

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”

$$\sum_{i=1}^{N^2} u_i$$

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

Remember It is **very** important to be able to interpret the effect of any estimated regression coefficient

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

$$\hat{y} = b_0 - b_1 X$$

$$\frac{d\hat{y}}{dX} = b_1$$

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

$$d\hat{y} = b_1 dX$$

(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)

StataSE 10.1 C:\pqn2\Lecture 1\marks_10.dta

File Edit Data Graphics Statistics User Window Help

Review Command

```

1 use "C:\pqn2\Lecture 1\marks_10.dta", clear
2 reg mark seminars
3 predict mhat
4 predict reshat, resid

```

Statistics/Data Analysis
Special Edition

Copyright 1984-2009
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (Fax)

Single-user Stata for windows perpetual license:
Serial numbers: 91920512665
Licensed to: Jonathan Wadsworth
Economics, Royal Holloway

Notes:
1. /vm# option or -set memory- 10.00 MB allocated to data
2. /v# option or -set maxvar- 5000 maximum variables

```

. use "C:\pqn2\Lecture 1\marks_10.dta", clear
. reg mark seminars

```

Source	SS	df	MS	Number of obs =
Model	10687.1183	1	10687.1183	176
Residual	36124.518	174	207.612173	F(1, 174) = 51.48
Total	46811.6364	175	267.495065	Prob > F = 0.0000

	mark	Coeff.	Std. Err.	t	P> t	[95% Conf. Interval]
seminars	2.187335	.3048676	7.17	0.000	1.58562	2.789049
_cons	23.5833	4.43775	5.76	0.000	16.82455	34.34205

```

. predict mhat
(option xb assumed; fitted values)
. predict reshat, resid

```

Command

C:\Data

StataSE 10.1 C:\pqn2\Lecture 1\marks_10.dta

File Edit Data Graphics Statistics User Window Help

Review Command

```

1 use "C:\pqn2\Lecture 1\marks_10.dta", clear
2 reg mark seminars
3 predict mhat
4 predict reshat, resid
5 su
6 .

```

variable	Obs	Mean	Std. Dev.	Min	Max
student	176	89.5	50.95096	2	177
seminars	176	14.11364	1.572696	1	18
mark	176	56.45455	16.35528	6	86
mhat	176	56.45455	7.814682	27.77061	64.95532
reshat	176	-5.42e-08	14.36753	-40.58065	40.73069

```

. su

```

	student	seminars	mark	mhat	reshat
1.	2	15	73	58.39332	14.60668
2.	3	17	59	62.76799	-3.767989
3.	4	7	42	40.89464	1.105358
4.	5	12	30	51.83131	-21.83132
5.	6	12	79	51.83131	27.16868
6.	7	16	47	60.58065	-13.58065
7.	8	9	32	45.26931	-13.26931
8.	9	17	79	62.76799	16.23201
9.	10	16	75	60.58065	14.41935
10.	11	16	68	60.58065	7.419346
11.	12	17	76	62.76799	13.23201
12.	13	41	54	54.01865	-13.01865
13.	14	18	53	64.95532	-11.95532
14.	15	14	61	56.20599	4.794015
15.	16	16	63	60.58065	2.419346
16.	17	13	58	54.01865	3.98135
17.	18	15	70	58.39332	11.60668
18.	19	17	82	62.76799	19.23201
19.	20	8	19	43.08198	-24.08198
20.	21	18	72	64.95532	7.044677
21.	22	9	43	45.26931	-2.269311

Command

C:\Data

Let's do some 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 $\hat{Y}_i = \hat{b}_0 - \hat{b}_1 X_i$

So that $\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{b}_0 - \hat{b}_1 X_i$

Summing over all N observations in the data set

$$\sum \hat{u}_i = \sum Y_i - \hat{b}_0 - \hat{b}_1 \sum X_i$$

and dividing by N

$$\frac{1}{N} \sum \hat{u}_i = \frac{1}{N} \sum Y_i - \hat{b}_0 - \hat{b}_1 \frac{1}{N} \sum X_i$$

Since the sum of any series divided by the sample size gives the mean, can write

$$\frac{1}{N} \sum \hat{u}_i = \frac{1}{N} \sum Y_i - \hat{b}_0 - \hat{b}_1 \frac{1}{N} \sum X_i$$

$$\bar{\hat{u}} = \bar{Y} - \hat{b}_0 - \hat{b}_1 \bar{X}$$

and since $\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$

$$\bar{\hat{u}} = \bar{Y} - (\bar{Y} - \hat{b}_1 \bar{X}) - \hat{b}_1 \bar{X}$$

$$\bar{\hat{u}} = \bar{Y} - \bar{Y} + \hat{b}_1 \bar{X} - \hat{b}_1 \bar{X} = 0$$

So the mean value of the OLS residuals is zero

(as any residual should be, since random and unpredictable by definition)

The 2nd useful property of OLS is that

$$\overline{\hat{Y}} = \bar{Y}$$

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

(so OLS predicts *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:

$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{b}_0 - \hat{b}_1 X_i$$

summing

$$\sum \hat{u}_i = \sum Y_i - \sum \hat{Y}_i$$

Dividing by N

$$\frac{1}{n} \sum \hat{u}_i = \frac{1}{n} \sum Y_i - \frac{1}{n} \sum \hat{Y}_i$$

We know from above that

$$\bar{\hat{u}} = \bar{Y} - \bar{\hat{Y}}$$

$$\bar{\hat{u}} = 0$$

so
$$\bar{\hat{Y}} = \bar{Y}$$

The 3rd useful result is that

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

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

Proof: See Problem Set 1

The 3rd useful result is that

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

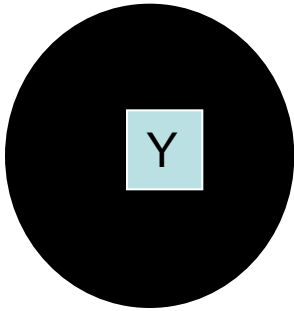
$$\begin{aligned}\text{Cov}(\hat{Y}, e) &= \text{Cov}([b_1 + b_2 X], e) = \text{Cov}(b_1, e) + \text{Cov}(b_2 X, e) \\ &= 0 + b_2 \text{Cov}(X, e) = b_2 \text{Cov}(X, [Y - b_1 - b_2 X]) \\ &= b_2 [\text{Cov}(X, Y) - \text{Cov}(X, b_1) - \text{Cov}(X, b_2 X)] \\ &= b_2 [\text{Cov}(X, Y) - b_2 \text{Cov}(X,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

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

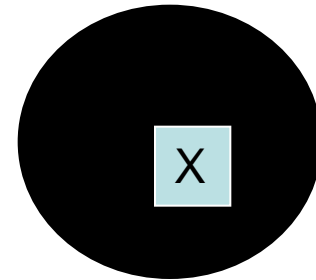
GOODNESS OF FIT

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

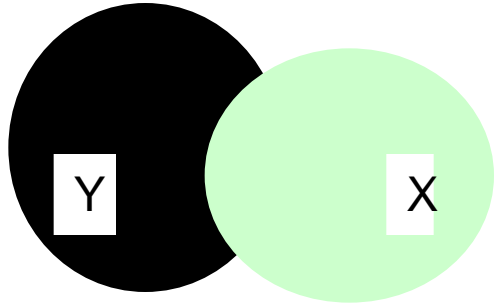
Think of the dispersion of all possible y values (the variation in Y) being represented by a circle



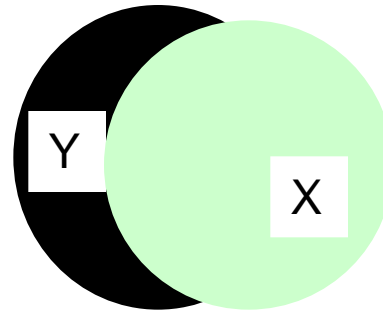
And similarly the dispersion in the range of x values



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

$$\hat{u}_i = Y_i - \hat{Y}_i \quad \Rightarrow \quad Y_i = \hat{Y}_i + \hat{u}_i \quad (1)$$

Using the rules on covariances (see problem set 0) we know that

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(\hat{Y} + \hat{u}) = \text{Var}(\hat{Y}) + \text{Var}(\hat{u}) + 2\text{Cov}(\hat{Y}, \hat{u}) \\ &= \text{Var}(\hat{Y}) + \text{Var}(\hat{u}) \end{aligned}$$

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 $\text{Var}(\hat{u})$

\hat{Y}

So we use the ratio

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

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

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

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

$$\bar{\hat{Y}} = \bar{Y} \qquad \bar{\hat{u}} = \mathbf{0}$$

Then we have $\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum \hat{u}^2$

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

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

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

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

Sub. In for
$$Y = \hat{Y} + u$$

(actual value = predicted value plus residual)

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

GOODNESS OF FIT

Expand the covariance terms

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

(since already proved $\text{Cov}(\hat{Y}, \hat{u}) = 0$)

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

$$\frac{\sqrt{\text{Var}(\hat{Y}) \text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y) \text{Var}(\hat{Y})}} = \frac{\sqrt{\text{Var}(\hat{Y})}}{\sqrt{\text{Var}(Y)}}$$

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