Econometrics, Lecture 1
Introduction and the bivariate Linear Regression Model, LRM

Ron Smith: R.Smith@bbk.ac.uk
EMS, Birkbeck, University of London

Autumn 2020

Schedule

- Watch Video lecture for Monday, send any questions to Ron
- live collaborate session: Ron Monday 6pm.
- Tutorial Classes live collaborate sessions: Walter, Thursday 15.30-1700 & 19.30-21.00. Same material. starts week 2
- Watch Video lectures for Friday
- live collaborate session: Ron Friday 6pm.
- Do practical exercises using a statistical program on a computer
- Do next weeks tutorial class exercise.
- Do not get left behind.

Resources on Moodle Tiles 1

- WELCOME: General module information.
- YOUR STUDIES: General College Information
- READING LISTS AND MODULE NOTES. There are a detailed set of notes here which covers everything in the Autumn Lectures.
- EXERCISES for TUTORIALS and PRACTICALS
  - You can use any program you like there are instructions how to use some programs.
  - Cross section data from Gapminder. Y_i is life expectancy in country i and X_i is log international dollar per-capita income. Great source of data.
  - Time series on financial data 1870-2016 from Shiller

Resources on Moodle Tiles 2

- USEFUL SUPPLEMENTARY MATERIAL. We put stuff here.
- ASSESSMENT
  - January Exam:
  - Project due May: There is advice on doing projects,
  - June Exam,
- WEEK 1, etc. Tutorial Classes start in week 2.
  - Will have videos, recordings of collaborate classes etc.

Econometrics involves

- asking a question. It may be about the future, appropriate policy etc
- obtaining data, ask FRED or Gapminder:
  - cross-section, Y_i, $i = 1, 2, ..., N$; time-series, Y_t, $t = 1, 2, ..., T$; or panel, Y_{it}.
  - asking what does it measure? how was it constructed?
- using economic theory and background information
  - asking what do you know about the "data generation process"?
- together with statistical methods for estimation and inference (testing)
- and a computer package
- to construct empirical models, to answer the question.

Purpose

- The purpose of the exercise -the question you are trying to answer - is central.
- You would use different models for different purposes such as
  - forecasting,
  - policy analysis,
  - causal analysis,
  - testing hypotheses etc.
- You need to integrate all the elements.
- You need to practice.
- Our starting statistical model is the linear regression model, LRM.

Linear Regression Model, LRM

Explain a dependent variable Y by independent variables, X. A cross-section bivariate LRM is:

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \; i = 1, 2, ..., N.$$

A set of N equations one for each observation, the variables may be non-linear functions, e.g. logs.
We wish to obtain estimates of

- the intercept, $\widehat{\beta}_1$
- the slope $\widehat{\beta}_2$
- the predicted values: $\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i$
- the residuals: $\widehat{u}_i = Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i$.

Notice we distinguish between the true, unknown values, like β_1, and our estimates from the sample $\widehat{\beta}_1$, which are random variables.

First Practical

- Y_i life expectancy in country $LE_i : i,\ N = 189$
- X_i log per capita GDP in international constant dollars: $\ln PCGDP_i$
- Estimates (standard errors)

$$\begin{array}{lcccccc} LE_i = & 26.91 & +4.99 & \ln PCGDP_i & +\widehat{u}_i & R^2 = 0.71 \\ & (2.16) & (0.23) & & & SER = 3.83 \end{array}$$

Efficiency 1

Regulators of monopoly industries, like water, often have to decide whether to allow price increases or whether the firms are inefficient and could reduce costs.

- Data: on output for firm i, Q_i, its employment E_i, and capital stock, K_i, $i = 1, 2, ..., N$.
- Economic Model, Production Function

$$Q = F(K, E)$$

- Need functional form: Cobb-Douglas Production Function,

$$Q_i = AK_i^{\beta_1} E_i^{\beta_2}$$

- But it doesn't hold exactly, firms differ in efficiency, call u_i a measure of efficiency

$$Q_i = AK_i^{\beta_1} E_i^{\beta_2} e^{u_i}$$

Efficiency 2

- Statistical methods LRM, get it linear by taking logs
- Empirical model

$$\begin{aligned} \ln Q_i &= \ln A + \beta_1 \ln K_i + \beta_2 \ln E_i + u_i, \\ q_i &= a + \beta_1 k_i + \beta_2 e_i + u_i \end{aligned}$$

- Lots of ways to test for constant returns to scale $H_0 : \beta_1 + \beta_2 = 1$. One way: make a coefficient which is $(\beta_1 + \beta_2 - 1)$.
- Reparameterise: subtract $\ln E_i$ from both sides

$$q_i - e_i = a + \beta_1 k_i + (\beta_2 - 1)e_i + u_i$$

- add and subtract $\beta_1 e_i$

$$\begin{aligned} q_i - e_i &= a + \beta_1 k_i \{-\beta_1 e_i + \beta_1 e_i\} + (\beta_2 - 1)e_i + u_i \\ q_i - e_i &= a + \beta_1 (k_i - e_i) + (\beta_1 + \beta_2 - 1)e_i + u_i \end{aligned}$$

Bivariate regression

- A time-series equation is of the form:

$$Y_t = \beta_1 + \beta_2 X_t + u_t, \ \ t = 1, 2, ..., T. \tag{1}$$

- We want to obtain estimators of β_j, denoted $\widehat{\beta}_j$, $j = 1, 2.$ in equations like (1).
- Estimators are formulae which tell you how to calculate an estimate from data for a particular sample.
- We will use 3 procedures:
  - (a) method of moments,
  - (b) least squares and
  - (c) maximum likelihood assuming normal errors.
- Here the 3 procedures give the same answer. This is not generally the case.
- To obtain estimators we need to make some assumptions.

Assumptions about errors

- The errors have expected value (population mean) zero

$$E(u_t) = 0, \tag{2}$$

  The intercept will pick up any non-zero mean.
- The errors have constant variance, are homoskedastic.

$$E(u_t^2) = \sigma^2, \tag{3}$$

  If the assumption fails and the variance is not constant the errors are heteroskedastic.
- The errors have no serial correlation or autocorrelation

$$E(u_t u_{t-i}) = 0, i \neq 0, \tag{4}$$

Assumptions about independent variables

- We also require that the X_t vary so we can measure their effect and that they are not related to the u_t.
- The lack of relationship may arise because the explanatory variables in X are either
  - (a) non stochastic
  - (b) exogenous, distributed independently of the errors u_t or
  - (c) pre-determined, uncorrelated with the errors u_t.
- All of these imply that $E(u_t) = 0$ and $E(X_t u_t) = 0$. Note independence is a much stronger assumption than uncorrelated.
- Only require exogeneity for estimation not prediction.

Method of Moments 1

- MoM estimators find $\widehat{\beta}_1$ and $\widehat{\beta}_2$, to make our assumptions: $E(u_t) = 0$ and $E(X_t u_t) = 0$ hold for their sample equivalents.
- The sample equivalent of the errors u_t are the residuals $\widehat{u}_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_t$.
- The sample equivalent of the expected value is the mean
- So

The sample equivalent of $E(u_t) = 0$ is $T^{-1} \sum_t \widehat{u}_t = 0$, and
The sample equivalent of $E(X_t u_t) = 0$ is $T^{-1} \sum_t X_t \widehat{u}_t = 0$.

- Find $\widehat{\beta}_1$ and $\widehat{\beta}_2$, to make
  - $T^{-1} \sum_t \widehat{u}_t = 0$ the mean (sum) of the residuals zero, and
  - $T^{-1} \sum_t X_t \widehat{u}_t = 0$ the residuals are uncorrelated with X_t or x_t.

Method of Moments 2

Now

$$
\begin{aligned}
T^{-1}\sum_t \widehat{u}_t &= T^{-1}\sum_t \left(Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_t\right) \\
T^{-1}\sum_t \widehat{u}_t &= T^{-1}\left(\sum_t Y_t - T\widehat{\beta}_1 - \widehat{\beta}_2 \sum_t X_t\right) \\
&= T^{-1}\sum_t Y_t - \widehat{\beta}_1 - T^{-1}\sum_t \widehat{\beta}_2 X_t \\
T^{-1}\sum_t \widehat{u}_t &= \overline{Y} - \widehat{\beta}_1 - \widehat{\beta}_2 \overline{X}
\end{aligned}
$$

So the moment condition $T^{-1}\sum_t \widehat{u}_t = 0$ implies

$$\overline{Y} - \widehat{\beta}_1 - \widehat{\beta}_2 \overline{X} = 0$$

Rearranging gives

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X}.$$

Method of moments 3

Use

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X}$$

to rewrite (1) in terms of the estimates

$$\begin{aligned}
Y_t &= \widehat{\beta}_1 + \widehat{\beta}_2 X_t + \widehat{u} \\
Y_t &= (\overline{Y} - \widehat{\beta}_2 \overline{X}) + \widehat{\beta}_2 X_t + \widehat{u}_t \\
Y_t - \overline{Y} &= \widehat{\beta}_2 (X_t - \overline{X}) + \widehat{u}_t \\
y_t &= \widehat{\beta}_2 x_t + \widehat{u}_t
\end{aligned}$$

Where we define

$$\begin{aligned}
y_t &= Y_t - \overline{Y} \\
x_t &= X_t - \overline{X}.
\end{aligned}$$

Working with deviations from the mean, gets rid of the intercept and makes the algebra easier.

Method of moments 4

The second moment condition $T^{-1}\sum_t X_t \widehat{u}_t = 0$ is equivalent to $T^{-1}\sum_t x_t \widehat{u}_t = 0$ and

$$\begin{aligned} T^{-1}\sum_t x_t \widehat{u}_t &= T^{-1}\sum_t x_t(y_t - \widehat{\beta}_2 x_t) \\ &= (T^{-1}\sum_t x_t y_t) - \widehat{\beta}_2(T^{-1}\sum_t x_t^2) = 0 \end{aligned}$$

$$(T^{-1}\sum_t x_t y_t) = \widehat{\beta}_2(T^{-1}\sum_t x_t^2)$$

solving for $\widehat{\beta}_2$

$$\begin{aligned} \widehat{\beta}_2 &= \frac{(T^{-1}\sum_t x_t y_t)}{(T^{-1}\sum_t x_t^2)} \\ &= \frac{\sum_t (X_t - \overline{X})(Y_t - \overline{Y})/T}{\sum_t (X_t - \overline{X})^2/T} = \frac{Cov(X_t Y_t)}{Var(X_t)}, \end{aligned}$$

as long as $Var(X_t) \neq 0$. So we now have estimators for $\widehat{\beta}_2$ and $\widehat{\beta}_1$.

Least Squares

The least squares estimator minimises $S = \sum_t \widehat{u}_t^2$

$$
\begin{aligned}
S &= \sum_t (y_t - \widehat{\beta}_2 x_t)^2 \\
&= \sum_t y_t^2 + \widehat{\beta}_2^2 \sum_t x_t^2 - 2\widehat{\beta}_2 \sum_t x_t y_t \\
\frac{\partial S}{\partial \widehat{\beta}} &= 2\widehat{\beta}_2 \sum_t x_t^2 - 2 \sum_t x_t y_t = 0 \\
\widehat{\beta}_2 &= \frac{\left(\sum_t x_t y_t\right)}{\left(\sum_t x_t^2\right)}.
\end{aligned}
$$

The same as before. Note the second derivative is $2\sum_t x_t^2 > 0$, so it is a minimum.

Properties of the estimator: unbiased 1

- For simplicity, we will write β for β_2 below. If the expected value of the random variable $\widehat{\beta}$ (it is different in every sample) equals its true value
$$E(\widehat{\beta}) = \beta$$
then $\widehat{\beta}$ is said to be unbiased.
- Since
$$\widehat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t(\beta x_t + u_t)}{\sum x_t^2} = \beta + \frac{\sum x_t u_t}{\sum x_t^2}$$
- Then
$$\widehat{\beta} - \beta = \frac{\sum x_t u_t}{\sum x_t^2} \tag{5}$$

Properties of the estimator: unbiased 2

- We want to show that

$$E\left(\widehat{\beta}-\beta\right)=E\left(\frac{\sum x_t u_t}{\sum x_t^2}\right)=0$$

- If x_t and u_t are independent, we can write $E(AB)=E(A)E(B)$:

$$E\left\{\frac{\sum x_t u_t}{\sum x_t^2}\right\}=E\left\{\frac{\sum x_t}{\sum x_t^2}\right\}E(u_t)$$

  and $E(u_t)=0$.

- So $E\left(\widehat{\beta}-\beta\right)=0$ and $E(\widehat{\beta})=\beta$, and it is unbiased.

Properties of the estimator: variance 1

Since $\widehat{\beta}$ is unbiased we can replace $E(\widehat{\beta})$ by β, and using (5)

$$V(\widehat{\beta}) = E(\widehat{\beta} - E(\beta))^2 = E(\widehat{\beta} - \beta)^2 = E(\frac{\sum x_t u_t}{\sum x_t^2})^2 \qquad (6)$$

treating x_t as fixed

$$E(\frac{\sum x_t u_t}{\sum x_t^2})^2 = \frac{1}{\left(\sum x_t^2\right)^2} E\left(\sum x_t u_t\right)^2$$

Expand $E\left(\sum x_t u_t\right)^2$ as

$$\begin{aligned}
&E(x_1 u_1 + x_2 u_2 ... + x_T u_T)(x_1 u_1 + x_2 u_2 ... + x_T u_T) \\
&E(x_1^2 u_1^2 + x_2^2 u_2^2 + ... + x_T^2 u_T^2 + 2x_1 u_1 x_2 u_2 + ...) \\
&x_1^2 \sigma^2 + x_2^2 \sigma^2 + ... + x_T^2 \sigma^2 + 0 + ... \\
&\sigma^2 \sum x_t^2
\end{aligned}$$

Properties of the estimator: variance 2

- Using $E(u_t^2) = \sigma^2$ and $E(u_t u_{t-i}) = 0$.
- So substituting for $E\left(\sum x_t u_t\right)^2$

$$V(\widehat{\beta}) = \frac{1}{\left(\sum x_t^2\right)^2}\left(\sigma^2 \sum x_t^2\right) \tag{7}$$
$$= \frac{\sigma^2}{\sum x_t^2} = \frac{\sigma^2}{T\sigma_X^2}$$

Where $\sigma_X^2 = \sum_t (X_t - \overline{X})^2 / T$

- Note that $\sum x_t^2$ rises with T, the sample size, so $V(\widehat{\beta})$ and the standard error of $\widehat{\beta}$ fall with T, as in the case of the standard error of a mean.

Residuals

Returning to the original notation, the residuals are

$$\widehat{u}_t = y_t - \widehat{\beta} x_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_t.$$

We prove later that the unbiased estimator of σ^2 is

$$s^2 = \sum \widehat{u}_t^2 / (T - 2)$$

because we estimate two parameters $\widehat{\beta}_1$ and $\widehat{\beta}_2$. Our estimator for the standard error of $\widehat{\beta}_2$ is the square root of $V(\widehat{\beta}_2)$ with σ replaced by s :

$$se(\widehat{\beta}_2) = s / \sqrt{\sum x_t^2}.$$

Multiple regression 1

with k explanatory variables takes the form

$$Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} + u_t$$

where $X_{1t} = 1$ all t. Doing the algebra for multiple regression in scalar form is very messy. It is much more convenient to use vectors or matrices.
This can be written in vector form as:

$$Y_t = \beta' X_t + u_t$$

where β and X_t are $k \times 1$ vectors.
β' is the transpose of β a $1 \times k$ vector. So $\beta' X_t$ is $(1 \times k) \times (k \times 1)$ conformable and a scalar.

Multiple regression 1

In matrix form

$$Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} + u_t$$

is

$$\underset{T \times 1}{y} = \underset{T \times k}{X} \underset{k \times 1}{\beta} + \underset{T \times 1}{u}$$

where y and u are $T \times 1$ vectors and X is a $T \times k$ matrix.
WRITE THE DIMENSIONS BELOW
For the bivariate regression, this is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ .. \\ Y_T \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ .. & .. \\ 1 & X_T \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ .. \\ u_T \end{bmatrix}.$$

What next?

- This time
  - We set out the LRM and some assumptions,
  - derived estimators of β by method of moments and least squares,
  - showed the estimator is unbiased,
  - derived its variance and standard errors
- Next time we will do the same things in matrix algebra.
- In the notes there is some revision material at the end of section 1, to remind you about taking derivatives with matrices. Check you are familiar with that.
- There is also some more on the bivariate case to show how the scalar and matrix cases match.