ECON4150 - Introductory Econometrics

Lecture 4: Linear Regression with One Regressor

Monique de Haan
(moniqued@econ.uio.no)

Stock and Watson Chapter 4

Lecture outline

- The OLS estimators
    - The effect of class size on test scores
- The Least Squares Assumptions
    - $E(u_i|X_i) = 0$
    - (X_i, Y_i) are *i.i.d*
    - Large outliers are unlikely
- Properties of the OLS estimators
    - unbiasedness
    - consistency
    - large sample distribution
- The compulsory term paper

The OLS estimators

Question of interest: What is the effect of a change in X_i on Y_i?

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Last week we derived the OLS estimators of β_0 and β_1:

$$\boxed{\widehat{\beta_0} = \overline{Y} - \hat{\beta}_1 \overline{X}}$$

$$\boxed{\widehat{\beta_1} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)} = \frac{s_{xy}}{s_x^2}},$$

OLS estimates: The effect of class size on test scores

Question of interest: What is the effect of a change in class size on test scores?

$$TestScore_i = \beta_0 + \beta_1 ClassSize_i + u_i$$

```
. regress test_score class_size, robust

Linear regression                               Number of obs     =        420
                                                  F(1, 418)         =      19.26
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.0512
                                                  Root MSE          =     18.581
```

test_score	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
class_size	**-2.279808**	**.5194892**	**-4.39**	**0.000**	**-3.300945**	**-1.258671**
_cons	**698.933**	**10.36436**	**67.44**	**0.000**	**678.5602**	**719.3057**

$$\widehat{TestScore}_i = 698.93 - 2.28 \cdot ClassSize_i$$

The Least Squares assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Under what assumptions does the method of ordinary least squares provide appropriate estimators of β_0 and β_0?

Under what assumptions does the method of ordinary least squares provide an appropriate estimator of the effect of class size on test scores?

The Least Squares assumptions:

Assumption 1: The conditional mean of u_i given X_i is zero

$$E(u_i | X_i) = 0$$

Assumption 2: (Y_i, X_i) for $i = 1, ..., n$ are independently and identically distributed ($i.i.d$)

Assumption 3: Large outliers are unlikely

$$0 < E\left(X_i^4\right) < \infty \quad \& \quad 0 < E\left(Y_i^4\right) < \infty$$

The Least Squares assumptions: Assumption 1

$$E\left(u_i|X_i\right) = 0$$

The first OLS assumption states that:

All other factors that affect the dependent variable Y_i (contained in u_i) are unrelated to X_i in the sense that, given a value of X_i, the mean of these other factors equals zero.

In the class size example:

All the other factors affecting test scores should be unrelated to class size in the sense that, given a value of class size, the mean of these other factors equals zero.

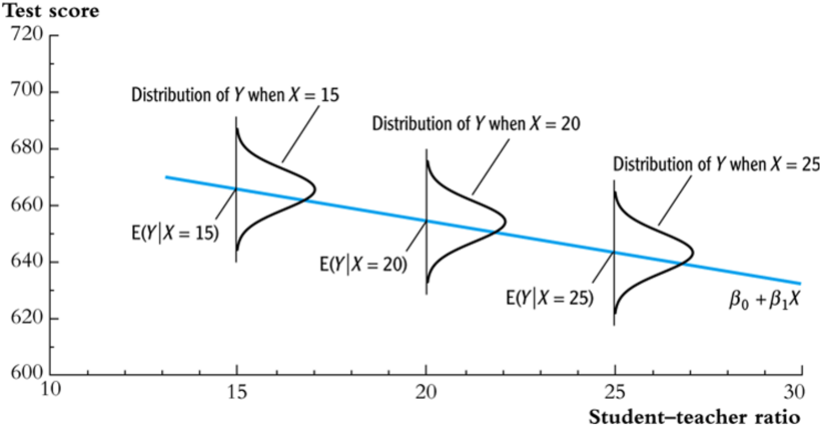
The Least Squares assumptions: Assumption 1

The first OLS assumption can also be written as:

$$
\begin{aligned}
E(Y_i|X_i) &= E(\beta_0 + \beta_1 X_i + u_i|X_i) \\
& \qquad \textit{Expectation rules} \\
&= \beta_0 + \beta_1 E(X_i|X_i) + E(u_i|X_i) \\
& \qquad ASS\#1: \; E(u_i|X_i) = 0 \\
&= \beta_0 + \beta_1 X_i
\end{aligned}
$$

The Least Squares assumptions: Assumption 1

$$E\left(Y_i|X_i\right) = \beta_0 + \beta_1$$



The Least Squares assumptions: Assumption 1

Example of a violation of assumption 1:

Suppose that

- districts which wealthy inhabitants have small classes and good teachers
  - these districts have a lot of money which they can use to hire more and better teachers
- districts with poor inhabitants have large classes and bad teachers.
  - These districts have little money and can hire only few and not very good teachers

In this case class size is related to teacher quality.

Since teacher quality likely affects test scores it is contained in u_i.

This implies a violation of assumption 1:

$$E\left(u_i | ClassSize_i = small\right) \neq E\left(u_i | ClassSize_i = large\right) \neq 0$$

The Least Squares assumptions: Assumption 2

$$(Y_i, X_i) \text{ for } i = 1, ..., n \text{ are } i.i.d$$

- If the sample is drawn by simple random sampling assumption 2 will hold

Example: What is effect of mother's education (X_i) on child's education (Y_i)

Example of simple random sampling:

- randomly draw sample of mother's with information on her education and the education of *one randomly selected* child
- (Y_i, X_i) for $i = 1, ..., n$ are $i.i.d$

Example of a violation of simple random sampling

- randomly draw sample of mothers with information on her education and the education of *all* of her children.
- (Y_i, X_i) for $i = 1, ..., n$ are NOT $i.i.d$
- Observations on children from the same mother are not *independent*!

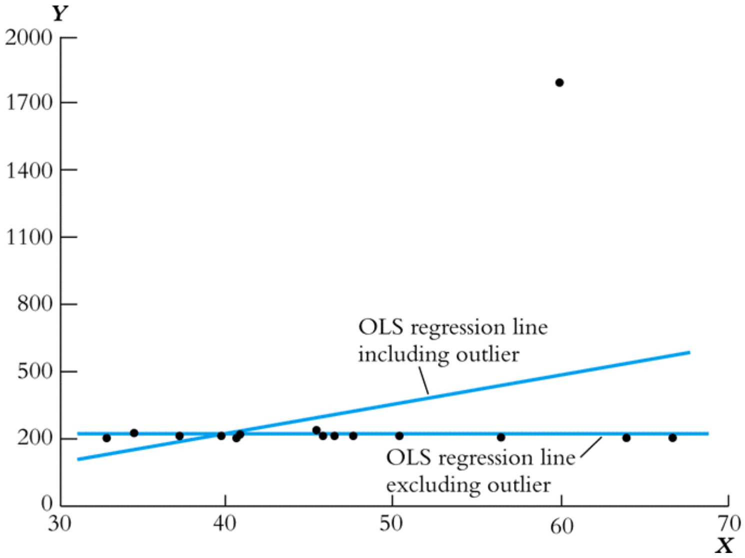
The Least Squares assumptions: Assumption 3

Large outliers are unlikely

$$0 < E\left(X_i^4\right) < \infty \quad \& \quad 0 < E\left(Y_i^4\right) < \infty$$

- Outliers are observations that have values far outside the usual range of the data
- Large outliers can make OLS regression results misleading
- Another way to state assumption is that X and Y have finite kurtosis.
- Assumption is necessary to justify the large sample approximation to the sampling distribution of the OLS estimators

The Least Squares assumptions: Assumption 3



Use of the Least Squares assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Assumption 1: $E(u_i|X_i) = 0$

Assumption 2: (Y_i, X_i) for $i = 1, ..., n$ are $i.i.d$

Assumption 3: Large outliers are unlikely

If the 3 least squares assumptions hold the OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$

- Are unbiased estimators of β_0 and β_1
- Are consistent estimators of β_0 and β_1
- Have a jointly normal sampling distribution

Properties of the OLS estimator: unbiasedness

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \overline{Y} = \beta_0 + \beta_1 X_i + \overline{u}$$

$$E\left[\widehat{\beta}_1\right] = E\left[\frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

substitute for $Y_i, \overline{Y}$

$$= E\left[\frac{\sum_{i=1}^n (X_i - \overline{X})\left(\beta_0 + \beta_1 X_i + u_i - (\beta_0 + \beta_1 \overline{X} + \overline{u})\right)}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

rewrite (β_0 *drops out*)

$$= E\left[\frac{\sum_{i=1}^n (X_i - \overline{X})\left(\beta_1 (X_i - \overline{X}) + (u_i - \overline{u})\right)}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

rewrite & use expectation rules

$$= E\left[\frac{\beta_1 \sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right] + E\left[\frac{\sum_{i=1}^n (X_i - \overline{X})(u_i - \overline{u})}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

Properties of the OLS estimator: unbiasedness

.

$$E\left[\widehat{\beta}_1\right] = E\left[\frac{\beta_1 \sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right] + E\left[\frac{\sum_{i=1}^n (X_i - \overline{X})(u_i - \overline{u})}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

take β_1 out of 1st expectation
Algebra trick

$$= \beta_1 + E\left[\frac{\sum_{i=1}^n (X_i - \overline{X}) u_i}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

Law of iterated expectations

$$= \beta_1 + E\left[\frac{\sum_{i=1}^n (X_i - \overline{X}) E[u_i|X_i]}{\sum_{i=1}^n (X_i - \overline{X})(X_i - \overline{X})}\right]$$

$$E\left[\widehat{\beta}_1\right] = \beta_1 \quad \textit{if} \quad E\left[u_i|X_i\right] = 0$$

Algebra trick

$$
\begin{aligned}
\sum_{i=1}^{n}\left(X_i-\overline{X}\right)(u_i-\overline{u}) &= \sum_{i=1}^{n}X_iu_i-\sum_{i=1}^{n}X_i\overline{u}-\sum_{i=1}^{n}\overline{X}u_i+\sum_{i=1}^{n}\overline{X}\overline{u} \\
&= \sum_{i=1}^{n}X_iu_i-n\cdot\left(\tfrac{1}{n}\sum_{i=1}^{n}X_i\right)\overline{u}-\sum_{i=1}^{n}\overline{X}u_i+n\overline{X}\overline{u} \\
&= \sum_{i=1}^{n}X_iu_i-n\overline{X}\overline{u}+\sum_{i=1}^{n}\overline{X}u_i+n\overline{X}\overline{u} \\
&= \sum_{i=1}^{n}X_iu_i-\sum_{i=1}^{n}\overline{X}u_i \\
&= \sum_{i=1}^{n}\left(X_i-\overline{X}\right)u_i
\end{aligned}
$$

Consistency

Consistency: $\widehat{\beta}_1 \xrightarrow{p} \beta_1$ *or* $plim\ \widehat{\beta}_1 = \beta_1$

$$Plim\ \widehat{\beta}_1 = plim\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})}\right)$$

$$= \frac{Plim\ \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{Plim\ \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})} = \frac{Plim\ s_{XY}}{Plim\ s_X^2}$$

law of large numbers
OLS assumptions 2 *and* 3

$$= \frac{Cov(X_i, Y_i)}{Var(X_i)}$$

substitute for Y_i

$$= \frac{Cov(X_i, \beta_0 + \beta_1 X_i + u_i)}{Var(X_i)}$$

see Key Concept 2.3

$$= \frac{\beta_1 Var(X_i) + Cov(X_i, u_i)}{Var(X_i)}$$

Consistency

$$
\begin{aligned}
Plim\ \widehat{\beta}_1 &= \frac{\beta_1 Var(X_i) + Cov(X_i, u_i)}{Var(X_i)} \\
&= \beta_1 \frac{Var(X_i)}{Var(X_i)} + \frac{Cov(X_i, u_i)}{Var(X_i)}
\end{aligned}
$$

substitute covariance expression

$$= \beta_1 + \frac{E[(X_i - \mu_X)(u_i - \mu_u)]}{Var(X_i)}$$

algebra trick

$$= \beta_1 + \frac{E[(X_i - \mu_X)u_i]}{Var(X_i)}$$

Law of iterated expectations

$$= \beta_1 + \frac{E[(X_i - \mu_X)E[u_i|X_i]]}{Var(X_i)}$$

so

$$Plim\ \widehat{\beta}_1 = \beta_1 \quad \text{if} \quad E[u_i|X_i] = 0$$

Unbiasedness vs Consistency

- Unbiasedness & consistency both rely on $E\left[u_i|X_i\right] = 0$
- Unbiasedness implies that $E\left[\widehat{\beta}_1\right] = \beta_1$ for a given sample size n
- Consistency implies that the sampling distribution becomes more and more tightly distributed around β_1 if the sample size n becomes larger and larger.

Consistency: A simulation example

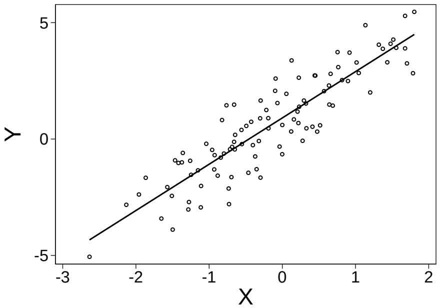
- Lets create a data set with 100 observations
- $X_i \sim N(0,1)$
- $u_i \sim N(0,1)$
- We define Y to depend on X as: $Y_i = 1 + 2X_i + u_i$

```
set obs 100
gen x=rnormal()
gen y=1+2*x+rnormal()
```

```
. sum y x
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	**100**	**.6123606**	**2.211365**	**-5.05828**	**5.462746**
x	**100**	**-.1479108**	**.9928607**	**-2.633841**	**1.80305**

A simulation example




```
. regress y x

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =  385.45
       Model |  385.987671     1  385.987671           Prob > F      =  0.0000
    Residual |  98.1357149    98  1.00138485           R-squared     =  0.7973
-------------+------------------------------           Adj R-squared =  0.7952
       Total |  484.123386    99  4.89013521           Root MSE      =  1.0007

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.988753   .1012965    19.63   0.000     1.787733    2.189772
       _cons |   .9065187   .1011847     8.96   0.000      .705721    1.107316
------------------------------------------------------------------------------
```

A simulation example n=100

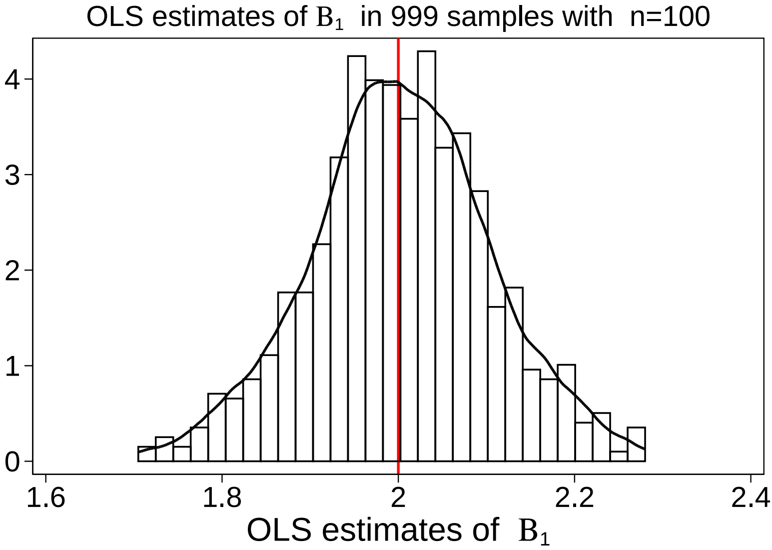
We can create 999 of these data sets with 100 observations and use OLS to estimate

$$Y_i = \beta_0 + \beta_1 + u_i$$

```
1 . program define ols, rclass
    1.  drop _all
    2.  set obs 100
    3.  gen x=invnorm(uniform())
    4.  gen y=1+2*x+invnorm(uniform())
    5.  regress y x
    6. end

2 .
3 . simulate _b, reps(999)  nodots : ols

      command:  ols


4 . sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	999	1.997521	.1018595	1.67569	2.308795
_b_cons	999	1.003246	.1019056	.6844429	1.285363

A simulation example n=100

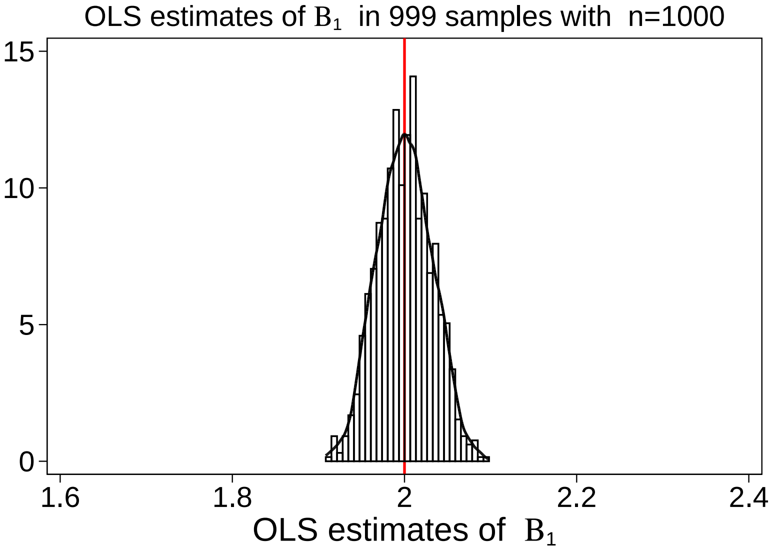


A simulation example n=1000

```
1 . program define ols, rclass
    1.  drop _all
    2.  set obs 1000
    3.  gen x=invnorm(uniform())
    4.  gen y=1+2*x+invnorm(uniform())
    5.  regress y x
    6. end

2 .
3 . simulate _b, reps(999)  nodots : ols

      command:  ols


4 . sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	**999**	**2.000035**	**.030417**	**1.908725**	**2.112585**
_b_cons	**999**	**1.000791**	**.0311526**	**.8970624**	**1.088724**

A simulation example n=1000

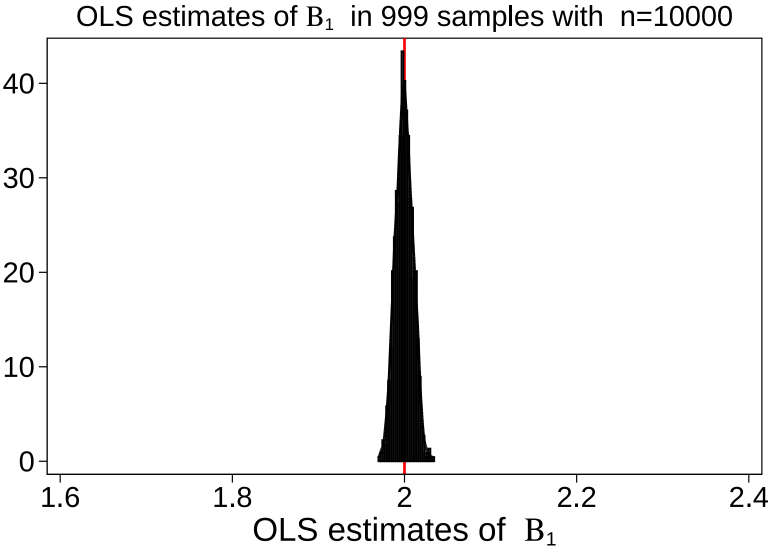


A simulation example n=10000

```
1 . program define ols, rclass
    1.  drop _all
    2.  set obs 10000
    3.  gen x=invnorm(uniform())
    4.  gen y=1+2*x+invnorm(uniform())
    5.  regress y x
    6. end

2 .
3 . simulate _b, reps(999) nodots: ols

      command:  ols


4 . sum
```

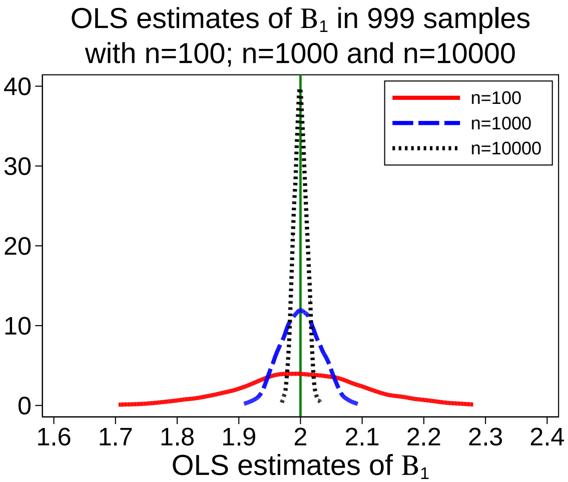
Variable	Obs	Mean	Std. Dev.	Min	Max
_b_x	**999**	**1.999748**	**.0099715**	**1.969678**	**2.034566**
_b_cons	**999**	**1.000391**	**.0100135**	**.9699681**	**1.033458**

A simulation example n=10000



Consistency of the OLS estimator of $\widehat{\beta}_1$

True model : $Y_i = 1 + 2X_i + u_i$, *Estimated model* : $Y_i = \beta_0 + \beta_1 X_i + u_i$



Sampling distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$

We discussed the sampling distribution of the sample average $\overline{Y}$:

- sampling distribution is complicated for small n, but if $Y_1, ..., Y_n$ are i.i.d. we know that

$$E\left(\overline{Y}\right) = \mu_Y$$

- By the Central Limit theorem the large sample distribution can be approximated by the normal distribution:

$$\overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

If the 3 least squares assumptions hold we can make similar statements about the OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$

Large-sample distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$

- Technically the Central Limit theorem concerns the large sample distribution of averages (like $\overline{Y}$)
- Examining the formulas of the OLS estimators shows that these are functions of sample averages:

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

$$\widehat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})}$$

- It turns out that the Central Limit theorem also applies to these functions of sample averages.

Sampling distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$

If the first least squares assumption holds:

- The OLS estimators are unbiased which implies that (for any sample size n)

$$E\left(\widehat{\beta}_0\right) = \beta_0 \quad \textit{and} \quad E\left(\widehat{\beta}_1\right) = \beta_1$$

In addition, if all 3 least squares assumptions hold

- The Central Limit theorem implies that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are approximately jointly normally distributed in large samples:

$$\widehat{\beta}_0 \sim N\left(\beta_0, \sigma^2_{\widehat{\beta}_0}\right)$$

$$\widehat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\widehat{\beta}_1}\right)$$

Large-sample distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$

In large samples

$$\widehat{\beta}_0 \sim N\left(\beta_0, \sigma^2_{\widehat{\beta}_0}\right)$$

$$\widehat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\widehat{\beta}_1}\right)$$

where it can be shown that

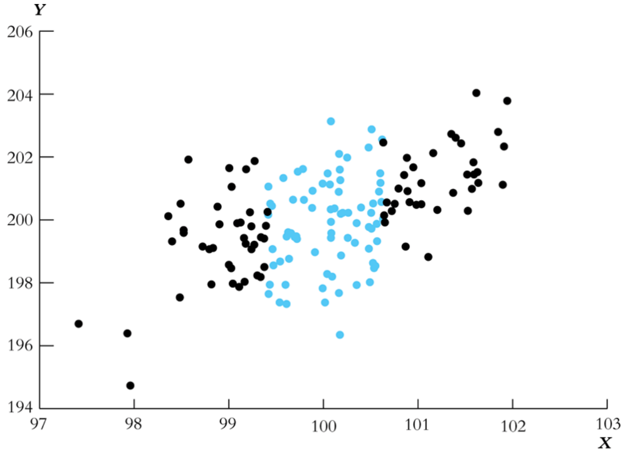
$$\sigma^2_{\widehat{\beta}_0} = \frac{1}{n}\frac{Var(H_i u_i)}{\left[E\left(H_i^2\right)\right]^2} \qquad with\ H_i = 1 - \left[\frac{\mu_X}{E\left(X_i^2\right)}\right] X_i$$

$$\sigma^2_{\widehat{\beta}_1} = \frac{1}{n}\frac{Var[(X_i - \mu_X)u_i]}{[Var(X_i)]^2}$$

Expression for $\sigma^2_{\widehat{\beta}_1}$ shows that the larger the variation in the regressor X_i the smaller the variance of $\widehat{\beta}_1$

Large-sample distribution of $\widehat{\beta}_0$ and $\widehat{\beta}_1$

- When $Var(X_i)$ is low, it is difficult to obtain an accurate estimate of the effect of X on Y which implies that $Var\left(\widehat{\beta}_1\right) = \sigma^2_{\widehat{\beta}_1}$ is high.
- If there is more variation in X, then there is more information in the data that you can use to fit the regression line.



Compulsory term paper

- There is an extensive theoretical literature about the trade-off between child quality and quantity (Becker & Lewis (1973), Becker & Tomes (1976)).
- Idea behind these models is that if parents have more children, investing a certain amount in per-child quality is more expensive, than if they have fewer children.
- When there is an increase in the number of children parents will, for a given budget constraint, lower the investment in per-child quality.
- This reasoning predicts that there is a negative effect of the number of children in a family on the educational attainment of these children.

Compulsory term paper

- A negative relationship between family size and educational achievements is however not necessarily proof of a negative effect of the number of children.
- The number of children is a choice variable of the parents.
- It might be that certain characteristics of parents, such as their preferences, their abilities or their educational attainments, affect both the number of children as well as the educational attainments of those children.

Compulsory term paper

- In the term paper you are going to investigate the following research question.

What is the causal effect of family size on children's years of completed schooling?

- This research question can be addressed by using the data set WLSfamily.dta.
- Data set contains information about 14,127 children from 5000 families.
- The data set can be downloaded from the course website site.
- The document WLSfamily.pdf gives more information about the data set and the variables contained in the data set.

Compulsory term paper

The term paper should consist of the following sections:

- Introduction
- Empirical approach
- Data
- Results
- Conclusion
- References
- Appendix with Stata code & output

The term paper should be at most 10 pages including tables and figures (but excluding the stata code and output).

The quality (and not the quantity) of the content of the term paper will determine your grade.

Compulsory term paper

Important dates

- 29 January 2018– Hand-out of term paper
- 23 March 2018 – Hand-in of term paper on Fronter
- 13 April 2018 – Notification of grade (pass/fail)
- 23 April 2018 – Hand-in of improved term paper for those who failed
- 4 May 2018– Everyone is informed about final grade for term paper