Econometric Methods I & Quantitative and Statistical Methods I
Problem Set 1 Solutions

Barcelona School of Economics, 2025–2026 Academic Year

Due date: October 8th, 12:00 PM (noon)

Group members: AB, CD, EF, GH.

January 15, 2026

General setup

Consider the classical linear regression model discussed in class. We have

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underset{(n\times K)}{\boldsymbol{X}}\underset{(K\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}\,, \tag{1}$$

where the dimensions are shown under the arrays. The model includes a constant (*i.e.*, the first column of $\boldsymbol{X}$ is a vector of ones). Assume that Assumptions 1–4 hold throughout the problem set. Let us define the OLS estimator of $\boldsymbol{\beta}$ as

$$\boldsymbol{b} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\,, \tag{2}$$

and the fitted values $\widehat{\boldsymbol{y}}$ and OLS residuals $\boldsymbol{e}$ as

$$\widehat{\boldsymbol{y}} \equiv \boldsymbol{X}\boldsymbol{b}\,, \tag{3}$$

$$\boldsymbol{e} \equiv \boldsymbol{y} - \widehat{\boldsymbol{y}}\,, \tag{4}$$

where $\widehat{\boldsymbol{y}} = (\widehat{y}_1, \ldots, \widehat{y}_n)'$ and $\boldsymbol{e} = (e_1, \ldots, e_n)'$. The projection matrix $\boldsymbol{P}$ and the annihilator matrix $\boldsymbol{M}$ are

$$\boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,, \tag{5}$$

$$\boldsymbol{M} \equiv \boldsymbol{I}_n - \boldsymbol{P}\,. \tag{6}$$

To answer the questions in Section 2, the following might be useful:

1. *Definition*: A square matrix $\boldsymbol{A}$ is **idempotent** if $\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}$.

2. *Definition*: A square matrix $\boldsymbol{B}$ is **symmetric** if it is equal to its transpose: $\boldsymbol{B} = \boldsymbol{B}'$.

3. For two n-vectors of observations $\boldsymbol{u} = (u_1, \ldots, u_n)'$ and $\boldsymbol{v} = (v_1, \ldots, v_n)'$ of the random variables U and V, we estimate their covariance as $\widehat{\text{Cov}}(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{n-1}\sum_{i=1}^{n}\left[(u_i - \overline{u})\cdot(v_i - \overline{v})\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} u_i \cdot v_i - n\cdot\overline{u}\cdot\overline{v}\right]$, where $\overline{u} = \frac{1}{n}\sum_{i=1}^{n} u_i$ and $\overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$.

1 (Conditional) expectations (6 points)

1. (2 points) Prove that Assumptions 2 and 4 imply

$$\mathbb{C}\text{ov}(\varepsilon_i, \varepsilon_j | \boldsymbol{X}) = 0 \quad \text{for } i, j = 1, \ldots, n\,, i \neq j\,. \tag{7}$$

Solution: Here comes the solution.

2. (2 points) Prove that Assumptions 2 and 4 imply

$$\mathbb{V}\text{ar}(\varepsilon_i) = \sigma^2 \quad \text{for } i = 1, \ldots, n\,, \tag{8}$$

$$\mathbb{C}\text{ov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{for } i, j = 1, \ldots, n\,, i \neq j. \tag{9}$$

Solution: Here comes the solution.

3. (2 points) Prove that under random sampling, $\mathbb{E}(\varepsilon_i \varepsilon_j | \boldsymbol{X}) = \mathbb{E}(\varepsilon_i | \boldsymbol{x}_i)\, \mathbb{E}(\varepsilon_j | \boldsymbol{x}_j)$ for $i \neq j$.

Solution: Here comes the solution.

2 OLS: algebraic properties (14 points)

1. (1 point) Prove that $\boldsymbol{P}$ is symmetric and idempotent.

Solution: Here comes the solution.

2. (1 point) Prove that $\boldsymbol{M}$ is symmetric and idempotent.

Solution: Here comes the solution.

3. (1 point) Prove that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$.

Solution: Here comes the solution.

4. (1 point) Prove that $\widehat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}$. (This is why $\boldsymbol{P}$ is often called the "hat matrix", as it "puts a hat on $\boldsymbol{y}$").

Solution: Here comes the solution.

5. (1 point) Prove that $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$.

Solution: Here comes the solution.

6. (1 point) Prove that $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y}$. (This is why $\boldsymbol{M}$ is often called the "residual maker matrix").

Solution: Here comes the solution.

7. (1 point) Prove that $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{\varepsilon}$.

Solution: Here comes the solution.

8. (1 point) Prove that the sum of squared residuals can be written as $\boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$.

Solution: Here comes the solution.

9. (1 point) Prove that the average of the residuals is zero: $\frac{1}{n}\sum_{i=1}^{n} e_i = 0$.

Solution: Here comes the solution.

10. (2 points) Prove that the *sample covariance* between the kth regressor and the residuals is zero: $\widehat{\text{Cov}}(\boldsymbol{X}_{[1:n,k]}, \boldsymbol{e}) = 0$ for $k = 1, \ldots, K$, where $\boldsymbol{X}_{[1:n,k]}$ denotes the kth column of the data matrix $\boldsymbol{X}$.

Solution: Here comes the solution.

11. (1 point) Prove that the *sample covariance* between the fitted values and the residuals is zero: $\widehat{\text{Cov}}(\widehat{\boldsymbol{y}}, \boldsymbol{e}) = 0$.

> **Solution:** Here comes the solution.

12. (1 point) Prove that the estimated regression line passes through the sample mean of the dependent variable and the sample mean of the regressors. In other words, let $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$, and you need to prove that $\overline{y} = \overline{\boldsymbol{x}}'\boldsymbol{b}$.

> **Solution:** Here comes the solution.

13. (1 point) Prove that the mean of the fitted values equals the mean of the observed values of the dependent variable: $\frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i$.

> **Solution:** Here comes the solution.

3 Non-singular linear transformation of regressors (2 bonus points)

Consider a simple linear regression model given by

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i\,, \tag{10}$$

where y_i corresponds to the electricity consumption of household i measured in kilowatthours (kW h) in a given year, and x_i is the average summer temperature measured in Celsius (centigrade) degrees in the block where household i is located. Imagine you estimate the model using OLS, and obtain $\boldsymbol{b}$, $\widehat{\boldsymbol{y}}$ and $\boldsymbol{e}$.

Now imagine that your coauthor prefers measuring temperature in Fahrenheit degrees, and tells you that $f = 32 + 1.8 \cdot c$ is the formula to convert c degrees Celsius into f degrees Fahrenheit.

1. (1 point) Find the transformation that gives you the new data matrix $\widetilde{\boldsymbol{X}}$, corresponding your coauthor's preference. In other words, find the (2×2) matrix $\boldsymbol{Q}$ such that $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{Q}$ measures temperature in Fahrenheit.

> **Solution:** Here comes the solution.

2. (1 point) Now you re-estimate the model with $\widetilde{\boldsymbol{X}}$ in place of $\boldsymbol{X}$. Let $\widetilde{\boldsymbol{b}}$ denote the new OLS estimate, $\widetilde{\boldsymbol{y}}$ the new fitted values, and $\widetilde{\boldsymbol{e}}$ the new residuals. How are they related to the original $\boldsymbol{b}$, $\widehat{\boldsymbol{y}}$ and $\boldsymbol{e}$? Interpret your findings.

> **Solution:** Here comes the solution.