

- 1 Introduction to the module
- 2 Aims and Objectives
- 3 Assessment
- 4 Revision from Term 1
- 5 When does OLS not work?

Why Econometrics?

4 Monitor discussion groups

- # Reading

Look at the aims and objectives!

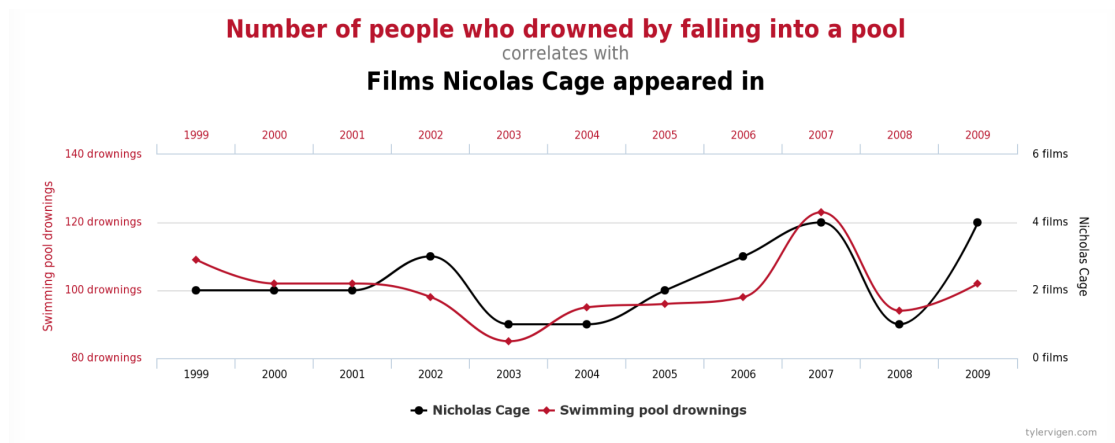
How do you think you will receive feedback?

- Topic 7): Stationarity and Cointegration
- Topic 8): Forecasting models
- Topic 9): Revision

Your questions please....?

What are we trying to do?

We want to find out if there is a relationship between two variables of interest



<http://tylervigen.com/spurious-correlations>

What is econometrics?

"The application of statistical methods to the analysis of economic phenomena" Tintner, 1953

A way of quantifying the relationship of various variables of interest

$$Drownings = \alpha + \beta_1 \text{Nicolas Cage films} + \varepsilon$$

Interesting in estimating the values of the two parameters α and β_1

- Can never know their true values so always an estimate!

Let's look at what some of these mean...

1: Computational Cost

As economists always making cost benefit decisions...

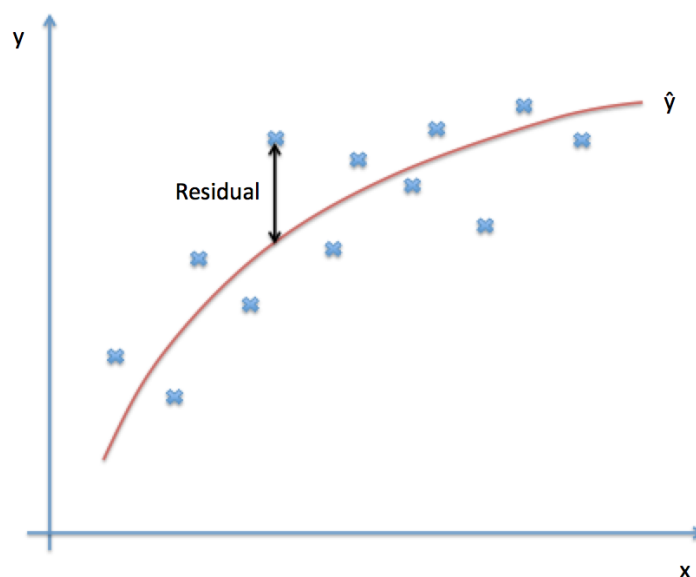
Computational cost and ease is something that needs to be taken into account when selecting an estimator

Not as much of a problem as it used to be

- Will not be a problem for the methods on this course

Can still be troublesome for certain methods

2: Least Squares



Estimating values of the dependent variable (y), based on the independent variables in the dataset (x)

These estimated values ($\hat{y}$) can be subtracted from the actual values of y to produce **residuals** ($y - \hat{y}$)

3: Unbiasedness

We use an estimator to calculate an estimate of β for our sample

Do not know the true population β and if we take different samples our estimate of β will change

What would happen if we keep resampling and trying to estimate the value of β ?

Will form a distribution of these β estimates

- This distributed is called a sampling distribution

Sampling Distribution

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60

$$\text{Mean} = \frac{15}{5} = 3$$

Sampling Distribution

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60

Mean= 58

Sampling Distribution

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60

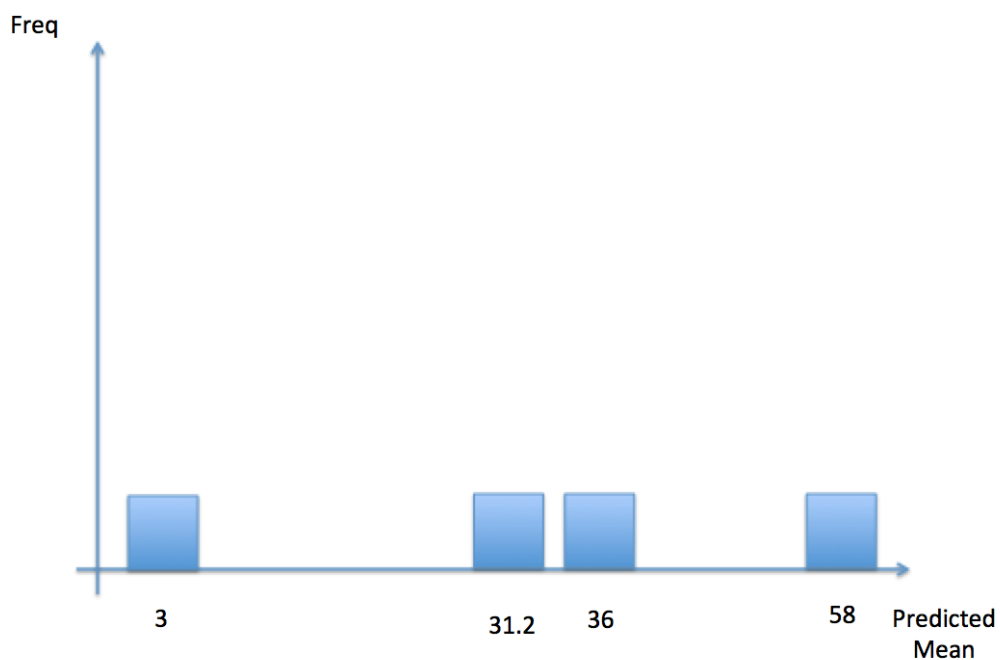
Mean= 36

Sampling Distribution

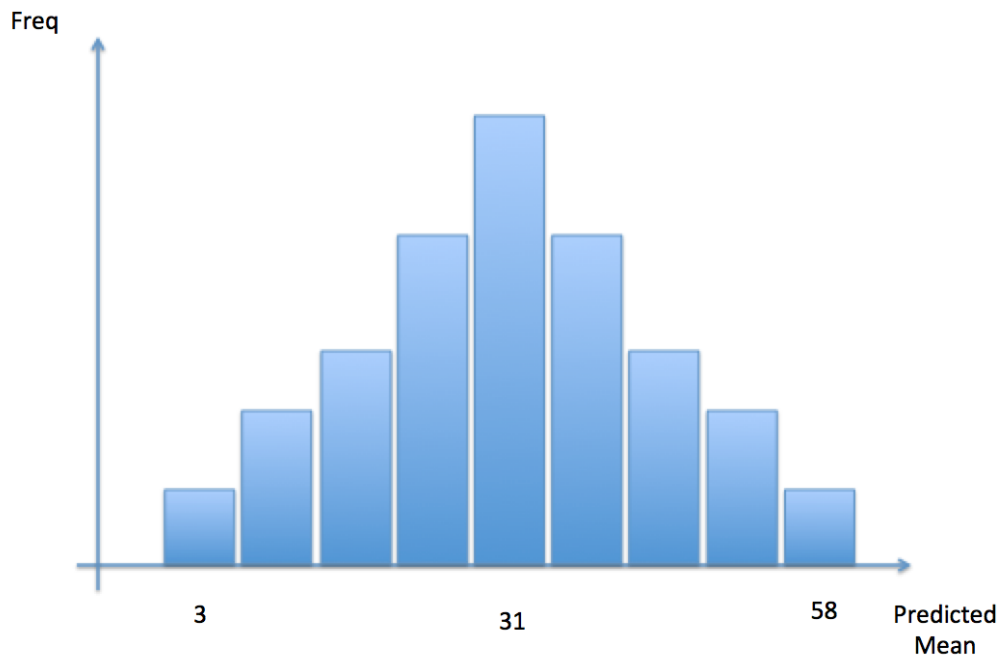
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60

Mean= 31.2

Sampling Distribution



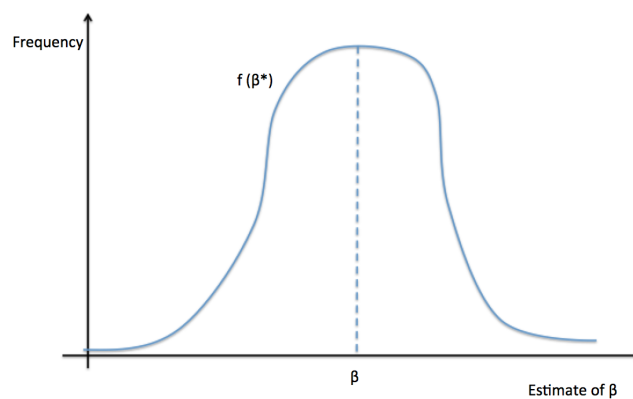
Sampling Distribution



Sampling Distribution

This shows the sampling distribution of the estimator $f(\beta)$

What is the frequency of this value of β , when the sample is re-estimated 2,000 times



Sampling Distribution

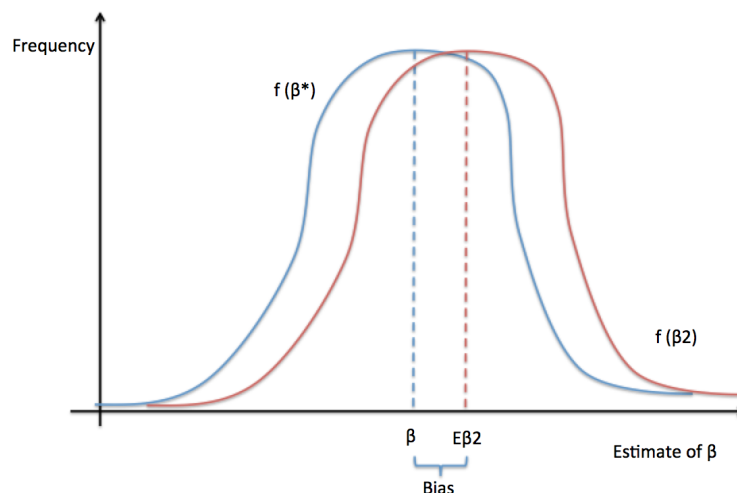
These are key in economics - 'good' properties of sampling distributions are important!

Sampling Distribution

What properties do we want this sampling distribution to have?

- We want it to be unbiased!

Which means the mean of the sampling distribution for an estimator lies over the true value of β



3: Unbiasedness

The unbiasedness property does not mean that our estimate of β will be the true value of β

Again, it does not guarantee that $\hat{\beta} = \beta$

What does it mean?

Unbiased

Unbiased property says that with repeat sampling an infinite number of times we would get the **true value of β on average!**

Best Unbiased

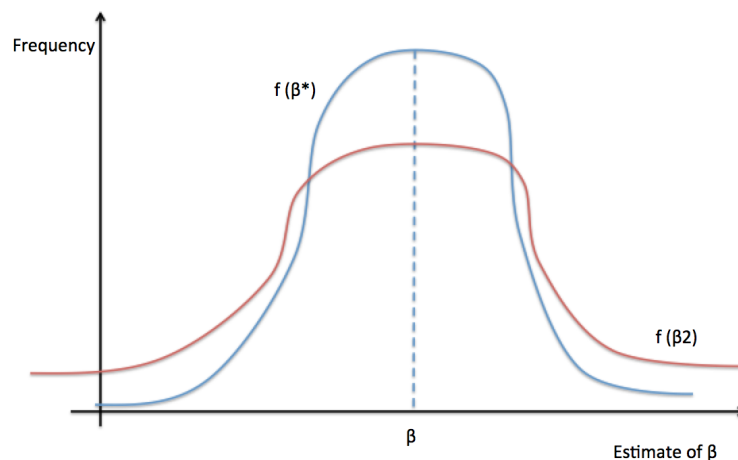
Sometimes it is impossible to find an unbiased estimator

But when you can find one... *Can normally find more than one!*

In this case try and find the **best unbiased**

- i.e. *the one with the smallest variance*

Will be the most efficient estimator out of all the unbiased ones



Best Unbiased

Why is a small variance a desirable property?

Both would have a mean of β

- One would range more widely
 - This is less desirable

What is the Best Unbiased Linear Estimator (BLUE)?

OLS

OLS is best (small variance) unbiased (sampling distribution mean is β) linear estimator

4: Mean Square Error (MSE)

Last properties we should talk about...

- is Mean Squared Error (MSE)

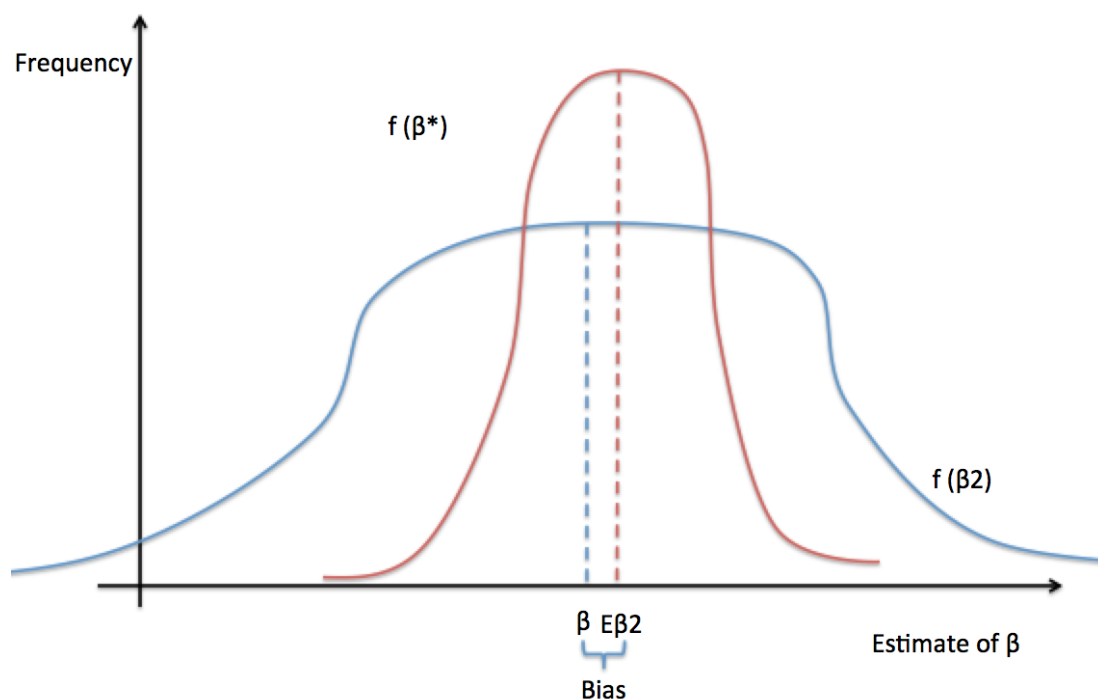
We have seen that best unbiased is a nice condition to have

Can sometimes play **too important a role**

Restricts us to just estimators that are unbiased

- *But a slightly biased estimator may have a much small variance*

4: Mean Square Error



Here $f(\beta)$ is not centered over the true value of β

But its variance is much smaller than $f(\beta_2)$ which is

4: Mean Square Error (MSE)

Most researches would prefer $f(\beta)$ over $f(\beta_2)$

- **Would prefer the biased estimator**

So best unbiased is not always the model that we want!

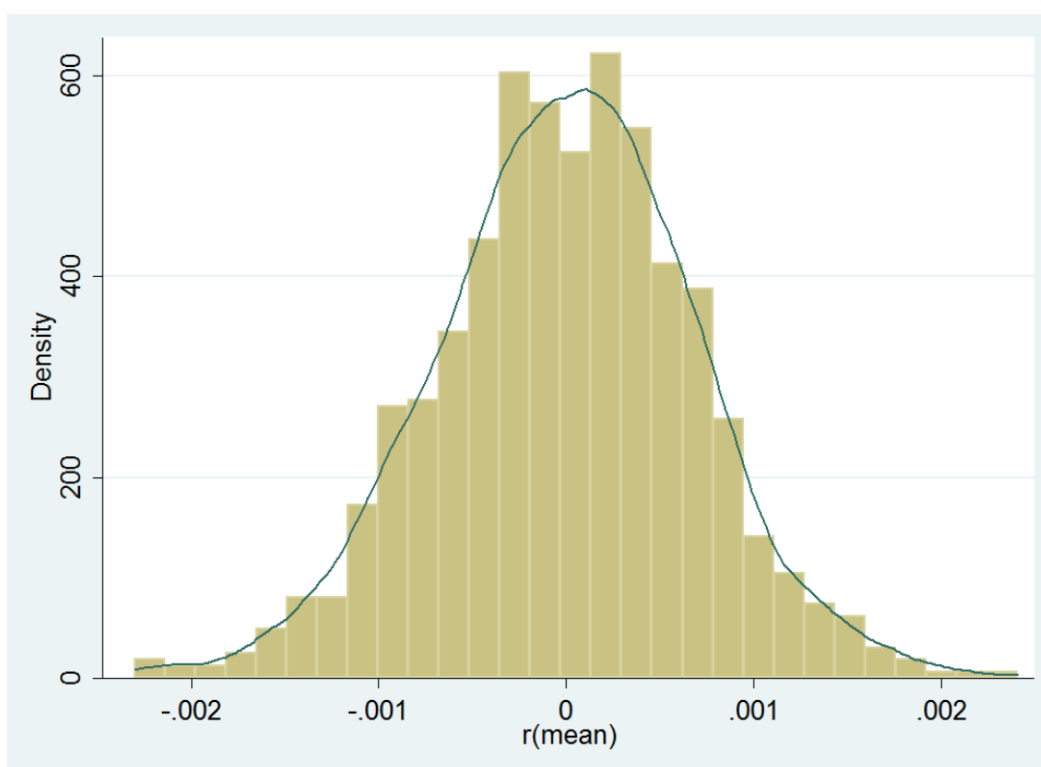
Trade off between low variance and low bias is formalised by a weighted average of the bias and the variance

MSE

When an estimator is chosen so as to minimise an equally weighted average of the variance and the square of the bias, the estimator is said to be chosen the MSE criterion

Not normally adopted unless the best unbiased produces estimates with large variances!

Sampling Distributions and Estimators



Recap

Most of what econometricians do is either finding "good" estimates for unknown numbers, or testing hypotheses

The sampling distribution concept creates the following underlying logic for these two activities

Underlying Idea 1

Using a formula $f(\beta)$ to produce an estimate of β can be conceptualized as the econometrician shutting his or her eyes and obtaining an estimate of β by reaching blindly into the sampling distribution of $f(\beta)$ to obtain a single number

Here β could be a parameter or a value to be forecast

Recap

Because of Idea 1, choosing between $f(\beta)$ and a competing formula $f_2(\beta)$ comes down to the following:

Underlying Idea 2

Would you prefer to produce your estimate of β by reaching blindly into the sampling distribution of $f(\beta)$ or by reaching blindly into the sampling distribution of $f_2(\beta)$?



Recap

Underlying Idea 3

Because of Idea 2, desirable properties of an estimator $f(\beta)$ are defined in terms of its sampling distribution

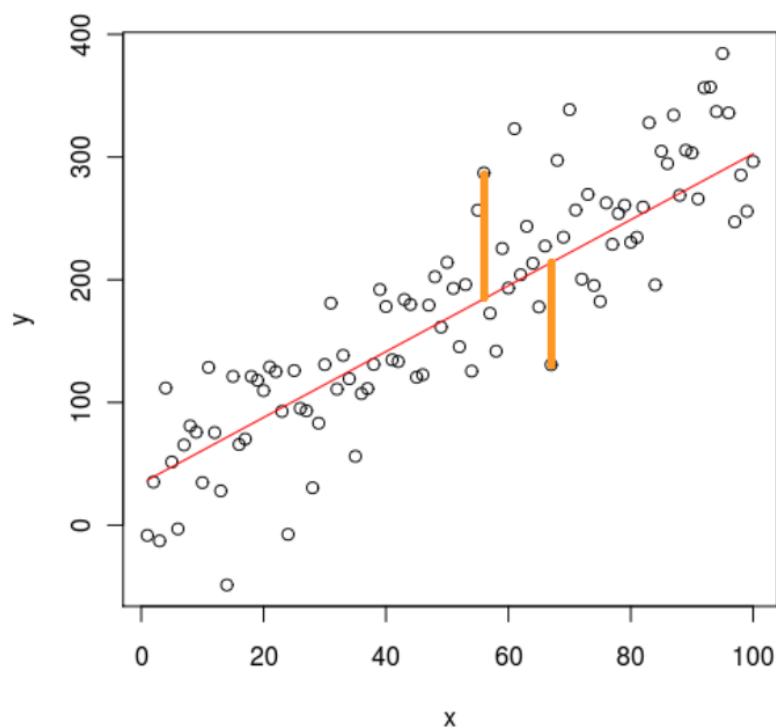
For example, $f(\beta)$ is unbiased if the mean of its sampling distribution equals the number β being estimated

This explains why econometricians spend so much algebraic energy figuring out sampling distribution properties, such as mean and variance

Underlying Idea 4

The properties of the sampling distribution of an estimator $f(\beta)$ depend on the process generating the data, so an estimator can be a good one in one context but a bad one in another

When we move from one topic to another we are moving from one data generating process to another, necessitating a re-examination of the sampling distribution properties of familiar estimators, and development of new estimators designed to have "better" sampling distribution properties



- 1 Linear
 - Y should be a linear function of X variables and ε
- 2 Expected value of ε is zero for all observations
 - $E(\varepsilon_i) = 0$
- 3 Homoskedasticity
 - The conditional variance of the error term is constant in all x and over time
- 4 Error term is independently distributed and not correlated
 - $Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) = 0 \text{ } i \neq j$
- 5 No perfect collinearity between regressors
- 6 Randomly sampled from the same population
 - *Gauss-Markov assumptions*

If this is not the case there is a biased intercept problem

Means there should be no autocorrelation

5: No Perfect Collinearity in Regressors

Different regressors used in the model cannot contain exactly the same information

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- Given this we cannot have: $X_1 = \delta_0 + \delta_1 X_2$
- So X_1 and X_2 cannot share exactly the same information

E.g. house prices

$$Price = \alpha + \beta_1 SQmeter + \beta_2 SQfoot + \varepsilon$$

Perfect collinearity as SQmeter and SQfoot provide the same information

Beyond OLS

What happens if these assumptions break down?

Then OLS will not necessarily be the best model to use

This is what we will base most of this course on

Starting with models that have non-continuous dependent variables

- Which **can** make linear models poor estimators

- 1 Be able to review the limitations of simple linear regression analysis and the situations where it is not the most appropriate method to use;
 - What assumptions need to hold for OLS to be appropriate?
 - What do we mean by be appropriate?
- 2 Have developed an awareness of which models are appropriate to use when an OLS regression may not be robust, and be able to apply these models correctly;
- 3 Command an experienced understanding of the empirical approach to economics, and be able to undertake a robust piece of econometric analysis;
- 4 Be able to explain the covered econometric concepts clearly in concise non-technical language
 - Do you know what the key words mean?

Your questions please....?

Gauss-Markov assumptions - a set of assumptions that if hold ensure that the best linear unbiased estimator (BLUE) of the coefficients is given by the ordinary least squares (OLS) estimator