Erasmus
School of
Economics

# Econometrics I

## Lecture 1: Simple regression & multiple regression

Annika Camehl

November - December, 2021

# Today

1. Linear regression
2. Properties of OLS
3. Regression with multiple explanatory variables
4. Properties of multiple regression
5. Adding or deleting variables

# Linear regression

# Basic econometric model

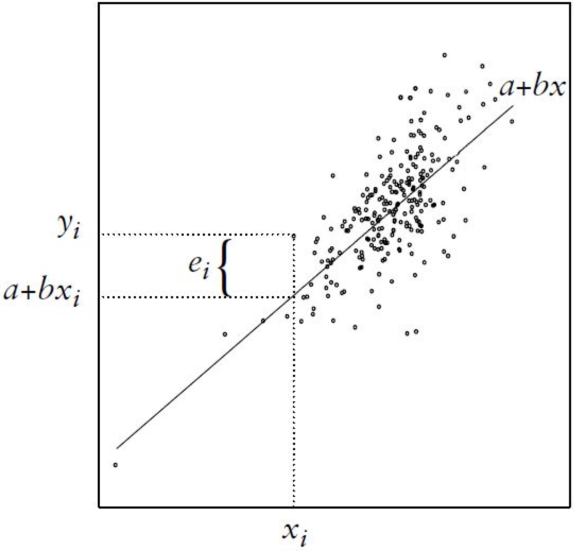
## Simple regression model (review)

$$\underbrace{y_i}_{\text{Data}} = \underbrace{a + bx_i}_{\text{intercept + slope}\times\text{regressor}} + \underbrace{e_i}_{\text{Deviations}}$$

for $i = 1, .., n$, $n$ observations
Problem: given data $(y_i, x_i)$, $i = 1, \ldots, n \rightarrow$ Find best values of $a$ & $b$.
Interpretation of $e_i$:

$$e_i = y_i - a - bx_i$$

# Graphical representation



# Ordinary Least Squares [OLS]

How to find (estimate) $a \& b$ given data?

$$y_i = a + bx_i + e_i$$

Idea: Small values of $e_i$ (close to zero) are preferred
$\rightarrow$ Minimize sum of squared $e_i$ (=OLS)

$$\min_{a,b} S(a,b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Calculating the first derivatives and setting these to zero yields:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ and}$$
$$a = \bar{y} - b\bar{x}$$

# Example 1.1: Bank Wages

# (Statistical) properties

**Question**

How to judge whether OLS is a good method?

# (Statistical) properties

**Question**

How to judge whether OLS is a good method?

Answer depends on the "true" relationship between $y$ and $x$

To analyze properties of OLS we need to

- define the true (unknown) relationship
  also known as the data generating process [DGP]:
  $$y_i = \alpha + \beta x_i + \varepsilon_i$$

where $\alpha$ and $\beta$ are unknown and $\varepsilon_i$ is "pure" random variation

# DGP: Assumptions

A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$

# DGP: Assumptions

A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$

A2 $\varepsilon_i$ are random with $\mathrm{E}[\varepsilon_i] = 0$

# DGP: Assumptions

A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$

A2 $\varepsilon_i$ are random with $\mathrm{E}[\varepsilon_i] = 0$

A3 $\mathrm{E}[\varepsilon_i^2] = \mathrm{Var}(\varepsilon_i) = \sigma^2$ (homoskedasticity)

# DGP: Assumptions

A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$

A2 $\varepsilon_i$ are random with $\mathrm{E}[\varepsilon_i] = 0$

A3 $\mathrm{E}[\varepsilon_i^2] = \mathrm{Var}(\varepsilon_i) = \sigma^2$ (homoskedasticity)

A4 $\mathrm{E}[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$ (no autocorrelation)

# DGP: Assumptions

A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$

A2 $\varepsilon_i$ are random with $\mathrm{E}[\varepsilon_i] = 0$

A3 $\mathrm{E}[\varepsilon_i^2] = \mathrm{Var}(\varepsilon_i) = \sigma^2$ (homoskedasticity)

A4 $\mathrm{E}[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$ (no autocorrelation)

A5 $\alpha, \beta, \sigma^2$ are fixed and unknown

# DGP: Assumptions

A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$

A2 $\varepsilon_i$ are random with $\mathrm{E}[\varepsilon_i] = 0$

A3 $\mathrm{E}[\varepsilon_i^2] = \mathrm{Var}(\varepsilon_i) = \sigma^2$ (homoskedasticity)

A4 $\mathrm{E}[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$ (no autocorrelation)

A5 $\alpha, \beta, \sigma^2$ are fixed and unknown

A6 $y_i = \alpha + \beta x_i + \varepsilon_i$ (linear model)

# DGP: Assumptions

- A1 $x_i$ are fixed (non-random) with $\sum(x_i - \bar{x})^2 > 0$
- A2 $\varepsilon_i$ are random with $\mathrm{E}[\varepsilon_i] = 0$
- A3 $\mathrm{E}[\varepsilon_i^2] = \mathrm{Var}(\varepsilon_i) = \sigma^2$ (homoskedasticity)
- A4 $\mathrm{E}[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$ (no autocorrelation)
- A5 $\alpha, \beta, \sigma^2$ are fixed and unknown
- A6 $y_i = \alpha + \beta x_i + \varepsilon_i$ (linear model)
- A7 $\varepsilon_i \sim N(0, \sigma^2)$

# OLS is BLUE

**Gauss-Markov Theorem**

If A1-A6 is satisfied then OLS is BLUE

BLUE: Best Linear Unbiased Estimator

- U: Unbiased, $E[a] = \alpha$, $E[b] = \beta$
- L: Linear, $a$ and $b$ are linear functions of $y_i$, $i = 1, \ldots, n$
- B: Best, the estimator has smallest variance in it's class (of linear unbiased estimators)
  It holds that:
  for any LUE $\hat{\alpha}$ & $\hat{\beta}$: $\text{Var}(\hat{\beta}) \geq \text{Var}(b)$ and $\text{Var}(\hat{\alpha}) \geq \text{Var}(a)$
  $\rightarrow$ OLS is *efficient*

**Uniformly Minimum Variance Unbiased**

If A1-A<u>7</u> hold then OLS is BUE (=UMVU, Uniformly Minimum Variance Unbiased)

# Simulation exercise illustrating unbiasedness

# Estimating $\sigma^2$

Need to estimate the variance of the true disturbances

$$\varepsilon_i = y_i - \alpha - \beta x_i,$$

which are unobserved, but can be estimated by the residuals

$$e_i = y_i - a - b x_i, \qquad \text{with } a \text{ and } b \text{ the OLS estimates}$$

Residual has mean 0, so sample variance equals

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2$$

However:

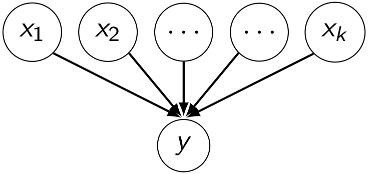
- $\hat{\sigma}^2$ is biased: $\mathrm{E}[\hat{\sigma}^2] \neq \sigma^2$
- Unbiased estimate: $s^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2$ (degrees of freedom correction)
- $s$ is called the *standard error of regression*

# Multiple regression

# Multiple explanatory variables

Before: regression with one explanatory variable: $x \rightarrow y$

Dependent variable usually depends on many variables:

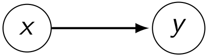



A simple regression on one variable measures so-called "total effect"

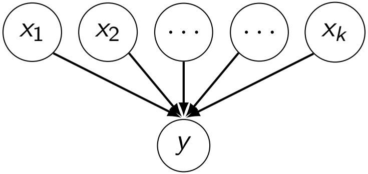
## Example

$\log(\text{Salary}) = \alpha + \beta\text{Gender} + \varepsilon \rightarrow$ OLS gives total gender difference.

# Multiple explanatory variables

Before: regression with one explanatory variable: 

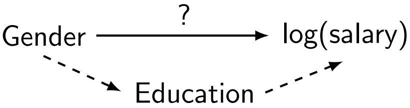

Dependent variable usually depends on many variables:




A simple regression on one variable measures so-called "total effect"

## Example

log(Salary) $= \alpha + \beta$Gender $+ \varepsilon \rightarrow$ OLS gives total gender difference.
However, it may be explained by other things



# Multiple regression model

Model with multiple variables for $i = 1, .., n$

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

Rewrite in matrix notation (note: we define $x_{i1} = 1$)

$$y_i = (x_{i1}, x_{i2}, \ldots, x_{ik}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \varepsilon_i = x_i' \beta + \varepsilon_i$$

Collect all observations

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ & & & \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} \beta + \varepsilon = X\beta + \varepsilon$$

# Parameter estimation: OLS

Least squares follows the same principle:
$\rightarrow$ Choose estimator $b$ of $\beta$ so that $e = y - Xb$ is small

$$\min_b S(b) = \sum_{i=1}^{n} e_i^2 = e'e = (y - Xb)'(y - Xb)$$

# Parameter estimation: OLS

Least squares follows the same principle:
$\rightarrow$ Choose estimator $b$ of $\beta$ so that $e = y - Xb$ is small

$$\min_b S(b) = \sum_{i=1}^{n} e_i^2 = e'e = (y - Xb)'(y - Xb)$$

Solve minimization

- Solve $\frac{\partial S(b)}{\partial b} = 0$
- Above gives $k$ equations in $k$ unknowns:

$$X'Xb = X'y$$

- Solution:

$$b = (X'X)^{-1}X'y$$

# Example 1.1: Bank Wages

# Geometric interpretation of OLS

- $y$ (and $\hat{y}$) and the columns of $X$ are vectors in $\mathbb{R}^n$
- The columns (variables) of $X$ span a ($k$-dimensional) plane in $\mathbb{R}^n$
- $\hat{y}$ is a vector in the plane spanned by the columns of $X$
- $e$ is the difference between $y$ and $\hat{y}$
- OLS: minimize $e'e = ||e||$=(length of $e$)
  Solution:
  - $\hat{y}$ is the projection of $y$ on the $X$-plane
  - $e$ should be orthogonal to $X$-plane: $X'e = 0$

OLS solution:

$$0 = X'e = X'(y - Xb) = X'y - X'Xb \Rightarrow b = (X'X)^{-1}X'y$$

# Two projection matrices

1. Matrix $H$: projects a vector onto the $X$-plane

$$\hat{y} = Xb = X(X'X)^{-1}X'y = Hy$$

   $\rightarrow$ H gives fit (hat-matrix)

2. Matrix $M$: projects $\perp$ (orthogonal) to $X$-plane

$$e = y - \hat{y} = (I - X(X'X)^{-1}X')y = (I - H)y = My$$

   $\rightarrow$ M gives residuals (residual maker)

Properties of projection matrices:

1. Idempotent ($H = HH$ & $M = MM$)
2. Symmetric ($H = H'$ & $M = M'$)
3. $H + M = I$
4. $MH = HM = 0$

# Generalization of assumptions

- A1 $X$ non-random, $\text{rank}(X) = k$ (no multicollinearity)
- A2 $\varepsilon$ random, $\text{E}[\varepsilon] = 0$
- A3 $\text{E}[\varepsilon_i^2] = \text{Var}[\varepsilon_i] = \sigma^2$ (homoskedasticity)
- A4 $\text{E}[\varepsilon_i \varepsilon_j] = 0$ for $i \neq j$ (no correlation)
- A5 $\beta$ and $\sigma^2$ unknown, fixed
- A6 $y = X\beta + \varepsilon$ (linear model)
- A7 $\varepsilon$ has a normal distribution

A3+A4: $\text{Var}(\varepsilon) = \text{E}[\varepsilon\varepsilon'] = \sigma^2 I$

# Properties

If A1-A7 true:

1. $y \sim N(X\beta, \sigma^2 I)$
2. $b \overset{A1}{=} (X'X)^{-1}X'y \overset{A6}{=} (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$
3. $b$ is unbiased

$$\mathrm{E}[b] \overset{A1,A6}{=} \mathrm{E}[\beta + (X'X)^{-1}X'\varepsilon] \overset{A1,A5}{=} \beta + (X'X)^{-1}X'\mathrm{E}[\varepsilon] \overset{A2}{=} \beta$$

4. $\mathrm{Var}(b) = \mathrm{E}[(b - \mathrm{E}[b])(b - \mathrm{E}[b])']$

$$\begin{aligned}
&\overset{A1,2,5,6}{=} \mathrm{E}[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)'] = \mathrm{E}[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\
&\overset{A1}{=} (X'X)^{-1}X'\mathrm{E}[\varepsilon\varepsilon']X(X'X)^{-1} \overset{A3,A4}{=} (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
&\overset{A1}{=} \sigma^2(X'X)^{-1}
\end{aligned}$$

5. Gauss-Markov theorem: $b$ is BLUE (needs A1-A6)
   For any Linear Unbiased Estimator [LUE] $\hat{\beta}$ it holds that:
   $\text{Var}(\hat{\beta}) - \text{Var}(b)$ is positive semidefinite

   Therefore

   - For all $k \times 1$ vectors $c$: $c'[\text{Var}(\hat{\beta}) - \text{Var}(b)]c \geq 0$
   - Especially: $\text{Var}(\hat{\beta}_j) \geq \text{Var}(b_j)$ for all $j = 1, \ldots, k$
     (choose $c$ a unit vector)

# Estimation of $\sigma^2$

- $\sigma^2 = \mathrm{E}[\varepsilon_i^2]$
- unbiased estimator for the variance: $s^2 = \frac{1}{n-k} e'e$

Terminology

- degrees of freedom $n - k$
- $s$: standard error of regression
- $se(b_j) = \sqrt{\widehat{\mathrm{Var}}(b_j)} = s\sqrt{((X'X)^{-1})_{jj}}$: standard error of $b_j$

# Evaluating model quality

Model quality depends on match between fitted values ($\hat{y}_i = x_i' b$) and true values ($y_i$).

Three sums of squares

1. Total Sum of Squares = SST = $\sum_{i=1}^{n} (y_i - \bar{y})^2$
2. Explained Sum of Squares = SSE = $\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2$
3. Residual Sum of Squares = SSR = $\sum_{i=1}^{n} (e_i - \bar{e})^2$

The fit of the model is measured by the Coefficient of Determination $= R^2$

- $R^2 = \frac{\text{SSE}}{\text{SST}}$
- $R^2$ is a relative measure (% explained variance)
- If the model contains a constant (intercept) it holds that
  1. $\bar{\hat{y}} = \bar{y}$ and $\bar{e} = 0$
  2. SST=SSE+SSR
  3. $R^2 = \frac{\text{SST-SSR}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}$

# Adding variables and the $R^2$ in multiple regression

Recall the definition of the $R^2$

$$R^2 = \frac{\text{SSExplained}}{\text{SSTotal}} = \frac{\hat{y}' N \hat{y}}{y' N y} \underset{\substack{\text{if intercept} \\ \text{in model}}}{=} 1 - \frac{e'e}{y' N y}$$

where $\hat{y} = Xb$, $e = y - Xb$, $N$: matrix that gives deviation from mean

Note

- If no. explanatory variables ($k$) $\uparrow$, $R^2 \uparrow$
  (simple reason is that $e'e \downarrow$)
- Adjust $R^2$ for this:

$$\overline{R^2} = \text{adjusted } R^2 = 1 - \frac{e'e/(n-k)}{y'Ny/(n-1)}$$

- You could say that an additional variable is only "worthwhile" if $\overline{R^2}$ increases

Let’s do a small quiz

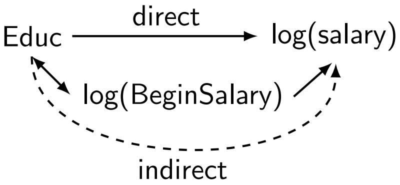
## Adding or deleting variables

# Parameter interpretation in multiple regression

Consider two models:

$$\log \text{Salary} = 1.647 + 0.023 \times \text{Educ} + 0.869 \times \log \text{BeginSalary} + e_{\text{full}}$$

$$\log \text{Salary} = 9.062 + 0.096 \times \text{Educ} + e_{\text{restricted}}$$




- Estimate partial effect (0.023) (direct effect)
  - Keeping other variables fixed
- Estimate "total" effect (0.096) (direct + indirect)
  - Also include side effects through other variables
- We are interested in measuring the impact of a variable ($x$) in isolation
- Compare this to "mathematics" $y = f(x_1, x_2)$ with $x_2 = h(x_1)$

$$\frac{dy}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{dh}{dx_1} = \text{direct} + \text{indirect}$$

# Partial Regression

Given $y = X_1 b_1 + X_2 b_2 + e$, suppose we want the partial effect of $X_1$ on $y$ (keeping $X_2$ fixed).

Partial Regression:

1. Regress $y$ on $X_2$ and (each column of) $X_1$ on $X_2$ $\rightarrow$ Clean $X_1$ and $y$ for their correlation with $X_2$
2. Regress $M_2 y$ on $M_2 X_1$ $\rightarrow$ Look at the relationship between the cleaned variables $\rightarrow$ Call the resulting OLS estimate $b_*$

# Result of Frisch-Waugh

**Theorem: Frisch-Waugh**

$b_*$ is precisely the partial effect $X_1 \to y$ obtained by regressing $y$ on $X_1$ *and* $X_2$. That is

$$b_* = b_1$$

and

$$e_* = e$$

# Omitted variable

Suppose the true model (DGP) is

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

where $X_1$ $(n \times (k-g))$ and $X_2$ $(n \times g)$.
But we use only $X_1$ in the model. Define $b_R$ as our estimate for $\beta_1$

$$\begin{aligned} b_R &= (X_1'X_1)^{-1}X_1'y \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\epsilon \end{aligned}$$

# Consequences of omitted variable

Omitted variable bias:

$$E(b_R) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 = \beta_1 + P\beta_2$$

Smaller variances:

$$\text{Var}(b_1) - \text{Var}(b_R) = \underbrace{P\text{Var}(b_2)P'}_{\text{pos.sem.def}}$$

# Redundant variable

Suppose the true model (DGP) is

$$y = X_1\beta_1 + \varepsilon$$

where $X_1$ $(n \times (k - g))$.
But we include $X_1$ and $X_2$ in the model. The model neglects that $\beta_2 = 0$:

$$b_1 = b_R - (X_1'X_1)^{-1}X_1'X_2b_2$$

# Consequences of redundant variable

Unbiased:

$$E(b_1) = \beta_1$$

and

$$E(b_2) = \beta_2 = 0$$

Inefficient:

$$\text{Var}(b_1) - \text{Var}(b_R) = \underbrace{P\text{Var}(b_2)P'}_{\text{pos.sem.def}}$$

# Summary omitted and redundant variables

| | $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ | $y = X_1\beta_1 + \varepsilon$ |
|---|---|---|
| $y = X_1 b_R + e_R$ | $b_R$ biased,<br>smaller variance than $b_1$ | $b_R$ BLU |
| $y = X_1 b_1 + X_2 b_2 + e$ | $b_1$ unbiased<br>larger variance than $b_R$ | $b_1$ unbiased,<br>not efficient |