}} = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\theta} g(Z_i; \hat{\theta}) \nabla_{\theta'} g(Z_i; \hat{\theta})' \end{cases}$$

⊛ **GMM–type**: $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, (\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\right)$

$\implies$ Var estimation by $\hat{\mathbf{V}}(\hat{\theta}) = \left(\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{W}}\hat{\mathbf{G}}(\hat{\mathbf{G}}'\hat{\mathbf{W}}\hat{\mathbf{G}})^{-1}\right)$, where

$$\begin{cases} \hat{\mathbf{G}} = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\theta'} g(Z_i; \hat{\theta}) \\ \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} g(Z_i; \hat{\theta}) g(Z_i; \hat{\theta})' \end{cases} \tag{2.6}$$

**Corollary 2.1.** Under the same conditions as in **(AN–GMM)**, if $\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$, then $\hat{\mathbf{V}}(\hat{\theta}) \xrightarrow{p} \mathbf{V}$.

**Remark.** Similar result holds for **(AN–MLE)**.

# 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