Statistics Learning Theory:Generalized Method of Moments

2019.03.17

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

$$Var(\hat{\beta}|X) = E[(X'X)^{-1}X'ee'X(X'X)^{-1}|X]$$

$$= (X'X)^{-1}X'DX(X'X)^{-1}$$

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 $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$ then

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

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

Which is coincide with OLS estimator

Solution of (b).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 restoreshomoscedasticity. The estimator can be defined as

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

We can derive the asympototic distribution as part(a) by changing $m(x, \beta) = \frac{x_i^{\beta}}{\sigma}$

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

Solution of (b).d

Where

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

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

Therefore

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

Solution of Problem(c)

Let f_n be a sequence of functions on $S \subset R$ such that $\sup_{x \in S} |f_n(x) - f(x)| \to 0$

1. Show that $\sup_{x \in S} f_n(x) \to \sup_{x \in S} f(x)$
2. Show that $\inf_{x \in S} f_n(x) \to \inf_{x \in S} f(x)$

Solution of Problem(c)

Since $\sup_{x\in S}|f_n(x) - f(x)| \to 0$, 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 $\frac{-\varepsilon}{2} < f(x) - f(x) < \frac{\varepsilon}{2} \quad \forall x \in S$
(a): By the definition of supremum, there exists $x_{1n}, x_{2n} \in S$ such that

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

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

Solution of Problem(c)

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

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

Solution of Problem(c)

(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.

Solution of (d)

By the definition , we have that

$$\sqrt{n}(\tilde{\theta}-\theta_0)=\sqrt{n}(\tilde{\theta}-\bar{\theta})+\sqrt{n}(\bar{\theta}-\theta_0)$$

$$=-\sqrt{n}(\frac{\partial^2 Q_n(\bar{\theta})}{\partial\theta^2})^{-1}(\frac{\partial Q_n(\bar{\theta})}{\partial\theta})+(\sqrt{n}(\bar{\theta}-\theta_0)$$

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

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

Solution of (d)

Substitute the above, we have that

$$\sqrt{n}(\tilde{\theta} - \theta_0)$$

$$= -\sqrt{n}(\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2})^{-1}(\frac{\partial Q_n(\bar{\theta})}{\partial \theta}) + \sqrt{n}(\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2})^{-1}(\frac{\partial Q_n(\bar{\theta})}{\partial \theta} - \frac{\partial Q_n(\theta_0)}{\partial \theta})$$

$$= -\sqrt{n}(\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2})^{-1}\frac{\partial Q_n(\theta_0)}{\partial \theta}$$

$$+\sqrt{n}\frac{\partial Q_n(\bar{\theta})}{\partial \theta}((\frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2})^{-1} - (\frac{\partial^2 Q_n(\bar{\theta})}{\partial \theta^2})^{-1})$$

$$= (I) + (II)$$

Solution of (d)

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

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

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

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

Since both

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

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

Solution of (d)

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

$$|\sqrt{n}\frac{\partial Q_n(\bar{\theta})}{\partial \theta}|$$

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

$$\leq \sqrt{n}|\frac{\partial Q_n(\theta_0)}{\partial \theta}| \leq \frac{\partial^2 Q_n(\theta_B)}{\partial \theta^2}(\bar{\theta} - \theta_0)|$$

$$= O_p(1) + O_p(1) = O_p(1)$$

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 β_1?
2. One may estimate β_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 β us k-dimenstional 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 β(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 β(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

$$0 = \frac{\partial J_n(\hat{\beta})}{\partial \beta} = 2n(\frac{\partial \bar{g}(\hat{\beta})}{\partial \beta'})' W_n \bar{g}(\hat{\beta})$$

$$= -2n(\frac{1}{n}X'Z)W_n(\frac{1}{n}Z'(y - X\hat{\beta}))$$

Solbing 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

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

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

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

GMM:estimator

Suppose $W_n \to^p W$ for positive definite symmetric W. By CMT

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

where $V_W = (Q'WQ)^{-1}Q'W\Sigma WQ(Q'WQ)^{-1}$
Asympototic 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: 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. Θ *is compact*
2. $W_n \to^p W$ *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

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

GMM: Consistency proof

GMM:Asymptotic normality

Theorem

Suppose

1. $\theta_0 \in int\Theta$
2. $g(w,\theta)$ *is twice continuously differentiable in a neighborhood* $\mathcal{N}$ *of* θ_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

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

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

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

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

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

GMM: Optimal weight proof

By definition, the cross term would be zero:

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

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

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 $\hat{\theta}^j_W \geq \hat{\theta}^j_{\Omega^{-1}}$

GMM: Estimation of Ω

Optimal weight

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

is unknown and should be estimated

By taking sample analog, Ω 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 Ω

Theorem

Suppose

1. $\hat{\theta}_W \to^p \theta_0$
2. $\sup_{\theta \in \mathcal{N}} |\frac{1}{n} \sum_{i=1}^n g(w_i, \theta) g(w_i, \theta)' - E[g(w, \theta) g(w, \theta)']| \to^p 0$
3. $E[g(w, \theta) g(w, \theta)']$ *is continuous at* θ_0

Then

$$\hat{\Omega} \to^p \Omega$$

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

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

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

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

$$J \rightarrow^{d} \mathcal{X}^2(dimg - dim\theta) \text{ under } H_0$$

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

Test statistic:proof for linear case

PS5

Parameter hypothesis

Parameter hypothesis

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

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

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

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

$$W \rightarrow^{d} \mathcal{X}^2(dimr) \;\; under \;\; H_0$$

$$W \rightarrow \infty \;\; under \;\; H_1$$

GMM distance statistic

We can also consider likelihood ratio type test
In thsi 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

$$LR \to^{d} \mathcal{X}^2(dimr) \quad under \quad H_0$$

$$LR \to \infty \quad under \quad H_1$$

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

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

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: GMM:Conditional moment restriction

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

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

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

$$\tilde{\theta} = (\sum \sigma_i^{-2} x_i x_i')^{-1} (\sum \sigma_i^{-2} x_i y_i)$$

$$= \theta_0 + (\frac{1}{n} \sum \sigma_i^{-2} x_i x_i')^{-1} (\frac{1}{n} \sum \sigma_i^{-2} x_i e_i)$$

$$\rightarrow^p \theta_0 + E[\sigma_i^{-2} x_i x_i']^{-1} E[\sigma_i^{-2} x_i e_i]$$

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 θ_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
Whcih $a(\cdot)$(also called intruments) 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

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

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

GMM: Optimal instruments proof

Pick any $a(\cdot)$. Let

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

$$a_i = a(x_i)$$

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

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

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

$$V_a - V_*$$

$$= V_a V_a^{-1} V_a - V_*$$

$$E[m_i m_i^{*'}]^{-1} E[m_i m_i'] E[m^{*_i} m_i']^{-1} - E[m_i^* m_i^{*'}]^{-1}$$

$$E[RR']$$

$$\geq 0$$

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

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

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

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

where

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

and $\hat{\theta}_*$ is asympototically 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

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

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. Feauble 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

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

Optimal IV GMME solves

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

Thus

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