Lecture 6. The Classical Linear Regression Model

Simple linear regression model

$$Y = \alpha + \beta X$$

Assumptions in previous lecture

Assumption 1: u **is a random variable with** $E(u) = 0$

Assumption 2: The probability distribution of the random error u **is independent of** X

This is implied by

Assumption 2': X is a deterministic, i.e. non-stochastic, variable

In other words: the observed values $X_i, i = 1,\ldots,n$ can be treated as n constants.

In practice assumption 2' only holds in special cases, e.g. if X is (calendar) time.

In most cases X is also the outcome of some random experiment, e.g. in macroeconomics an important relation is that between aggregate consumption Y and aggregate income X. Both variables are random variables associated with the random experiment that determines the state of the economy in a year.

The random experiment associated with the linear regression model is the determination of Y given the value of X. It is not important how X is determined, as long as the assumptions of the regression model (until now assumptions 1 and 2) are satisfied.

Hence we have two random experiments: one that determines X and one that determines Y given the outcome of the first experiment.

We are only interested in the second random experiment.

If we have data $Y_i, X_i, i = 1,\ldots,n$, i.e. n observations on a dependent variable Y and an independent variable X, we consider these as outcomes of the n random experiments

(1) $$Y_i = \alpha + \beta X_i + u_i$$

with $i = 1,\ldots,n$.

These random experiments are as follows:

1. **Treat $X_i, i = 1,\ldots,n$ as n constants (it does not matter how they are determined).**
2. **For each i, u_i is a draw from a probability distribution that does not depend on X_i (assumption 2) and has mean 0 (assumption 1).**
3. **Y_i is determined as in (1).**

If we choose a particular distribution for u_i, e.g. the normal (mean 0) distribution, then we can write a computer program that generates datasets of size n. This will give repeated samples of size n that are outcomes of random experiment (1).
The Ordinary Least Squares (OLS) solutions to fitting a straight line in a scatterdiagram are

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

In the Simple Linear Regression random experiment these solutions are estimators of the parameters, here regression coefficients, α, β.

These estimators have a sampling distribution that as usual can be used to

- **Evaluate the quality of the estimators**
- **Find confidence intervals**
- **Perform hypothesis tests**

Sampling distribution: Distribution of estimators in repeated samples of size n.

We can use the computer to find the sampling distribution (see above for description of the random experiment)

Let

$\alpha = 1$

$\beta = 3$

$X_i, \quad i = 1, \ldots, n$ **are uniform [0,1] random numbers**

u_i **has a standard normal (mean 0, variance 1) distribution, and all** u_i**'s are independent and have the same distribution**

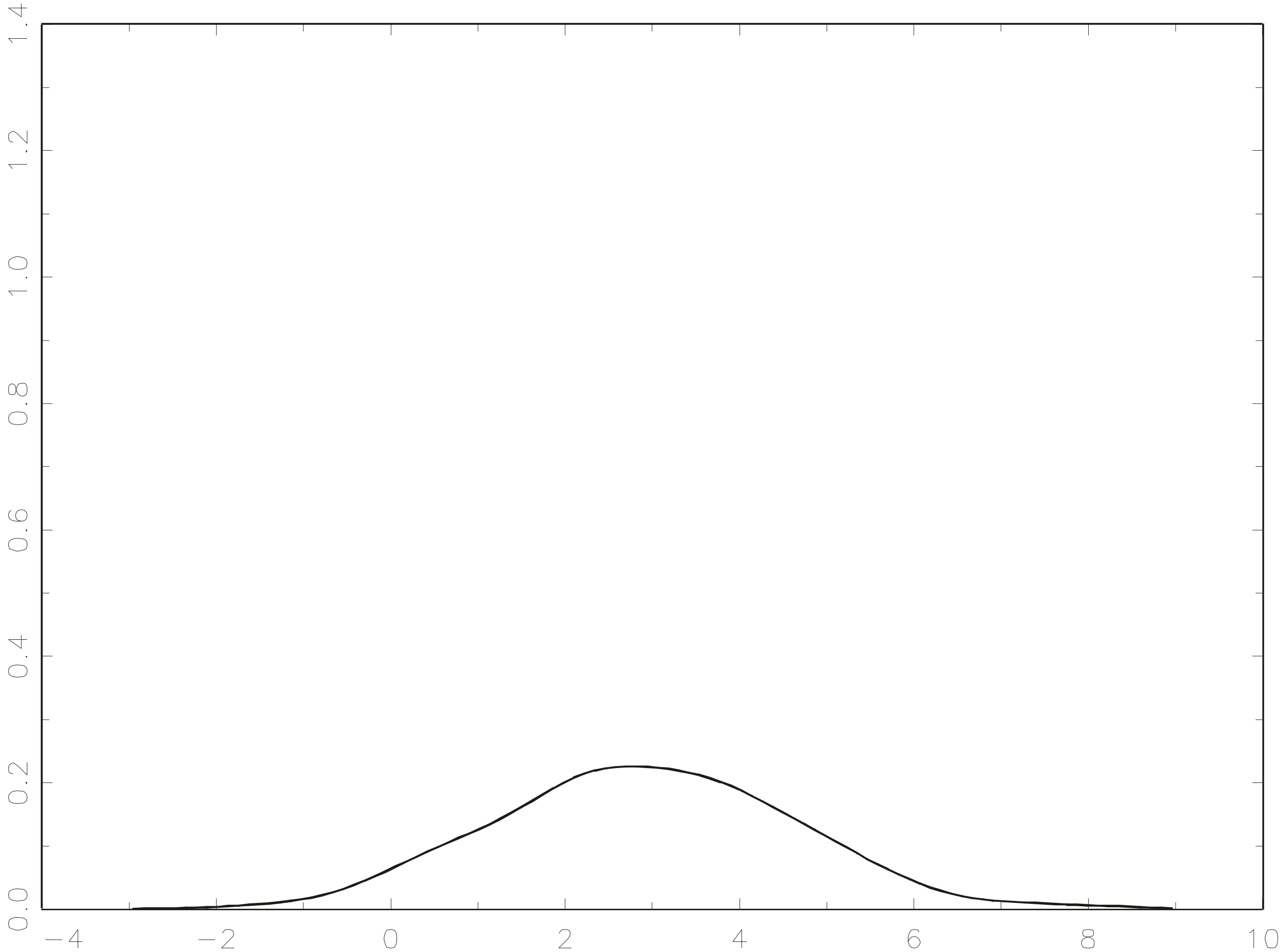
This completely specifies the Linear Regression random experiment.

The graphs give the sampling distribution of $\hat{\beta}$ for $n = 10$ and 100.

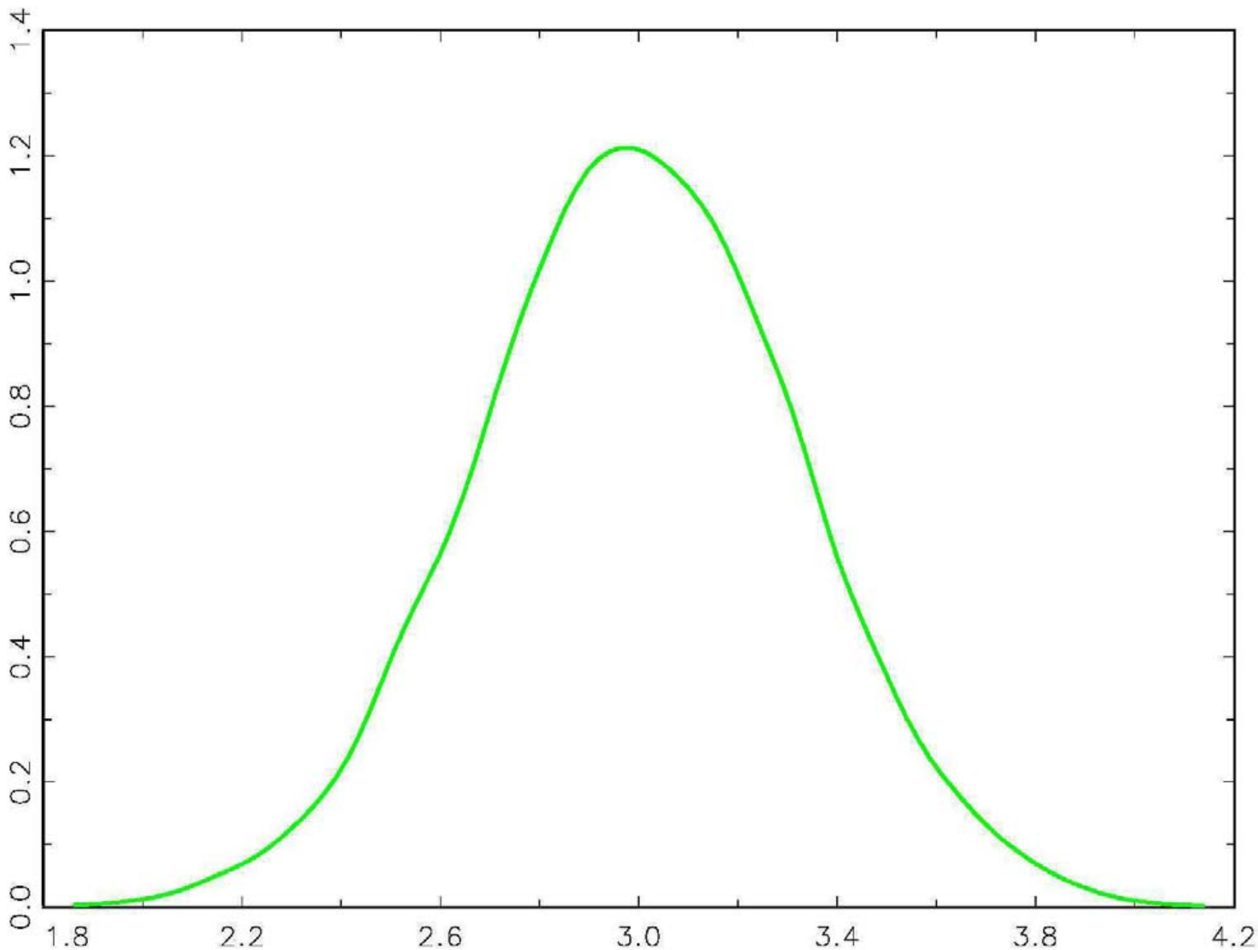
A second set of graphs gives the sampling distribution if we assume that the u_i's are independent and have the same distribution, but that distribution is the uniform $[-\sqrt{3}, \sqrt{3}]$ distribution (also has mean 0 and variance 1).

Which is more relevant: the normal or uniform distribution?

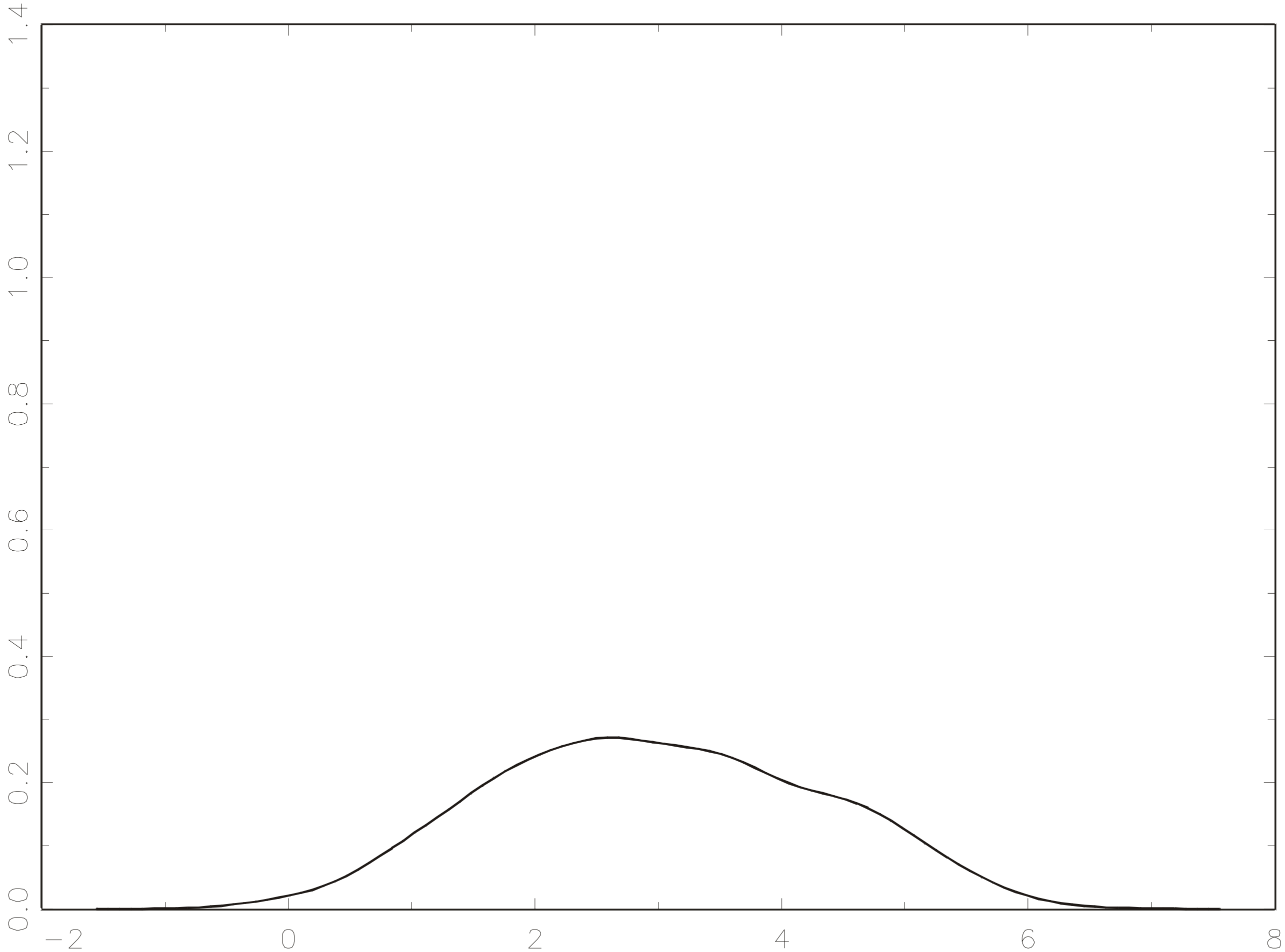
Sampling distr. of $\hat{\beta}$, $n = 10$, u standard normal



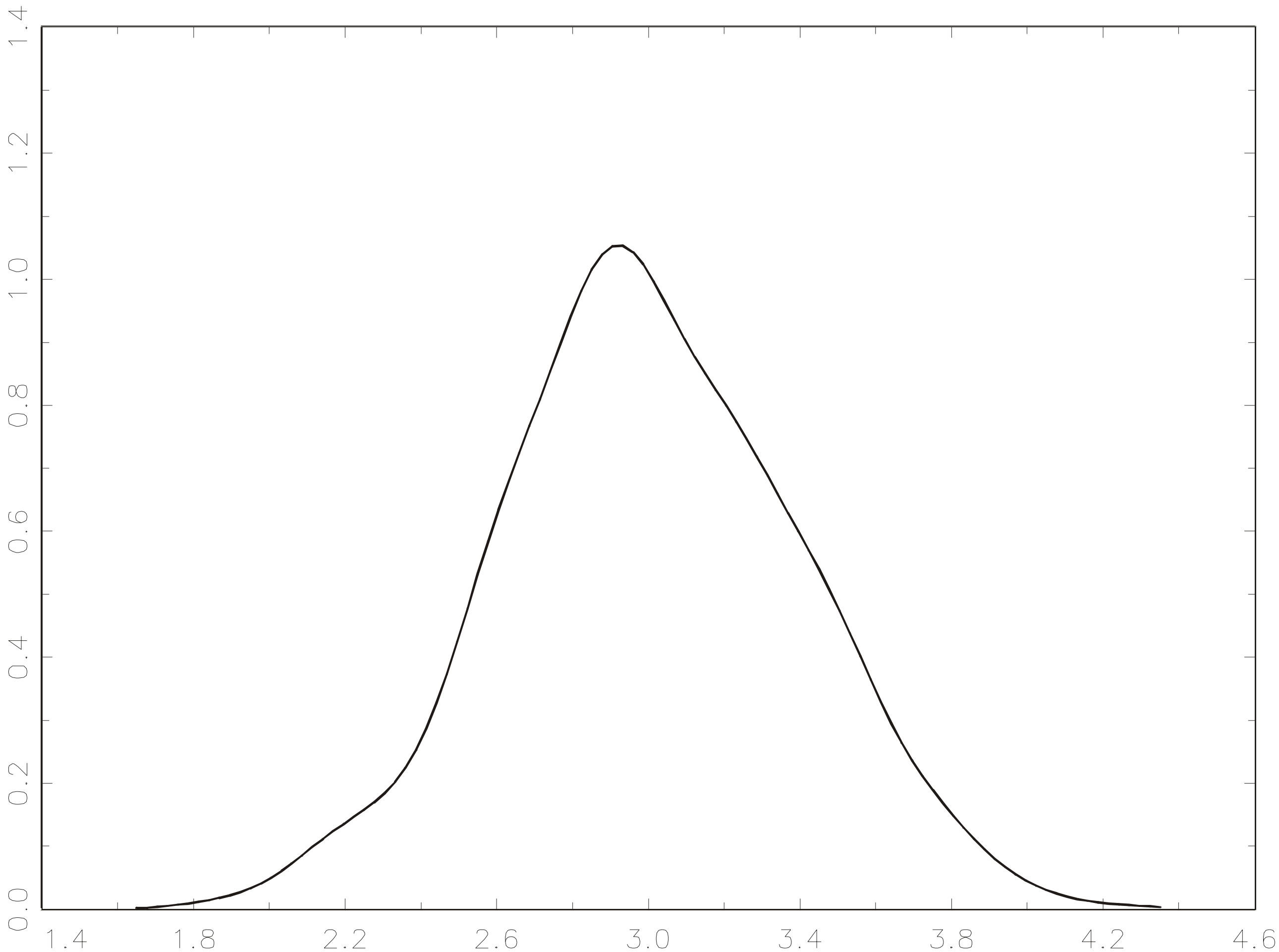
Sampling distr. of $\hat{\beta}$, $n = 100$, u standard normal



Sampling distr. of $\hat{\beta}$, $n = 10$, u uniform



Sampling distr. of $\hat{\beta}$, $n = 100$, u uniform



If $n = 10, \quad 100$ the sampling distribution is close to normal even if the u_i's have a uniform distribution. Variance is larger if $n = 10$.

As in the coin tossing experiment, we can derive the sampling distribution of $\hat{\alpha}, \hat{\beta}$ by using the assumptions instead of using the computer to generate samples.

Consider the linear regression model with $\alpha = 0$.

Then

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^{2}} = \sum_{i=1}^{n} W_i Y_i$$

with

$$W_i = \frac{X_i}{\sum_{j=1}^{n} X_j^2}$$

From the Linear Regression model

$$Y_i = \beta X_i + u_i$$

Substitution gives, using $\sum_{i=1}^{n} W_i X_i = 1$**,**

$$\hat{\beta} = \sum_{i=1}^{n} W_i Y_i = \sum_{i=1}^{n} W_i (\beta X_i + u_i) = \beta + \sum_{i=1}^{n} W_i u_i$$

Because the W_i can be treated as constants (assumption 2'), we have by assumption 1

$$E(\hat{\beta}) = \beta + E\left(\sum_{i=1}^{n} W_i u_i\right) = \beta + \sum_{i=1}^{n} W_i E(u_i) = \beta$$

Conclusion: The OLS estimator of β is unbiased

The same conclusion holds if the Linear Regression model has an intercept.

Hence, the OLS estimators $\hat{\alpha}, \hat{\beta}$ **of** α, β **are unbiased estimators.**

Compare with the sampling distribution in the computer experiment.

Next step is to derive the sampling variance of the OLS estimators.

That derivation is simpler if we make two additional assumptions

Assumption 3 (Homoskedasticity)
All $u_i's$ have the same variance

$$Var(u_i) = E(u_i^2) = \sigma^2$$

Assumption 4 (No serial correlation)
The random errors u_i and u_j are not correlated for all $i \neq j$

$$Cov(u_i, u_j) = E(u_i u_j) = 0$$

Discussion assumptions

Heteroskedasticity (Greek for equal dispersion) affects shape of scatterplot

Example: Education and late career income.

No serial correlation is important in time-series data. Most data are

- **Cross-section data: variables are for n individuals, households, firms, countries etc. in a particular time period**
- **Time-series data: variables are for one individual, firm, country etc. in n subsequent time periods (weeks, months, quarters, years)**
- **Panel data: combination of these two**

Remember error term u captures omitted variables. Most time-series are such that the observations in subsequent time periods are correlated. Same is true for omitted variables.
Hence in time-series data assumption 4 need not hold.

Serial correlation also affects shape of scatterplot.

The linear regression model

$$Y_i = \alpha + \beta X_i + u_i \quad i = 1, \ldots, n$$

where the random error term satisfies assumptions 1-4 is called the Classical Linear Regression (CLR) model.

Now we use assumptions 3 and 4 to derive the sampling variance of $\hat{\beta}$ (if we assume that $\alpha = 0$)

Remember

$$\hat{\beta} = \beta + \sum_{i=1}^{n} W_i u_i$$

Hence,

$$Var(\hat{\beta}) = E\left((\hat{\beta} - \beta)^2\right) = E\left[\left(\sum_{i=1}^{n} W_i u_i\right)^2\right] =$$

$$= E\left[\sum_{i=1}^{n} W_i^2 u_i^2 + 2\sum_{i<j}^{n} W_i W_j u_i u_j\right] =$$

$$= \sum_{i=1}^{n} W_i^2 E\left(u_i^2\right) + 2\sum_{i<j}^{n} W_i W_j E(u_i u_j)$$

The second term is 0 by assumption 4. By assumption 3 $E(u_i^2) = \sigma^2$.

Also by the definition of W_i, we have $\sum_{i=1}^{n} W_i^2 = \frac{1}{\sum_{i=1}^{n} X_i^2}$

Combining this we find

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}$$

Note

- **Variance smaller if σ^2 smaller**
- **Variance smaller if $\sum_{i=1}^{n} X_i^2$ larger**

How does this translate in scatterplot?

What happens to the variance of n is large?
Implication for sampling distribution?

Definition: An estimator with a sampling distribution that becomes a degenerate distribution in the population value of the parameter if the number of observations becomes large, is called consistent.

The OLS estimator $\hat{\beta}$ is consistent.

Using a similar argument as above we can derive

$$Var(\hat{\alpha}) = \frac{\sum_{i=1}^{n} X_i^2}{n\sum_{i=1}^{n}(X_i - \overline{X})^2}\sigma^2$$

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

$$Cov(\hat{\alpha}, \hat{\beta}) = -\frac{\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sigma^2$$

Moreover, the OLS estimators are consistent.