Lecture 2:The Simple Regression Model

1.Definition of the Simple Regression Model

Definition of the Simple Regression Model

Definition of the simple linear regression model

"Explains variable y in terms of variable x"

Intercept

Slope parameter

$$y = \beta_0 + \beta_1 x + u$$

Dependent variable,
explained variable,
response variable,…

Independent variable,
explanatory variable,
regressor,…

Error term,
disturbance,
unobservables,…

Definition of the Simple Regression Model

Interpretation of the simple linear regression model

"Studies how y varies with changes in x:"

$$\frac{\Delta y}{\Delta x} = \beta_1 \quad \text{as long as} \quad \frac{\Delta u}{\Delta x} = 0$$

By how much does the dependent variable change if the independent variable is increased by one unit?

Interpretation only correct if all other things remain equal when the independent variable is increased by one unit

The simple linear regression model is rarely applicable in prac-tice but its discussion is useful for pedagogical reasons

Definition of the Simple Regression Model

Example: Soybean yield and fertilizer

$$yield = \beta_0 + \beta_1 fertilizer + u$$

Rainfall,
land quality,
presence of parasites, …

Measures the effect of fertilizer on yield, holding all other factors fixed

Example: A simple wage equation

$$wage = \beta_0 + \beta_1 educ + u$$

Labor force experience,
tenure with current employer,
work ethic, intelligence, …

Measures the change in hourly wage given another year of education, holding all other factors fixed

Definition of the Simple Regression Model

When is there a causal interpretation?

Conditional mean independence assumption

$$E(u|x) = 0$$

The explanatory variable must not contain information about the mean of the unobserved factors

Example: wage equation

$$wage = \beta_0 + \beta_1 educ + u$$

e.g. intelligence …

The conditional mean independence assumption is unlikely to hold because individuals with more education will also be more intelligent on average.

Definition of the Simple Regression Model

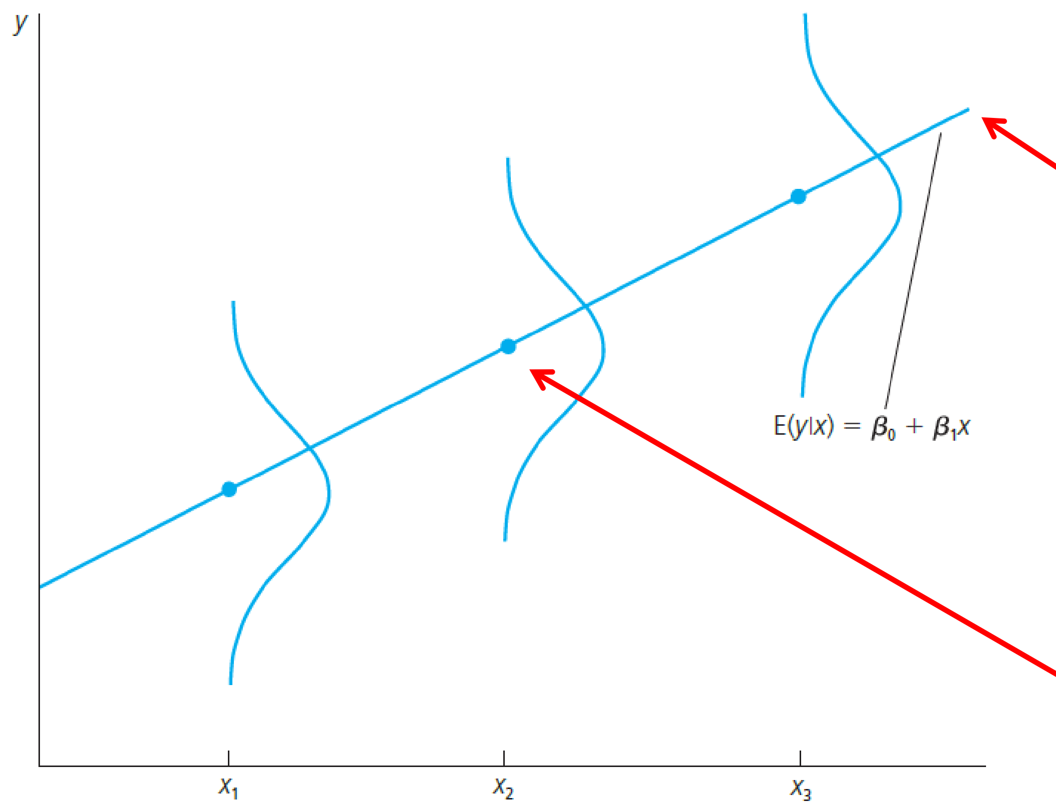
Population regression function (PFR)

- The conditional mean independence assumption implies that

$$
\begin{aligned}
E(y|x) &= E(\beta_0 + \beta_1 x + u|x) \\
&= \beta_0 + \beta_1 x + E(u|x) \\
&= \beta_0 + \beta_1 x
\end{aligned}
$$

- This means that the average value of the dependent variable can be expressed as a linear function of the explanatory variable

Definition of the Simple Regression Model



Population regression function

For individuals with $x = x_2$, the average value of y is $\beta_0 + \beta_1 x_2$

2.Deriving OLS Estimates

Deriving OLS Estimates

Deriving the ordinary least squares estimates

In order to estimate the regression model one needs data

A random sample of n observations

(x_1, y_1) First observation

(x_2, y_2) Second observation

(x_3, y_3) Third observation

$\vdots$

(x_n, y_n) n-th observation

$$\{(x_i, y_i) : i = 1, \ldots, n\}$$

Value of the explanatory variable of the i-th observation

Value of the dependent variable of the i-th observation

Deriving OLS Estimates

What does “as good as possible” mean?

Regression residuals

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

Minimize sum of squared regression residuals

$$\min \sum_{i=1}^{n} \hat{u}_i^2 \quad \rightarrow \quad \hat{\beta}_0, \hat{\beta}_1$$

Ordinary Least Squares (OLS) estimates

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Deriving OLS Estimates

To derive the OLS estimates we need to realize that our main assumption of $E(u|x) = E(u) = 0$ also implies that

$$Cov(x, u) = E(xu) = 0$$

Why? Remember from basic probability that

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

Deriving OLS Estimates continued

We can write our 2 restrictions just in terms of x, y, β_0 and β_1, since $u = y - \beta_0 - \beta_1 x$

$$E(y - b_0 - b_1 x) = 0$$

$$E[x(y - b_0 - b_1 x)] = 0$$

These are called moment restrictions

The method of moments approach to estimation implies imposing the population moment restrictions on the sample moments.

What does this mean? Recall that for E(X), the mean of a population distribution, a sample estimator of E(X) is simply the arithmetic mean of the sample.

More Derivation of OLS

Given the definition of a sample mean, and properties of summation, we can rewrite the first condition as follows

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x},$$

or

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\sum_{i=1}^{n} x_i \left(y_i - \left(\bar{y} - \hat{\beta}_1 \bar{x} \right) - \hat{\beta}_1 x_i \right) = 0$$

$$\sum_{i=1}^{n} x_i \left(y_i - \bar{y} \right) = \hat{\beta}_1 \sum_{i=1}^{n} x_i \left(x_i - \bar{x} \right)$$

$$\sum_{i=1}^{n} \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right) = \hat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2$$

Alternate approach to derivation

Given the intuitive idea of fitting a line, we can set up a formal minimization problem

That is, we want to choose our parameters such that we minimize the following:

$$\sum_{i=1}^{n} (\hat{u}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

- If one uses calculus to solve the minimization problem for the two parameters you obtain the following first order conditions, which are the same as we obtained before, multiplied by *n*

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\sum_{i=1}^{n} x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

Summary of OLS slope estimate

The slope estimate is the sample covariance between *x* and *y* divided by the sample variance of *x*

If *x* and *y* are positively correlated, the slope will be positive

If *x* and *y* are negatively correlated, the slope will be negative

Only need *x* to vary in our sample

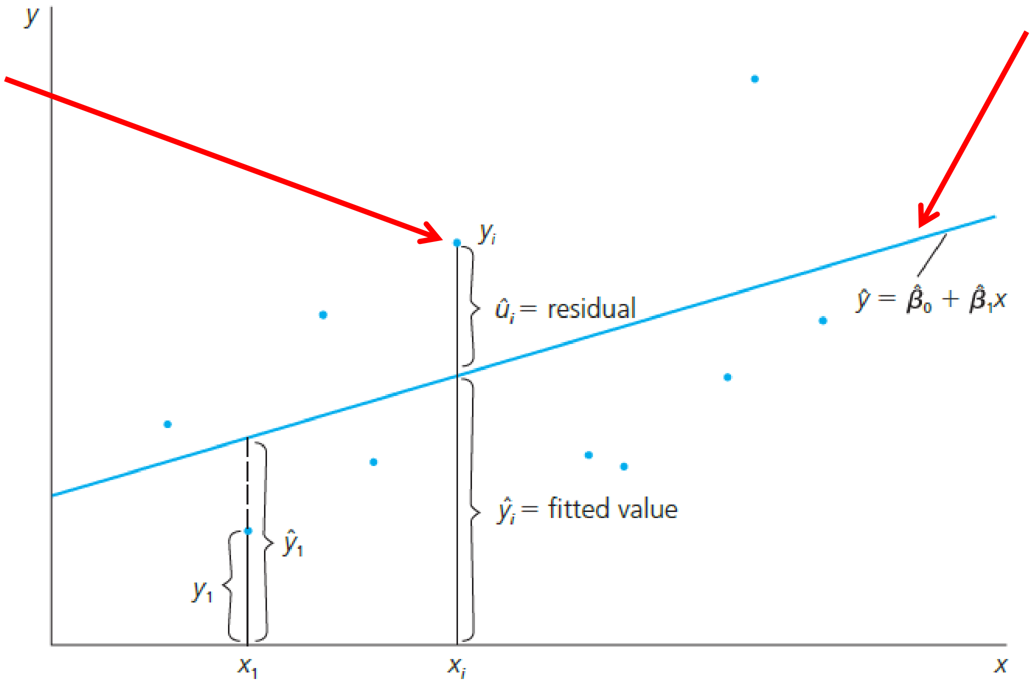
$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$$

Deriving OLS Estimates

Fit as good as possible a regression line through the data points:

For example, the i-th data point (x_i, y_i)

Fitted regression line



Deriving OLS Estimates

CEO Salary and return on equity

$$salary = \beta_0 + \beta_1 roe + u$$

Salary in thousands of dollars

Average return on equity of the CEO's firm

Fitted regression

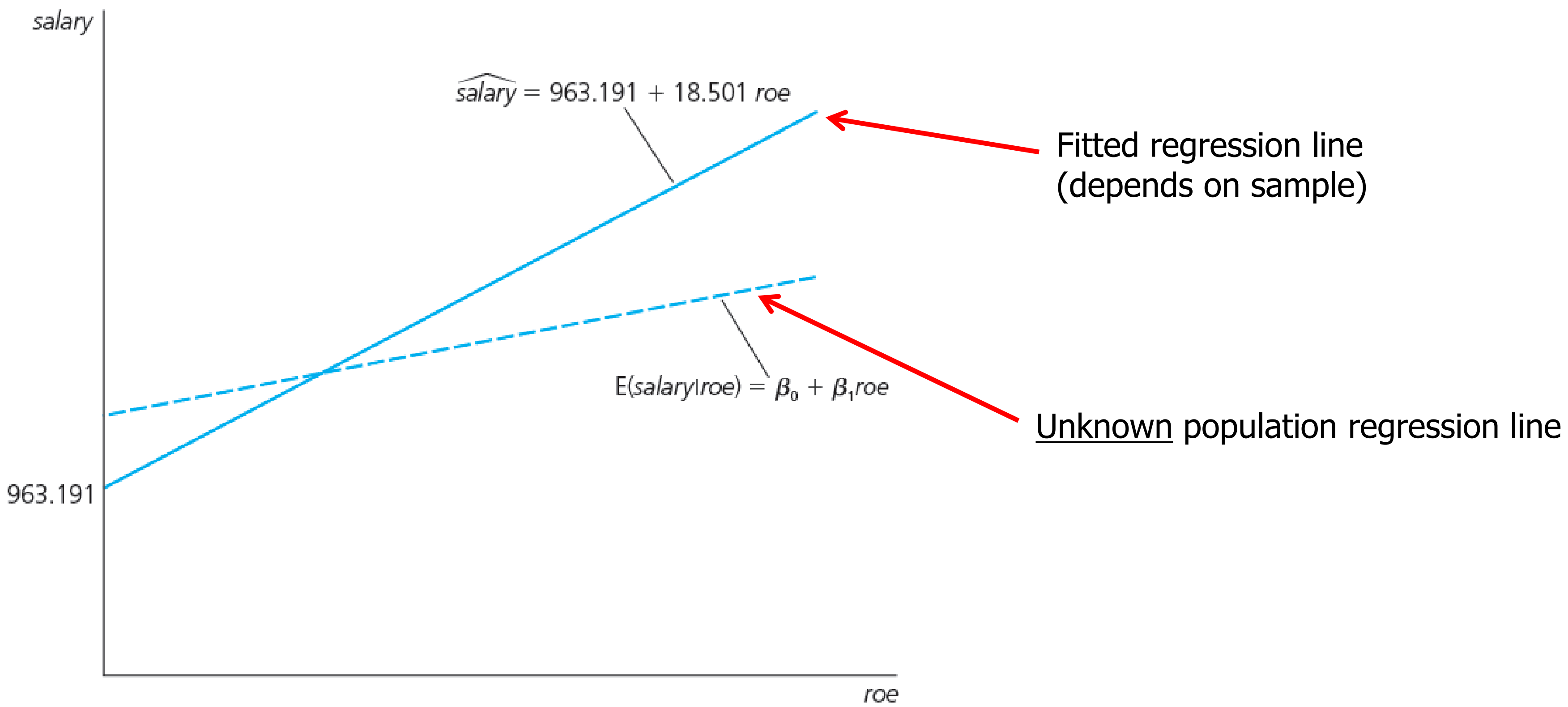
$$\widehat{salary} = 963.191 + 18.501\ roe$$

Intercept

If the return on equity increases by 1 percent, then salary is predicted to change by $18,501

Causal interpretation?

Deriving OLS Estimates



Deriving OLS Estimates

Wage and education $$wage = \beta_0 + \beta_1 educ + u$$

Hourly wage in dollars

Years of education

Fitted regression

$$\widehat{wage} = -0.90 + 0.54 \ educ$$

Intercept

In the sample, one more year of education was associated with an increase in hourly wage by $0.54

Causal interpretation?

Deriving OLS Estimates

Voting outcomes and campaign expenditures (two parties)

$$voteA = \beta_0 + \beta_1 shareA + u$$

Percentage of vote for candidate A

Percentage of campaign expenditures candidate A

Fitted regression

$$\widehat{voteA} = 26.81 + 0.464 \; shareA$$

Intercept

If candidate A's share of spending increases by one percentage point, he or she receives 0.464 percentage points more of the total vote

Causal interpretation?

3.Properties of OLS on any sample of data

Properties of OLS on any sample of data

Fitted values and residuals

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Fitted or predicted values

$$\hat{u}_i = y_i - \hat{y}_i$$

Deviations from regression line (= residuals)

Algebraic properties of OLS regression

$$\sum_{i=1}^{n} \hat{u}_i = 0$$

Deviations from regression line sum up to zero

$$\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

Covariance between deviations and regressors is zero

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Sample averages of y and x lie on regression line

Properties of OLS on any sample of data

x_i

TABLE 2.1 Fitted Values and Residuals for the First 15 CEOs

obsno	roe	salary	salaryhat	uhat
1	14.1	1095	1224.058	–129.0581
2	10.9	1001	1164.854	–163.8542
3	23.5	1122	1397.969	–275.9692
4	5.9	578	1072.348	–494.3484
5	13.8	1368	1218.508	149.4923
6	20.0	1145	1333.215	–188.2151
7	16.4	1078	1266.611	–188.6108
8	16.3	1094	1264.761	–170.7606
9	10.5	1237	1157.454	79.54626
10	26.3	833	1449.773	–616.7726
11	25.9	567	1442.372	–875.3721
12	26.8	933	1459.023	–526.0231
13	14.8	1339	1237.009	101.9911
14	22.3	937	1375.768	–438.7678
15	56.3	2011	2004.808	6.191895

y_i

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

$$\widehat{u}_i = y_i - \widehat{y}_i$$

For example, CEO number 12‘s salary was $526,023 lower than predicted using the the information on his firm‘s return on equity

Properties of OLS on any sample of data

Goodness-of-Fit

"How well does the explanatory variable explain the dependent variable?"

Measures of Variation

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$SSR \equiv \sum_{i=1}^{n} \hat{u}_i^{\,2}$$

<u>Total sum of squares</u>, represents total variation in the dependent variable

<u>Explained sum of squares</u>, represents variation explained by regression

<u>Residual sum of squares</u>, represents variation <u>not</u> explained by regression

Properties of OLS on any sample of data

Decomposition of total variation

$$SST = SSE + SSR$$

Total variation Explained part Unexplained part

Goodness-of-fit measure (R-squared)

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

R-squared measures the fraction of the total variation that is explained by the regression

Proof that SST = SSE + SSR

$$\sum (y_i - \bar{y})^2 = \sum [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2$$

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

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

$$= \text{SSR} + 2\sum \hat{u}_i (\hat{y}_i - \bar{y}) + \text{SSE}$$

and we know that $\sum \hat{u}_i (\hat{y}_i - \bar{y}) = 0$

4.Units of Measurement and Functional Form

Units of Measurement and Functional Form

CEO Salary and return on equity

$$\widehat{salary} = 963.191 + 18.501\ roe$$

The regression explains only 1.3% of the total variation in salaries

Voting outcomes and campaign expenditures

$$n = 209, \quad R^2 = 0.0132$$

$$\widehat{voteA} = 26.81 + 0.464\ shareA$$

Caution: A high R-squared does not necessarily mean that the regression has a causal interpretation!

The regression explains 85.6% of the total variation in election outcomes

$$n = 173, \quad R^2 = 0.856$$

Units of Measurement and Functional Form

Incorporating nonlinearities: Semi-logarithmic form

Regression of log wages on years of education

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

Natural logarithm of wage

This changes the interpretation of the regression coefficient:

$$\beta_1 = \frac{\Delta \log(wage)}{\Delta educ} = \frac{1}{wage} \cdot \frac{\Delta wage}{\Delta educ} = \frac{\frac{\Delta wage}{wage}}{\Delta educ}$$

<u>Percentage</u> change of wage

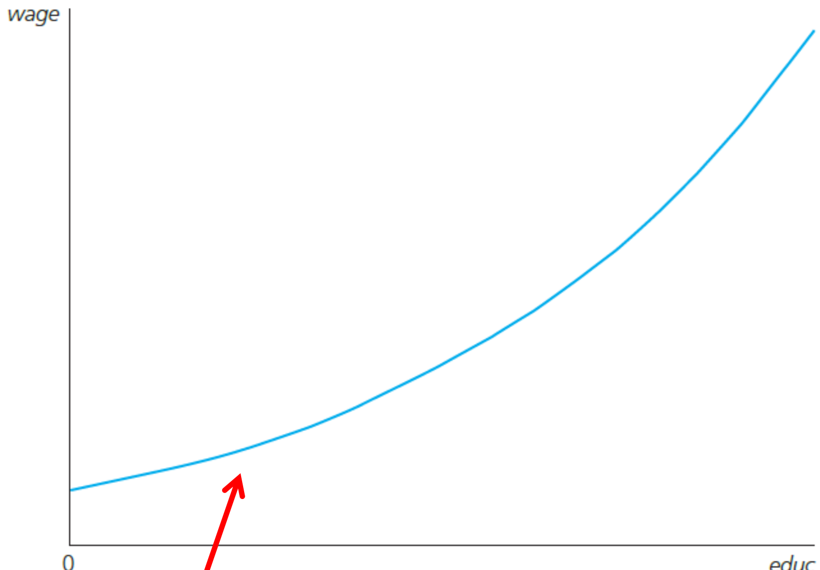
… if years of education are increased by one year

Units of Measurement and Functional Form

Fitted regression

$$\widehat{\log}(wage) = 0.584 + 0.083\ educ$$

The wage increases by 8.3% for
every additional year of education
(= return to another year of education)



Growth rate of wage is 8.3%
per year of education

For example:

$$\frac{\frac{\Delta wage}{wage}}{\Delta educ} = \frac{\frac{+0.83\$}{10\$}}{+1\ year} = 0.083 = +8.3\%$$

Units of Measurement and Functional Form

Incorporating nonlinearities: Log-logarithmic form

CEO salary and firm sales

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + u$$

Natural logarithm of CEO salary

Natural logarithm of his/her firm‘s sales

This changes the interpretation of the regression coefficient:

$$\beta_1 = \frac{\Delta \log(salary)}{\Delta \log(sales)} = \frac{\frac{\Delta salary}{salary}}{\frac{\Delta sales}{sales}}$$

<u>Percentage</u> change of salary

… if sales increase <u>by 1%</u>

<u>Logarithmic changes are always percentage changes</u>

Units of Measurement and Functional Form

CEO salary and firm sales: fitted regression

$$\widehat{\log}(salary) = 4.822 + 0.257 \log(sales)$$

+ 1% sales; + 0.257% salary

For example:

$$\frac{\frac{\Delta salary}{salary}}{\frac{\Delta sales}{sales}} = \frac{\frac{+2{,}570\$}{1{,}000{,}000\$}}{\frac{+10{,}000{,}000\$}{1{,}000{,}000{,}000\$}} = \frac{+0.257\%\ \text{salary}}{+1\%\ \text{sales}} = 0.257$$

The log-log form postulates a constant elasticity model, whereas the semi-log form assumes a semi-elasticity model

5.Expected Values and Variance of the OLS Estimators

Unbiasedness of OLS

Expected values and variances of the OLS estimators

The estimated regression coefficients are random variables because they are calculated from a random sample

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Data is random and depends on particular sample that has been drawn

The question is what the estimators will estimate on average and how large their variability in repeated samples is

$$E(\hat{\beta}_0) = ?, \; E(\hat{\beta}_1) = ? \qquad Var(\hat{\beta}_0) = ?, \; Var(\hat{\beta}_1) = ?$$

Unbiasedness of OLS

Standard assumptions for the linear regression model

Assumption SLR.1 (Linear in parameters)

$$y = \beta_0 + \beta_1 x + u$$

In the population, the relationship between y and x is linear

Assumption SLR.2 (Random sampling)

$$\{(x_i, y_i) : i = 1, \ldots, n\}$$

The data is a random sample drawn from the population

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Each data point therefore follows the population equation

Unbiasedness of OLS

Discussion of random sampling: Wage and education

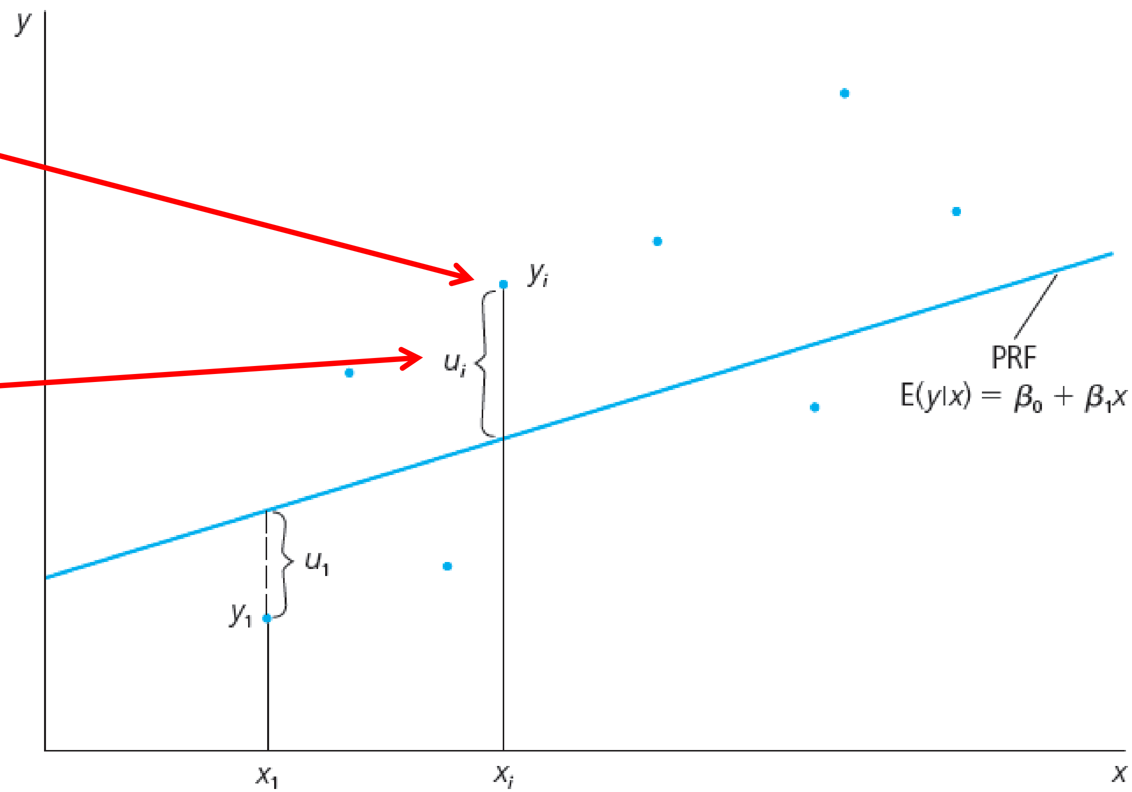
- The population consists, for example, of all workers of country A
- In the population, a linear relationship between wages (or log wages) and years of education holds
- Draw completely randomly a worker from the population
- The wage and the years of education of the worker drawn are random because one does not know beforehand which worker is drawn
- Throw back worker into population and repeat random draw n times
- The wages and years of education of the sampled workers are used to estimate the linear relationship between wages and education

Unbiasedness of OLS

The values drawn for the i-th worker (x_i, y_i)

The implied deviation from the population relationship for the i-th worker:

$$u_i = y_i - \beta_0 - \beta_1 x_i$$



Unbiasedness of OLS

Assumptions for the linear regression model (cont.)

Assumption SLR.3 (Sample variation in the explanatory variable)

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$$

The values of the explanatory variables are not all the same (otherwise it would be impossible to stu-dy how different values of the explanatory variable lead to different values of the dependent variable)

Assumption SLR.4 (Zero conditional mean)

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

The value of the explanatory variable must contain no information about the mean of the unobserved factors

Unbiasedness of OLS

Theorem 2.1 (Unbiasedness of OLS)

$$SLR.1 - SLR.4 \quad \Rightarrow \quad E(\widehat{\beta}_0) = \beta_0,\ E(\widehat{\beta}_1) = \beta_1$$

Interpretation of unbiasedness

- The estimated coefficients may be smaller or larger, depending on the sample that is the result of a random draw
- However, on average, they will be equal to the values that charac-terize the true relationship between y and x in the population
- “On average” means if sampling was repeated, i.e. if drawing the random sample and doing the estimation was repeated many times
- In a given sample, estimates may differ considerably from true values

Unbiasedness of OLS

Assume the population model is linear in parameters as $y = \beta_0 + \beta_1 x + u$

Assume we can use a random sample of size n, $\{(x_i, y_i): i=1, 2, \ldots, n\}$, from the population model. Thus we can write the sample model

$y_i = \beta_0 + \beta_1 x_i + u_i$

Assume $E(u|x) = 0$ and thus $E(u_i|x_i) = 0$

Assume there is variation in the x_i

In order to think about unbiasedness, we need to rewrite our estimator in terms of the population parameter

Start with a simple rewrite of the formula as

Unbiasedness of OLS (cont)

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{s_x^2}, \text{ where}$$

$$s_x^2 \equiv \sum (x_i - \bar{x})^2$$

$$\sum (x_i - \bar{x}) y_i = \sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i) =$$

$$\sum (x_i - \bar{x}) \beta_0 + \sum (x_i - \bar{x}) \beta_1 x_i$$

$$+ \sum (x_i - \bar{x}) u_i =$$

$$\beta_0 \sum (x_i - \bar{x}) + \beta_1 \sum (x_i - \bar{x}) x_i$$

$$+ \sum (x_i - \bar{x}) u_i$$

Unbiasedness of OLS (cont)

$$\sum (x_i - \bar{x}) = 0,$$

$$\sum (x_i - \bar{x})x_i = \sum (x_i - \bar{x})^2$$

so, the numerator can be rewritten as

$\beta_1 s_x^2 + \sum (x_i - \bar{x})u_i$, and thus

$$\hat{\beta}_1 = \beta_1 + \frac{\sum (x_i - \bar{x})u_i}{s_x^2}$$

let $d_i = (x_i - \bar{x})$, so that

$\hat{\beta}_i = \beta_1 + \left(1/s_x^2 \right) \sum d_i u_i$, then

$$E(\hat{\beta}_1) = \beta_1 + \left(1/s_x^2 \right) \sum d_i E(u_i) = \beta_1$$

Unbiasedness Summary

The OLS estimates of β_1 and β_0 are unbiased

Proof of unbiasedness depends on our 4 assumptions – if any assumption fails, then OLS is not necessarily unbiased

Remember unbiasedness is a description of the estimator – in a given sample we may be “near” or “far” from the true parameter

Variances of the OLS estimators

Variances of the OLS estimators

- Depending on the sample, the estimates will be nearer or farther away from the true population values
- How far can we expect our estimates to be away from the true population values on average (= sampling variability)?
- Sampling variability is measured by the estimator's variances

$$Var(\widehat{\beta}_0),\ Var(\widehat{\beta}_1)$$

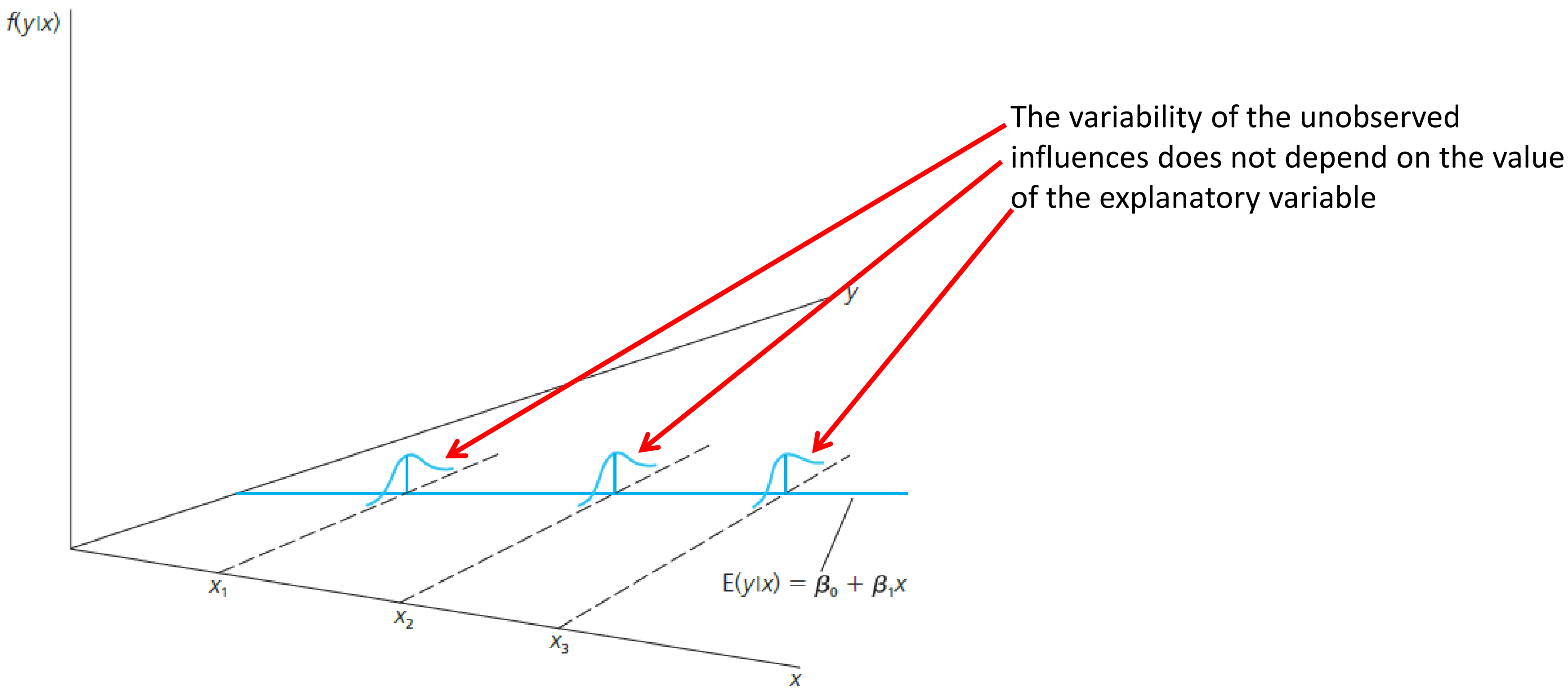
Assumption SLR.5 (Homoskedasticity)

$$Var(u_i|x_i) = \sigma^2$$

The value of the explanatory variable must contain no information about the <u>variability</u> of the unobserved factors

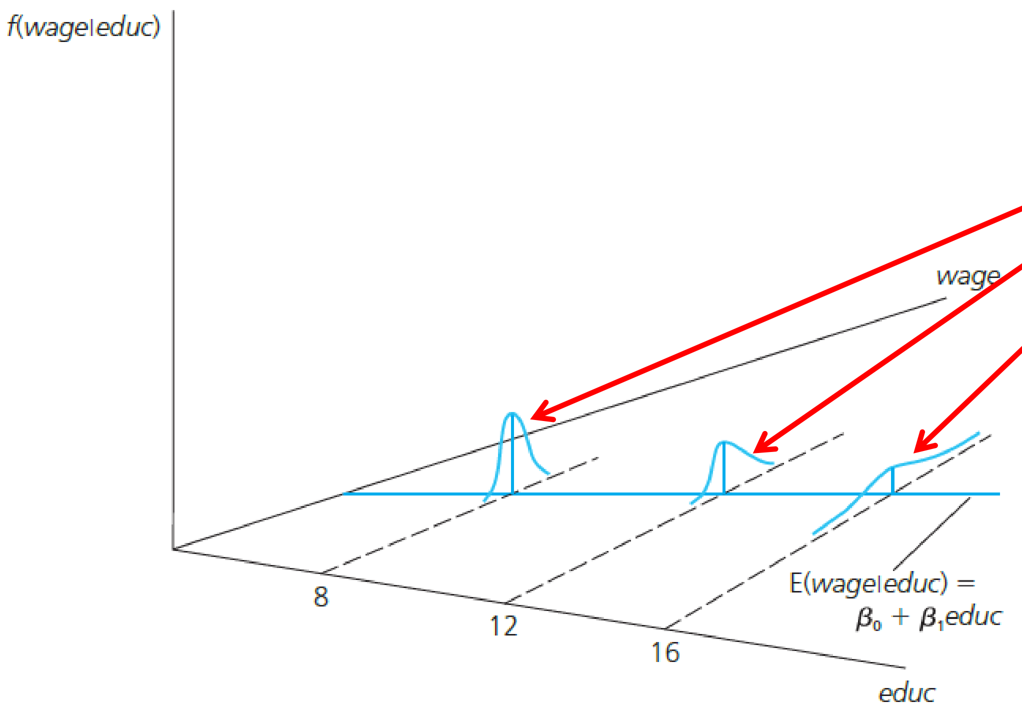
Variances of the OLS estimators

Graphical illustration of homoskedasticity



Variances of the OLS estimators

An example for heteroskedasticity: Wage and education



The variance of the unobserved determinants of wages increases with the level of education

Variances of the OLS estimators

Theorem 2.2 (Variances of the OLS estimators)

Under assumptions SLR.1 – SLR.5:

$$Var(\widehat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{SST_x}$$

$$Var(\widehat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{SST_x}$$

Conclusion:

- The sampling variability of the estimated regression coefficients will be the higher, the larger the variability of the unobserved factors, and the lower, the higher the variation in the explanatory variable

Variance of OLS (cont)

$$Var\left(\hat{\beta}_1\right) = Var\left(\beta_1 + \left(\frac{1}{s_x^2}\right)\sum d_i u_i\right) =$$

$$\left(\frac{1}{s_x^2}\right)^2 Var\left(\sum d_i u_i\right) = \left(\frac{1}{s_x^2}\right)^2 \sum d_i^2 Var\left(u_i\right)$$

$$= \left(\frac{1}{s_x^2}\right)^2 \sum d_i^2 \sigma^2 = \sigma^2 \left(\frac{1}{s_x^2}\right)^2 \sum d_i^2 =$$

$$\sigma^2 \left(\frac{1}{s_x^2}\right)^2 s_x^2 = \frac{\sigma^2}{s_x^2} = Var\left(\hat{\beta}_1\right)$$

Variance of OLS Summary

The larger the error variance, σ^2, the larger the variance of the slope estimate

The larger the variability in the x_i, the smaller the variance of the slope estimate

As a result, a larger sample size should decrease the variance of the slope estimate

Problem that the error variance is unknown

Estimating the error variance

Estimating the error variance

$$Var(u_i|x_i) = \sigma^2 = Var(u_i)$$

The variance of u does not depend on x, i.e. equal to the unconditional variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2$$

One could estimate the variance of the errors by calculating the variance of the residuals in the sample; unfortunately this estimate would be biased

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$$

An unbiased estimate of the error variance can be obtained by substracting the number of estimated regression coefficients
from the number of observations

Estimating the error variance

Theorem 2.3 (Unbiasedness of the error variance)

$$SLR.1 - SLR.5 \quad \Rightarrow \quad E(\hat{\sigma}^2) = \sigma^2$$

Calculation of standard errors for regression coefficients

$$se(\hat{\beta}_1) = \sqrt{\widehat{Var}(\hat{\beta}_1)} = \sqrt{\hat{\sigma}^2 / SST_x}$$

Plug in $\hat{\sigma}^2$ for the unknown σ^2

$$se(\hat{\beta}_0) = \sqrt{\widehat{Var}(\hat{\beta}_0)} = \sqrt{\hat{\sigma}^2 n^{-1} \sum_{i=1}^{n} x_i^2 / SST_x}$$

The estimated standard deviations of the regression coefficients are called “standard errors.”
<u>They measure how precisely the regression coefficients are estimated.</u>

Estimating the Error Variance

We don't know what the error variance, σ^2, is, because we don't observe the errors, u_i

What we observe are the residuals, $\hat{u}_i$

We can use the residuals to form an estimate of the error variance

Estimating the error variance (cont)

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$= \left(\beta_0 + \beta_1 x_i + u_i\right) - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$= u_i - \left(\hat{\beta}_0 - \beta_0\right) - \left(\hat{\beta}_1 - \beta_1\right)$$

Then, an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{(n-2)} \sum \hat{u}_i^2 = SSR/(n-2)$$

Estimating the error variance (cont)

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \text{Standard error of the regression}$$

recall that $\text{sd}\left(\hat{\beta}\right) = \sigma \big/ s_x$

if we substitute $\hat{\sigma}$ for σ then we have

the standard error of $\hat{\beta}_1$,

$$\text{se}\left(\hat{\beta}_1\right) = \hat{\sigma} \Big/ \left(\sum \left(x_i - \bar{x}\right)^2\right)^{1/2}$$