Lecture 1: Statistical Review

R.G. Pierse

1 What is Econometrics?

Some definitions:

> The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. *T. Haavelmo (1944)*

> ...econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference.
> *P.A. Samuelson, T.C. Koopmans and J.R.N. Stone (1954)*

> Econometrics, the result of a certain outlook on the role of economics, consists of the application of mathematical statistics to economic data to lend empirical support to the models constructed by mathematical economics and to obtain numerical results. *G. Tintner (1968)*

> Econometrics is concerned with the empirical determination of economic laws. *H. Theil (1971)*

2 Econometric Methodology

1. **Statement of economic theory or hypothesis**

   e.g. Keynesian consumption function:

$$C = f(Y) \quad , \quad 0 < \frac{\partial C}{\partial Y} < 1$$

2. **Specification of mathematical model**

$$C = a + bY \qquad 0 < b < 1$$

This is a precise mathematical representation of the economic theory. It may make additional assumptions that are not a necessary part of the theory: e.g. in this case a *linear* functional form is assumed. The mathematical model is *exact* or *deterministic*.

3. **Specification of econometric model**

$$C = a + bY + u$$

u is a disturbance or error term. It is a random variable with well-defined properties. Unlike the mathematical model, the econometric model is probabilistic or *stochastic*.

4. **Obtaining data** This involves making further assumptions e.g. which measures of consumption and income to use, whether variables should be in real or nominal terms etc. A plot of the data will show that in practice no exact relationship will hold between the measured variables.

5. **Estimation of parameters of the model** Regression analysis is used to obtain *estimates* of the parameters a and b in the econometric model. These estimates are generally denoted using hats as $\widehat{a}$ and $\widehat{b}$.

6. **Hypothesis testing** Having estimated a and b we can then test statistical hypotheses about them e.g. is the marginal propensity to consume *significantly* different from unity?

7. **Forecasting and Prediction** Having estimated the parameters a and b also allows us to use the model for prediction or forecasting of consumption for a given value of income using the equation

$$\widehat{C} = \widehat{a} + \widehat{b}Y$$

8. **Using the model for policy analysis** The model can be used to answer policy questions e.g. what is the effect on consumption of cutting taxes (and thereby raising disposable income) by 5%?

3 The Summation Operator

3.1 The definition of the summation operator

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \tag{3.1}$$

3.2 Properties of the summation operator

3.2.1 Sum of a constant k

$$\sum_{i=1}^{n} k = nk \tag{3.2}$$

3.2.2 Product of a variable with a constant

$$\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i \tag{3.3}$$

3.2.3 Adding sums of two variables

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \tag{3.4}$$

3.2.4 The square of a sum of variables

$$\begin{aligned} \left(\sum_{i=1}^{n} x_i\right)^2 &= \sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j = \sum_{i=1}^{n} x_i^2 + \sum_{j\neq i}\sum_{i=1}^{n} x_i x_j \\ &= \sum_{i=1}^{n} x_i^2 + 2\sum_{j>i}\sum_{i=1}^{n} x_i x_j \end{aligned} \tag{3.5}$$

3.3 Things you can't do with summations

3.3.1 Multiplying sums (note the *inequality*)

$$\sum_{i=1}^{n} x_i * y_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \tag{3.6}$$

3.3.2 Dividing sums (note the *inequality*)

$$\sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} \tag{3.7}$$

4 Probability

The probability of an event in the *relative frequency* of that event occurring

$$0 \leq P(x) \leq 1$$

The sum of the probabilities over all possible events is equal to one.

Throwing a single die, the probability of achieving any of the possible values $1, 2, 3, 4, 5, 6$ is the same and equals $\frac{1}{6}$. Throwing a pair of dice, the possible scores are $2 - 12$ and their probabilities are as follows:

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

5 Random Variables

A variable whose value is determined probabilistically is called a random variable. Random variables can either be *discrete* (taking only a finite number of values) or *continuous* (taking any value in some range). The score from throwing a pair of dice is an example of a discrete random variable. The height of a person is an example of a continuous random variable.

6 Probability Density Functions (PDFs)

6.1 Probability density function (continuous random variable x)

$$f(x) \geq 0$$

$$\int_a^b f(x)dx = P(a \leq x \leq b) \tag{6.1}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

6.2 Probability density function (discrete random variable x)

$$f(x) = P(x = x_i) \geq 0 \qquad i = 1, \cdots, n$$

$$\sum_{i=1}^{n} P(x = x_i) = 1$$

6.3 Cumulative distribution function (continuous random variable x)

$$F(a) = \int_{-\infty}^{a} f(x)dx = P(x \leq a)$$

6.4 Joint probability density functions (continuous variables x and y)

$$f(x, y) \geq 0$$

$$\int_{c}^{d} \int_{a}^{b} f(x, y)\ dx\ dy = P(a \leq x \leq b, c \leq y \leq d) \tag{6.2}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\ dx\ dy = 1$$

6.5 Marginal density functions (continuous variables x and y)

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\ dy$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\ dx$$

6.6 Conditional density functions

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

7 Independence

If two random variables x and y are independent then

$$\begin{aligned} f(x,y) &= f(x)\ f(y) \\ f(y|x) &= f(y) \\ f(x|y) &= f(x) \end{aligned} \tag{7.1}$$

8 The Algebra of Expectations

8.1 Expected Value (continuous random variable)

$$E(X) = \mu_x = \int_{-\infty}^{\infty} f(x)\ x\ dx \tag{8.1}$$

8.2 Expected Value (discrete random variable)

$$E(X) = \mu_x = \sum_{i=1}^{n} P(x_i)\ x_i \tag{8.2}$$

8.3 Properties of the Expectations operator

8.3.1 Expectation of a fixed constant k

$$E(k) = k \tag{8.3}$$

8.3.2 Expectation of a product of a variable with a constant

$$E(kX) = kE(X) \tag{8.4}$$

8.3.3 Adding expectations of two variables

$$E\left(X+Y\right) = E(X) + E(Y) \tag{8.5}$$

8.3.4 Expectation of a function of a discrete random variable

$$E(g(X)) = \sum_{i=1}^{n} g(x_i)\ P(x_i) \tag{8.6}$$

8.4 Expectation of a function of a continuous random variable

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\ f(x)\ dx \tag{8.7}$$

8.5 Things you can't do with the expectations operator

8.5.1 Multiplying expectations (note the *inequality*)

$$E(X * Y) \neq E(X) * E(Y) \tag{8.8}$$

8.5.2 Dividing expectations (note the *inequality*)

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)} \tag{8.9}$$

8.5.3 The sum of squares is not equal to the square of the sums

$$E(X^2) \neq (E(X))^2 \tag{8.10}$$

9 Variance

$$\begin{aligned} Var(X) &= E(X - E(X))^2 = E(X - \mu_x)^2 \\ &= E(X^2) - \mu_x^2 \end{aligned} \tag{9.1}$$

9.1 Variance (continuous random variable)

$$Var(X) = \int_{-\infty}^{\infty} (X - \mu_x)^2\ f(x)\ dx \tag{9.2}$$

9.2 Variance (discrete random variable)

$$Var(X) = \sum_{i=1}^{n} (X_i - \mu_x)^2\ P(x_i) \tag{9.3}$$

10 Covariance

$$\begin{aligned} Cov(X,Y) &= E(X - E(X))(Y - E(Y)) = E(X - \mu_x)(Y - \mu_y) \\ &= E(X\ Y) - \mu_x \mu_y \end{aligned} \tag{10.1}$$

10.1 Covariance (continuous random variable)

$$Cov(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - \mu_x)\ (Y - \mu_y)\ f(x,y)\ dx\ dy \tag{10.2}$$

10.2 Covariance (discrete random variable)

$$Cov(X,Y) = \sum_{j=1}^{n} \sum_{i=1}^{n} (X_i - \mu_x)\ (Y_j - \mu_y)\ f(x,y) \tag{10.3}$$

If X and Y are independent random variables, then their covariance is zero. The reverse is not true.

10.3 Properties of Variance

10.3.1 Variance of a fixed constant k

$$Var(k) = 0$$

10.3.2 Variance of a product of a variable with a constant

$$Var(kX) = k^2 Var(X) \tag{10.4}$$

10.3.3 Variance of the sum of two random variables

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$