Business School, Brunel University
MSc. EC5501/5509 Modelling Financial Decisions and Markets/Introduction to Quantitative Methods
Prof. Menelaos Karanasos (Room SS269, Tel. 01895265284)

Lecture Notes 2

1. Estimation

The probability distributions serve as models for the underlying processes that produce our observed data. The goal of statistical inference in econometrics is to use the principles of mathematical statistics to combine these theoretical distributions and the observed data into an empirical model of the economy. The classical theory of statistical inference centers on rules for using the sampled data effectively.

A sample of n observations on one (or more) variable x is a *random sample* if the n observations are drawn independently from the same population, or probability density $f(x, \theta)$, where θ denotes the population parameters (i.e. the parameters of interest). Data are generally drawn in one of two settings. *A cross section* is a sample of a number of observational units all drawn at the same point in time. *A time series* is a set of observations drawn on the same observational unit at a number of (usually evenly spaced) points in time. Our objective is to use the sample data to infer the value of a parameter or set of paramaters, θ.

An **estimator** is a rule or strategy for using the data to estimate the parameters. An *estimator is a random variable*, thus it has a distribution characterized by an expectation and a variance. The estimator corresponding to a specific sample is called an **estimate** and it is a number. In other words, an estimator is simply an algebraic function of a potential sample of data; once the sample is drawn,this function creates an actual numerical estimate. Econometric theory focuses not on the estimate itself, but on the estimator because the justification of an estimate computed from a particular sample rests on a justification of the estimation method. So, the search for good estimators constitutes much of econometrics. Estimators are compared on the basis of a variety of attributes.

Finite sample properties of estimators are those attributes that can be compared regardless of the sample size.

- An estimator $\overset{\wedge}{\theta}$ of a parameter, θ, is **unbiased** if the mean of its sampling distribution is θ, i.e.

$$E\left(\overset{\wedge}{\theta}\right) = \theta \text{ or } E\left(\overset{\wedge}{\theta} - \theta\right) = \text{Bias}\left(\overset{\wedge}{\theta}\right) = 0.$$

  If θ is a vector of parameters, the estimator is unbiased if the expected value of every element of $\overset{\wedge}{\theta}$ equals the corresponding element of θ.

- An unbiased esimator $\overset{\wedge}{\theta}_1$ is **more efficient** than another unbiased estimator $\overset{\wedge}{\theta}_2$, if the variance of the sampling distribution of $\overset{\wedge}{\theta}_1$ is less than that of $\overset{\wedge}{\theta}_2$. That is

$$V\left(\overset{\wedge}{\theta}_1\right) < V\left(\overset{\wedge}{\theta}_2\right).$$

The above discussion was restricted to unbiased estimators. However, a biased estimator might be preferred to an unbiased estimator with a much larger variance. A criterion that recognizes this possible tradeoff is the mean squared error.

- The **mean squared error** of an estimator is

$$\text{MSE}\left(\overset{\wedge}{\theta}\right) = E\left[\left(\overset{\wedge}{\theta} - \theta\right)^2\right] = V\left(\overset{\wedge}{\theta}\right) + \left[\text{Bias}\left(\overset{\wedge}{\theta}\right)\right]^2,$$

  if θ is a scalar. The estimator with the smaller MSE is preferred.

- An estimator is **minimum variance linear unbiased** (MVLUE) or **best linear unbiased** (BLUE) if it is a linear function of the data and has minimum variance among all the linear unbiased estimators.

Some estimation problems involve characteristics that are not known in finite samples. In these instances, estimators are compared on the basis of their large sample, or *asymptotic properties.*

- Let x_n be a random variable (r.v) indexed by the size of a sample. x_n converges in probability to a constant, c, if

$$\lim_{n\to\infty} \text{Prob}\left(|x_n - c| > \varepsilon\right) = 0,$$

for any positive ε. *Convergence in probability* can also be expressed as

$$\text{plim } x_n = c \Leftrightarrow x_n \xrightarrow{P} c$$

- A special case of convergence in probability is *convergence in mean square* or convergence in quadratic mean. x_n converges in mean square to a constant, c, if
$$\lim_{n\to\infty} E(x_n) = c \text{ and } \lim_{n\to\infty} V(x_n) = 0.$$
Convergence in mean square implies convergence in probability, the converse is not true.

- An estimator $\overset{\wedge}{\theta}_n$ of a parameter θ is a **consistent estimator** if and only if
$$\text{plim}\overset{\wedge}{\theta}_n = \theta.$$

(Note: an *asymptotic distribution* is a distribution that is used to approximate the true finite sample distribution of a random variable.)

1.1. The Method of Least Squares

The method of least squares was first introduced by Legendre in 1805 and Gauss in 1809 in the context of astronomical measurements. The principle of least squares involves minimising the sum of squared errors.

Consider a random sample, y_i, $i = 1, ..n$, drawn from some population with mean μ and variance σ^2. We can write:

$$\begin{aligned} y_i &= E(y_i) + \varepsilon_i, \text{ where } \varepsilon_i \sim iid\left(0, \sigma^2\right), \\ \text{or } y_i &= \mu + \varepsilon_i. \end{aligned}$$

Thus

$$\min_{\mu} \sum_{i=1}^{n} \varepsilon_i^2$$

gives the least-squares estimator (OLS) of the population mean

$$\widehat{\mu} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} y_i,$$

which is the sample mean. In this case, we say that the least squares estimate for the population mean is obtained by regressing y on a constant.

Regression analysis is concerned with describing and evaluating the relationship between a given variable (dependent) and one or more other variables (independent). The term regression was coined by Sir Francis Galton (1822-1911) who was studying the relationship between the height of children and the height of parents. He observed that although tall (short) parents had tall (short) children, there was a tendency for children's heights to converge toward the average.

1.2. The Method of Maximum Likelihood

The maximum likelihood method of estimation was formulated by Fisher in a series of papers in the 1920's and 30's.

Consider a random sample, y_i, $i = 1, ..n$, drawn from a normal distribution with mean μ and variance σ^2, i.e.

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - \mu)^2}{2\sigma^2} \right\}.$$

Suppose that we are interested in estimating the population mean μ. The likelihood function (L) is the joint probability density function of the y_i's (times a constant which plays no role in the estimation). Since the sample is independent we can write

$$L \equiv f(y_1, y_2, ..., y_n) = \prod_{i=1}^{n} f(y_i).$$

The principle of maximum likelihood involves maximising the likelihood function with respect to the papameters of interest. It has to be noted that the logarithm of the likelihood function represents the same information. So we can maximise the log-likelihood funcion (LL):

$$\begin{aligned} \max_{\mu} LL &= \sum_{i=1}^{n} \log f(y_i) \Rightarrow \\ \max_{\mu} LL &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 \end{aligned}$$

and obtain the maximum-likelihood estimator (MLE) of the population mean

$$\widehat{\mu} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} y_i.$$

Observe that the above, as in the case of least squares, is the sample mean.