Applied Econometrics

Derek Li

Contents

1 Introduction

Econometrics are quantitative methods of analyzing and interpreting economic data, which need to combine economic theory, math and statistics, data, and statistical or econometrics software.

We focus on estimating economic relationships, testing hypothesis involving economic behavior, and forecasting the behavior of economic variables.

1.1 Econometric Methodology

- Ask a question - statement of theory or hypothesis
- Specification of economic model
- Specification of econometric model
- Collection of Data
- Estimation of the econometric model
- Hypothesis testing
- Prediction or forecasting

1.2 Data Types

There are different data structures: cross-section, time series, repeated cross-section, and panel data.

1.2.1 Cross-Section

Cross-section consists of a sample of individuals, households, firms, countries, etc, taken at a given point in time. Observations are generally independent draws from the population. It is commonly indexed by i as x_i.

1.2.2 Time Series

Time series consists of observations on a variable or several variables over time. Observations are almost never independent of each other. It is commonly indexed by t as x_t.

1.2.3 Repeated Cross-Section

Repeated cross-section consists of two or more cross-sectional data in different points in time, and is different units in different periods. It is commonly index by it as x_{it}.

1.2.4 Panel Data

Panel data consists of a time series for each cross-sectional unit in the data set. Observations are independent among units and dependent over time for each unit. It is commonly index by it as x_{it}.

2 Review of Statistics

Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. Assume a random sample $\{x_1, \cdots, x_n\}$, i.e., identically and independently distributed (i.i.d.).

2.1 Estimator

Definition 2.1 (Statistic)**.** A statistic is a function of the data.

Definition 2.2 (Estimator)**.** An estimator is a statistic that is used to estimate the parameter of interest.

Take μ as an example, the proposed estimator is sample average

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

2.2 Sampling Distribution

We have

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$

i.e., $\overline{X}_n$ is an unbiased estimator of μ. Besides,

$$\mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} x_i\right] \overset{\text{independet}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[x_i\right] = \frac{\sigma^2}{n}.$$

Definition 2.3 (Consistency)**.** Let W_n be an estimator of θ based on a sample $Y_1, \cdots, Y_n$. Then W_n is a consistent estimator of θ is for every $\varepsilon > 0$,

$$P(|W_n - \theta| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

As commonly stated, an estimator is called consistent when its sampling distribution becomes more and more concentrated around the parameters of interest as the sample size increases. Note that $\overline{X}_n$ is a consistent estimator of μ.

Theorem 2.1. If $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then

$$Y_1 + Y_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\mathrm{Cov}(Y_1, Y_2)),$$

By theorem, we have the sampling distribution

$$\overline{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}),$$

provided $X \sim \mathcal{N}(\mu, \sigma^2)$.

2.3 Confidence Interval

Theorem 2.2. If Y has $\mathbb{E}[Y] = \mu, \mathrm{Var}[Y] = \sigma^2$, then $Z = \frac{Y-\mu}{\sigma}$ is such that $\mathbb{E}[Z] = 0, \mathrm{Var}[Z] = 1$.

By theorem, we have

$$Z = \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1).$$

Therefore,

$$\begin{aligned} 1 - \alpha &= P\left(-z_{\frac{\alpha}{2}} \leqslant Z \leqslant z_{\frac{\alpha}{2}}\right) = P\left(-z_{\frac{\alpha}{2}} \leqslant \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant z_{\frac{\alpha}{2}}\right) \\ &= P\left(\overline{X}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \overline{X}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right), \end{aligned}$$

i.e., $(1 - \alpha)\%$ confidence interval is

$$\left[\overline{X}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

2.4 Proves of Some Theorems and Results (optional)

Theorem 2.3 (Markov's Inequality)**.** If X is a nonnegative random variable and $a > 0$, then

$$P(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}.$$

Proof. Since X is a nonnegative random variable, we have

$$\begin{aligned}\mathbb{E}[X] = \int_0^\infty xf(x)\mathrm{d}x = \mathbb{E}[X] &= \int_0^a xf(x)\mathrm{d}x + \int_a^\infty xf(x)\mathrm{d}x \\ \geqslant \int_a^\infty xf(x)\mathrm{d}x &\geqslant \int_a^\infty af(x)\mathrm{d}x = a\int_a^\infty f(x)\mathrm{d}x = aP(X \geqslant a).\end{aligned}$$

Hence

$$P(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}.$$

□

Theorem 2.4 (Chebyshev Inequality)**.** For any $b > 0$,

$$P(|X - \mathbb{E}[X]| \geqslant b) \leqslant \frac{\mathrm{Var}[X]}{b^2}.$$

Proof. By Markov's Inequality, we have

$$P((X - \mathbb{E}[X])^2 \geqslant b^2) \leqslant \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{b^2} = \frac{\mathrm{Var}[X]}{b^2}.$$

Therefore,

$$P(|X - \mathbb{E}[X]| \geqslant b) \leqslant \frac{\mathrm{Var}[X]}{b^2}.$$

□

Theorem 2.5 (Weak Law of Large Numbers)**.** Let $X_1, \cdots, X_n$ be a sequence of independent random variables with $\mathbb{E}[X_i] = \mu$. Let $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, then

$$\forall \varepsilon > 0, \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = 0.$$

Proof. By Chebyshev Inequality, for all $\varepsilon > 0$,

$$P\left(\left|\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]\right| > \varepsilon\right) \leqslant \frac{\mathrm{Var}\left[\overline{X}\right]}{\varepsilon^2} \Leftrightarrow 0 \leqslant P(|\overline{X}_n - \mu| > \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon}.$$

Since

$$\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon} = 0,$$

then

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = 0.$$

□

It follows that $\overline{X}_n$ is a consistent estimator of μ.