Econometrics II

(Thu., 8:50-10:20)

Room # 4 (法経講義棟)

- The prerequisites of this class are **Special Lectures in Economics (Statistical Analysis), 経済学特論（統計解析）** (last semester) and **Econometrics I (エコノメトリックス I)** (graduate level, last semester).

TA Session

by Sakamoto (阪本 諒 r.sk2015.g [at] gmail.com¿)

and Kiku (鞠 露瑤 juluyao [at] yahoo.com)

From	**Oct. 13, 2021**
	PM16:50 – 18:20
zoom ミーティング ID：	**235 834 5146**
パスコード：	**TA2021**

- **Download the lecture notes from the following websites:**

http://www2.econ.osaka-u.ac.jp/˜tanizaki/class/2021/econome2/

http://stat.econ.osaka-u.ac.jp/˜tanizaki/class/2021/econome2/

Contents

1 Maximum Likelihood Estimation (MLE, 最尤(さいゆう)法) — Review **7**

2 Qualitative Dependent Variable (質的従属変数) **30**

2.1 Discrete Choice Model (離散選択モデル) 31

2.1.1 Binary Choice Model (二値選択モデル) 31

2.2 Limited Dependent Variable Model (制限従属変数モデル) 52

2.3 Count Data Model (計数データモデル) 61

2.4 Applications 69

3 Panel Data **95**

3.1 GLS — Review 95

3.2 Panel Model Basic 99

3.2.1 Fixed Effect Model (固定効果モデル) 100
3.2.2 Random Effect Model (ランダム効果モデル) 108
3.3 Hausman's Specification Error (特定化誤差) Test 112
3.4 Choice of Fixed Effect Model or Random Effect Model 114
3.4.1 The Case where $\boldsymbol{X}$ is Correlated with $\boldsymbol{u}$ — Review 114
3.4.2 Fixed Effect Model or Random Effect Model 116
3.5 Applications 118

4 Generalized Method of Moments (GMM, 一般化積率法) 123

4.1 Method of Moments (MM, 積率法) 123
4.2 Generalized Method of Moments (GMM, 一般化積率法) 131
4.3 Generalized Method of Moments (GMM, 一般化積率法) II — Non-linear Case — 153

5 Time Series Analysis (時系列分析) 175

5.1 Introduction 175

5.2 Autoregressive Model (自己回帰モデル or AR モデル) 180

5.3 MA Model 209

5.4 ARMA Model 224

5.5 ARIMA Model 229

5.6 SARIMA Model 230

5.7 Optimal Prediction 230

5.8 Identification 233

5.9 Example of SARIMA using Consumption Data 236

6 Unit Root (単位根) and Cointegration (共和分) 241

6.1 Unit Root (単位根) Test (Dickey-Fuller (DF) Test) 241

6.2 Serially Correlated Errors 277

6.2.1 Augmented Dickey-Fuller (ADF) Test 278
6.3 Cointegration (共和分) 281
6.4 Testing Cointegration 298
6.4.1 Engle-Granger Test 298

1 Maximum Likelihood Estimation (MLE, 最尤法) — Review

1. We have random variables X_1, X_2, $\cdots$, X_n, which are assumed to be mutually independently and identically distributed.

2. The distribution function of $\{X_i\}_{i=1}^n$ is $f(x;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$ and $\theta = (\mu, \Sigma)$.

   Note that X is a vector of random variables and x is a vector of their realizations (i.e., observed data).

   Likelihood function $L(\cdot)$ is defined as $L(\theta; x) = f(x;\theta)$.

   Note that $f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$ when X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed.

The maximum likelihood estimator (MLE) of θ is θ such that:

$$\max_{\theta} \; L(\theta; X). \qquad \Longleftrightarrow \qquad \max_{\theta} \; \log L(\theta; X).$$

MLE satisfies the following two conditions:

(a) $\dfrac{\partial \log L(\theta; X)}{\partial \theta} = 0.$

(b) $\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

3. **Fisher's information matrix (フィッシャーの情報行列)** is defined as:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big),$$

where we have the following equality:

$$-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big) = \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\Big) = \mathrm{V}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big)$$

Proof of the above equality:

$$\int L(\theta; x)\mathrm{d}x = 1$$

Take a derivative with respect to θ.

$$\int \frac{\partial L(\theta; x)}{\partial \theta}\mathrm{d}x = 0$$

(We assume that (i) the domain of x does not depend on θ and (ii) the derivative $\dfrac{\partial L(\theta; x)}{\partial \theta}$ exists.)

Rewriting the above equation, we obtain:

$$\int \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x)\mathrm{d}x = 0,$$

i.e.,

$$\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0.$$

Again, differentiating the above with respect to θ, we obtain:

$$\begin{aligned}
&\int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial L(\theta; x)}{\partial'\theta}\mathrm{d}x \\
&\quad = \int \frac{\partial^2 \log L(\theta; x)}{\partial\theta\partial\theta'} L(\theta; x)\mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial\theta} \frac{\partial \log L(\theta; x)}{\partial\theta'} L(\theta; x)\mathrm{d}x \\
&\quad = \mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\Big) + \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\Big) = 0.
\end{aligned}$$

Therefore, we can derive the following equality:

$$-\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) = \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta} \frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right),$$

where the second equality utilizes $\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right) = 0$.

4. **Cramer-Rao Lower Bound (クラメール・ラオの下限)**: $(I(\theta))^{-1}$

Suppose that an unbiased estimator of θ is given by $s(X)$.

Then, we have the following:

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1}$$

Proof:

The expectation of $s(X)$ is:

$$\mathrm{E}(s(X)) = \int s(x)L(\theta; x)\mathrm{d}x.$$

Differentiating the above with respect to θ,

$$\begin{aligned}\frac{\partial \mathrm{E}(s(X))}{\partial \theta'} &= \int s(x)\frac{\partial L(\theta; x)}{\partial \theta'}\mathrm{d}x = \int s(x)\frac{\partial \log L(\theta; x)}{\partial \theta'}L(\theta; x)\mathrm{d}x \\ &= \mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)\end{aligned}$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$
\begin{aligned}
\left(\frac{\partial \mathrm{E}(s(X))}{\partial\theta}\right)^2 = \left(\mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial\theta}\right)\right)^2 &= \rho^2 \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right) \\
&\leq \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right),
\end{aligned}
$$

where ρ denotes the correlation coefficient between $s(X)$ and $\dfrac{\partial \log L(\theta; X)}{\partial\theta}$, i.e.,

$$
\rho = \frac{\mathrm{Cov}\left(s(X), \dfrac{\partial \log L(\theta; X)}{\partial\theta}\right)}{\sqrt{\mathrm{V}\left(s(X)\right)}\sqrt{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial\theta}\right)}}.
$$

Note that $|\rho| \leq 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 \leq \mathrm{V}(s(X))\ \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$\mathrm{V}(s(X)) \geq \frac{\left(\dfrac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2}{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $\mathrm{E}(s(X)) = \theta$,

$$\mathrm{V}(s(X)) \geq \frac{1}{-\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$\begin{aligned} I(\theta) &= -\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) \\ &= \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right). \end{aligned}$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

5. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ.

As n goes to infinity, we have the following result:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right),$$

where it is assumed that $\lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)$ converges.

That is, when n is large, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $\mathrm{V}(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

Practically, we utilize the following approximated distribution:

$$\tilde{\theta} \sim N\left(\theta, (I(\tilde{\theta}))^{-1}\right).$$

Then, we can obtain the significance test and the confidence interval for θ

6. **Central Limit Theorem:** Let X_1, X_2, $\cdots$, X_n be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for $i = 1, 2, \cdots, n$.

   Define $\overline{X} = (1/n) \sum_{i=1}^{n} X_i$.

   Then, the central limit theorem is given by:

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1).$$

   Note that $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{V}(\overline{X}) = \sigma^2/n$.