

# Statistics Learning Theory: Generalized Method of Moments

2019.03.23

## GLS estimator

Suppose the linear regression model

$$y_i = x_i' \beta + e_i$$

1.  $E[e_i|x_i] = 0$
2.  $\{(y_i, x_i)\}$  iid
3.  $E[y_i^2] < \infty$
4.  $E|x_i|^2 < \infty$
5.  $E[x_i x_i']$  is positive definite

## GLS estimator

Since

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

We have

$$\begin{aligned}\text{Var}(\hat{\beta}|X) &= E[(X'X)^{-1}X'ee'X(X'X)^{-1}|X] \\ &= (X'X)^{-1}X'DX(X'X)^{-1}\end{aligned}$$

Where

$$D = E[ee'|X] = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

where

$$\sigma_i^2 = E[e_i^2|x_i]$$

# Gauss Markov Theorem

## Theorem

*In linear regression model, the best unbiased linear estimator is*

$$\tilde{\beta} = (X'D^{-1}X)^{-1}X'D^{-1}y$$

Where  $\tilde{\beta}$  is called generalized least square (GLS) The best means that

$$Var(\tilde{\beta}) - Var(\hat{\beta})$$

is positive semi-definite

## Homoscedastic

if  $\sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma$  then

$$D^{-1} = \frac{1}{\sigma^2} I$$

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

Which coincide with OLS estimator

## Exercise 1: Nonlinear regression model

Consider the following model

$$y_i = x_i^\beta + e_i \quad E[e_i|x_i] = 0 \quad x_i > 0 \text{ is scalar}$$

We aim to solve

1. Write down the asymptotic distribution of the NLLS estimator  $\hat{\beta}$ .
2. Let  $\theta = \sqrt{\beta}$ . Find the 95% asymptotic confidence interval for  $\theta$ .
3. Let  $m(c) = E[y_i|x_i = c]$  be the conditional mean at  $c$ . Find the 95% asymptotic confidence interval for  $m(c)$ .
4. Suppose  $\sigma_i^2 = E[e_i^2|x_i]$  is known. Find a better estimator than the NLLS estimator  $\hat{\beta}$ . Explain briefly why it is better.

## Nonlinear regression model:(a)

In the part, we want to apply the general theorem for asymptotic distribution, under the suitable conditions, we know that

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow^d N(0, H^{-1}\Sigma H^{-1})$$

Where

$$H = E\left[\frac{\partial m(x, \theta_0)}{\partial \theta} \frac{\partial m(x, \theta_0)}{\partial \theta'}\right]$$

$$\Sigma = E\left[e^2 \frac{\partial m(x, \theta_0)}{\partial \theta} \frac{\partial m(x, \theta_0)}{\partial \theta'}\right]$$

Since

$$\frac{\partial m(x, \beta_0)}{\partial \beta} = x^{\beta_0} \log x$$

Therefore,

$$\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow^d$$

$$N(0, (E[x^{2\beta_0}(\log x)^2])^{-1} E[e^2 x^{2\beta_0}(\log x)^2] E[x^{2\beta_0}(\log x)^2]^{-1})$$

## Nonlinear regression model:(b)

Using the delta method, that is the Taylor expansion

$$\theta = f(\hat{\beta}) = f(\beta_0) + \frac{\partial f(\beta)}{\partial \beta}(\tilde{\beta})(\hat{\beta} - \beta_0)$$

where  $\tilde{\beta}$  is on the line between  $\hat{\beta}$  and  $\beta_0$ , therefore

$$\sqrt{n}(\theta - \theta_0) = \frac{\partial f(\beta)}{\partial \beta}(\tilde{\beta})\sqrt{n}(\hat{\beta} - \beta_0)$$

The consistency of  $\hat{\beta}$  implies  $\tilde{\beta} \xrightarrow{p} \beta_0$ , therefore

$$\sqrt{n}(\theta - \theta_0) \rightarrow^d N(0, \frac{1}{4\beta_0} V)$$

where

$$V = E[x^{2\beta_0}(\log x)^2])^{-1} E[e^2[x^{2\beta_0}(\log x)^2] E[x^{2\beta_0}(\log x)^2]]^{-1}$$



## Nonlinear regression model:(b)

Therefore, the 95 % confidence interval can be constructed as

$$[\hat{\theta} - z_{0.025} \sqrt{\frac{\hat{V}}{4\hat{\beta}_n}}, \hat{\theta} + z_{0.025} \sqrt{\frac{\hat{V}}{4\hat{\beta}_n}}]$$

where

$$\hat{V} = \left(\frac{1}{n} \sum_{i=1}^n x_i^{2\hat{\beta}} (\log x_i)^2\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \hat{e}_i^2 x_i^{2\hat{\beta}} (\log x_i)^2\right) \left(\frac{1}{n} \sum_{i=1}^n x_i^{2\hat{\beta}} (\log x_i)^2\right)^{-1}$$

$$\hat{e}_i = y_i - x_i^{\hat{\beta}}$$

## Nonlinear regression model:(c)

Since  $m(c) = c^\beta$  and

$$\frac{\partial c^\beta}{\partial \beta} = c^\beta \log c$$

Using the delta method, we can get

$$\sqrt{n}(m(\hat{c}) - m(c)) \rightarrow^d N(0, c^{2\beta}(\log c)^2 V)$$

## Nonlinear regression model:(d)

Consider the GLS-type NLLS in the presence of known heteroscedasticity

$$\frac{y_i}{\sigma_i} = \frac{x_i^\beta}{\sigma_i} + u_i \quad u_i = \frac{e_i}{\sigma}$$

Therefore,  $E[u_i^2|x_i] = 1$  which restores homoscedasticity. The estimator can be defined as

$$\hat{\beta}_{GLS} = \arg \min_{\beta} \sum_{i=1}^n \left( \frac{y_i}{\sigma_i} - \frac{x_i^\beta}{\sigma} \right)^2$$

We can derive the asymptotic distribution as part(a) by changing

$$m(x, \beta) = \frac{x_i^\beta}{\sigma}$$

$$\sqrt{n}(\hat{\beta}_{GLS} - \beta) \rightarrow^d N(0, H^{-1}\Sigma H^{-1})$$

## Nonlinear regression model:(d)

Where

$$H = E\left[\frac{x_i^{2\beta_0}(\log x_i)^2}{\sigma_i^2}\right]$$

$$\Sigma = E\left[\frac{x_i^{2\beta_0}(\log x_i)^2}{\sigma_i^2}\right]$$

Therefore

$$V_{GLS} = (E\left[\frac{x_i^{2\beta_0}(\log x_i)^2}{\sigma_i^2}\right])^{-1}$$

## Exercise 2

Let  $f_n$  be a sequence of functions on  $S \in \mathbb{R}$  such that

$$\sup_{x \in S} |f_n(x) - f(x)| \rightarrow 0$$

(a): show that  $\sup_{x \in S} f_n(x) \rightarrow \sup_{x \in S} f(x)$

(b): show that  $\inf_{x \in S} f_n(x) \rightarrow \inf_{x \in S} f(x)$

## Exercise 2: (a)

Since  $\sup_{x \in S} |f_n(x) - f(x)| \rightarrow 0$ , therefore, given  $\varepsilon > 0$ , there exists  $N > 0$  such that  $n > N$  then

$$\sup_{x \in S} |f_n(x) - f(x)| < \frac{\varepsilon}{2}$$

which implies that  $-\frac{\varepsilon}{2} < f_n(x) - f(x) < \frac{\varepsilon}{2} \quad \forall x \in S$

(a): By the definition of supremum, there exists  $x_{1n}, x_{2n} \in S$  such that

$$f_n(x_{1n}) + \frac{\varepsilon}{2} > \sup_{x \in S} f_n(x)$$

$$f(x_{2n}) + \frac{\varepsilon}{2} > \sup_{x \in S} f(x)$$

## Exercies 2: (a)

Therefore, given  $n > N$ , we have

$$f_n(x_{1n}) + \frac{\varepsilon}{2} - f(x_{1n}) > \sup_{x \in S} f_n(x) - \sup_{x \in S} f(x) > f(x_{2n}) - f(x_{2n}) - \frac{\varepsilon}{2}$$

where

$$f_n(x_{1n}) + \frac{\varepsilon}{2} - f(x_{1n}) < \varepsilon$$

$$f(x_{2n}) - f(x_{2n}) - \frac{\varepsilon}{2} > -\varepsilon$$

Therefore

$$\sup_{x \in S} f_n(x) \rightarrow \sup_{x \in S} f(x)$$

## Exercise 2:(b)

(b) Using the similar technique, we have

$$\frac{-\varepsilon}{2} < f(x) - f(x) < \frac{\varepsilon}{2} \quad \forall x \in S$$

$$\inf_{x \in S} f_n(x) > f_n(x_{1n}) - \frac{\varepsilon}{2}$$

$$\inf_{x \in S} f(x) > f(x_{2n}) - \frac{\varepsilon}{2}$$

Therefore

$$f_n(x_{2n}) - f(x_{2n}) + \frac{\varepsilon}{2} > \inf_{x \in S} f_n(x) - \inf_{x \in S} f(x) > f_n(x_{1n}) - f(x_{1n}) - \frac{\varepsilon}{2}$$

It is a desired result.



## Exercise 3: linearized estimator

Let  $\hat{\theta} = \arg \max_{\theta \in \Theta} Q_n(\theta)$  be an extremum estimator of a scalar parameter  $\theta_0$ , and  $\bar{\theta}$  be a preliminary estimator of  $\theta_0$  such that  $\sqrt{n}(\bar{\theta} - \theta_0) = O_p(1)$ . Consider the following 'linearized' estimator of  $\theta_0$ :

$$\tilde{\theta} = \bar{\theta} - \left( \frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2} \right)^{-1} \left( \frac{\partial Q_n(\bar{\theta})}{\partial \theta} \right)$$

Find the asymptotic distribution of  $\tilde{\theta}$

## Exercise 3: Solution

By the definition , we have that

$$\begin{aligned}\sqrt{n}(\tilde{\theta} - \theta_0) &= \sqrt{n}(\tilde{\theta} - \bar{\theta}) + \sqrt{n}(\bar{\theta} - \theta_0) \\ &= -\sqrt{n}\left(\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2}\right)^{-1}\left(\frac{\partial Q_n(\bar{\theta})}{\partial \theta}\right) + (\sqrt{n}(\bar{\theta} - \theta_0))\end{aligned}$$

By Taylor Expansion, we can obtain that

$$\frac{\partial Q_n(\bar{\theta})}{\partial \theta} = \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2}(\bar{\theta} - \theta_0)$$

Therefore

$$\bar{\theta} - \theta_0 = \left(\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2}\right)^{-1}\left(\frac{\partial Q_n(\bar{\theta})}{\partial \theta} - \frac{\partial Q_n(\theta_0)}{\partial \theta}\right)$$

## Exercises 3: Solution

Substitute the above, we have that

$$\begin{aligned}& \sqrt{n}(\tilde{\theta} - \theta_0) \\&= -\sqrt{n}\left(\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2}\right)^{-1}\left(\frac{\partial Q_n(\bar{\theta})}{\partial \theta}\right) + \sqrt{n}\left(\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2}\right)^{-1}\left(\frac{\partial Q_n(\bar{\theta})}{\partial \theta} - \frac{\partial Q_n(\theta_0)}{\partial \theta}\right) \\&= -\sqrt{n}\left(\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2}\right)^{-1}\frac{\partial Q_n(\theta_0)}{\partial \theta} \\&\quad + \sqrt{n}\frac{\partial Q_n(\bar{\theta})}{\partial \theta}\left(\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2}\right)^{-1} - \left(\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2}\right)^{-1}) \\&= (I) + (II)\end{aligned}$$

## Exercies 3: Solution

Since  $\bar{\theta} \rightarrow^p \theta$ , we know that  $\theta_B \rightarrow^p \theta_0$ , therefore

$$(I) \rightarrow^d N(0, H(\theta_0)^{-2} \Sigma)$$

We claim that the second term would converge to zero in probability, since

$$(II) \leq \left| \sqrt{n} \frac{\partial Q_n(\bar{\theta})}{\partial \theta} \right| \left| \left( \frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2} \right)^{-1} - \left( \frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2} \right)^{-1} \right|$$

Since both

$$\left( \frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2} \right)^{-1} \rightarrow^p H(\theta_0)^{-1}$$

$$\left( \frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2} \right)^{-1} \rightarrow^p H(\theta_0)^{-1}$$

## Exercies 3: Solution

The second term in inequality would be  $o_p(1)$ , now, turning to the first term

$$\begin{aligned}& \left| \sqrt{n} \frac{\partial Q_n(\bar{\theta})}{\partial \theta} \right| \\&= \sqrt{n} \left| \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2} (\bar{\theta} - \theta_0) \right) \right| \\&\leq \sqrt{n} \left| \frac{\partial Q_n(\theta_0)}{\partial \theta} \right| \leq \frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2} (\bar{\theta} - \theta_0) \\&= O_p(1) + O_p(1) = O_p(1)\end{aligned}$$

Therefore,

$$(II) \leq o_p(1) O_p(1) = o_p(1)$$

# GMM: Introduction

GMM is one of the most popular estimation method in applied econometrics. GMM generalizes the classical method of moments estimator by allowing for models that have more equations than unknown parameters and are thus overidentified. GMM includes as special cases OLS, IV, multivariate regression, and 2SLS.

## GMM: : Linear case

Consider linear projection model

$$y_i = x'_{1i}\beta_1 + x_{2i}\beta_2 + e_i$$

$$E \begin{bmatrix} x_{1i}e_i \\ x_{2i}e_i \end{bmatrix} = 0$$

1. This model can be estimated by OLS.
2. Now suppose we know that a priori that  $\beta_2 = 0$ . Then model becomes

$$y_i = x'_{1i}\beta_1 + e_i$$

$$E \begin{bmatrix} x_{1i}e_i \\ x_{2i}e_i \end{bmatrix} = 0$$

## GMM: : Linear case

1. How do we estimate  $\beta_1$ ?
2. One may estimate  $\beta_1$  by OLS from  $y$  on  $x_1$  which utilize information  $E[x_{1i}e_i] = 0$
3. But this is not necessarily efficient because it does not use additional information  $E[x_{2i}e_i] = 0$
4. In this model, the number of parameters is  $\dim\beta_1 = k$  but the number of moment restrictions is  $\dim x_1 + \dim x_2 = k + r$
5. Such situation is called overidentified



# Moment restriction model

1. In general we consider

$$E[g(w_i, \beta)] = 0$$

where  $\beta$  is  $k$ -dimensional parameters and  $g$  is  $l$ -dimensional vector of functions with  $l \geq k$

2. Above example is

$$g(w_i, \beta) = x_i(y_i - x_{1i}'\beta_1)$$

where  $x_i = (x_{1i}', x_{2i}')'$

# GMM estimator

To generalize, consider linear model

$$y_i = x_i' \beta + e_i$$

$$E[z_i e_i] = 0$$

1. If  $\dim g = \dim \beta$  (called just identification), then we can apply method of moments that solves
$$\bar{g} = \frac{1}{n} \sum_{i=1}^n g(w_i, \beta) = 0$$
2. But if  $\dim g > \dim \beta$  (overidentified), we cannot solve this equation in general

# GMM estimator

1. For overidentified case, we can minimize weighted Euclidean norm

$$J_n(\beta) = n\bar{g}(\beta)'W_n\bar{g}(\beta)$$

where  $W_n$  is symmetric weight matrix

2. Minimizer of this object is called Generalized method of moments(GMM) estimator

$$\hat{\beta} = \arg \min_{\beta} J_n(\beta)$$

## GMM:estimator

In the linear model

$$\bar{g}(\beta) = \frac{1}{n} \sum_{i=1}^n z_i(y_i - x_i'\beta) = \frac{1}{n} Z'(y - X\beta)$$

So, FOC of  $\hat{\beta}$  is

$$\begin{aligned} 0 &= \frac{\partial J_n(\hat{\beta})}{\partial \beta} = 2n \left( \frac{\partial \bar{g}(\hat{\beta})}{\partial \beta'} \right)' W_n \bar{g}(\hat{\beta}) \\ &= -2n \left( \frac{1}{n} X'Z \right) W_n \left( \frac{1}{n} Z'(y - X\hat{\beta}) \right) \end{aligned}$$

Solving for  $\hat{\beta}$  yields

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

## GMM: Asymptotic distribution of $\hat{\beta}$

Note that

$$\hat{\beta} = \beta + (X'ZW_nZ'X)^{-1}X'ZW_nZ'e$$

and thus

$$\sqrt{n}(\hat{\beta} - \beta) = [(\frac{1}{n}X'Z)W_n(\frac{1}{n}Z'X)]^{-1}(\frac{1}{n}X'Z)W_n(\frac{1}{\sqrt{n}}Z'e)$$

By LLN and CLT under certain assumption

$$\frac{1}{n}Z'X \rightarrow^p E[z_i x_i'] := Q$$

$$\frac{1}{\sqrt{n}}Z'e \rightarrow^d N(0, \Sigma)$$

Where  $\Sigma = E[e_i^2 z_i z_i']$

## GMM:estimator

Suppose  $W_n \rightarrow^P W$  for positive definite symmetric  $W$ . By CMT

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow^d N(0, V_W)$$

where  $V_W = (Q'WQ)^{-1}Q'W\Sigma WQ(Q'WQ)^{-1}$

Asymptotic variance  $V_W$  depends on weight  $W$ . This is minimized by choosing

$$W^* = \Sigma^{-1}$$

which implies  $V_{W^*} = (Q'\Sigma^{-1}Q)^{-1}$

## GMM: Linear Case

To show that

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

We note that the asymptotic of  $\hat{\beta}_{GMM}$  is

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

We will prove the following statements

1. Let  $V_0$  be the matrix when  $W = \Omega^{-1}$ . Show that  $V_0 = (Q'\Omega Q)^{-1}$
2. We want to show that for any  $W$ ,  $V - V_0$  is positive semi-definite, to do this, finding matrices  $A$  and  $B$  such that  $V = A'\Omega A$  and  $V_0 = B'\Omega B$
3. Show that  $B'\Omega A = B'\Omega B$  and therefore that  $B'\Omega(A - B) = 0$
4. Use the expressions  $V = A'\Omega A$ ,  $A = B + (A - B)$  and  $B'\Omega(A - B) = 0$  to show that  $V \geq V_0$

## GMM: Linear Case

1 is straightforward:

$$V_0 = (Q'\Omega^{-1}Q)^{-1}Q'\Omega^{-1}\Omega\Omega^{-1}Q(Q'\Omega^{-1}Q)^{-1}$$

2 is

$$A = WQ(Q'WQ)^{-1}$$

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

3 is simply by plugging in

4

$$V = A'\Omega A$$

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

$$= V_0 + (A - B)'\Omega(A - B)$$

Which implies that  $V - V_0$  is positive-definite



## GMM: Estimation of $\Omega$

We don't know the true value of  $\Omega$ , how to estimate it? Take the model  $y_i = x_i'\beta + e_i$  with  $E(z_i e_i) = 0$ . Let  $\tilde{e}_i = y_i - x_i'\tilde{\beta}$ , where  $\tilde{\beta}$  is consistent for  $\beta$ . Define an estimate of the optimal GMM weight matrix

$$\hat{W} = \left( \frac{1}{n} \sum_{i=1}^n z_i z_i' \tilde{e}_i^2 \right)^{-1}$$

Show that  $\hat{W} \rightarrow^p \Omega^{-1}$  where  $\Omega = E[z_i z_i' e_i^2]$

## GMM: Estimation of $\Omega$

By continuous mapping theorem, we only need to show that

$$\frac{1}{n} \sum_{i=1}^n z_i z_i' \tilde{e}_i^2)^{-1} \rightarrow^p \Omega$$

## GMM: General case

Moment restriction model

$$E[g(w_i, \theta_0)] = 0$$

Here  $g$  could be nonlinear

GMM estimator

$$\hat{\theta}_W = \arg \min_{\theta \in \Theta} \bar{g}(\theta)' W_n \bar{g}(\theta)$$

# GMM: Consistency

## Theorem

*Suppose*

1.  $\Theta$  is compact
2.  $W_n \rightarrow^p W$  and  $W$  is symmetric and positive semi-definite
3.  $g(w, \theta)$  is almost surely continuous at each  $\theta \in \Theta$
4.  $E[\sup_{\theta \in \Theta} |g(w, \theta)|] < \infty$
5.  $WE[g(w, \theta)] = 0$  only if  $\theta = \theta_0$

*Then*

$$\hat{\theta}_W \rightarrow^p \theta_0$$

## Recall: Consistency Theorem

### Theorem (General Consistency Theorem)

*Suppose*

1.  $\Theta$  is compact
2.  $\sup_{\theta \in \Theta} |Q_n(\theta) - Q_*(\theta)| \rightarrow^P 0$  for some  $Q_* : \Theta \rightarrow R$
3.  $Q_*$  is continuous in  $\theta \in \Theta$
4.  $Q_*$  is uniquely maximized at  $\theta_0$

*Then*

$$\hat{\theta} \rightarrow^P \theta_0$$

## Recall: Uniform law of large number

### Theorem

1.  $\Theta$  is compact
2.  $g(z, \theta)$  is almost surely continuous at each  $\theta \in \Theta$
3. There is  $d(z)$  such that  $|g(z, \theta)| \leq d(z)$  for all  $\theta \in \Theta$  and almost every  $z$  and  $E[d(z)] < \infty$

Then

$$\sup_{\theta \in \Theta} |\bar{g}(\theta) - E[g(z, \theta)]| \xrightarrow{P} 0$$

## GMM: Consistency proof

It is sufficient to check the conditions 1 – 4 in general theorem.

Condition 1 is guaranteed by (i)

For Condition 2, let

$$Q_*(\theta) = -E[g(z, \theta)]' W E[g(z, \theta)]$$

By triangle inequality

$$\begin{aligned} & |Q_n(\theta) - Q_*(\theta)| \\ & \leq |\{\bar{g}(\theta) - E[g(z, \theta)]\}' \hat{W} \{\bar{g}(\theta) - E[g(z, \theta)]\}| \\ & \quad + |E[g(z, \theta)]' (\hat{W} + \hat{W}') \{\bar{g}(\theta) - E[g(z, \theta)]\}| \\ & \quad + |E[g(z, \theta)]' (\hat{W} - W) E[g(z, \theta)]| \\ & \quad T_1(\theta) + T_2(\theta) + T_3(\theta) \end{aligned}$$

## GMM: Consistency proof

By the uniform law of large numbers (ULLN)

$$\sup_{\theta \in \Theta} |\bar{g}(\theta) - E[g(z, \theta)]| \xrightarrow{P} 0$$

implies  $\sup_{\theta \in \Theta} T_1(\theta) \xrightarrow{P} 0$  and  $\sup_{\theta \in \Theta} T_2(\theta) \xrightarrow{P} 0$ . This condition is guaranteed by (i), (iii), (iv).

Also from (ii), we obtain  $\sup_{\theta \in \Theta} T_3(\theta) \xrightarrow{P} 0$ , therefore, the condition 2 is guaranteed.



## GMM: Consistency proof

Also ULLN presented above implies  $E[g(z, \theta)]$  is continuous in  $\theta$ , thus  $Q_n(\theta)$  is continuous in  $\theta$ . The condition 3 is verified.

Since  $E[g(z, \theta_0)] = 0$ , it holds  $Q_*(\theta_0) = 0$ , thus it is sufficient to show

$$Q_*(\theta) < 0$$

for all  $\theta \neq \theta_0$ . Since  $W$  is symmetric and positive semi-definite, we can take  $R$  such that

$$W = R'R$$

Any  $\theta \neq \theta_0$ . (v) implies

$$0 \neq WE[g(z, \theta)] = R'RE[g(z, \theta)]$$

Which implies  $RE[g(z, \theta)] \neq 0$  Therefore

$$Q_*(\theta) = -\{RE[g(z, \theta)]\}'\{RE[g(z, \theta)]\} < 0$$

The condition 4 is verified.

## Recall: General Asymptotic Normality Theorem: Basic Idea

Now consider asymptotic distribution of extremum estimator

$Q_n(\theta)$  = some objective function

$$\hat{\theta} = \arg \max_{\theta \in \Theta} Q_n(\theta)$$

Assume consistency  $\hat{\theta} \rightarrow^P \theta_0$

We want to derive asymptotic normal distribution in the form of

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow^d N(0, V)$$

The result can be used to construct confidence interval or to conduct hypothesis testing

# Recall: General Asymptotic Normality Theorem

## Theorem

*Suppose*

1.  $\hat{\theta} \rightarrow^P \theta_0$  and  $\theta_0 \in \text{int}\Theta$
2.  $Q_n(\theta)$  is twice continuously differentiable in a neighborhood  $\mathcal{N}$  of  $\theta_0$
3.  $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \rightarrow^d N(0, \Sigma)$
4. There exists  $H(\theta)$  that is continuous at  $\theta_0$ ,

$$\sup_{\theta \in \mathcal{N}} \left| \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} - H(\theta) \right| \rightarrow^P 0$$

*and  $H = H(\theta_0)$  is nonsingular*

*Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow^d N(0, H^{-1}\Sigma H^{-1})$$

## GMM: Asymptotic normality

Let

$$Q_n(\theta) = -\frac{1}{2} \left\{ \frac{1}{n} \sum_{i=1}^n g(z_i, \theta) \right\}' \hat{W} \left\{ \frac{1}{n} \sum_{i=1}^n g(z_i, \theta) \right\}$$

Let us apply general asymptotic normality theorem

# GMM: Asymptotic normality

## Theorem

*Suppose*

1.  $\theta_0 \in \text{int}\Theta$
2.  $g(w, \theta)$  is twice continuously differentiable in a neighborhood  $\mathcal{N}$  of  $\theta_0$  with probability one
3.  $E[|g(w, \theta_0)|^2] < \infty$  and  $E[\sup_{\theta \in \mathcal{N}} |\frac{\partial^2 g^{(j)}(z, \theta)}{\partial \theta \partial \theta'}|] < \infty$
4.  $G'WG$  is nonsingular where  $G = E[\frac{\partial g(w, \theta_0)}{\partial \theta'}]$

*Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow^d N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$$

## GMM: Asymptotic normality proof

It is enough to check conditions for general asymptotic normality theorem. Condition 1 is already verified. The condition 2 is satisfied by (ii).

In this case, condition 3 is verified as

$$\begin{aligned}\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} &= \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial g(z_i, \theta_0)}{\partial \theta'} \right)' \hat{W} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i, \theta_0) \right) \\ &\rightarrow^d N(0, G' W \Omega W G)\end{aligned}$$

Where convergence follows from LLN and CLT with (iii)-(iv)

## GMM: Optimal weight

Based on the asymptotic variance of  $\hat{\theta}_W$

$$V_W = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$$

we can find an optimal  $W$  to minimize the asymptotic variance in the positive definite sense

### Theorem

*For any  $W$  with nonsingular  $G'WG$ , it holds*

$$V_W - V_{\Omega^{-1}} \text{ is positive semi-definite}$$

where

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

## GMM: optimal weight proof

Let

$$A = WG(G'WG)^{-1}$$

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

Then

$$V_W = A'\Omega A$$

$$V_{\Omega^{-1}} = B'\Omega B$$

Note that

$$\begin{aligned} V_W &= (A - B + B)'\Omega(A - B + B) \\ &= (A - B)'\Omega(A - B) + V_{\Omega^{-1}} \\ &\quad + B'\Omega(A - B) + (A - B)'\Omega B \end{aligned}$$



## GMM: Optimal weight proof

By definition, the cross term would be zero:

$$\begin{aligned} & B' \Omega (A - B) \\ &= (G' \Omega^{-1} G)^{-1} G' \Omega^{-1} \Omega (W G (G' W G)^{-1} - \Omega^{-1} G (G' \Omega^{-1} G)^{-1}) \\ &= (G' \Omega^{-1} G)^{-1} - (G' \Omega^{-1} G)^{-1} = 0 \end{aligned}$$

Therefore

$$V_W - V_{\Omega^{-1}} = (A - B)' \Omega (A - B) \text{ is positive semi-definite}$$

## GMM: implication of positive semi-definiteness

If  $V_W - V_\Omega$  is positive semi-definite, then for any vector  $c \in R^{dimg}$

$$c' V_W c \geq c' V_{\Omega^{-1}} c$$

In particular

$$[V_W]_{(j,j)} \geq [V_{\Omega^{-1}}]_{(j,j)}$$

It means that the asymptotic variance of  $\hat{\theta}_W^j \geq \hat{\theta}_{\Omega^{-1}}^j$

## GMM: Estimation of $\Omega$

Optimal weight

$$\Omega^{-1} = E[g(w, \theta_0)g(w, \theta_0)']^{-1}$$

is unknown and should be estimated

By taking sample analog,  $\Omega$  can be estimated by

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n g(w_i, \hat{\theta}_W)g(w_i, \hat{\theta}_W)'$$

where  $\hat{\theta}_W$  is some 1st step GMME

# GMM: Consistency of $\Omega$

## Theorem

Suppose

1.  $\hat{\theta}_W \rightarrow^P \theta_0$
2.  $\sup_{\theta \in \mathcal{N}} |\frac{1}{n} \sum_{i=1}^n g(w_i, \theta)g(w_i, \theta)' - E[g(w, \theta)g(w, \theta)']| \rightarrow^P 0$
3.  $E[g(w, \theta)g(w, \theta)']$  is continuous at  $\theta_0$

Then

$$\hat{\Omega} \rightarrow^P \Omega$$

# Two-step GMM

1. Compute 1st-step GMME  $\hat{\theta}_W$  by some  $W_n$

$$\hat{\theta}_W = \arg \min_{\theta \in \Theta} \bar{g}(\theta)' W_n \bar{g}(\theta)$$

2. Compute  $\hat{\Omega}$
3. Compute two-step GMME by estimated optimal weight  $\hat{\Omega}^{-1}$

$$\hat{\theta}_{(2)} = \arg \min_{\theta \in \Theta} \bar{g}(\theta)' \hat{\Omega}^{-1} \bar{g}(\theta)$$

Note that :

1.  $\hat{\theta}_{(2)}$  is asymptotically more efficient than  $\hat{\theta}_W$
2. Step 2-3 may be repeated(called repeated GMM)

# GMM: Hypothesis testing

In this part, we would discuss

1. Specification test
2. Parameter hypothesis

## GMM: Specification test

Test validity of overidentified moment restrictions

$$H_0 : E[g(w, \theta)] = 0 \text{ for some } \theta \in \Theta$$

$$H_1 : E[g(w, \theta)] \neq 0 \text{ for all } \theta \in \Theta$$

## Test statistic

Test statistic(called J-statistic)

$$J = n\bar{g}(\hat{\theta})\hat{\Omega}^{-1}\bar{g}(\hat{\theta})$$

Intuition: If the model is correct,  $\bar{g}(\hat{\theta})$  has to be close to zero and  $J$  also has to be close to zero

Asymptotic distribution

$$J \rightarrow^d \chi^2(\dim g - \dim \theta) \text{ under } H_0$$

$$J \rightarrow^p 0 \text{ under } H_1$$



Test statistic: proof for linear case

PS5

# Parameter hypothesis

Parameter hypothesis

$$H_0 : r(\theta_0) = 0 \quad \text{vs.} \quad H_1 : r(\theta_0) \neq 0$$

Recall that

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow^d N(0, V)$$

where

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

## $r(\hat{\theta})$ distribution: delta method

By expanding  $r(\hat{\theta})$  around  $\hat{\theta} = \theta_0$

$$r(\hat{\theta}) = r(\theta_0) + \frac{\partial r(\tilde{\theta})}{\partial \theta'}(\hat{\theta} - \theta_0)$$

for some  $\tilde{\theta} \in [\hat{\theta}, \theta_0]$  and thus

$$\sqrt{n}(r(\hat{\theta}) - r(\theta_0)) \rightarrow^d N(0, R'VR)$$

where

$$R' = \frac{\partial r(\theta_0)}{\partial \theta'}$$

## Wald statistic

Based on the asymptotic distribution of  $r(\hat{\theta})$ , we can use the Wald statistic

$$W = nr(\hat{\theta})'[\hat{R}'\hat{V}\hat{R}]^{-1}r(\hat{\theta})$$

Asymptotic distribution

$$W \rightarrow^d \chi^2(\dim r) \text{ under } H_0$$

$$W \rightarrow \infty \text{ under } H_1$$

## GMM distance statistic

We can also consider likelihood ratio type test

In this case, the GMM objective function

$$J(\theta) = n\bar{g}(\theta)' \hat{\Omega}^{-1} \bar{g}(\theta)$$

plays the role as likelihood function

GMM distance statistic

$$LR = J(\tilde{\theta}) - J(\hat{\theta})$$

Asymptotic distribution

$$LR \rightarrow^d \chi^2(\dim r) \text{ under } H_0$$

$$LR \rightarrow \infty \text{ under } H_1$$

Asymptotically equivalent to Wald statistic

# Interval estimation

1. Wald-type(or t-value) confidence interval

$$CI_W = [\hat{\theta}^{(j)} - z_{\frac{\alpha}{2}} \sqrt{\frac{[\hat{V}]_{jj}}{n}}, \hat{\theta}^{(j)} + z_{\frac{\alpha}{2}} \sqrt{\frac{[\hat{V}]_{jj}}{n}}]$$

2. LR-type confidence interval

$$CI_{LR} = \{c : LR(c) \leq \chi_{\alpha}^2(1)\}$$

## Wald vs. LR

1.  $CI_W$  is computationally cheaper than  $CI_{LR}$
2. However the shape of  $CI_{LR}$  is more flexible.
3. LR test is invariant to the functional form of nonlinear hypothesis  $H_0 : r(\theta_0) = 0$  but Wald is not invariant(e.g.  $r(\theta_0) = \theta_0^{(1)}\theta_0^{(2)} - 1$  and  $r(\theta_0) = \theta_0^{(1)} - \frac{1}{\theta_0^{(2)}}$ ).

## GMM: Conditional moment restriction

So far, we consider unconditional moment restrictions

$$E[g(w_i, \theta_0)] = 0$$

Example: Linear projection model

$$y_i = x_i' \theta_0 + e_i$$

$$E[x_i e_i] = 0$$

which implies unconditional moment restriction

$$E[g(w_i, \theta_0)] = E[x_i(y_i - x_i' \theta_0)] = 0$$



## GMM: Conditional moment restriction

This model is just-identified and method of moments estimator is

$$\hat{\theta} = \left( \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \sum_{i=1}^n x_i y_i \right)$$

which coincides with OLS estimator

By GMM theory,  $\hat{\theta}$  is asymptotically semiparametric efficient regardless of heteroskedasticity.

## GLS estimator

How about generalized least squares estimator?

Letting  $\sigma_i^2 = E[e_i^2|x_i]$ , GLSE is

$$\begin{aligned}\tilde{\theta} &= (\sum \sigma_i^{-2} x_i x_i')^{-1} (\sum \sigma_i^{-2} x_i y_i) \\ &= \theta_0 + (\frac{1}{n} \sum \sigma_i^{-2} x_i x_i')^{-1} (\frac{1}{n} \sum \sigma_i^{-2} x_i e_i) \\ &\rightarrow^p \theta_0 + E[\sigma_i^{-2} x_i x_i']^{-1} E[\sigma_i^{-2} x_i e_i]\end{aligned}$$

In the projection model, we only assume  $E[x_i e_i] = 0$  which does not necessarily imply  $E[\sigma_i^{-2} x_i e_i] = 0$ , therefore, GLSE may be inconsistent in the projection model.

# GMM: linear regression model

Linear regression model

$$y_i = x_i' \theta_0 + e_i$$

$$E[e_i | x_i] = 0$$

This model implies conditional moment restriction which is stronger condition than unconditional moment restriction

$$E[h(w_i, \theta_0 | x_i)] = E[y_i - x_i' \theta_0 | x_i] = 0$$

## GMM: linear regression model

Conditional moment restriction  $E[e_i|x_i] = 0$  implies infinitely many unconditional moment restrictions in the form of  $E[a(x_i)e_i] = 0$   
Under  $E[e_i|x_i] = 0$ , GLSE is consistent therefore, we can write down the conditional moment restriction

$$E[h(w_i, \theta_0)|x_i] = 0$$

For simplicity, assume  $\dim(h) = 1$   
How can we estimate  $\theta_o$  efficiently

## GMM: Conditional moment restrictions

$$E[h(w_i, \theta_0)|x_i] = 0$$

imply

$$E[a(x_i)h(w_i, \theta_0)] = 0 \text{ for any } a(\cdot)$$

Based on unconditional moment restrictions, we can do GMM  
Which  $a(\cdot)$  (also called instruments) should be chosen?

# Optimal instruments

Pick some  $a(\cdot)$ . If assumptions for GMM are satisfied for the model  $E[a(x_i)h(w_i, \theta_0)] = 0$ , asymptotic variance of GMME is

$$asy.var(\hat{\theta}_a) = (G'_a \Omega_a^{-1} G_a)^{-1}$$

where

$$G_a = E[a(x_i) \frac{\partial h(w_i, \theta_0)}{\partial \theta'}]$$

$$\Omega_a = E[a(x_i)a(x_i)'h(w_i, \theta_0)^2]$$

We choose  $a(\cdot)$  to minimize  $asy.var(\hat{\theta}_a)$  in positive semi-definite sense

# GMM: Optimal instruments

## Theorem

For any  $a(\cdots)$ , it holds

$$(G'_a \Omega_a^{-1} G_a)^{-1} \geq (G'_* \Omega^{-1} G_*)^{-1}$$

Where

$$a^*(x_i) = E\left[\frac{\partial h(w_i, \theta_0)}{\partial \theta} \middle| x_i\right] E[h(w_i, \theta_0)^2 | x_i]^{-1}$$

$a^*(x_i)$  is called optimal IV

## GMM: Optimal instruments proof

Pick any  $a(\cdot)$ . Let

$$h_i = h(w_i, \theta_0)$$

$$a_i = a(x_i)$$

$$H_i = E\left[\frac{\partial h(w_i, \theta_0)}{\partial \theta'} \mid x_i\right]$$

$$V_i = E[h(w_i, \theta_0)^2 \mid x_i]$$

Then

$$V_a^{-1} = G_a' \Omega_a^{-1} G_a = E[a_i H_i]' E[a_i V_i a_i']^{-1} E[a_i H_i]$$

$$V_*^{-1} = G_*' \Omega_*^{-1} G_* = E[a_i^* H_i]' E[a_i^* H_i] E[a_i^* V_i a_i^{*'}]^{-1} E[a_i^* H_i] \\ E[H_i' V_i^{-1} H_i]$$



## GMM: Optimal instruments proof

Now define

$$m_i = G_a' \Omega_a^{-1} a_i h_i$$

$$m_i^* = a_i^* h_i$$

Then

$$E[m_i m_i'] = G_a' \Omega_a^{-1} G_a = V_a^{-1}$$

$$E[m_i^* m_i^{*'}] = E[a_i^* V_i a_i^{*'}] = V_*^{-1}$$

$$E[m_i m_i^{*'}] = G_a' \Omega_a^{-1} E[a_i V_i a_i^{*'}] = V_a^{-1}$$

## GMM: Optimal instruments proof

Therefore, the difference in variance is written as

$$\begin{aligned} & V_a - V_* \\ &= V_a V_a^{-1} V_a - V_* \\ &= E[m_i m_i^{*'}]^{-1} E[m_i m_i'] E[m_i^{*'} m_i']^{-1} - E[m_i^* m_i^{*'}]^{-1} \\ & \quad E[RR'] \\ & \geq 0 \end{aligned}$$

where

$$R = E[m_i m_i^{*'}]^{-1} \{m_i - E[m_i m_i^{*'}] E[m_i m_i^{*'}]^{-1} m_i^*\}$$

## GMM: Optimal IV estimator

Optimal IV is

$$a^*(x_i) = E\left[\frac{\partial h(w_i, \theta_0)}{\partial \theta} \middle| x_i\right] E[h(w_i, \theta_0)^2 | x_i]^{-1}$$

Note that

$$\dim(a^*(x_i)) = \dim(\theta)$$

Thus optimal IV estimator is defined as method of moments estimator

$$\frac{1}{n} \sum_{i=1}^n a^*(x_i) h(w_i, \hat{\theta}_*) = 0$$

## GMM: Optimal IV estimator

Suppose the GMM assumptions hold true then

$$\sqrt{n}(\hat{\theta}_* - \theta_0) \rightarrow^d N(0, V_*)$$

where

$$V_* = E\left[E\left[\frac{\partial h(w_i, \theta_0)}{\partial \theta} \middle| x_i\right] E[h(w_i, \theta_0)^2 \middle| x_i]^{-1} E\left[\frac{\partial h(w_i, \theta_0)}{\partial \theta} \middle| x_i\right]\right]^{-1}$$

and  $\hat{\theta}_*$  is asymptotically semiparametric efficient for conditional moment restriction model.

## GMM: Optimal IV estimator

Since optimal IV is unknown, the optimal IV estimator is infeasible. To estimate  $a^*(x_i)$ , we need to estimate conditional moments

$$E\left[\frac{\partial h(w_i, \theta_0)}{\partial \theta} | x_i\right]$$

and

$$E[h(w_i, \theta_0)^2 | x_i]$$

There are estimated by nonparametric regression, e.g. kernel regression.

## GMM: feasible IV estimator

1. For each  $c$ ,  $a^*(c)$  is nonparametrically estimated, we can compute  $\hat{a}(x_1) \cdots \hat{a}(x_n)$
2. Feasible optimal IV estimator solves

$$\frac{1}{b} \sum_{i=1}^n \hat{a}^*(x_i) h(w_i, \tilde{\theta}_*) = 0$$

$\tilde{\theta}$  is asymptotically equivalent to infeasible version  $\hat{\theta}_*$

## Optimal IV: Linear regress model

$$\begin{aligned}a^*(x_i) &= -x_i E[e_i^2 | x_i]^{-1} \\ &= -\sigma_i^{-2} x_i\end{aligned}$$

Optimal IV GMME solves

$$0 = \frac{1}{n} \sum_{i=1}^n a^*(x_i) h(w_i, \hat{\theta}_*)$$

Thus

$$\hat{\theta}_* = \left( \sum_{i=1}^n \sigma_i^{-2} x_i x_i' \right)^{-1} \left( \sum_{i=1}^n \sigma_i^{-2} x_i y_i \right)$$