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