Econometrics

Professor Seppo Pynnönen
Department of Mathematics and Statistics
University of Vaasa

with amendments and corrections
by Bernd Pape

1. Introduction

1.1 Econometrics

Econometrics is a discipline of statistics, specialized for using and developing mathematical and statistical tools for empirical estimation of economic relationships, testing economic theories, making economic predictions, and evaluating government and business policy.

Data: Nonexperimental (observational)

Major tool: Regression analysis (in wide sense)

1.2 Steps in Econometric Analysis

(a) Economic model

For example, consumption function

(1) $$C = f(Y, W),$$

where C is consumption, Y is income and W wealth.

(b) Econometric model

Usually the economic model does not specify exactly the functional form. In addition the economic model is assumed to be exact in the simplified world satisfying the simplifying assumptions.

The task of econometrics is to turn the economic model to a operational one. Usually this amounts to a linear approximation,

$$C = \beta_0 + \beta_1 Y + \beta_2 W + u, \quad (2)$$

where β_0, β_1, and β_2 are parameters of the model, to be estimated from the data, and u is (unobservable) random *error* or *disturbance* term, which determines the stochastic properties of the model.

Another popular specification in economics is the log-log model.

(3)

$$\log(C) = \beta_0 + \beta_1 \log(Y) + \beta_2 \log(W) + u.$$

where log is the natural logarithm. In (3) parameters β_1 and β_2 are *elasticities*. Coefficient β_1 is the income elasticity of consumption and β_2 is the wealth elasticity of consumption. (Can you prove this?)

Remark 1.1: Although the same symbols are used, parameters β_0, β_1, and β_2 in (2) and (3) are different. The same is true with the error term u.

Remark 1.2: Model (3) can be again consideed technically in the framework of linear models by defining $c = \log(C)$, $y = \log(Y)$, and $w = \log(W)$ so that (3) becomes

$$c = \beta_0 + \beta_1 y + \beta_2 w + u,$$

which is again a linear model.

A third possibility could be

$$\log(C) = \beta_0 + \beta_1 Y + \beta_2 W + u. \tag{4}$$

Can you figure out how e.g. β_1 can be interpreted here?

1.3 Types of Economic Data

(a) Cross-sectional

Data collected at given point of time. E.g. a sample of households or firms, from each of which are a number of variables like turnover, operating margin, market value of shares, etc., are measured.

From econometric point of view it is important that the observations consist a *random sample* from the underlying *population*.

(b) Time Series Data

A time series consist of observations on a variable(s) over time. Typical examples are daily share prices, interest rates, CPI values.

An important additional feature over cross-sectional data is the *ordering* of the observations, which may convey important information.

An additional feature is *data frequency* which may require special attention.

(c) Pooled Cross-sections

Both time series and cross-section features. An example is a data set where a number of firms are randomly selected, say in 1990, and another sample is selected in 2000. If in both samples same features are measured, combining both years form a pooled cross-section data set.

Pooled cross-section data is analyzed much the same way as usual cross-section data. However, many times it is important to pay special attention to the fact that there are 10 years in between. Usually the interest is whether there are some important changes between the time points. Statistical tools are usually the same as those used for analysis of differences between two independently sampled populations.

(d) Panel Data

Panel data (longitudinal data) consists of time series for each cross-sectional member in the data set. That is one has series of history from each individual.

1.4 Causality, Ceteris Paribus

Causality: Cause and effect $x \to y$

Ceteris Paribus: "Holding other relevant factors fixed".