USC, Fall 2016, Economics 513

Lecture 1: Empirical economic relations

What is econometrics?

Econometrics is concerned with the measurement of economic relations.

So we need to know

- What is an economic relation and why do we want to measure it?
- How do we measure such a relation?

Definition: An economic relation is a relation between economic variables.

Examples:

- Production function: relation between output of firm and inputs of labor, capital, materials.
- Engel curve: relation between expenditure on a commodity and household income.
- Hedonic relation: relation between price and attributes of a good or service.
- Phillips curve: relation between inflation and unemployment rates.
- Earnings function: relation between earnings and education, work experience.

All these relations can be expressed as mathematical functions

$$y = f(x_1, \ldots, x_K)$$

with y the dependent variable in the relation and $x_1, \ldots, x_K$ the independent variables.

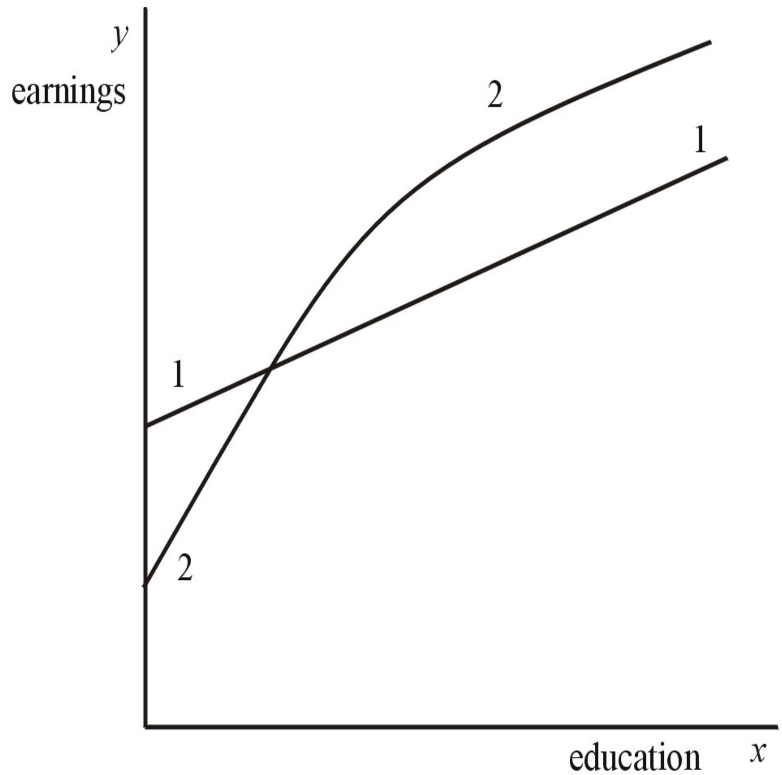
Example: Earnings function

1: straight line, linear relation $y = \beta_0 + \beta_1 x$

2: non-linear relation

Economic theory usually does not specify functional form, but it may sign derivatives.

From the National Longitudinal Survey of Youth (NLSY) we obtain (usual) weekly earnings and years of education for a sample of 935 individuals



y_i = logarithm of usual weekly earnings of i

x_i = years of education of i

We plot the 935 pairs (y_i, x_i) in a diagram that is called a scatterplot (see figure)

Note

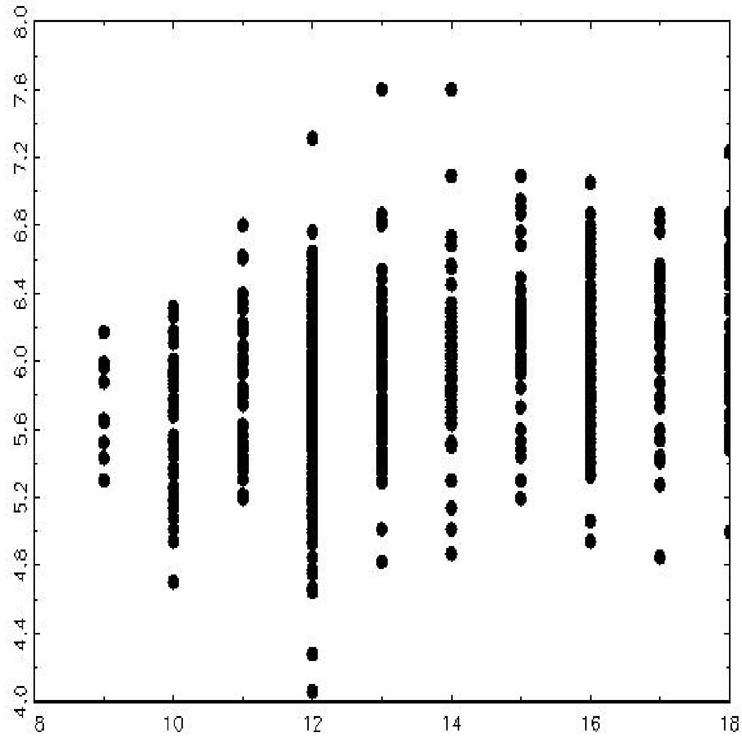
- Relation is not a simple mathematical function.
- Relation is not even a mathematical function, because for same level of education we may have 2 or more levels of expenditure.

First attempt to measure an economic relation seems a total failure!

The problem is that economic relations hold ceteris paribus, i.e. holding all other relevant variables constant.

In example earnings also depends on work experience, ability etc.

The exact relation is



$$y = f(x_1, x_2, \ldots, x_M)$$

but in the scatterplot $x_2, \ldots, x_M$ are omitted or assumed to be constant.

The scatterplot is the two-dimensional projection of a relation involving many variables.

Example: Earnings function

$$y_i = \beta_0 + \beta_1 \underbrace{x_{i1}}_{\text{education}} + \underbrace{\beta_2 x_{i2} + \ldots + \beta_M x_{iM}}_{\text{other relevant vars}}$$

Even if relation between earnings and education is linear, the observed values of these variables satisfy

$$y_i = \beta_0 + \beta_1 x_{i1} + e_i$$

with $e_i = \beta_2 x_{i2} + \ldots + \beta_M x_{iM}$ the contribution of the unobserved, but relevant variables.

Why do we want to measure economic relations?

- Testing predictions of economic theories. Example: economic theory predicts that price has a negative effect on demand and a positive effect on supply. This is not easy to confirm with observed data.
- Estimation of the causal effect of a policy/program/intervention.
- Prediction. Examples: predicting GDP in future years or recommendations by Amazon or Netflix.

We will see that the concerns when measuring economic relations are different between prediction and testing theory/causal effect.

Fitting a straight line

Assume that the relation between y and x_1 is indeed linear ceteris paribus.

How do we measure it, i.e. how do we measure β_0, β_1 ?

Idea: choose β_0, β_1 such that the line fits the observations as well as possible. Quality of fit is measured by size of the deviations $e_1, \ldots, e_n$

$$e_i = y_i - \beta_0 - \beta_1 x_{i1} \ , \qquad i = 1, \ldots, n$$

Deviations $e_1, \ldots, e_n$ can be positive or negative.

As overall measures of fit we may consider

- Sum of absolute deviations $\sum_{i=1}^{n} |e_i|$
- Sum of squared deviations $\sum_{i=1}^{n} e_i^2$

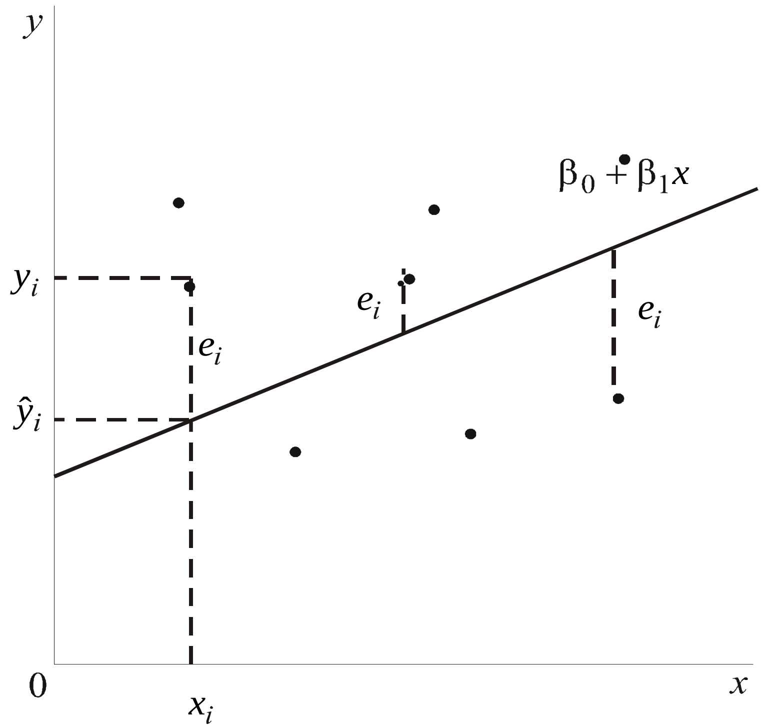
For reasons that will become clear later on, we prefer (for now) the sum of squared deviations.

We obtain β_0, β_1 as the solution to

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} e_i^2$$

Define (I omit subscript 1 on x)

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$



We obtain β_0, β_1 by solving the first-order conditions

$$\frac{\partial S}{\partial \beta_0}(\beta_0, \beta_1) = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial S}{\partial \beta_1}(\beta_0, \beta_1) = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0$$

These first-order conditions are called the *normal equations*.

Solution:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

with $\overline{y}, \overline{x}$ the sample average of y, x.

This method to obtain is called *(Ordinary) Least Squares* (OLS)

Some properties of the OLS solution

Define OLS residuals by

$$e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

and the fitted or predicted value of y by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

From the first normal equation

$$\sum_{i=1}^{n} e_i = \frac{1}{n}\sum_{i=1}^{n} e_i = \bar{e} = 0$$

and from the second

$$\frac{1}{n}\sum_{i=1}^{n} x_i e_i = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(e_i - \bar{e}) = 0$$

Conclusion: sample average of e_i and the covariance of x_i and e_i are 0.

From the solution for $\hat{\beta}_0$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{\hat{y}}$$

Hence the sample average of the fitted values is equal to that of y_i.

Example: Earnings function

OLS solution for the best fitting linear relation between earnings and education is

$$\hat{\beta}_0 = 5.045 \qquad \hat{\beta}_1 = 0.06673$$

Interpretation of slope coefficient: return to one year of additional education is 6.7%.

Fitting a linear function

Consider a linear relation with K variables

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K$$

Terminology

- y dependent variable, variable to be explained, regressand, left-hand side variable, outcome variable.
- x_k is the k-th independent variable, explanatory variable, regressor, covariate, right-hand side variable.

With n observations we have

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + e_i\,, \quad i = 1, \ldots, n$$

Matrix notation

n-vectors

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

$n \times (K+1)$ matrix

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1K} \\ \vdots & & & \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nK} \end{bmatrix}$$

$K + 1$ vector

$$\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_K \end{bmatrix}$$

Hence we can write

$$y = X\beta + e$$

and the sum of squared residuals in matrix notation is

$$\begin{aligned} S(\beta) = &\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_K x_{iK})^2 = \\ = &(y - X\beta)'(y - X\beta) \end{aligned}$$

The OLS solution is to minimize

$$S(\beta) = y'y - 2y'X\beta + \beta' X' X\beta$$

The first-order condition for minimum

$$\frac{\partial S}{\partial \beta}(\beta) = -2X'y + 2X'X\beta = 0$$

Hence, the OLS estimator $\hat{\beta}$ satisfies the normal equations

$$X'X\hat{\beta} = X'y$$

and hence if $X'X$ has full-rank, then

$$\hat{\beta} = (X'X)^{-1}X'y$$

Appendix

The linear relation for the n observations

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + e_i\,, \quad i = 1, \ldots, n$$

are organized as

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \cdots + \beta_K x_{1K} + e_1 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \cdots + \beta_K x_{nK} + e_n \end{aligned}$$

Using the definitions of the $n \times 1$ vectors y, e, the $n \times (K+1)$ matrix X and the $K \times 1$ vector β, we find using the definition of matrix multiplication on e.g. pp. 790-791 in Appendix D of Wooldridge (W) that this is the same as

$$y = X\beta + e$$

The sum of squared residuals

$$S(\beta) = \sum_{i=1}^{n} e_i^2 = e'e$$

with e' the transpose of the vector e defined on p. 791 of W and we again use the definition of matrix multiplication. Because $e = y - X\beta$ we have

$$S(\beta) = (y - X\beta)'(y - X\beta) = (y' - \beta' X')(y - X\beta)$$

where we use the properties of the transpose as given on pp. 791-792 of W. Now we use the properties of matrix multiplication (check the order of all the terms of the extreme right expression)

$$S(\beta) = (y' - \beta' X')(y - X\beta) = (y' - \beta' X')y - (y' - \beta' X')X\beta = y'y - \beta' X' y - y' X\beta + \beta' X' X\beta$$

Now $\beta' X' y$ is an $(1 \times (K+1)) \times ((K+1) \times n) \times (n \times 1)$ so the final result is a 1×1 matrix which is just a number or, as we call it in matrix algebra, a scalar. The transpose of a scalar is equal to that scalar, i.e. the transpose has no effect. Therefore

$$\beta' X' y = (\beta' X' y)' = y' X\beta$$

using the properties of the transpose. With this result we obtain

$$S(\beta) = y'y - 2y' X\beta + \beta' X' X\beta$$

$S(\beta)$ is a function of β. A function of several arguments (here $K+1$ arguments) can be minimized. As in the case of a function of one argument we compute the derivative and set the derivative equal to 0. With a function of several arguments we compute the vector of partial derivatives. The partial derivative with respect to say β_0 is obtained by differentiating $S(\beta)$ with respect to β_0 treating $\beta_1, \ldots, \beta_K$ as constants. In Appendix D.6 of W you find the two results we need. You can verify them by expressing $S(\beta)$ in summation notation. We will just state the results. There are $K+1$ partial derivatives with respect to β. We organize them as a *column* vector, i.e. a $(K+1) \times 1$ matrix. In W there are organized as a row vector. That is mathematically more appropriate but let us stick with organizing them as a column. In that case

$$\frac{\partial 2y'X\beta}{\partial \beta} = 2X'y$$

and

$$\frac{\partial \beta' X' X \beta}{\partial \beta} = 2X'X\beta$$

With these results we obtain the first order condition. As will be discussed later under a certain assumption the square and symmetric (see W) matrix $X'X$ has an inverse, i.e. there is a matrix $(X'X)^{-1}$ such that

$$X'X(X'X)^{-1} = (X'X)^{-1}X'X = I$$

with I the $(K+1) \times (K+1)$ identity matrix (see definition D.5 in W). Note that $X'X(X'X)^{-1} = (X'X)^{-1}X'X$ which is not true in general for matrices. The identity matrix plays the role of the number 1 in scalar algebra. With the inverse we can solve the first order condition for β to obtain the OLS solution.