Lecture 3

What do we know now?

The world is not a straight line, but we may be able to approximate economic relationships by a straight line.

If so then can use the idea of Ordinary Least Squares (OLS) which gives the best straight line (the best fit to the data) by "minimising the sum of squared residuals"

\[ \sum_{i=1}^{N} u_i^2 \]

If we do this then the equations that give the OLS estimate of the intercept and slope of the straight line are

\[ \hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \]

\[ \hat{b}_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \]
Today

Go over how to interpret the meaning of an OLS estimate

Look at some algebra that gives some more intuition about what OLS is doing

Come up with a summary statistic that tells us how closely the OLS straight line captures the real world variation

Find out why OLS has such a good reputation as an estimation technique
So

\[ \hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} \quad \hat{b}_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \]

are how the computer determines the size of the intercept and the slope respectively in an OLS regression.

The OLS equations give a nice, clear intuitive meaning about the influence of the variable X on the size of the slope, since it shows that:

i) the greater the covariance between X and Y, the larger the (absolute value of) the slope

ii) the smaller the variance of X, the larger the (absolute value of) the slope
It is equally important to be able to interpret the effect of an estimated regression coefficient.

Given OLS essentially passes a straight line through the data, then given

\[
\frac{dY}{dX} = b_1
\]

So the OLS estimate of the slope will give an estimate of the unit change in the dependent variable \(y\) following a unit change in the level of the explanatory variable.

(so you need to be aware of the units of measurement of your variables in order to be able to interpret what the OLS coefficient is telling you.
Remember every time you estimate an OLS regression there will be a predicted value (forecast)

\[ \hat{y} = b_0 - b_1 X \]

and a residual for each observation

\[ \hat{u}_i = Y_i - \hat{Y}_i \]
Should get used to summarising and plotting the predicted and residual values
Because hardly any the values lie exactly on the predicted straight line, there are both under and over predictions.

This is reflected in the graph of residuals which are scattered above and below zero.

(Note **both** the the forecast values and the residual values are measured in terms of marks gained)
Let’s play darts

No. Let’s do some more algebra
PROPERTIES OF OLS

Using the fact that for any individual observation, \( i \), the ols residual is the difference between the actual and predicted value

\[
\hat{u}_i = Y_i - \hat{Y}_i
\]

Sub. in

\[
\hat{Y}_i = b_0 - b_1 X_i
\]

So that

\[
u_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i
\]

Summing over all \( N \) observations in the data set

\[
\sum \hat{u}_i = \sum Y_i - b_0 - b_1 \sum X_i
\]

and dividing by \( N \)

\[
\frac{1}{N} \sum \hat{u}_i = \frac{1}{N} \sum Y_i - b_0 - b_1 \frac{1}{N} \sum X_i
\]
Since the sum of any series divided by the sample size gives the mean, can write

\[ \frac{1}{N} \sum u_i = \frac{1}{N} \sum Y_i - b_0 - b_1 \frac{1}{N} \sum X_i \]

\[ \bar{u} = \bar{Y} - b_0 - b_1 \bar{X} \]

and since

\[ b_0 = \bar{Y} - b_1 \bar{X} \]

\[ \bar{u} = \bar{Y} - (\bar{Y} - b_1 \bar{X}) - b_1 \bar{X} \]

\[ \bar{u} = \bar{Y} - \bar{Y} + b_1 \bar{X} - b_1 \bar{X} = 0 \]

So the mean value of the OLS residuals is zero

(as any residual should be, since random and unpredictable by definition)
The 2\textsuperscript{nd} useful property of OLS is that

\[ \bar{\hat{Y}} = \bar{Y} \]

the mean of the OLS predicted values equals the mean of the actual values in the data

(so OLS predicts \textit{average} behaviour in the data set – another useful property)

This also means that the OLS regression line passes through the mean of the dependent variable
Proof:
\[ u_i = Y_i - \hat{Y}_i \]

summing
\[ \sum u_i = \sum Y_i - \sum \hat{Y}_i \]

Dividing by N
\[ \frac{1}{n} \sum u_i = \frac{1}{n} \sum Y_i - \frac{1}{n} \sum \hat{Y}_i \]

We know from above that
\[ u = \bar{Y} - \bar{\hat{Y}} \]

so
\[ \bar{\hat{Y}} = \bar{Y} \]
The 3\textsuperscript{rd} useful result is that

\[ \hat{\text{Cov}}(\hat{Y}, u) = 0 \]

the covariance between the fitted values of $Y$ and the residuals must be zero

Proof: See Problem Set 1
Now we know how to summarise the relationships in the data using the OLS method, we next need a summary measure of “how well” the estimated OLS line fits the data.

Think of the dispersion of all possible y values (the variation in Y) being represented by a circle.

And similarly the dispersion in the range of x values.
The more the circles overlap the more the variation in the X data explains the variation in Y

Little overlap in values so X not explain much of variation in Y

Large overlap in values so X variable explains much of variation in Y
To derive a statistical measure which does much the same thing remember that
\[ \hat{u}_i = Y_i - \hat{Y}_i \implies Y_i = \hat{Y}_i + \hat{u}_i \quad (1) \]

Using the rules on covariances (see problem set 0) we know that
\[
\text{Var}(Y) = \text{Var}(\hat{Y} + u) = \text{Var}(\hat{Y}) + \text{Var}(u) + 2\text{Cov}(\hat{Y}, u) \\
= \text{Var}(\hat{Y}) + \text{Var}(u)
\]

So the variation in the variable of interest, \( \text{var}(Y) \), is explained by either the variation in the variables included in the OLS model,
\[ \text{Var}(\hat{Y}) \]

or by variation in the residual \( \text{Var}(\hat{u}) \).
So we use the ratio

$$R^2 = \frac{\text{var}(Y)}{\text{var}(Y)}$$

As a measure of how well the model fits the data. ($R^2$ is also known as the coefficient of determination)

So $R^2$ measures the % of variation in the dependent variable explained by the model.

If the model explains all the variation in $y$ then the ratio equals 1

If the model explains none of the variation then the ratio $= 0$

So the closer the ratio is to one the better the fit.
It is more common however to use one further algebraic adjustment. Given (1) says that

\[ \hat{u}_i = Y_i - \hat{Y}_i \implies Y_i = \hat{Y}_i + \hat{u}_i \]

It follows that

\[ \text{Var}(Y) = \text{Var}(\hat{Y}) + \text{Var}(u) \]

Can write this as

\[
\frac{1}{n} \sum (Y - \bar{Y})^2 = \frac{1}{n} \sum (\hat{Y} - \bar{Y})^2 + \frac{1}{n} \sum (\hat{u} - \bar{u})^2
\]

The \(1/n\) is common to both sides, so can cancel out and using the results that

\[ \hat{\bar{Y}} = \bar{Y} \quad \hat{\bar{u}} = 0 \]

Then we have

\[
\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum \hat{u}^2
\]
\[ \sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum u \]

The left side of the equation is the sum of the squared deviations of \( Y \) about its sample mean.

This is called the **Total Sum of Squares**.

The right hand side consists of the sum of squared deviations of the predictions around the sample mean (the **Explained Sum of Squares**)

and the **Residual Sum of Squares**

\[ TSS = ESS + RSS \]

From this can have an alternative definition of the goodness of fit

\[ R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \]
Can see from above that it must hold that
\[ 0 \leq R^2 \leq 1 \]

when \( ESS = 0 \), then \( R^2 = 0 \)
(and model explains none of the variation in the dependent variable)

when \( ESS = TSS \), then \( R^2 = 1 \)
(and model explains all of the variation in the dependent variable)

In general the \( R^2 \) lies between these two extremes.

You will find that
for cross-section data (ie samples of individuals, firms etc) the \( R^2 \) are typically in the region of 0.2

for time-series data (ie samples of aggregate (whole economy) data measured at different points in time) the \( R^2 \) are typically in the region of 0.9
Look at the $R^2$ in the two regressions. The first uses time series data, the second uses cross-section data.
GOODNESS OF FIT

So the $R^2$ measures the proportion of variance in the dependent variable explained by the model.

Another useful interpretation of the $R^2$ is that it equals the square of the correlation coefficient between the actual and predicted values of $Y$.

Proof: We know the formula for the correlation coefficient:

$$ r_{Y,\hat{Y}} = \frac{\text{Cov}(Y, \hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} $$

Sub. In for

$$ Y = \hat{Y} + u $$

(actual value = predicted value plus residual)

$$ r_{Y,\hat{Y}} = \frac{\text{Cov}([\hat{Y} + u], \hat{Y})}{\sqrt{\text{Var}(Y) \cdot \text{Var}(\hat{Y})}} $$
GOODNESS OF FIT

Expand the covariance terms

\[
\frac{\hat{\text{Cov}}(\hat{Y} + u, \hat{Y})}{\sqrt{\text{Var}(Y) \, \text{Var}(\hat{Y})}} = \frac{\hat{\text{Cov}}(\hat{Y}, \hat{Y}) + \hat{\text{Cov}}(u, \hat{Y})}{\sqrt{\text{Var}(Y) \, \text{Var}(\hat{Y})}}
\]

\[
= \frac{\text{Var}(\hat{Y})}{\sqrt{\text{Var}(Y) \, \text{Var}(\hat{Y})}} \quad \text{(since already proved \( \hat{\text{Cov}}(\hat{Y}, u) = 0 \))}
\]

And can always write any variance term as square root of the product

\[
= \frac{\sqrt{\text{Var}(\hat{Y}) \, \text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y) \, \text{Var}(\hat{Y})}}
\]
GOODNESS OF FIT

Cancelling terms

\[
\frac{\sqrt{\text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y)}} \cdot \frac{\sqrt{\text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y)}} = \frac{\sqrt{\text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y)}}
\]

so

\[r_{xy} = \sqrt{R^2}\]

Thus the correlation coefficient is the square root of $R^2$.

Eg $R^2 = 0.25$ implies correlation coefficient between $Y$ variable & $X$ variable (or between $Y$ and predicted values) = $\sqrt{0.25} = 0.5$
So while we would like the $R^2$ to be as high as possible you can only compare $R^2$ in models with the SAME dependent ($Y$) variable
The 3rd useful result is that

\[ \text{Cov}(\hat{Y}, \hat{u}) = 0 \]

\[ \text{Cov}(\hat{Y}, e) = \text{Cov}([b_1 + b_2 X], e) = \text{Cov}(b_1, e) + \text{Cov}(b_2 X, e) \]
\[ = 0 + b_2 \text{Cov}(X, e) = b_2 \text{Cov}(X, [Y - b_1 - b_2 X]) \]
\[ = b_2 \left[ \text{Cov}(X, Y) - \text{Cov}(X, b_1) - \text{Cov}(X, b_2 X) \right] \]
\[ = b_2 \left[ \text{Cov}(X, Y) - b_2 \text{Cov}(X, X) \right] \]
\[ = b_2 \left[ \text{Cov}(X, Y) - \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \text{Var}(X) \right] = 0 \]

the covariance between the fitted values of \( Y \) and the residuals must be zero.