LECTURE 3

Introduction to Econometrics

INTRODUCTION TO LINEAR REGRESSION ANALYSIS II

Dali Laxton

October 8, 2021

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

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

- We found the formulae for the estimates:

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

REVISION: THE PREVIOUS LECTURE

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

- 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

- 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:

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

or in a simplified notation:

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

OLS - DERIVATION UNDER MATRIX NOTATION(OPTIONAL)

- We have to find

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

- FOC:

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

- This gives us

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

Meaning of regression coefficient

- Consider the multivariate model

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

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

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

- Meaning of β_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 a 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:

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

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

1. Linearity: the regression model is linear in the parameters (coefficients)
2. Random sampling: the data is a random sample drawn from the population and each data point follows the population equation
3. No perfect collinearity: the values of explanatory variables are not all the same and no explanatory variable is a perfect linear function of any other explanatory variable(s)
4. Zero conditional mean: values of explanatory variables must contain no information about the mean of the unobserved factors - explanatory variables are uncorrelated with the error term
5. Homoskedasticity: the error term has a constant variance
6. Normality of the error term: the error term is normally distributed

1. Linearity in parameters

The regression model is linear in coefficients.

- 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
- Regression models are **linear in parameters**, but they do not need to be linear in variables

2. Random Sampling

The data is a random sample drawn from the population and each data point follows the population equation.

- Discussion during last class

3. No Perfect Collinearity

The values of explanatory variables are not all the same and 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 or 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
  - OLS estimation cannot be conducted
  - Example: we include dummy variables for men and women together with the intercept

3. No Perfect Collinearity

- **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?)
  - 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 $(X_i)^2$

4. Before Zero Conditional Mean

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

- On average, we make no"mistakes"

- This assumption is satisfied as long as there is an **intercept** included in the equation

4. Zero Conditional Mean

All explanatory variables are uncorrelated with the error term.

- Notation: $E[x_i \varepsilon_i] = 0$ or $E[\mathbf{X}^j \boldsymbol{\varepsilon}] = \mathbf{0}$
- If an explanatory variable and the error term were correlated with each other, the OLS estimates would be likely to attribute some of the variation in y to the x when it actually came from the error term
- Example: Impact of skipping classes on exam scores:

$$score = \beta_0 + \beta_1 skipped + u,$$

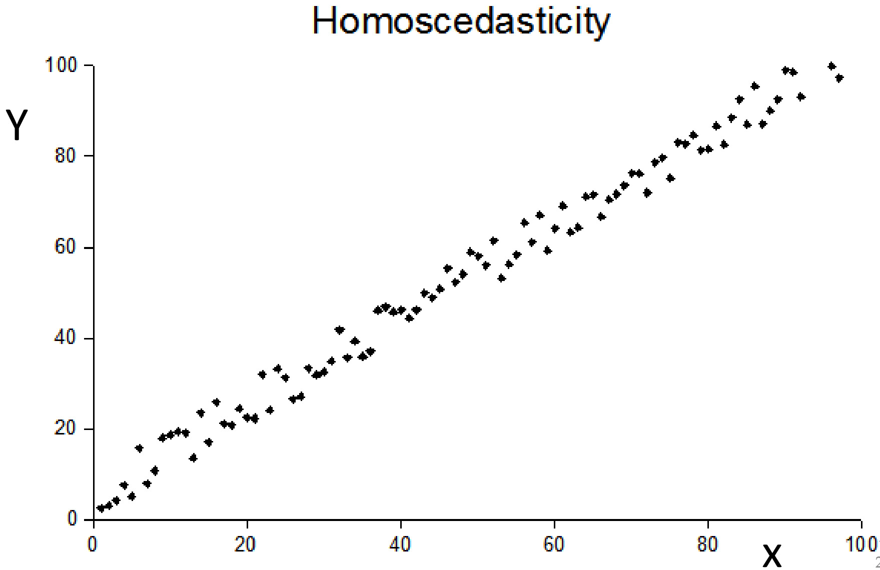
  Motivated students are less likely to skip classes → negative correlation between *skipped* and error term
- Leads to biased and inconsistent estimates
- We will solve this problem using IV approach

5. HOMOSKEDASTICITY

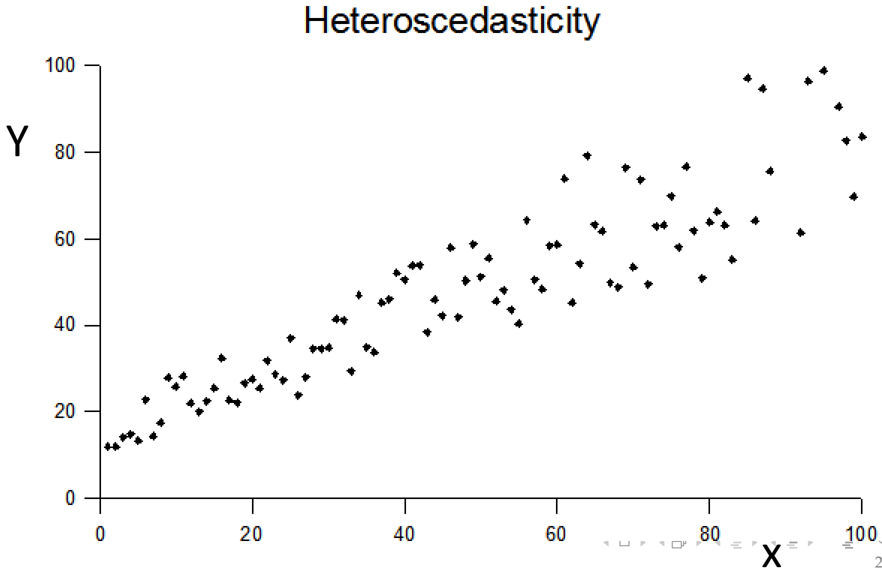
The error term has a constant variance - $Var(\varepsilon_i|X_i) = \sigma^2$

- 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

5. HOMOSKEDASTICITY - GRAPHICAL REPRESENTATION



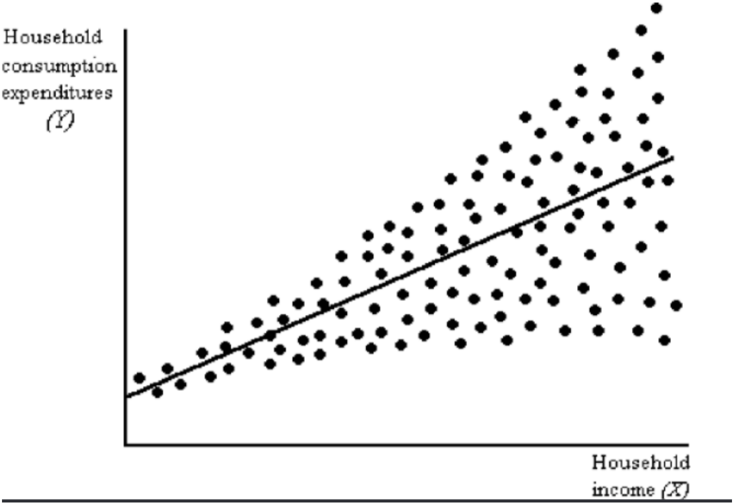
Graphical representation



5. HOMOSKEDASTICITY

- 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



6. Normality of the error term

The error term is normally distributed.

- This is an empirical question
- Normality of the error term is inherited by the estimate $\hat{\beta}$
- Knowing the distribution of the estimate allows us to find its confidence intervals and to test hypotheses about coefficients

Properties of the OLS estimate

- OLS estimate is defined by the formula

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

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

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

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

EXPECTED VALUE OF THE OLS ESTIMATOR

- Under the assumptions 1-4, OLS is **unbiased**: $E(\hat{\beta}) = \beta$
- The estimated coefficients may be smaller or larger, depending on the sample

- However, on average, they will be equal to the true parameters

- NOTE: in a given sample, estimates may differ considerably from true values

VARIANCE OF THE OLS ESTIMATOR

- Under the assumptions1-5 , OLS is **efficient** :

$$Var\left[\widehat{\beta}\right] = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

- The error variance (σ^2): increases the variance of an estimator
- The variation in explanatory variable reduces the variance of the estimator

Gauss-Markov Theorem

Under the assumptions 1 - 5, the OLS estimator of $\boldsymbol{\beta}$ *is the best linear unbiased estimator (BLUE) of the regression coefficients*

- NOTE: assumption 6, normality, is not needed for this theorem
- Gauss-Markov Theorem 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 (OPTIONAL)

- We show:

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

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

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

VARIANCE OF THE OLS ESTIMATE (OPTIONAL)

- We show:

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

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

Normality of the OLS estimate

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

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

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

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

$$\widehat{\beta}_j \sim N\left(\beta_j, Var(\hat{\beta}_j)\right)$$

- 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, because $\widehat{\beta}$ is asymptotically normal anyway

Consistency of the OLS estimate

- When no explanatory variables are correlated with the error term (Assumption 4), OLS estimate is consistent:

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

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

Consistency of the OLS estimate

- As long as the OLS estimate of $\hat{\boldsymbol{\beta}}$ 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

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

is the most important one to satisfy in regression models

SUMMARY

- We expressed the multivariate OLS model in matrix notation $\mathbf{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ and we found the formula of the estimate:
- We listed the classical assumptions of regression models:
  - model linear in parameters, random sampling, explanatory variables linearly independent
  - (normally distributed) error term with zero mean and constant variance
  - 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)