Econometrics I

TA Session 8

Jukina HATAKEYAMA *

July 14, 2024

Contents

*E-mail: u868710a@ecs.osaka-u.ac.jp

1 GLS(cont'd)

Reiew of GLS

Consider the case of the linear heteroscedastic model.

$$y = Xb + u, \tag{1}$$

where $y \in \mathbb{R}^{n\times 1}$, $X \in \mathbb{R}^{n\times K}$. There are important assumptions as follows.

GH1 $\mathbb{E}[u|X] = 0$.

GH2 $\mathrm{Var}(u|X) = \Omega \succ 0$.

GH3 $X'X \succ 0$.

Under the assumptions GH1-GH3, the OLS estimator is NOT a BLUE. In place of the OLSE, GLSE is the unbiased linear estimator with the minimal variance.

$$\hat{b}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{2}$$

(End of review)

1.1 GLS–the F Statistic

Suppose the case that we want to test $H_0 : Rb = r$, where $R \in \mathcal{M}_{q\times K}(\mathbb{R})$, $\text{rank}(R) = q \le K$ and $R\hat{b}_{GLS} \sim N(Rb, \sigma^2 R(X'\Omega^{-1}X)^{-1})R')$. similarly to the case of the OLS, we can derive the Wald criterion such that:

$$W_0 = \frac{(R\hat{b}_{GLS} - r)'\{R(X'\Omega^{-1}X)^{-1}R'\}^{-1}(R\hat{b}_{GLS} - r)}{\sigma^2} \sim \chi^2(q), \tag{3}$$

Since $(y^* - X^*\hat{b}_{GLS})'(y^* - X^*\hat{b}_{GLS})/\sigma^2 \sim \chi^2(n-K)$, we have:

$$\frac{(y - X\hat{b}_{GLS})'\Omega^{-1}(y - X\hat{b}_{GLS})}{\sigma^2} \sim \chi^2(n-K).$$

Here, we can say $\hat{b}_{GLS} \perp y - X\hat{b}_{GLS}$ like in the case of the OLS. Therefore, we have the F distribution because it is defined as the portion of the two independent random variables which follow χ^2 distribution:

$$f = \frac{(R\hat{b}_{GLS} - r)'\{R(X'\Omega^{-1}X)^{-1}R'\}^{-1}(R\hat{b}_{GLS} - r)/q}{(y - X\hat{b}_{GLS})'\Omega^{-1}(y - X\hat{b}_{GLS})/(n-K)} \sim F(q, n-K). \tag{4}$$

In addition, we can denote the F statistic as follows by reperesnting $\dot{\hat{u}}_{GLS}$ as the constrained GLS residuals.

$$f_0 = \frac{\dot{\hat{u}}'_{GLS}\Omega^{-1}\dot{\hat{u}}_{GLS} - \hat{u}'_{GLS}\Omega^{-1}\hat{u}_{GLS}/q}{\hat{u}'_{GLS}\Omega^{-1}\hat{u}_{GLS}/(n-K)} \tag{5}$$

Review of the F Distribution for OLS and COLS

Suppose the case of OLS and COLS. In the case of the COLS, the test statistic f_0 is rewritten by the simple representation which we use the residual of OLS and COLS:

$$\begin{aligned} f_0 &= \frac{(R\hat{b} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b} - r)/K - 1}{\hat{u}'\hat{u}/(n-K)} \\ &= \frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}/q}{\hat{u}'\hat{u}/(n-K)}, \end{aligned} \tag{6}$$

where $\hat{u}$ represents the residual of OLS and $\tilde{u}$ is the residual of COLS.

(End of review)

2 M-estimation

An estimator $\hat{\theta}$ is called an extremum estimator if there is a scalar objective function $Q_n(\theta)$ such that

$$\hat{\theta} \text{ maximizes } Q_n(\theta) \text{ subject to } \theta \in \Theta, \tag{7}$$

where $\Theta \in \mathbb{R}^p$ is the parameter space, the set of possible parameter values. The objective function $Q_n(\theta)$ depends not only on θ but also the sample $(w_1, w_2, \cdots, w_n)$, where w_i is the i th observation and n is the sample size. The maximum likelihood method and the generalized method of moments(GMM) estimators are particular extremum estimators. Although we do not prove, the extremum estimator can be derived under some general conditions[1].

[1] Please see Hayashi(2000):446-447 for details.

One of the extremum estimators explained in Econometrics I is M-estimator. The objective function of it is a sample average:

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(w_i;\theta), \tag{8}$$

where $m(w_i;\theta)$ is a real-valued function of w_i and θ. An example of this estimator is MLE (explained in this session).

3 Introductory Topics of the ML Method

3.1 Maximum Likelihood Estimation

Suppose that $X_1, X_2, \cdots, X_n$: i.i.d. random variables. Here, $f\theta(x_i)$ implies the probability density function of X, where θ is a parameter. Then, the maximum likelihood estimator maximizes the likelihood function defined as $l(\theta) = \prod_{i=1}^{n} f\theta(x_i)$. We can rewrite the likelihood by taking logarithm to the likelihood function as follows:

$$\mathbb{L}_n(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \log[f\theta(x_i)]. \tag{9}$$

This is called as a log-likelihood function of X. The maximum likelihood estimator is of θ satisfies the following condition.

Definition 3.1. We say that $\hat{\theta} = \hat{\theta}(X)$ is a MLE of θ if it satisfies the following condition:

$$\hat{\theta} = \arg\max_{\theta} \mathbb{L}_n(\theta).$$

In other words, MLE satisfies the following conditions as follows:

$$\frac{\partial \mathbb{L}_n(\theta)}{\partial\theta} = 0, \quad \frac{\partial^2 \mathbb{L}_n(\theta)}{\partial\theta\partial\theta'} \prec 0. \tag{10}$$

3.2 The Fisher Information

In this subsection, we establish a remarkable inequality called the Cramér–Rao lower bound which gives a lower bound on the variance of any unbiased estimate. Assume that the log likelihood function is continuously twice differentiable and the integral of the log likelihood function is also continuously differentiated twice. Then, we begin with the identity such that $\int f\theta(x)d\lambda(x) = 1$, where $\lambda(x)$ is a Lebesgue measure and take a derivation with respect to θ:

$$\frac{\partial}{\partial\theta}\int f_\theta(x)d\lambda(x) = \int \frac{\partial f_\theta(x)/\partial\theta}{f_\theta(x)} f_\theta(x)d\lambda(x) = \int \frac{\partial \log\, f_\theta(x)}{\partial\theta} f_\theta(x)d\lambda(x) \tag{11}$$

$$= \mathbb{E}[\frac{\partial \log\, f_\theta(x)}{\partial\theta}] = 0. \tag{12}$$

That is, the expectation of the random variable $\frac{\partial \log f_\theta(x)}{\partial\theta}$ is equal to 0. By the product derivative of (11) with respect to θ' again, we can derive:

$$\int \frac{\partial^2 \log\ f_\theta(x)}{\partial\theta\partial\theta'} f_\theta(x) d\lambda(x) + \underbrace{\int \frac{\partial \log f_\theta(x)}{\partial\theta}\frac{\partial \log\ f_\theta(x)}{\partial\theta'} f_\theta(x) d\lambda(x)}_{I(\theta)} = 0.$$

The second term of the (LHS) of this equation can be rewritten as an expectation. We call this expectation **the Fisher information** and denote it by $\mathrm{I}(\theta)$ as follows:

$$\mathrm{I}(\theta) = \mathrm{Var}[\nabla_\theta \log f_\theta(x)] = -\mathbb{E}[\nabla^2_{\theta\theta'} \log f_\theta(x)], \tag{13}$$

because of (12) and the formula of the variance, $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$. In other words, we use following relationships to derive (13).

$$\underbrace{\int \frac{\partial \log\ f_\theta(x)}{\partial\theta}\frac{\partial \log\ f_\theta(x)}{\partial\theta'} f_\theta(x) d\lambda(x)}_{I(\theta)} = -\int \frac{\partial^2 \log\ f_\theta(x)}{\partial\theta\partial\theta'} f_\theta(x) d\lambda(x) := -\mathbb{E}[\nabla^2_{\theta\theta'} \log f_\theta(x)],$$

$$\underbrace{\int \frac{\partial \log\ f_\theta(x)}{\partial\theta}\frac{\partial \log\ f_\theta(x)}{\partial\theta'} f_\theta(x) d\lambda(x)}_{I(\theta)} = \mathbb{E}[(\nabla_\theta \log f_\theta(x)])^2] = \mathrm{Var}[\nabla_\theta \log f_\theta(x)].$$

Definition 3.2. The Fisher information matrix is defined as

$$I(\theta) = -\mathbb{E}[\nabla^2_{\theta\theta'} \log\ f_\theta(X_i)] \tag{14}$$

and we have the equalities

$$I(\theta) = \mathbb{E}[\nabla_\theta \log\ f_\theta(X_i) \nabla_{\theta'} \log\ f_\theta(X_i)] = \mathrm{Var}[\nabla_\theta \log\ f_\theta(X_i)]. \tag{15}$$

3.3 The Cramér–Rao Lower Bound

Note that the following important function is called the score function.

$$\nabla_\theta l_\theta(X) = \nabla_\theta \log\ f_\theta(x)$$

Theorem 3.3. Suppose that an unbiased estimator is given by $f(X)$. Then, we can establish a following relationship:

$$\mathrm{Var}(f(X)) \geq \frac{1}{-\mathbb{E}[\nabla^2_{\theta\theta} l_\theta(x)]} = I(\theta)^{-1}. \tag{16}$$

Proof. At first, taking the derivative with respect to θ to the expectation of $f(X)$ as follows:

$$\begin{aligned}\nabla_\theta \mathbb{E}[f(X)] &= \int f(x) \nabla_\theta f\theta(x) d\lambda(x) \\ &= \int f(x) \nabla_\theta log(f\theta(x)) f\theta(x) d\lambda(x) \\ &= \mathrm{Cov}(f(X), \nabla_\theta l_\theta(X)). \end{aligned} \tag{17}$$

Next, we explain the reason why we can establish (17).

Proof. By definition, we can rewrite $\mathrm{Cov}(f(X), \nabla_\theta l_\theta(X))$ as follows:

$$\begin{aligned}\mathrm{Cov}(f(X), \nabla_\theta l_\theta(X)) &= \int [f(x) - \mathbb{E}[f(x)]][\nabla_\theta l_\theta(x) - \mathbb{E}[\nabla_\theta l_\theta(x)]] f\theta(x) d\lambda(x) \\ &= \int [f(x) - \theta][\nabla_\theta l_\theta(x) - 0] f\theta(x) d\lambda(x) \\ &= \int f(x) \nabla_\theta l_\theta(x) f\theta(x) d\lambda(x), \end{aligned} \tag{18}$$

because we have the following relationship which is a special case of the formula of the covariance:

$$\mathbb{E}[(X - \mu_x)(Y - \mu_y)] = \mathbb{E}[XY - \mu_x Y] = \mathbb{E}[XY] - \mu_x \mathbb{E}[Y] = \mathbb{E}[XY]. \tag{19}$$

□

In the one dimensional case, we can rewrite (17),

$$\begin{aligned}(\nabla_\theta \mathbb{E}[f(X)])^2 &= [\mathrm{Cov}(f(X), \nabla_\theta l_\theta(X))]^2 \\ &= \rho^2 \mathrm{Var}(f(X)) \mathrm{Var}(\nabla_\theta l_\theta(X)) \\ &\leq \mathrm{Var}(f(X)) \mathrm{Var}(\nabla_\theta l_\theta(X)), \end{aligned} \tag{20}$$

with ρ, which implies the correlation between $f(X)$ and $\nabla_\theta l_\theta(X)$. Remind that we can say

$$[\mathrm{Cov}(f(X), \nabla_\theta l_\theta(X))]^2 = \rho^2 \mathrm{Var}(f(X)) \mathrm{Var}(\nabla_\theta l_\theta(X)),$$

by the definition of the correlation coefficient. Since $|\rho| \leq 1$, we have (20). or equivalently, we have following inequality because we have $\mathbb{E}[f(X)] = \theta$ and the derivative of this relationship w.r.t. θ is 1.

$$\mathrm{Var}(f(X)) \geq \frac{1}{-\mathbb{E}[\nabla^2_{\theta\theta} l_\theta(X)]} = I(\theta)^{-1} \tag{21}$$

This inequality also holds for multi dimensional cases. The Cramér–Rao lower bound is a lower bound on the variance of estimators. □

3.4 Example of the ML Method

Suppose the case of the random variable $X \sim N_{\mathbb{R}}(0, \sigma^2)$. The likelihood of the each observed variable x_i $(i = 1, 2, \cdots, n)$ is given as follows:

$$f\theta(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x_i - \mu)^2}{2\sigma^2}] \tag{22}$$

By taking the logarithm, the above equation is rewritten as follows:

$$\log f\theta(x_i) = \frac{1}{2} \log 2\pi - \log\sigma - \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Recall that we must minimize $\sum_{i=1}^{n} \log f\theta(x_i)$ such that:

$$\sum_{i=1}^{n} \log f\theta(x_i) = (constant) - n\log\sigma - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}.$$

Therefore, when we estimate μ, the first order condition is given as follows:

$$\frac{d\sum_{i=1}^{n}(x_i-\mu)^2}{d\mu} = -2\sum_{i=1}^{n}(x_i-\mu) = 0,$$

and $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

4 R Excecise

4.1 BPG Test and White Estimator

Here[2], we study how to test the heteroscedasticity and correct it. Usually, we do not know whether the model is heteroscedastic. In this case, we apply the Breusch=Pagan=Godfrey(BPG) test.

The BPG Test

Suppose that the variance of the regression model is represented as follows:

$$\sigma_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_p Z_{pi} \quad \text{where } i = 1, 2, \cdots, n.$$

The null hypothesis is $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$.

The lmtest package is useful to perform the BPG test. After checking the the model has heteroscedasticity, we estimate the White heteroscedasticity consistent estimator.

The White Estimator

In the heteroscedasticity model, the OLS estimator is still unbiased, consistent, and asymptotically normally distributed. Then, the variance-covariance matrix of the OLSE is given as follows:

$$\mathrm{Var}(\hat{b}) = (X'X)^{-1} X' \Omega X (X'X)^{-1}.$$

Here, we estimate $X'\Omega X$ by using the residuals of OLS. Then, we can estimate the asymptotic variance of OLSE[a].

$$\text{Est. Asy. Var}(\hat{b}) = (X'X)^{-1} \sum_{i=1}^{n} (x_i' x_i \hat{u_i}^2)(X'X)^{-1}.$$

By using this method, we can use the Z statistic for the hypothesis test.

[a]More details of these topics is explained in Greene(2011), Chapter9

4.2 Result of Estimation

In this class, we use the data of the population in Japan and the budgets for the libraries in each prefectures. Consider the following regression model:

$$(Budget)_i = \alpha + \beta (Population)_i + u_i,$$

where i represents the order of the prefectures in the JIS X 0401 code. The data is uploaded in Prof. Okumura's(Univ. Mie) web page [3]. Usually, the White estimator is applied to the large sample. Although we cannot say this example is the large sample case definitely, we introduce this method for reference. R code is given as follows. The lmtest and the sandwich package is useful to the estimation.

[2]These topics are not the coverages of the Econometrics I classes.

[3]奥村晴彦「奥村晴彦研究室ホームページ–いろいろな都道府県別データ」https://oku.edu.mie-u.ac.jp/ okumura/stat/prefectures.html. (最終閲覧日 2019/6/7)

```
library(lmtest)
library(sandwich)
#Today, we check the reg. model has the heterosce. or not.
#After testing the model, we introduce the White correction.

#Usually, the White estimator is applied to the large sample.
#Although we cannot say this example is the large samples definitely,
#we introduce this method for reference.

#The data is uploaded in Prof. Okumura's(Univ. Mie) web page,
#Pop is the population of each prefectures in Japan.
#Y2009 is the budget for the libraries in each pref.

kenmei = c北海道("", 青森県"", 岩手県"", 宮城県"", 秋田県"",山形県
"", 福島県"", 茨城県"", 栃木県"", 群馬県"", 埼玉県"",千葉県
"", 東京都"", 神奈川県"", 新潟県"", 富山県"", 石川県"",福井県
"", 山梨県"", 長野県"", 岐阜県"", 静岡県"", 愛知県"",三重県
"", 滋賀県"", 京都府"", 大阪府"", 兵庫県"", 奈良県"",和歌山県
"", 鳥取県"", 島根県"", 岡山県"", 広島県"", 山口県"",徳島県
"", 香川県"", 愛媛県"", 高知県"", 福岡県"", 佐賀県"",長崎県
"", 熊本県"", 大分県"", 宮崎県"", 鹿児島県"", 沖縄県"")

population = c(5506419, 1373339, 1330147, 2348165, 1085997,
1168924, 2029064, 2969770, 2007683, 2008068, 7194556,
6216289, 13159388, 9048331, 2374450, 1093247, 1169788,
806314, 863075, 2152449,2080773, 3765007, 7410719,
1854724, 1410777, 2636092, 8865245, 5588133, 1400728,
1002198, 588667, 717397, 1945276, 2860750, 1451338,
785491, 995842, 1431493, 764456, 5071968, 849788,
1426779, 1817426, 1196529, 1135233, 1706242, 1392818)

Y2009 = c(39971, 57946, 32853, 38735, 33457, 26141,
38478, 33233, 30389, 40300, 68214, 104662, 319651,
56857, 50000, 41562, 33388, 76735, 45842, 40191,
50872, 97551, 82558, 37100, 67390, 52984, 120266,
28197,58327, 58623, 102056, 53138, 174482, 45093,
50449, 32305, 30263, 23816, 25615, 70935, 53032,
64209, 34625, 47917, 46304, 44273, 23288)

dataset<-data.frame(Pop=population,lib=Y2009)
lib.lm<-lm(lib~Pop,data=dataset)

bptest(lib.lm)
#H_0:homosce. / p value is 0.0002403; H_0 is rejected.

#We can estimate and test White(1980)'s estimator
#by using the sandwich package.
coeftest(lib.lm, df = Inf, vcov = vcovHC(lib.lm, type = "HC0"))
```

The result of the estimation is given as follows.

Table 1: The Z test

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	2.8231e+04	1.0803e+04	2.6134	0.008966**
Population	1.1381e-02	4.8396e-03	2.3516	0.018692*

Signif. codes: 0 ‘ ***’ 0.001 ‘ ** ’ 0.01 ‘ * ’ 0.05 ‘ . ’ 0.1 ‘ ’ 1