Week 1: Nov. 3, 2022

Note Writer: Yu-Chieh Kuo†

†Department of Information Management, National Taiwan University

What is Econometrics

- Part of statistics.
- The statistical methods motivated by problems in economics and social science.

Philosophy

- Persona $\Longleftrightarrow$ Shadow in your mind $\Leftarrow$ Psychoanalysis.
- Mind $\Longleftrightarrow$ Body.
- Observers, organizer (Society, human) $\Longleftrightarrow$ Being observed environment + everything.
- Tightness for muscle causes the anxiety and tiredness.
- Duality $\Longleftrightarrow$ Non-dual $\Rightarrow$ Generate cutoffs in each line.

Science

- Science is the study of observation.
- Data are quantified by observations. Things are interpreted as sequences of random variables or vectors. For example, we may have the observation

$$\left\{ \begin{pmatrix} y_1 \\ x_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ x_2 \\ z_2 \end{pmatrix}, \cdots, \begin{pmatrix} y_i \\ x_i \\ z_i \end{pmatrix}, \cdots \right\},$$

  where i denotes individuals.

  The sequences follow some distributions, for example,

$$\text{Prob}\left(\left\{ \begin{pmatrix} y_1 \\ x_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ x_2 \\ z_2 \end{pmatrix}, \cdots, \begin{pmatrix} y_i \\ x_i \\ z_i \end{pmatrix}, \cdots \right\} \right).$$

  We obtain a special case: the individual observations are identically independently distributed *(i.i.d.)*

$$\text{Prob}\begin{pmatrix} y_1 \\ x_1 \\ z_1 \end{pmatrix}, \text{Prob}\begin{pmatrix} y_2 \\ x_2 \\ z_2 \end{pmatrix}, \cdots, \text{Prob}\begin{pmatrix} y_i \\ x_i \\ z_i \end{pmatrix}, \cdots$$

- Data is the observed, realized sequences. For example,

$$\left\{\begin{pmatrix} y_1 \\ x_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ x_2 \\ z_2 \end{pmatrix}, \cdots, \begin{pmatrix} y_i \\ x_i \\ z_i \end{pmatrix}, \cdots, \begin{pmatrix} y_n \\ x_n \\ z_n \end{pmatrix}\right\} = \left\{\begin{pmatrix} y_i \\ x_i \\ z_i \end{pmatrix}\right\}_{i=1}^{n},$$

  where n denotes the sample size or the number of observations.

Types of Data

Single-indexed y_i, x_i: i denotes individuals (Cross-sectional data)
y_t, x_t: t denotes discrete/continuous time (Time-series data)

Multiple-indexed y_{it}, x_{it}: (Panel data)
y_{ijt}, x_{ijt}: (Multi-dimensional panel spatial data)

Model

Model is the probability distribution over a sequence of random variables or vectors. For example,

- We have a sequence $y_1, y_2, \cdots, y_i, \cdots, y_n$. For each i, $y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$. Then, we estimate μ_i, σ_i^2 for all i.

- We have two sequence $y_1, y_2, \cdots, y_i, \cdots, y_n$ and $x_1, x_2, \cdots, x_i, \cdots, x_n$. For all i, the model represents

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

  where $e_i \sim \mathcal{N}(0, \sigma^2)$. We then want to estimate β, where β is the parameters of interest.

  Note that y_i, x_i and β are $1 \times 1, 1 \times k$, and $k \times 1$ vectors.

Micro-foundations

Whenever an (econometrics) model is derived from an economic model (optimization problem), we say that the model has a micro-foundation.

Parameters in econometrics model are functions of the primitive parameters in utility function or production functions, etc.

Exogeneity and Endogeneity

- Exogenous (given) variables are variables determined outside the world.

- Endogenous variables are variables determined inside the model. Typical description of endogeneity in econometrics textbooks represents

$$\mathbb{E}[x_i e_i] \neq 0.$$

  This statement indicates that there are some variables in x_i that are determined by conditions or equations not in the current model. A typical solution described in the textbook is to add equations to complete the model.

Topics

The following classes will cover topics including

1. Unbiased and consistent conditions (as the sample size goes to infinity).

2. Constrained estimation.

3. Shrinkage estimation (biased and inconsistent). Why we need to discuss such a biased and inconsistent estimator is that there is a trade-off between variance (Cramer Rao lower bound) and bias. We concern the prediction error (prediction performance, in other words) in some model. A simple description is

$$\mathbb{E}[y_i - \hat{y}_i]^2 = |bias|^2 + Variance.$$

4. Asymptotic theory (also known as large-sampe theory, discussing the properties when the sample size goes to infinity).

The textbook refers to Bruce Hanson, Econometrics.

Consistent Estimation

Least square

Criterion or objective functions satisfy the following form

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i;\beta))^2 \text{ or } \mathbb{E}[y_i - f(x_i;\beta)]^2,$$

where β is a $k\times 1$ vector and denotes the parameters of interest. k is the number of parameters.

Suppose we have data

$$\begin{array}{ccccc} y_1 & \cdots & y_i & \cdots & y_n \\ x_1 & \cdots & x_i & \cdots & x_n \\ k\times 1 & & k\times 1 & & k\times 1, \end{array}$$

we denote the conditional expectation of y by $\mu_i \equiv \mathbb{E}[y_i \mid x_i]$, define $\varepsilon_i \equiv y_i - \mathbb{E}[y_i \mid x_i]$, and impose an assumption $\mathbb{E}[\varepsilon_i \mid x_i] = 0$.

Suppose $g(x_i)$ are some functions of x_i, we want to minimize $\mathbb{E}[y_i - g(x_i)]^2$, which can be expanded to

$$\begin{aligned} \mathbb{E}[y_i - g(x_i)]^2 &= \mathbb{E}[y_i - \mu_i + \mu_i - g(x_i)]^2 \\ &= \mathbb{E}\Big[(y_i - \mu_i)^2 + (\mu_i - g(x_i))^2 + 2(y_i - \mu_i)(\mu_i - g(x_i))\Big] \\ &= \mathbb{E}\Big[\varepsilon^2\Big] + \mathbb{E}[\mu_i - g(x_i)]^2 \\ &\geq 0. \end{aligned}$$

The minimizer here is to choose $g(x_i) = \mu_i = \mathbb{E}[y_1 \mid x_i]$.

Now we turn into another scenario. Suppose we have the following minimization problem

$$\min_{\beta} \mathbb{E}\Big[y_i - x_i'\beta\Big]^2 \equiv Q(\beta),$$

and we define $\beta_0 \equiv \arg\min \mathbb{E}\left[y_i - x_i'\beta\right]^2$. The FOC of $Q(\beta)$ gives

$$\frac{\partial Q}{\partial \beta}\,\mathbb{E}\left[2x_i(y_i - x_i'\beta_0)\right] = 0.$$

Here, β_0 is obtained by $\beta_0 = \mathbb{E}\left[x_i x_i'\right]^{-1} \mathbb{E}[x_i y_i]$ Now, we define $e_i \equiv y_i - x_i'\beta_0$, we must automatically hold the result

$$\begin{aligned}
\mathbb{E}[x_i e_i] &= \mathbb{E}\left[x_i(y_i - x_i'\beta_0)\right] \\
&= \mathbb{E}\left[x_i y_i - x_i x_i'\,\mathbb{E}\left[x_i x_i'\right]^{-1}\mathbb{E}[x_i y_i]\right] \quad (\text{Assume } \mathbb{E}\left[x_i x_i'\right] \text{ is invertible}) \\
&= 0. \quad \textbf{why???????}
\end{aligned}$$

However, $x_i'\beta$ may not be the true $\mathbb{E}[y_i \mid x_i]$.

In general, we can summarize the above problem as

$$\beta_0 = \arg\min_{\beta} \mathbb{E}[y_i - f(x_i;\beta)]^2.$$

If we define $e_i \equiv y_i - f(x_i;\beta)$, then $\mathbb{E}\left[\frac{\partial f_i}{\partial \beta} e_i\right] = 0$.

In advance, suppose we have the following problem

$$Q_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2,$$

we have

$$\hat{\beta} \equiv \arg\min_{\beta} Q_n(\beta) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i'\right).$$

Law of large number

Suppose $z_1, \cdots, z_n$ are *i.i.d.*, then

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} z_i &\xrightarrow{p} \mathbb{E}[z_i] \\
\frac{1}{n}\sum_{i=1}^{n} x_i x_i' &\xrightarrow{p} \mathbb{E}\left[x_i x_i'\right] \\
\frac{1}{n}\sum_{i=1}^{n} x_i y_i' &\xrightarrow{p} \mathbb{E}\left[x_i y_i'\right].
\end{aligned}$$

That is,

$$\hat{\beta} \xrightarrow{p} \beta_0.$$

Remarks on the true model

- We may define the true parameters as β_0, then we have $\mathbb{E}[x_i e_i] = 0$
- Another way to start the econometrics problem is to define $y_i = x_i'\beta_0 + e_i$ with an assumption $\mathbb{E}[x_i e_i] = 0$.

  If $\mathbb{E}[x_i e_i] \neq 0$, we need instruments.

Asymptotic Theory

(This section refers to the Chapter 6 of the Bruce Hansen's econometrics textbook.)

To discuss the asymptotic properties, we need to define the limit firstly.

Definition. Suppose we have a non-random sequence of numbers $\{a_1, a_2, \cdots, a_n, \cdots\}$,

$$a_n \to a \text{ as } n \to \infty \quad \Longleftrightarrow \quad \lim_{n\to\infty} a_n = a.$$

Clearly,

$$\forall\, \delta > 0, \exists\, n_\delta < \infty \text{ s.t. } |a_n - a| < \delta \quad \forall n > n_\delta.$$

□

Convergence in probability

If z_n converges in probability to z as $n \to \infty$, we say

$$z_n \xrightarrow{p} z \quad \Longleftrightarrow \quad \operatorname*{plim}_{n\to\infty} z_n = z.$$

Clearly,

$$\forall\, \delta > 0, \lim_{n\to\infty} \text{Prob}\,(|z_n - z| \le \delta) = 1.$$

(Some notation states $\lim_{n\to\infty} \text{Prob}\,(|z_n - z| > \delta) = 0$)

Almost sure convergence

We denote z_n converging almost surely to z as $n \to \infty$ by $z_n \xrightarrow{a.s.} z$. Clearly,

$$\forall\, \delta > 0, \text{Prob}\left(\lim_{n\to\infty} |z_n - z| \le \delta\right) = 1.$$

Note that the almost sure convergence implies convergence in probability.

Following the conception of convergence above, we can now introduce the law of large number:

Weak Law of Large Number (WLLN) $\frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathbb{E}[y_i]$. The data y_i is i.i.d. here.

Strong Law of Large Number (SLLN) $\frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{a.s.} \mathbb{E}[y_i]$. The data y_i is i.i.d. here.

Convergence in distribution

Given $z_1, z_2, \cdots, z_n$ as a sequence of random variables or vectors, and $F_1(z), F_2(z), \cdots, F_n(z)$ are probability distributions. If z_n converges in distribution to z, says $z_n \xrightarrow{d} z$, it gives $F_n(z) \to F(z)$ point-wisely (for all continuous point of $F(\cdot)$).

Central limit theorem

Given an i.i.d. sequence of random variables $\{y_1, y_2, \cdots, y_n\}$ and the true expectation $\mathbb{E}[y_i] = \mu$, we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}[y_i - \mu]^2\right).$$

Remark. The key conception is *independent* in i.i.d. □

Remark. The re-scale coefficient $\sqrt{n}$ aims at decreasing the convergence speed in distribution to maintain the randomness. $\frac{1}{n}$ might be too fast. □

In the case of linear least square, we obtain

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

and it implies

$$\begin{aligned}\hat{\beta} - \beta_0 &= \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right) \quad \textbf{why??????}\\ &\xrightarrow{p} \mathbb{E}\left[x_i x_i'\right]\mathbb{E}[x_i e_i]\\ &\xrightarrow{p} 0.\end{aligned}$$

After re-scaling by $\sqrt{n}$, it alters to

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} x_i e_i\right)$$

and the latter part derives to

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i e_i - \mathbb{E}[x_i e_i]\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[x_i x_i' e_i^2\right]\right) \quad \text{since } \mathbb{E}[x_i e_i] = 0.$$

Hence,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[x_i x_i'\right]^{-1}\mathbb{E}\left[x_i x_i' e_i^2\right]\mathbb{E}\left[x_i x_i'\right]^{-1}\right).$$