Costas Meghir

- Econometrics – The analysis of data using specialised Statistical techniques for
  - Testing Economic Hypotheses
  - Measuring objects of interest such as demand and Supply curves
  - Assessing the effectiveness of Business Strategy such as advertising
  - Evaluating the impact of Public Policy (New Deal , Education Maintenance Allowance)
  - Providing expert advice to court cases on issues such as monopoly power
  - Constructing simulation models for assessing the impact of future public policy – such as the introduction of the Working Families Tax Credit

Example 1

- Measuring the returns to education
  - Public Policy towards education is predicated on the assumption that education has economic value. Education is supposed to promote earnings growth and as a result overall economic growth.
  - A simple cost benefit calculation may compare funds advanced for an extra year of education C to the stream of earnings. This requires measuring the **causal** effect of extra education on earnings.

Causality

- We often observe that two variables are correlated.
  - Examples:
    - Parental Income is correlated with child's education.
    - Pupil performance is correlated with the performance of peers.
    - Advertising is correlated with firm cash flow.
    - Health and Income are correlated.
- However this does not establish causal relationships.

- If a variable Y is causally related to X, then changing X will **LEAD** to a change in **Y.**
  - For example: Increasing VAT may cause a reduction of demand.
  - Correlation my not be due to causal relationship:
    - Part or the whole correlation may be induced by both variables depending on some common factor and does not imply causality.
    - For example: Brighter people have more education AND earn more. The question is how much of the increased in earnings is **caused** by the increased education.

– For example successful companies may carry out a lot of advertising – This does not imply that the advertising caused the success; it may well be that success allows them greater flexibility in advertising. This in turn my **cause** more success.

- The course in its more advanced phase will deal with the issue of causality and ways that we have of establishing and measuring causal relationships

The Regression Model

- The basic tool in Econometrics is the Regression Model
- Its simplest form is the two Variable regrression Model

$$Y_i = a + bX_i + u_i$$

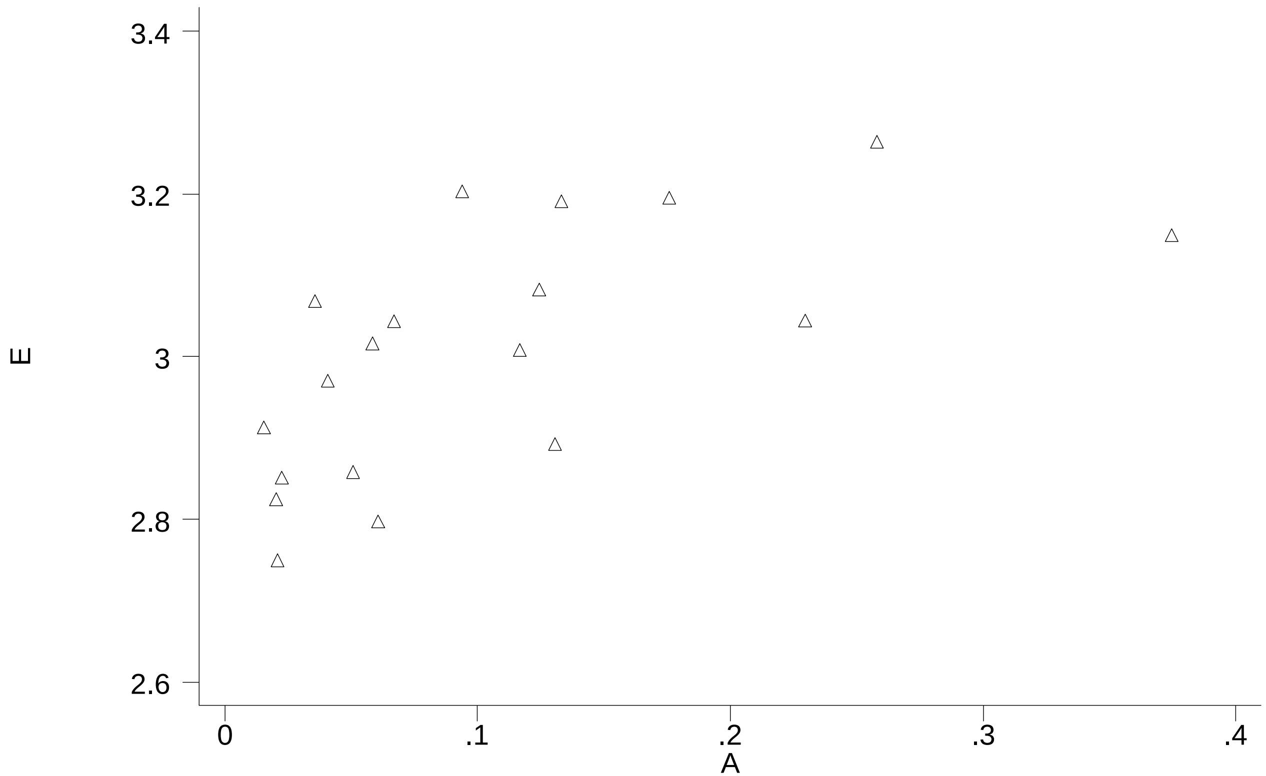
Explanation of Terms

- Y_i The DEPENDENT Variable
- X_i The *EXPLANATORY* variable
- u_i The ERROR term
- $i = (1, \ldots, N)$ The observation indicator. There are *N* observations

- The *Dependent variable Y* is the variable we are modeling
- The *Explanatory Variable X* is the variable of interest whose impact on *Y* we wish to measure
- The error term (*u*) reflects *all other factors* determining the dependent variable

- During most of the first term we will assume that *u* and *X are NOT* correlated
- This assumption will allow us to interpret the coefficient *b* as the effect of *X on Y*
- Note that $b = \frac{\partial Y_i}{\partial X_i}$
- This coefficient will be interpreted as the ceteris paribus *impact of a change in X on Y*
- Aim: To use data to *estimate the coefficients a* and *b*

Scatter Diagram



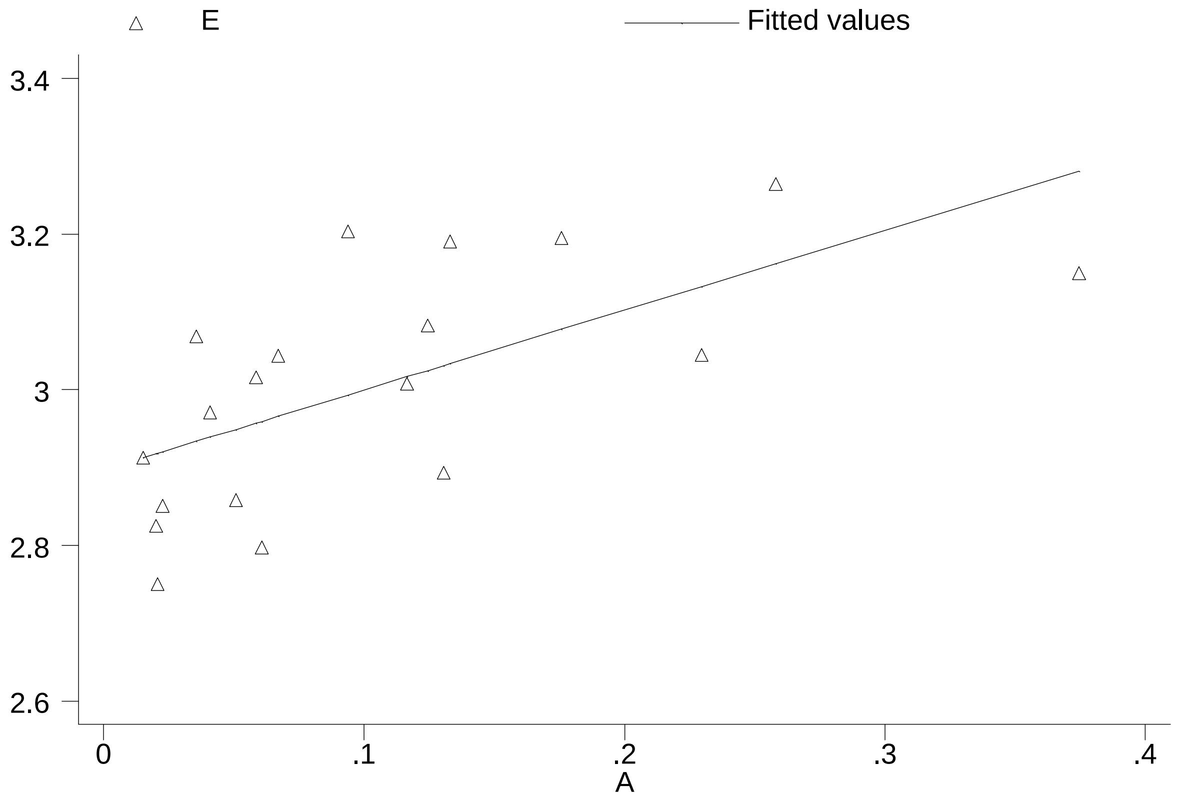
The Impact of Advertising on Sales

- Each point represents an observation on a pair of sales and advertising.
- The Key issue issues are:
  - Estimating the Coefficients of the regression line that fits this data best in the most efficient way possible.
  - Making inferences about the model based on these estimates
  - Using the model

The fitted Line

Intercept 2.9, Estimated Slope (estimated b) = 1.03

Estimated elasticity at the means of the data=0.037



The Fitted Regression Line

- The distance between any point and the fitted line is the estimated residual.
- This summarises the impact of *other factors* on sales.
- As we will see the chosen best line is fitted using the assumption that these other factors are not correlated with advertising.
- Thus the fitted line is constructed using this basic assumption of the approach.

More Advanced models

- In many occasions we will consider more elaborate models where a number of explanatory variables will be included.
- The regression models in this case will take the more general form:

$$Y_i = a + b_1 X_{1i} + b_2 X_{2i} + ... + b_k X_{ki} + u_i$$

- There are *k* explanatory variables and a total of *k+1* coefficients to estimate (including the intercept)
- Each coefficient represents the *ceteris paribus* effect of changing one variable.

Data Sources

- Time Series Data
  - Data on variables observed over time. Typically Macroeconomic measures such as GDP, Inflation, Prices, Exchange Rates, Interest Rates, etc.
  - Used to study and simulate macroeconomic relationships and to test macro hypotheses
- Financial Data
  - Data on share prices, bonds and other financial instruments at frequencies that range from minute to minute up to annual.
  - Used to study the working of financial markets and asset pricing

- Cross Section Survey Data
  - Data on Individuals, households or firms. Examples are data on expenditures, income, hours of work, household composition, assets, investments, employment etc.
  - Used to study household and firm behaviour when variation over time is not required
- Panel Data
  - Data on individual units followed over time
  - Used to study dynamic aspects of household and firm behaviour and to measure the impact of variables that vary predominantly over time.

An Example : Returns to Education and ability

$$\log wage_i = a + b(years\ of\ \ education_i) + u_i$$

Estimate this on a sample of individuals below median ability and then again for a sample of above median ability

. regress lnw ed if high==0

lnwp	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ed	.0511588	.002188	23.38	0.000	.0468702	.0554473
_cons	1.89473	.0058756	322.47	0.000	1.883213	1.906246

. regress lnw ed if high==1

lnwp	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
ed	.0766281	.0014497	52.86	0.000	.0737867	.0794694
_cons	1.967045	.0049728	395.56	0.000	1.957298	1.976792

Schooling level	Proportion High Ability by Schooling Level
Left at 15	0.32
Left at 16	0.46
Lower Voc	0.40
Higher Voc	0.63
Lower Second	0.63
Upper Second	0.87
University	0.96