Econometrics

Practical Session 7

Multiple Regression Analysis and Inference

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics

Theoretical Wrap-up

What have we learned so far?

What have we learned so far?

- A model with a single **regressor**:

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

What have we learned so far?

- A model with a single **regressor**:

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

- Ordinary Least Squares

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right)^2$$

What have we learned so far?

- A model with a single **regressor**:

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

- Ordinary Least Squares

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right)^2$$

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

Could the method work with multiple variables?

- A model with k variables and $p = k + 1$ parameters:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i$$

Could the method work with multiple variables?

- A model with k variables and $p = k + 1$ parameters:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i$$

- Ordinary Least Squares

$$\min_{\hat{\beta}_0,\ldots,\hat{\beta}_k} \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_k x_{ki}\right)\right)^2$$

Could the method work with multiple variables?

- A model with k variables and $p = k + 1$ parameters:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + u_i$$

- Ordinary Least Squares

$$\min_{\hat{\beta}_0,\ldots,\hat{\beta}_k} \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_k x_{ki}\right)\right)^2$$

- Or, more concisely and conveniently:

$$\min_{\hat{\boldsymbol{\beta}}} \quad \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}, \qquad \hat{\boldsymbol{u}}' = \begin{pmatrix} \hat{u}_1 & \hat{u}_2 & \ldots & \hat{u}_N \end{pmatrix}$$

Could the method work with multiple variables?

- Then:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

with:

$$\underset{(p\times 1)}{\hat{\boldsymbol{\beta}}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \quad \underset{(N\times p)}{\boldsymbol{X}} = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & x_{2N} & \dots & x_{kN} \end{pmatrix} \quad \underset{(N\times 1)}{\boldsymbol{y}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

- And the model could be written as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$

How to assess the model fit now?

How to assess the model fit now?

- **Including variables** in the model → reduces SSR → $R^2 = 1 - SSR/TSS$ **increases mechanically**

How to assess the model fit now?

- **Including variables** in the model → reduces SSR → $R^2 = 1 - \text{SSR}/\text{TSS}$ **increases mechanically**
- The **Adjusted** R^2 statistic penalizes the inclusion of additional variables that do not significantly improve the model fit:

$$\bar{R}^2 = 1 - \frac{n-1}{n-p}\frac{\text{SSR}}{\text{TSS}} = 1 - \frac{s_{\hat{u}}^2}{s_y^2} \quad \Rightarrow \quad \bar{R}^2 \in (-\infty, R^2)$$

How to assess the model fit now?

- **Including variables** in the model → reduces SSR → $R^2 = 1 - \text{SSR}/\text{TSS}$ **increases mechanically**
- The **Adjusted** R^2 statistic penalizes the inclusion of additional variables that do not significantly improve the model fit:

$$\bar{R}^2 = 1 - \frac{n-1}{n-p}\frac{\text{SSR}}{\text{TSS}} = 1 - \frac{s_{\hat{u}}^2}{s_y^2} \quad \Rightarrow \quad \bar{R}^2 \in \left(-\infty, R^2\right)$$

- The **Standard Error of the Regression (SER)** is identical to before:

$$\text{SER} = \sqrt{\frac{\text{SSR}}{n-p}}$$

Does the OLS method always work?

Does the OLS method always work?

- Yes... But it will not always produce **unbiased estimates**: $\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$

Does the OLS method always work?

- Yes... But it will not always produce **unbiased estimates**: $\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$
- That requires some assumptions

Does the OLS method always work?

- Yes... But it will not always produce **unbiased estimates**: $\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$
- That requires some assumptions
  1. **Zero conditional mean of the errors**: $\mathbb{E}[u_i | x_{1i}, x_{2i}, \dots, x_{ki}] = 0 \Rightarrow \text{corr}(u_i, \boldsymbol{x}_i) = 0$ and $\mathbb{E}[u_i] = 0, \forall i = 1, \dots, N$

Does the OLS method always work?

- Yes... But it will not always produce **unbiased estimates**: $\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$
- That requires some assumptions

1. **Zero conditional mean of the errors**: $\mathbb{E}[u_i | x_{1i}, x_{2i}, \ldots, x_{ki}] = 0 \Rightarrow \text{corr}(u_i, \boldsymbol{x}_i) = 0$ and $\mathbb{E}[u_i] = 0, \forall i = 1, \ldots, N$
2. **Random sampling**: $(x_{1i}, \ldots, x_{ki}, y_i), i = 1, \ldots, N$ are independent and identically distributed

Does the OLS method always work?

- Yes... But it will not always produce **unbiased estimates**: $\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$
- That requires some assumptions

1. **Zero conditional mean of the errors**: $\mathbb{E}[u_i | x_{1i}, x_{2i}, \ldots, x_{ki}] = 0 \Rightarrow \text{corr}(u_i, \boldsymbol{x}_i) = 0$ and $\mathbb{E}[u_i] = 0, \forall i = 1, \ldots, N$
2. **Random sampling**: $(x_{1i}, \ldots, x_{ki}, y_i), i = 1, \ldots, N$ are independent and identically distributed
3. **Large outliers are unlikely**: $0 < \mathbb{E}\left[x_{ji}^4\right] < \infty, \forall j = 1, \ldots, k; i = 1, \ldots, N$ and $0 < \mathbb{E}\left[y_i^4\right] < \infty, \forall i = 1, \ldots, N$

Does the OLS method always work?

- Yes... But it will not always produce **unbiased estimates**: $\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$
- That requires some assumptions

1. **Zero conditional mean of the errors**: $\mathbb{E}[u_i | x_{1i}, x_{2i}, ..., x_{ki}] = 0 \Rightarrow \text{corr}(u_i, \boldsymbol{x}_i) = 0$ and $\mathbb{E}[u_i] = 0, \forall i = 1, ..., N$
2. **Random sampling**: $(x_{1i}, ..., x_{ki}, y_i), i = 1, ..., N$ are independent and identically distributed
3. **Large outliers are unlikely**: $0 < \mathbb{E}\left[x_{ji}^4\right] < \infty, \forall j = 1, ..., k; i = 1, ..., N$ and $0 < \mathbb{E}\left[y_i^4\right] < \infty, \forall i = 1, ..., N$
4. **No perfect multicollinearity**: the columns of $\boldsymbol{X}$ must be linearly independent, $x_k \neq \lambda x_{-k}$

And are the OLS estimates reliable?

- Our estimates $\hat{\beta}$ **are sample-dependent** → any other sample would produce different values

And are the OLS estimates reliable?

- Our estimates $\hat{\beta}$ **are sample-dependent** → any other sample would produce different values
- The estimates are random variables → **sampling probability distribution**

$$\hat{\beta}_j \doteq \mathcal{N}\left(\beta_j, s_{\hat{\beta}_j}\right) \quad \Rightarrow \quad \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \doteq \mathcal{N}(0, 1)$$

And are the OLS estimates reliable?

- Our estimates $\hat{\beta}$ **are sample-dependent** → any other sample would produce different values
- The estimates are random variables → **sampling probability distribution**

$$\hat{\beta}_j \dot{\sim} \mathcal{N}\left(\beta_j, s_{\hat{\beta}_j}\right) \quad \Rightarrow \quad \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \dot{\sim} \mathcal{N}(0,1)$$

- This allows us to do **inference** → *given the sample-dependent $\hat{\beta}_j$ we got, how likely is it that the real value of β_j in the population is ...?*

And are the OLS estimates reliable?

- Our estimates $\hat{\beta}$ **are sample-dependent** → any other sample would produce different values
- The estimates are random variables → **sampling probability distribution**

$$\hat{\beta}_j \doteq \mathcal{N}\left(\beta_j, s_{\hat{\beta}_j}\right) \quad \Rightarrow \quad \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \doteq \mathcal{N}(0,1)$$

- This allows us to do **inference** → *given the sample-dependent $\hat{\beta}_j$ we got, how likely is it that the real value of β_j in the population is ...?*
- In particular, how likely is it that $\beta_j = 0$? → **statistical significance**

Problem: how do we determine $s_{\hat{\beta}_j}$?

Problem: how do we determine $s_{\hat{\beta}_j}$?

- Remember that:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Problem: how do we determine $s_{\hat{\beta}_j}$?

- Remember that:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- Assuming **homoskedasticity**: $\text{var}(u_i) = \sigma^2, \forall i = 1, \dots, N$ (i.e. $\mathbb{E}[\boldsymbol{u}\boldsymbol{u}'] = \sigma^2\boldsymbol{I}$):

$$s^2_{\hat{\boldsymbol{\beta}}} = \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Problem: how do we determine $s_{\hat{\beta}_j}$?

- Remember that:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- Assuming **homoskedasticity**: $\text{var}(u_i) = \sigma^2, \forall i = 1, \ldots, N$ (i.e. $\mathbb{E}[\boldsymbol{u}\boldsymbol{u}'] = \sigma^2\boldsymbol{I}$):

$$s^2_{\hat{\boldsymbol{\beta}}} = \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\right] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- Still... We don't know the **errors variance** σ^2 → use the **residuals variance**:

$$s^2_{\hat{\boldsymbol{u}}} = \frac{\text{SSR}}{N - p} = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{N - p}$$

Given $\hat{\beta}_j$, what are the most likely values of β_j?

Given $\hat{\beta}_j$, what are the most likely values of β_j?

- First, we need to determine how *confident* we want our answer to be → **level of significance α**

Given $\hat{\beta}_j$, what are the most likely values of β_j?

- First, we need to determine how *confident* we want our answer to be → **level of significance α**
- Now, we know that:

$$P\left[\Phi_{\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \leq \Phi_{1-\alpha/2}\right] = 1 - \alpha \Leftrightarrow$$

$$P\left[\hat{\beta}_j - s_{\hat{\beta}_j} \cdot \Phi_{\alpha/2} \geq \beta_j \geq \hat{\beta}_j - s_{\hat{\beta}_j} \cdot \Phi_{1-\alpha/2}\right] = 1 - \alpha$$

Given $\hat{\beta}_j$, what are the most likely values of β_j?

- First, we need to determine how *confident* we want our answer to be → **level of significance α**
- Now, we know that:

$$P\left[\Phi_{\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \leq \Phi_{1-\alpha/2}\right] = 1 - \alpha \Leftrightarrow$$

$$P\left[\hat{\beta}_j - s_{\hat{\beta}_j} \cdot \Phi_{\alpha/2} \geq \beta_j \geq \hat{\beta}_j - s_{\hat{\beta}_j} \cdot \Phi_{1-\alpha/2}\right] = 1 - \alpha$$

- Then, with $(1-\alpha)\%$ of probability, β_j relies in the **confidence interval**:

$$\mathrm{CI}_{\beta_j}^{1-\alpha} = \left\{\hat{\beta}_j \pm s_{\hat{\beta}_j} \cdot \Phi_{1-\alpha/2}\right\}$$

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- First, define again the **level of significance** α

Given $\hat{\beta}_j$, is it likely that $\beta_j = ...$?

- First, define again the **level of significance α**
- Second, define the hypothesis: $H_0 : \beta_j = ...$ → **the null hypothesis**

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- First, define again the **level of significance α**
- Second, define the hypothesis: $H_0 : \beta_j = \ldots$ → **the null hypothesis**
- Third, decide the **alternative hypothesis** $H_1 : \beta_j \neq \ldots$ or $H_1 : \beta_j \gtrless \ldots$

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- First, define again the **level of significance** α
- Second, define the hypothesis: $H_0 : \beta_j = \ldots$ → **the null hypothesis**
- Third, decide the **alternative hypothesis** $H_1 : \beta_j \neq \ldots$ or $H_1 : \beta_j \gtrless \ldots$
- Fourth, compute the t**-statistic**:

$$t = \frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \dot{\sim} \mathcal{N}(0,1) \qquad \longrightarrow \qquad \Phi(\cdot) \text{ denotes } \mathcal{N}(0,1) \text{ c.d.f.}$$

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- Finally, decide:

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- Finally, decide:

  1. **Compare with a critical value** for the level of significance

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- Finally, decide:

  1. **Compare with a critical value** for the level of significance
     - $z = \Phi^{-1}(1 - \alpha/2)$ for **bilateral tests**
     - $z = \Phi^{-1}(1 - \alpha)$ for **unilateral tests**

Given $\hat{\beta}_j$, is it likely that $\beta_j = \dots$?

- Finally, decide:
  1. **Compare with a critical value** for the level of significance
     - $z = \Phi^{-1}(1 - \alpha/2)$ for **bilateral tests**
     - $z = \Phi^{-1}(1 - \alpha)$ for **unilateral tests**
  2. **Compute the p-value** and compare with the level of significance

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- Finally, decide:
  1. **Compare with a critical value** for the level of significance
     - $z = \Phi^{-1}(1 - \alpha/2)$ for **bilateral tests**
     - $z = \Phi^{-1}(1 - \alpha)$ for **unilateral tests**
  2. **Compute the p-value** and compare with the level of significance
     - p-value $= 2 \cdot (1 - \Phi(|t|))$ for **bilateral tests**
     - p-value $= 1 - \Phi(|t|)$ for **unilateral tests**

Given $\hat{\beta}_j$, is it likely that $\beta_j = \ldots$?

- Finally, decide:
  1. **Compare with a critical value** for the level of significance
     - $z = \Phi^{-1}(1 - \alpha/2)$ for **bilateral tests**
     - $z = \Phi^{-1}(1 - \alpha)$ for **unilateral tests**
  2. **Compute the p-value** and compare with the level of significance
     - $p\text{-value} = 2 \cdot (1 - \Phi(|t|))$ for **bilateral tests**
     - $p\text{-value} = 1 - \Phi(|t|)$ for **unilateral tests**
- **Reject the null** hypothesis if either $|t| > z$ or $p\text{-value} < \alpha$

Exercises

Stock & Watson – Exercise 5.1

Suppose a researcher, using data on class size (CS) and average test scores from 50 third-grade classes, estimates the OLS regression:

$$\widehat{\text{TestScore}} = \underset{(23.5)}{640.3} - \underset{(2.02)}{4.93} \cdot \text{CS}, \quad R^2 = 0.11, \quad \text{SER} = 8.7.$$

a) Construct a 95% confidence interval for β_1, the regression slope coefficient.

b) Calculate the p-value for the two-sided test of the null hypothesis $H_0 : \beta_1 = 0$. Do you reject the null hypothesis at the 5% level? At the 1% level?

Stock & Watson – Exercise 5.1

$$\widehat{TestScore} = \underset{(23.5)}{640.3} - \underset{(2.02)}{4.93} \cdot CS, \quad R^2 = 0.11, \quad SER = 8.7.$$

c) Calculate the p-value for the two-sided test of the null hypothesis $H_0 : \beta_1 = -5.0$. Without doing any additional calculations, determine whether -5.0 is contained in the 95% confidence interval for β_1.

d) Construct a 90% confidence interval for β_0.

Stock & Watson – Exercise 5.5

In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to "regular" and "small" classes and given standardized tests at the end of the year. (Regular classes contained approximately 24 students, and small classes contained approximately 15 students.) Suppose, in the population, the standardized tests have a mean score of 925 points and a standard deviation of 75 points. Let *SmallClass* denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise. A regression of *TestScore* on *SmallClass* yields:

$$\widehat{\text{TestScore}} = \underset{(1.6)}{918.0} + \underset{(2.5)}{13.9} \cdot \text{SmallClass}, \quad R^2 = 0.01, \quad \text{SER} = 74.6.$$

Stock & Watson – Exercise 5.5

$$\widehat{TestScore} = \underset{(1.6)}{918.0} + \underset{(2.5)}{13.9} \cdot SmallClass, \quad R^2 = 0.01, \quad SER = 74.6.$$

a) Do small classes improve test scores? By how much? Is the effect large? Explain.

b) Is the estimated effect of class size on test scores statistically significant? Carry out a test at the 5% level.

c) Construct a 99% confidence interval for the effect of *SmallClass* on *TestScore*.

d) Does least squares assumption 1 plausibly hold for this regression? Explain.

Stock & Watson – Exercise 5.7

Suppose (Y_i, X_i) satisfy the least squares assumptions. A random sample of size $n = 250$ is drawn and yields:

$$\hat{Y} = \underset{(3.1)}{5.4} + \underset{(1.5)}{3.2} \cdot X_i, \quad R^2 = 0.26, \quad \text{SER} = 6.2.$$

a) Test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ at the 5% level.

b) Construct a 95% confidence interval for β_1.

c) Suppose you learned that Y_i and X_i were independent. Would you be surprised? Explain.

Stock & Watson – Exercise 5.7

d) Suppose Y_i and X_i are independent and many samples of size n = 250 are drawn, regressions estimated, and a) and b) answered. In what fraction of the samples would H_0 from a) be rejected? In what fraction of samples would the value $\beta_1 = 0$ be included in the confidence interval from b)?