# LECTURE 3

Introduction to Econometrics

# INTRODUCTION TO LINEAR REGRESSION ANALYSIS II

October 6, 2017

# Revision: the previous lecture

## Revision: the previous lecture

- (Desired) properties of an estimator:
  - An estimator is **unbiased** if the mean of its distribution is equal to the value of the parameter it is estimating

# REVISION: THE PREVIOUS LECTURE

- (Desired) properties of an estimator:
  - An estimator is **unbiased** if the mean of its distribution is equal to the value of the parameter it is estimating
  - An estimator is **consistent** if it converges to the value of the true parameter as the sample size increases

# Revision: the previous lecture

- (Desired) properties of an estimator:
  - An estimator is **unbiased** if the mean of its distribution is equal to the value of the parameter it is estimating
  - An estimator is **consistent** if it converges to the value of the true parameter as the sample size increases
  - An estimator is **efficient** if the variance of its sampling distribution is the smallest possible

# Revision: the previous lecture

# Revision: the previous lecture

- We explained the principle of OLS estimator: minimizing the sum of squared differences between the observation and the regression line $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

# Revision: the previous lecture

- We explained the principle of OLS estimator: minimizing the sum of squared differences between the observation and the regression line $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

- We found the formulae for the estimates:

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

## Revision: the previous lecture

- We explained that the stochastic error term must be present in a regression equation because of:

# Revision: the previous lecture

- We explained that the stochastic error term must be present in a regression equation because of:
  1. omission of many minor influences (unavailable data)

# Revision: the previous lecture

- We explained that the stochastic error term must be present in a regression equation because of:
  1. omission of many minor influences (unavailable data)
  2. measurement error

# Revision: the previous lecture

- We explained that the stochastic error term must be present in a regression equation because of:
  1. omission of many minor influences (unavailable data)
  2. measurement error
  3. possibly incorrect functional form

# REVISION: THE PREVIOUS LECTURE

- We explained that the stochastic error term must be present in a regression equation because of:
  1. omission of many minor influences (unavailable data)
  2. measurement error
  3. possibly incorrect functional form
  4. stochastic character of unpredictable human behavior

# Revision: the previous lecture

- We explained that the stochastic error term must be present in a regression equation because of:
    1. omission of many minor influences (unavailable data)
    2. measurement error
    3. possibly incorrect functional form
    4. stochastic character of unpredictable human behavior

- Remember that all of these factors are included in the error term and may alter its properties

# Revision: the previous lecture

- We explained that the stochastic error term must be present in a regression equation because of:
  1. omission of many minor influences (unavailable data)
  2. measurement error
  3. possibly incorrect functional form
  4. stochastic character of unpredictable human behavior

- Remember that all of these factors are included in the error term and may alter its properties

- The properties of the error term determine the properties of the estimates

## WARM-UP EXERCISE

- You receive a unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working). Obviously, you are very curious about what is the effect of experience on wages.

# Warm-up exercise

- You receive a unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working). Obviously, you are very curious about what is the effect of experience on wages.
- You run an OLS regression of monthly wage in CZK on the number of years of experience and obtain the following results:

$$\widehat{wage}_i = 14450 + 1135 \cdot exper_i$$

# Warm-up exercise

- You receive a unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working). Obviously, you are very curious about what is the effect of experience on wages.
- You run an OLS regression of monthly wage in CZK on the number of years of experience and obtain the following results:

$$\widehat{wage}_i = 14450 + 1135 \cdot exper_i$$

1. Interpret the meaning of the coefficient of $exper_i$.
2. Use the estimates to determine the average wage of a person with 1, 5, 20, and 40 years of experience.
3. Do the predicted wages seem realistic? Explain your answer.

# On today's lecture

# On today's lecture

- We will derive estimation formulas for multivariate OLS

## ON TODAY'S LECTURE

- We will derive estimation formulas for multivariate OLS
- We will list the assumptions about the error term and the explanatory variables that are required in classical regression models

# On today's lecture

- We will derive estimation formulas for multivariate OLS
- We will list the assumptions about the error term and the explanatory variables that are required in classical regression models
- We will show that under these assumptions, OLS is the best estimator available for regression models

# On today's lecture

- We will derive estimation formulas for multivariate OLS
- We will list the assumptions about the error term and the explanatory variables that are required in classical regression models
- We will show that under these assumptions, OLS is the best estimator available for regression models
- The rest of the course will mostly deal in one way or another with the question what to do when one of the classical assumptions is not met

# On today's lecture

- We will derive estimation formulas for multivariate OLS
- We will list the assumptions about the error term and the explanatory variables that are required in classical regression models
- We will show that under these assumptions, OLS is the best estimator available for regression models
- The rest of the course will mostly deal in one way or another with the question what to do when one of the classical assumptions is not met
- Readings:
  - Studenmund - chapter 4
  - Wooldridge - chapters 5, 8, 9, 12

# Ordinary Least Squares with several explanatory variables

## ORDINARY LEAST SQUARES WITH SEVERAL EXPLANATORY VARIABLES

- Usually, there are more than one explanatory variables in regression models

# Ordinary Least Squares with several explanatory variables

- Usually, there are more than one explanatory variables in regression models
- Multivariate model with $k$ explanatory variables:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

# Ordinary Least Squares with several explanatory variables

- Usually, there are more than one explanatory variables in regression models
- Multivariate model with $k$ explanatory variables:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

- For observations $1, 2, \ldots, n$, we have:

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_k x_{1k} + \varepsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_k x_{2k} + \varepsilon_2 \\
\vdots & \quad \vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk} + \varepsilon_n
\end{aligned}$$

# Matrix notation

- We can write in matrix form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

# Matrix notation

- We can write in matrix form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or in a simplified notation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

# OLS - DERIVATION UNDER MATRIX NOTATION

- We have to find

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\,\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

# OLS - DERIVATION UNDER MATRIX NOTATION

- We have to find

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\,\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

- FOC:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}: \qquad -\left(\mathbf{y}'\mathbf{X}\right)'-\mathbf{X}'\mathbf{y}+\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}}+(\mathbf{X}'\mathbf{X})'\widehat{\boldsymbol{\beta}} &= 0 \\ \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y}\end{aligned}$$

# OLS - derivation under matrix notation

- We have to find

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\,(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\,\mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

- FOC:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}: \qquad -\left(\mathbf{y}'\mathbf{X}\right)'-\mathbf{X}'\mathbf{y}+\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}}+(\mathbf{X}'\mathbf{X})'\widehat{\boldsymbol{\beta}} &= 0 \\ \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y}\end{aligned}$$

- This gives us

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

# Meaning of regression coefficient

# Meaning of regression coefficient

- Consider the multivariate model

$$Q = \beta_0 + \beta_1 P + \beta_2 P_s + \beta_3 Y + \varepsilon$$

estimated as $\widehat{Q} = 31.50 - 0.73P + 0.11P_s + 0.23Y$

$Q$ ... quantity demanded
$P$ ... commodity's price
$P_s$ ... price of substitute
$Y$ ... disposable income

# Meaning of regression coefficient

- Consider the multivariate model

$$Q = \beta_0 + \beta_1 P + \beta_2 P_s + \beta_3 Y + \varepsilon$$

estimated as $\widehat{Q} = 31.50 - 0.73P + 0.11P_s + 0.23Y$

$Q$ ... quantity demanded
$P$ ... commodity's price
$P_s$ ... price of substitute
$Y$ ... disposable income

- Meaning of $\beta_1$ is the impact of a one unit increase in $P$ on the dependent variable $Q$, **holding constant the other included independent variables** $P_s$ and $Y$

# MEANING OF REGRESSION COEFFICIENT

- Consider the multivariate model

$$Q = \beta_0 + \beta_1 P + \beta_2 P_s + \beta_3 Y + \varepsilon$$

estimated as $\widehat{Q} = 31.50 - 0.73P + 0.11P_s + 0.23Y$

| | | | |
|---|---|---|---|
| $Q$ | ... quantity demanded | $P_s$ | ... price of substitute |
| $P$ | ... commodity's price | $Y$ | ... disposable income |

- Meaning of $\beta_1$ is the impact of a one unit increase in $P$ on the dependent variable $Q$, **holding constant the other included independent variables** $P_s$ and $Y$
- When price increases by 1 unit (and price of substitute good and income remain the same), quantity demanded decreases by 0.73 units

# Exercise

- Remember the unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working).
- Because you realize that wages may not be linearly dependent on experience, you add an additional variable ${exper^2}_i$ into your model and you obtain the following results:

# Exercise

- Remember the unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working).
- Because you realize that wages may not be linearly dependent on experience, you add an additional variable ${exper^2}_i$ into your model and you obtain the following results:

$$\widehat{wage}_i = 14450 + 1160 \cdot exper_i - 25 \cdot {exper^2}_i$$

# Exercise

- Remember the unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working).
- Because you realize that wages may not be linearly dependent on experience, you add an additional variable $exper^2{}_i$ into your model and you obtain the following results:

$$\widehat{wage}_i = 14450 + 1160 \cdot exper_i - 25 \cdot exper^2{}_i$$

1. What is the overall impact of increasing the number of years of experience by 1 year?

# Exercise

- Remember the unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working).
- Because you realize that wages may not be linearly dependent on experience, you add an additional variable $exper^2{}_i$ into your model and you obtain the following results:

$$\widehat{wage}_i = 14450 + 1160 \cdot exper_i - 25 \cdot exper^2{}_i$$

1. What is the overall impact of increasing the number of years of experience by 1 year?
2. Use the estimates to determine the average wage of a person with 1, 5, 20, and 40 years of experience.

# Exercise

- Remember the unique dataset that includes wages of all citizens of Brno as well as their experience (number of years spent working).
- Because you realize that wages may not be linearly dependent on experience, you add an additional variable ${exper^2}_i$ into your model and you obtain the following results:

$$\widehat{wage}_i = 14450 + 1160 \cdot exper_i - 25 \cdot {exper^2}_i$$

1. What is the overall impact of increasing the number of years of experience by 1 year?
2. Use the estimates to determine the average wage of a person with 1, 5, 20, and 40 years of experience.
3. Do the predicted wages seem realistic now? Explain your answer.

# The Classical Assumptions

# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term

# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term
2. The error term has a zero population mean

# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term
2. The error term has a zero population mean
3. Observations of the error term are uncorrelated with each other

# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term
2. The error term has a zero population mean
3. Observations of the error term are uncorrelated with each other
4. The error term has a constant variance

# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term
2. The error term has a zero population mean
3. Observations of the error term are uncorrelated with each other
4. The error term has a constant variance
5. All explanatory variables are uncorrelated with the error term

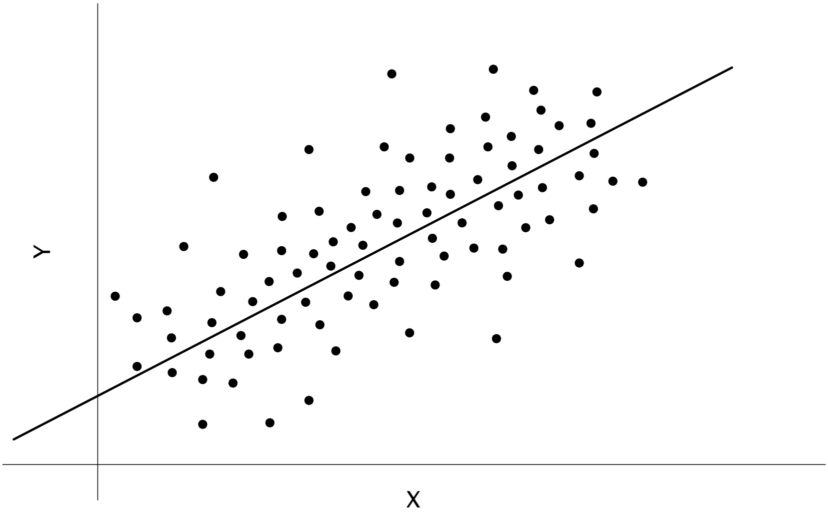
# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term
2. The error term has a zero population mean
3. Observations of the error term are uncorrelated with each other
4. The error term has a constant variance
5. All explanatory variables are uncorrelated with the error term
6. No explanatory variable is a perfect linear function of any other explanatory variable(s)

# The Classical Assumptions

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term
2. The error term has a zero population mean
3. Observations of the error term are uncorrelated with each other
4. The error term has a constant variance
5. All explanatory variables are uncorrelated with the error term
6. No explanatory variable is a perfect linear function of any other explanatory variable(s)
7. The error term is normally distributed

# Graphical representation



# 1. Linearity in coefficients

*The regression model is linear in the coefficients, is correctly specified, and has an additive error term.*

# 1. Linearity in coefficients

*The regression model is linear in the coefficients, is correctly specified, and has an additive error term.*

- Linearity in variables is not required

# 1. Linearity in coefficients

*The regression model is linear in the coefficients, is correctly specified, and has an additive error term.*

- Linearity in variables is not required
- Example: production function $Y = AK^{\beta_1}L^{\beta_2}$ for which we suppose $A = \exp^{\beta_0+\varepsilon}$ can be transformed so that

$$\ln Y = \beta_0 + \beta_1 \ln K + \beta_2 \ln L + \varepsilon$$

and the linearity in coefficients is restored

# 1. Linearity in coefficients

*The regression model is linear in the coefficients, is correctly specified, and has an additive error term.*

- Linearity in variables is not required
- Example: production function $Y = AK^{\beta_1}L^{\beta_2}$ for which we suppose $A = \exp^{\beta_0+\varepsilon}$ can be transformed so that

$$\ln Y = \beta_0 + \beta_1 \ln K + \beta_2 \ln L + \varepsilon$$

  and the linearity in coefficients is restored
- Note that it is the linearity in coefficients that allows us to rewrite the general regression model in matrix form

# Exercise

Which of the following models is/are linear?

- $y = \beta_0 + \beta_1 x + \varepsilon$
- $\ln y = \beta_0 + \beta_1 \ln x + \beta_2 \sqrt{z} + \varepsilon$
- $y = x^{\beta_1} + \varepsilon$

# Exercise

Which of the following models is/are linear?

- $y = \beta_0 + \beta_1 x + \varepsilon$ is a linear model
- $\ln y = \beta_0 + \beta_1 \ln x + \beta_2 \sqrt{z} + \varepsilon$ is a linear model
- $y = x^{\beta_1} + \varepsilon$ is NOT a linear model

# Exercise

Which of the following models is/are linear?

- $y = \beta_0 + \beta_1 x + \varepsilon$ is a linear model
- $\ln y = \beta_0 + \beta_1 \ln x + \beta_2 \sqrt{z} + \varepsilon$ is a linear model
- $y = x^{\beta_1} + \varepsilon$ is NOT a linear model
- Regression models are **linear in parameters**, but they do not need to be linear in variables

# 2. Zero mean of the error term

*The error term has a zero population mean.*

# 2. Zero mean of the error term

*The error term has a zero population mean.*

- Notation: $E[\varepsilon_i] = 0$ or $E[\boldsymbol{\varepsilon}] = \mathbf{0}$

# 2. Zero mean of the error term

*The error term has a zero population mean.*

- Notation: $E[\varepsilon_i] = 0$ or $E[\boldsymbol{\varepsilon}] = \mathbf{0}$
- Idea: observations are distributed around the regression line, the average of deviations is zero

# 2. Zero mean of the error term

*The error term has a zero population mean.*

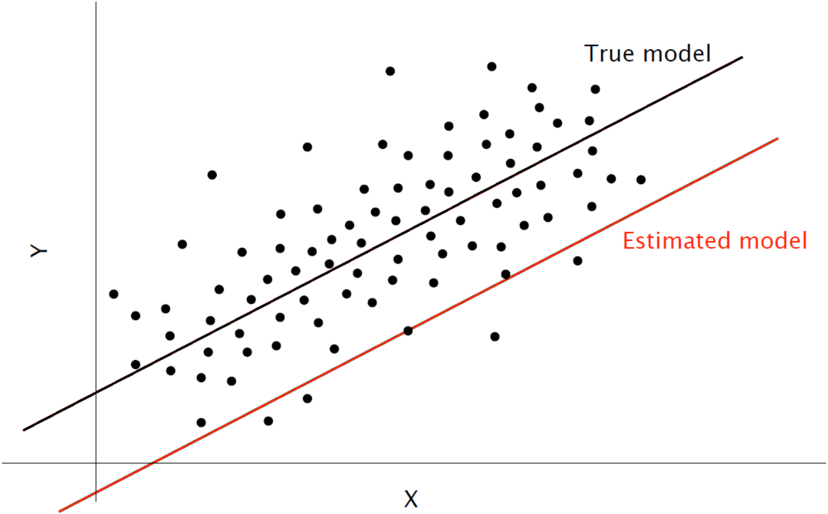
- Notation: $E[\varepsilon_i] = 0$ or $E[\boldsymbol{\varepsilon}] = \mathbf{0}$
- Idea: observations are distributed around the regression line, the average of deviations is zero
- In fact, the mean of $\varepsilon_i$ is forced to be zero by the existence of the intercept ($\beta_0$) in the equation

# 2. Zero mean of the error term

*The error term has a zero population mean.*

- Notation: $E[\varepsilon_i] = 0$ or $E[\boldsymbol{\varepsilon}] = \mathbf{0}$
- Idea: observations are distributed around the regression line, the average of deviations is zero
- In fact, the mean of $\varepsilon_i$ is forced to be zero by the existence of the intercept ($\beta_0$) in the equation
- Hence, this assumption is satisfied as long as there is an **intercept** included in the equation

# Graphical representation



# 3. Errors uncorrelated with each other

*Observations of the error term are uncorrelated with each other.*

# 3. Errors uncorrelated with each other

*Observations of the error term are uncorrelated with each other.*

- If there is a systematic correlation between one observation of the error term and another (serial correlation), it is more difficult for OLS to get precise estimates of the coefficients of the explanatory variables

# 3. Errors uncorrelated with each other

*Observations of the error term are uncorrelated with each other.*

- If there is a systematic correlation between one observation of the error term and another (serial correlation), it is more difficult for OLS to get precise estimates of the coefficients of the explanatory variables
- Technically: the OLS estimate will be consistent, but not efficient

# 3. Errors uncorrelated with each other

*Observations of the error term are uncorrelated with each other.*

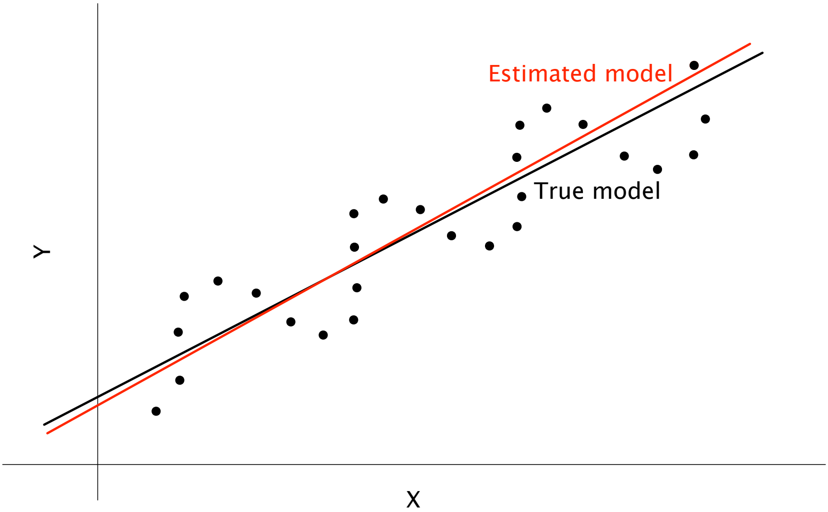
- If there is a systematic correlation between one observation of the error term and another (serial correlation), it is more difficult for OLS to get precise estimates of the coefficients of the explanatory variables
- Technically: the OLS estimate will be consistent, but not efficient
- Often happens in time series data, where a random shock in one time period affects the random shock in another time period

# 3. Errors uncorrelated with each other

*Observations of the error term are uncorrelated with each other.*

- If there is a systematic correlation between one observation of the error term and another (serial correlation), it is more difficult for OLS to get precise estimates of the coefficients of the explanatory variables
- Technically: the OLS estimate will be consistent, but not efficient
- Often happens in time series data, where a random shock in one time period affects the random shock in another time period
- We will solve this problem using Generalized Least Squares estimator

# Graphical representation



# 4. Constant variance of the error term

*The error term has a constant variance.*

# 4. Constant variance of the error term

*The error term has a constant variance.*

- This property is called *homoskedasticity*; if it is not satisfied, we talk about *heteroskedasticity*

# 4. Constant variance of the error term

*The error term has a constant variance.*

- This property is called *homoskedasticity*; if it is not satisfied, we talk about *heteroskedasticity*
- It states that each observation of the error is drawn from a distribution with the same variance and thus varies in the same manner around the regression line

# 4. Constant variance of the error term

*The error term has a constant variance.*

- This property is called *homoskedasticity*; if it is not satisfied, we talk about *heteroskedasticity*
- It states that each observation of the error is drawn from a distribution with the same variance and thus varies in the same manner around the regression line
- If the error term is heteroskedastic, it is more difficult for OLS to get precise estimates of the coefficients of the explanatory variables

# 4. Constant variance of the error term

*The error term has a constant variance.*

- This property is called *homoskedasticity*; if it is not satisfied, we talk about *heteroskedasticity*
- It states that each observation of the error is drawn from a distribution with the same variance and thus varies in the same manner around the regression line
- If the error term is heteroskedastic, it is more difficult for OLS to get precise estimates of the coefficients of the explanatory variables
- Technically: the OLS estimate will be consistent, but not efficient

## 4. Constant variance of the error term

- Heteroskedasticity is often present in cross-sectional data

# 4. Constant variance of the error term

- Heteroskedasticity is often present in cross-sectional data
- Example: Analysis of household consumption patterns

# 4. Constant variance of the error term

- Heteroskedasticity is often present in cross-sectional data
- Example: Analysis of household consumption patterns
  - Variance of the consumption of certain goods might be greater for higher-income households

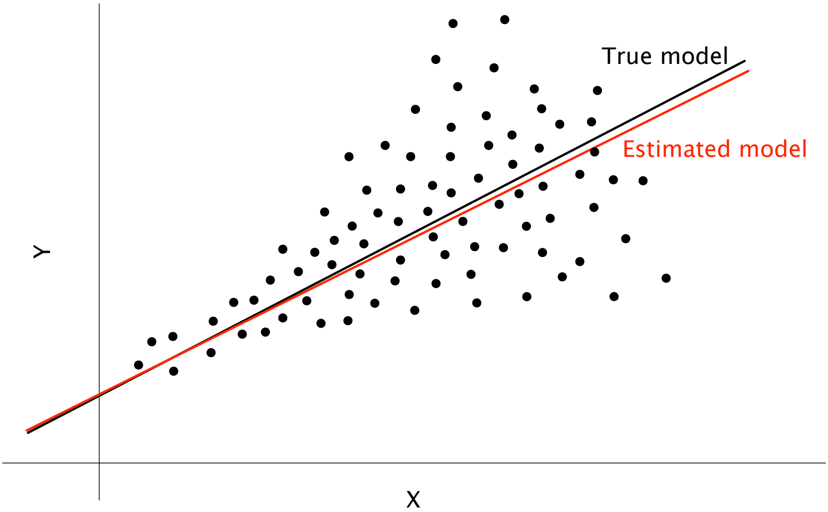
# 4. Constant variance of the error term

- Heteroskedasticity is often present in cross-sectional data
- Example: Analysis of household consumption patterns
  - Variance of the consumption of certain goods might be greater for higher-income households
  - These have more discretionary income than do lower-income households

# 4. Constant variance of the error term

- Heteroskedasticity is often present in cross-sectional data
- Example: Analysis of household consumption patterns
  - Variance of the consumption of certain goods might be greater for higher-income households
  - These have more discretionary income than do lower-income households
- We will solve this problem using Hull-White robust standard errors

# Graphical representation



# 3. No correlation + 4. Homoskedasticity

# 3. No correlation + 4. Homoskedasticity

- Notation:
  - no correlation: $corr(\varepsilon_i \varepsilon_j) \Rightarrow E[\varepsilon_i \varepsilon_j] = 0$ for each $i, j$
  - homoskedasticity: $E[\varepsilon_i^2] = \sigma^2$ for each $i$
- Matrix notation:

$$Var[\boldsymbol{\varepsilon}] = \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}$$

# 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

# 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}' \boldsymbol{\varepsilon}] = \mathbf{0}$

# 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}'\boldsymbol{\varepsilon}] = \mathbf{0}$
- If an explanatory variable and the error term were correlated with each other, the OLS estimates would be likely to attribute to the $x$ some of the variation in $y$ that actually came from the error term

## 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}'\boldsymbol{\varepsilon}] = \mathbf{0}$
- If an explanatory variable and the error term were correlated with each other, the OLS estimates would be likely to attribute to the $x$ some of the variation in $y$ that actually came from the error term
- Example: Analysis of household consumption patterns

# 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}'\boldsymbol{\varepsilon}] = \mathbf{0}$
- If an explanatory variable and the error term were correlated with each other, the OLS estimates would be likely to attribute to the $x$ some of the variation in $y$ that actually came from the error term
- Example: Analysis of household consumption patterns
  - Households with lower incomes may indicate higher consumption (because of shame)
  - Negative correlation between X and error term (measurement error higher for lower incomes)

## 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

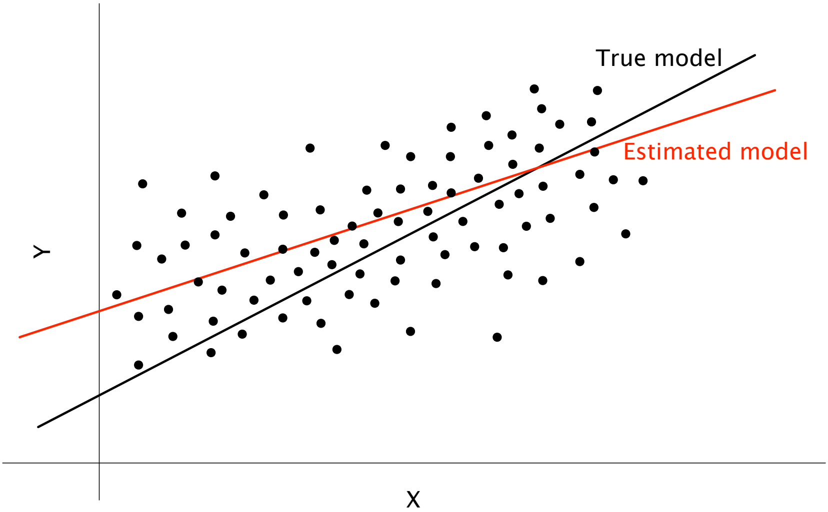
- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}'\boldsymbol{\varepsilon}] = \mathbf{0}$
- If an explanatory variable and the error term were correlated with each other, the OLS estimates would be likely to attribute to the $x$ some of the variation in $y$ that actually came from the error term
- Example: Analysis of household consumption patterns
  - Households with lower incomes may indicate higher consumption (because of shame)
  - Negative correlation between X and error term (measurement error higher for lower incomes)
- Leads to biased and inconsistent estimates

# 5. Variables uncorrelated with the error term

*All explanatory variables are uncorrelated with the error term.*

- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}'\boldsymbol{\varepsilon}] = \mathbf{0}$
- If an explanatory variable and the error term were correlated with each other, the OLS estimates would be likely to attribute to the $x$ some of the variation in $y$ that actually came from the error term
- Example: Analysis of household consumption patterns
  - Households with lower incomes may indicate higher consumption (because of shame)
  - Negative correlation between X and error term (measurement error higher for lower incomes)
- Leads to biased and inconsistent estimates
- We will solve this problem using IV approach

# Graphical representation



# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*
- Multicollinearity can be perfect of imperfect

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*
- Multicollinearity can be perfect of imperfect
- **Perfect multicollinearity**: one explanatory variable is an exact linear function of one or more other explanatory variables

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*
- Multicollinearity can be perfect of imperfect
- **Perfect multicollinearity**: one explanatory variable is an exact linear function of one or more other explanatory variables
  - In this case, the OLS model is incapable to distinguish one variable from the other

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*
- Multicollinearity can be perfect of imperfect
- **Perfect multicollinearity**: one explanatory variable is an exact linear function of one or more other explanatory variables
  - In this case, the OLS model is incapable to distinguish one variable from the other
  - Technical consequence: $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*
- Multicollinearity can be perfect of imperfect
- **Perfect multicollinearity**: one explanatory variable is an exact linear function of one or more other explanatory variables
  - In this case, the OLS model is incapable to distinguish one variable from the other
  - Technical consequence: $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist
  - OLS estimation cannot be conducted

# 6. Linearly independent variables

*No explanatory variable is a perfect linear function of any other explanatory variable(s).*

- If this condition does not hold, we talk about *(multi)collinearity*
- Multicollinearity can be perfect of imperfect
- **Perfect multicollinearity**: one explanatory variable is an exact linear function of one or more other explanatory variables
  - In this case, the OLS model is incapable to distinguish one variable from the other
  - Technical consequence: $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist
  - OLS estimation cannot be conducted
  - Example: we include dummy variables for men and women together with the intercept

# 6. Linearly independent variables

- **Imperfect multicollinearity**:

# 6. LINEARLY INDEPENDENT VARIABLES

- **Imperfect multicollinearity**:
  - There is a linear relationship between the variables, but there is some error in that relationship

# 6. Linearly independent variables

- **Imperfect multicollinearity**:
  - There is a linear relationship between the variables, but there is some error in that relationship
  - Example: we include two variables that proxy for individual health status

# 6. Linearly independent variables

- **Imperfect multicollinearity**:
  - There is a linear relationship between the variables, but there is some error in that relationship
  - Example: we include two variables that proxy for individual health status
- Consequences of multicollinearity:

# 6. Linearly independent variables

- **Imperfect multicollinearity**:
  - There is a linear relationship between the variables, but there is some error in that relationship
  - Example: we include two variables that proxy for individual health status
- Consequences of multicollinearity:
  - Estimated coefficients remain unbiased

# 6. Linearly independent variables

- **Imperfect multicollinearity**:
  - There is a linear relationship between the variables, but there is some error in that relationship
  - Example: we include two variables that proxy for individual health status
- Consequences of multicollinearity:
  - Estimated coefficients remain unbiased
  - But the standard errors of estimates are inflated - making the variable insignificant even though they might be significant

# 6. Linearly independent variables

- **Imperfect multicollinearity**:
  - There is a linear relationship between the variables, but there is some error in that relationship
  - Example: we include two variables that proxy for individual health status
- Consequences of multicollinearity:
  - Estimated coefficients remain unbiased
  - But the standard errors of estimates are inflated - making the variable insignificant even though they might be significant
- Solution: drop one of the variables

# EXERCISE

- Which of the following pairs of independent variables would violate the Assumption of no multicollinearity? (That is, which pairs of variables are perfect linear functions of each other?)

# Exercise

- Which of the following pairs of independent variables would violate the Assumption of no multicollinearity? (That is, which pairs of variables are perfect linear functions of each other?)
  - right shoe size and left shoe size (of students in the class)

# EXERCISE

- Which of the following pairs of independent variables would violate the Assumption of no multicollinearity? (That is, which pairs of variables are perfect linear functions of each other?)
  - right shoe size and left shoe size (of students in the class)
  - consumption and disposable income (in the United States over the last 30 years)

# Exercise

- Which of the following pairs of independent variables would violate the Assumption of no multicollinearity? (That is, which pairs of variables are perfect linear functions of each other?)
  - right shoe size and left shoe size (of students in the class)
  - consumption and disposable income (in the United States over the last 30 years)
  - $X_i$ and $2X_i$

# Exercise

- Which of the following pairs of independent variables would violate the Assumption of no multicollinearity? (That is, which pairs of variables are perfect linear functions of each other?)
  - right shoe size and left shoe size (of students in the class)
  - consumption and disposable income (in the United States over the last 30 years)
  - $X_i$ and $2X_i$
  - $X_i$ and $(X_i)^2$

# 7. Normality of the error term

*The error term is normally distributed.*

## 7. Normality of the error term

*The error term is normally distributed.*

- This assumption is optional, but usually it is invoked

# 7. Normality of the error term

*The error term is normally distributed.*

- This assumption is optional, but usually it is invoked
- Normality of the error term is inherited by the estimate $\widehat{\beta}$

# 7. Normality of the error term

*The error term is normally distributed.*

- This assumption is optional, but usually it is invoked
- Normality of the error term is inherited by the estimate $\widehat{\beta}$
- Knowing the distribution of the estimate allows us to find its confidence intervals and to test hypotheses about coefficients

# Properties of the OLS estimate

## PROPERTIES OF THE OLS ESTIMATE

- OLS estimate is defined by the formula

$$\widehat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \ ,$$

where $\mathbf{y} = \mathbf{X}\beta + \varepsilon$

# Properties of the OLS estimate

- OLS estimate is defined by the formula

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \ ,$$

  where $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

- Hence, it is dependent on the random variable $\boldsymbol{\varepsilon}$ and thus $\widehat{\boldsymbol{\beta}}$ is a random variable itself

# Properties of the OLS estimate

- OLS estimate is defined by the formula

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \ ,$$

  where $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

- Hence, it is dependent on the random variable $\boldsymbol{\varepsilon}$ and thus $\widehat{\boldsymbol{\beta}}$ is a random variable itself

- The properties of $\widehat{\boldsymbol{\beta}}$ are based on the properties of $\boldsymbol{\varepsilon}$

# Gauss-Markov Theorem

# Gauss-Markov Theorem

*Given Classical Assumptions 1. - 6., the OLS estimator of $\boldsymbol{\beta}$ is the minimum variance estimator from among the set of all linear unbiased estimators of $\boldsymbol{\beta}$.*

# Gauss-Markov Theorem

*Given Classical Assumptions 1. - 6., the OLS estimator of* $\boldsymbol{\beta}$ *is the minimum variance estimator from among the set of all linear unbiased estimators of* $\boldsymbol{\beta}$*.*

- Assumption 7., normality, is not needed for this theorem
- The theorem is also known as a stating: *"OLS is BLUE"*, where BLUE stands for "Best Linear Unbiased Estimator"

# Gauss-Markov Theorem

*Given Classical Assumptions 1. - 6., the OLS estimator of* $\boldsymbol{\beta}$ *is the minimum variance estimator from among the set of all linear unbiased estimators of* $\boldsymbol{\beta}$.

- Assumption 7., normality, is not needed for this theorem
- The theorem is also known as a stating: *"OLS is BLUE"*, where BLUE stands for "Best Linear Unbiased Estimator"
- It means that:

# Gauss-Markov Theorem

*Given Classical Assumptions 1. - 6., the OLS estimator of $\boldsymbol{\beta}$ is the minimum variance estimator from among the set of all linear unbiased estimators of $\boldsymbol{\beta}$.*

- Assumption 7., normality, is not needed for this theorem
- The theorem is also known as a stating: *"OLS is BLUE"*, where BLUE stands for "Best Linear Unbiased Estimator"
- It means that:
  - OLS is linear: $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{L}\mathbf{y}$ ,

# Gauss-Markov Theorem

*Given Classical Assumptions 1. - 6., the OLS estimator of $\boldsymbol{\beta}$ is the minimum variance estimator from among the set of all linear unbiased estimators of $\boldsymbol{\beta}$.*

- Assumption 7., normality, is not needed for this theorem
- The theorem is also known as a stating: *"OLS is BLUE"*, where BLUE stands for "Best Linear Unbiased Estimator"
- It means that:
  - OLS is linear: $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{L}\mathbf{y}$ ,
  - OLS is unbiased (see next slide)

# Gauss-Markov Theorem

*Given Classical Assumptions 1. - 6., the OLS estimator of $\boldsymbol{\beta}$ is the minimum variance estimator from among the set of all linear unbiased estimators of $\boldsymbol{\beta}$.*

- Assumption 7., normality, is not needed for this theorem
- The theorem is also known as a stating: *"OLS is BLUE"*, where BLUE stands for "Best Linear Unbiased Estimator"
- It means that:
  - OLS is linear: $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{L}\mathbf{y}$ ,
  - OLS is unbiased (see next slide)
  - OLS has the minimum variance of all unbiased estimators (it is efficient)

# Expected value of the OLS estimate

# Expected value of the OLS estimate

- We show:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) = \\
&= \underbrace{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}
\end{aligned}
$$

# Expected value of the OLS estimate

- We show:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right) = \\ &= \underbrace{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

$$\begin{aligned}E\left[\widehat{\boldsymbol{\beta}}\right] &= E\left[\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] = E\left[\boldsymbol{\beta}\right] + E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] = \\ &= \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\underbrace{E\left[\boldsymbol{\varepsilon}\right]}_{\mathbf{0}} = \boldsymbol{\beta}\end{aligned}$$

# Expected value of the OLS estimate

- We show:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right) = \\
&= \underbrace{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}
\end{aligned}
$$

$$
\begin{aligned}
E\left[\widehat{\boldsymbol{\beta}}\right] &= E\left[\boldsymbol{\beta}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] = E\left[\boldsymbol{\beta}\right] + E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] = \\
&= \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\underbrace{E\left[\boldsymbol{\varepsilon}\right]}_{\mathbf{0}} = \boldsymbol{\beta}
\end{aligned}
$$

- Since $E\left[\widehat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta},$ OLS is unbiased

# Variance of the OLS estimate

# Variance of the OLS estimate

- We show:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

# Variance of the OLS estimate

- We show:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$\begin{aligned} Var\left[\widehat{\boldsymbol{\beta}}\right] &= Var\left[\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] = \\ &= Var(\boldsymbol{\beta}) + Var\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] = \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \cdot Var\left[\boldsymbol{\varepsilon}\right] \cdot \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]' = \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \cdot \underbrace{Var\left[\boldsymbol{\varepsilon}\right]}_{\sigma^2\mathbf{I}} \cdot \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned}$$

# Normality of the OLS estimate

## Normality of the OLS estimate

- When we assume that $\varepsilon_i \sim N(0, \sigma^2)$, we can see that

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

  is also normally distributed (it is a linear combination of normally distributed variables)

## Normality of the OLS estimate

- When we assume that $\varepsilon_i \sim N(0, \sigma^2)$, we can see that

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

  is also normally distributed (it is a linear combination of normally distributed variables)

- Hence, we say that $\widehat{\boldsymbol{\beta}}$ is jointly normal:

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

# Normality of the OLS estimate

- When we assume that $\varepsilon_i \sim N(0, \sigma^2)$, we can see that

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\boldsymbol{\varepsilon}$$

  is also normally distributed (it is a linear combination of normally distributed variables)

- Hence, we say that $\widehat{\boldsymbol{\beta}}$ is jointly normal:

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

- This will help us to test hypotheses about regression coefficients (see next lecture)

# Normality of the OLS estimate

- When we assume that $\varepsilon_i \sim N(0, \sigma^2)$, we can see that

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

  is also normally distributed (it is a linear combination of normally distributed variables)
- Hence, we say that $\widehat{\boldsymbol{\beta}}$ is jointly normal:

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

- This will help us to test hypotheses about regression coefficients (see next lecture)
- Note that the normality of errors is not required for large samples, be-cause $\widehat{\boldsymbol{\beta}}$ is asymptotically normal anyway

# Consistency of the OLS estimate

# Consistency of the OLS estimate

- When no explanatory variables are correlated with the error term (Assumption 5.), OLS estimate is consistent:

# Consistency of the OLS estimate

- When no explanatory variables are correlated with the error term (Assumption 5.), OLS estimate is consistent:

$$E\left[\mathbf{X}'\boldsymbol{\varepsilon}\right] = \mathbf{0} \qquad \Rightarrow \qquad \widehat{\boldsymbol{\beta}} \overset{n\to\infty}{\longrightarrow} \boldsymbol{\beta}$$

# Consistency of the OLS estimate

- When no explanatory variables are correlated with the error term (Assumption 5.), OLS estimate is consistent:

$$E\left[\mathbf{X}'\boldsymbol{\varepsilon}\right] = \mathbf{0} \qquad \Rightarrow \qquad \widehat{\boldsymbol{\beta}} \overset{n\to\infty}{\longrightarrow} \boldsymbol{\beta}$$

- In other words: as the number of observations increases, the estimate converges to the true value of the coefficient

# Consistency of the OLS estimate

- When no explanatory variables are correlated with the error term (Assumption 5.), OLS estimate is consistent:

$$E\left[\mathbf{X}'\varepsilon\right] = \mathbf{0} \qquad \Rightarrow \qquad \widehat{\boldsymbol{\beta}} \overset{n\to\infty}{\longrightarrow} \boldsymbol{\beta}$$

- In other words: as the number of observations increases, the estimate converges to the true value of the coefficient

- **Consistency is the most important property of any estimate!!!**

# CONSISTENCY OF THE OLS ESTIMATE

- As long as the OLS estimate of $\widehat{\beta}$ is consistent, the residuals are consistent estimates of the error term

# Consistency of the OLS estimate

- As long as the OLS estimate of $\widehat{\beta}$ is consistent, the residuals are consistent estimates of the error term
- If we have consistent estimates of the error term, we can test if it satisfies the classical assumptions

# Consistency of the OLS estimate

- As long as the OLS estimate of $\widehat{\beta}$ is consistent, the residuals are consistent estimates of the error term
- If we have consistent estimates of the error term, we can test if it satisfies the classical assumptions
- Moreover, possible deviations from the classical model can be corrected

# Consistency of the OLS estimate

- As long as the OLS estimate of $\widehat{\beta}$ is consistent, the residuals are consistent estimates of the error term
- If we have consistent estimates of the error term, we can test if it satisfies the classical assumptions
- Moreover, possible deviations from the classical model can be corrected
- As a consequence, the assumption of zero correlation between explanatory variables and the error term

$$E\left[\mathbf{X}'\boldsymbol{\varepsilon}\right] = \mathbf{0}$$

is the most important one to satisfy in regression models

# Summary

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

## SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:

# SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate: $$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent

# SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation

## SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate: $$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation
  - no correlation between error term and explanatory variables

# Summary

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation
  - no correlation between error term and explanatory variables
- We showed that if these assumptions hold, OLS estimate is

# SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation
  - no correlation between error term and explanatory variables
- We showed that if these assumptions hold, OLS estimate is
  - consistent (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)

# SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation
  - no correlation between error term and explanatory variables
- We showed that if these assumptions hold, OLS estimate is
  - consistent (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)
  - unbiased (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)

# Summary

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation
  - no correlation between error term and explanatory variables
- We showed that if these assumptions hold, OLS estimate is
  - consistent (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)
  - unbiased (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)
  - efficient (if homoskedasticity and no autocorrelation of $\boldsymbol{\varepsilon}$)

# Summary

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$
- We listed the classical assumptions of regression models:
  - model linear in parameters, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance, no serial autocorrelation
  - no correlation between error term and explanatory variables
- We showed that if these assumptions hold, OLS estimate is
  - consistent (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)
  - unbiased (if no correlation between $\mathbf{X}$ and $\boldsymbol{\varepsilon}$)
  - efficient (if homoskedasticity and no autocorrelation of $\boldsymbol{\varepsilon}$)
  - normally distributed (if $\boldsymbol{\varepsilon}$ normally distributed)