DSE - Data-Driven Economic Analysis
Econometrics Module

Lecture 2 - Regression and OLS estimator

Michele De Nadai
michele.denadai@unimi.it

Trimester II, 2023

UNIVERSITÀ DEGLI STUDI DI MILANO

Simple linear regression

- Assume linearity of the CEF of y on w and q:

  $$y = \beta_0 + \beta_1 w + \beta_2 q + e,$$

  then $E[e|w, q] = 0$.

- More compactly we can write

  $$y = \mathbf{x}'\boldsymbol{\beta} + e$$

  where $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ w \\ q \end{pmatrix}$.

CEF as Best Predictor

- CEF can be seen as the **best predictor** of y given $\mathbf{x}$ under the mean squared loss function.
- Prediction error: $y - g(\mathbf{x})$ for any predictor function $g(\mathbf{x})$
- Best predictor: Minimise

$$E[(y - g(\mathbf{x}))^2]$$

- Answer: CEF is the best predictor. If $E[y^2] < \infty$, then for any predictor $g(\mathbf{x})$,

$$E[(y - g(\mathbf{x}))^2] \geqslant E[(y - m(\mathbf{x}))^2]$$

where $m(\mathbf{x}) = E[y|\mathbf{x}] = \text{CEF}$.

Estimating a linear CEF

- The linear CEF is a function of $\boldsymbol{\beta}$, which minimizes

$$E[(y - \mathbf{x}'\boldsymbol{\beta})^2]$$

- The result of the optimization problem is then:

$$\boldsymbol{\beta} = (E[\mathbf{x}\mathbf{x}'])^{-1}E[\mathbf{x}y],$$

  which is a function of the population moments.

- Sample: Suppose that we have observations $(y_i, \mathbf{x}_i)$, $i = 1, ..., n$.
- Random draw from the joint distribution of $(y, \mathbf{x})$.
- How can we estimate $\boldsymbol{\beta}$?

Moment Estimator

- We *estimate* population moments $E[\cdot]$ by replacing the unknown expectation with $\frac{1}{n}\sum_{i=1}^{n}[\cdot]$.
- This estimator is called the *moment* estimator.
- Any estimator is a random variable. So does the moment estimator.
- Justified by the **Law of Large Numbers** (LLN): For a random variable y such that $E[|y|] < \infty$, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} y_i \underset{p}{\to} E[y],$$

where y_i is the i^{th} realization of y and $\underset{p}{\to}$ denotes the convergence in probability.

Least Squares Estimator

- $\boldsymbol{\beta}$ solves the population criterion:

$$\boldsymbol{\beta} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^k}{\arg\min}\, \mathrm{E}\left((y - \mathbf{x}'\boldsymbol{\gamma})^2\right)$$

- $\widehat{\boldsymbol{\beta}}$ solves the sample criterion:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\gamma} \in \mathbb{R}^k}{\arg\min}\, \frac{1}{n} \sum_{i}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\gamma})^2$$

- Solution: the (ordinary) least squares (OLS) estimator

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i y_i$$

Solving for LS with One Regressor and no Constant

- Consider $k = 1$ so that β is a scalar.

$$SSE_n(\beta) = \sum_{i=1}^{n} (y_i - x_i\beta)^2$$

- Multiplying/Dividing SSE_n with $1/n$ does not matter (why?).
- Differenciate SSE_n with respect to β to get

$$-2 \sum_{i=1}^{n} x_i(y_i - x_i\beta)$$

- First order conditions imply

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Least Squares Residual

- Fitted value: $\widehat{y}_i = \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$
- Residual: $\widehat{e}_i = y_i - \widehat{y}_i$
- Note:
  - e_i: projection error, unobservable (population concept)
  - $\widehat{e}_i$: residual, by-product of OLS estimation
- Properties of $\widehat{e}_i$:
  - $\sum_{i=1}^n \mathbf{x}_i \widehat{e}_i = \mathbf{0}$.
  - $\sum_{i=1}^n \widehat{e}_i = 0$ if $\mathbf{x}_i$ contains a constant.

Matrix Form

- $\mathbf{y}$: $n \times 1$ vector, $\mathbf{X}$: $n \times k$ matrix, $\mathbf{e}$: $n \times 1$ vector

$$\begin{matrix} y_1 = \mathbf{x}_1'\beta + e_1 \\ y_2 = \mathbf{x}_2'\beta + e_2 \\ \vdots \\ y_n = \mathbf{x}_n'\beta + e_n \end{matrix} \quad \Rightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

- The OLS estimator

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

- Residual: $\widehat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$

Projection Matrix and Leverage

- Projection matrix: $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
- Properties:
  1. $\mathbf{PX} = \mathbf{X}$.
  2. Symmetric: $\mathbf{P} = \mathbf{P}'$
  3. Idempotent: $\mathbf{PP} = \mathbf{P}$
  4. $\mathbf{Py} = \widehat{\mathbf{y}}$
  5. $\mathbf{P}$ is positive semi-definite.
- The i^{th} diagonal element of $\mathbf{P}$: Leverage of the i^{th} observation

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{x}_i$$

- Properties:
  1. $0 \leqslant h_{ii} \leqslant 1$
  2. $\sum_{i=1}^{n} h_{ii} = \mathrm{tr}(\mathbf{P}) = k$

Orthogonal Projection Matrix

- Orthogonal projection matrix: $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$
- Properties:
  1. $\mathbf{MX} = 0$
  2. Symmetric: $\mathbf{M} = \mathbf{M}'$
  3. Idempotent: $\mathbf{MM} = \mathbf{M}$
  4. $\widehat{\mathbf{e}} = \mathbf{My} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = \mathbf{Me}$

Estimation of Error Variance

- Population parameter: $\sigma^2 = E[e_i^2] = Var(e_i)$

- Infeasible moment estimator: $\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\mathbf{e}'\mathbf{e}$

- Alternative moment estimator: $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i^2 = \frac{1}{n}\widehat{\mathbf{e}}'\widehat{\mathbf{e}}$
- Implication:

$$
\begin{aligned}
\widetilde{\sigma}^2 - \widehat{\sigma}^2 &= \frac{1}{n}\mathbf{e}'\mathbf{e} - \frac{1}{n}\mathbf{e}'\mathbf{M}\mathbf{e} \\
&= \frac{1}{n}\mathbf{e}'\mathbf{P}\mathbf{e} \geqslant 0
\end{aligned}
$$

- $\widehat{\sigma}^2$ is always (weakly) smaller than $\widetilde{\sigma}^2$.

Normal Regression Model

- Model:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + e_i$$

- Assume $e_i | \mathbf{x}_i \sim \mathrm{N}(0, \sigma^2)$.
- Implies

$$y_i | \mathbf{x}_i \sim \mathrm{N}(\mathbf{x}_i' \boldsymbol{\beta}, \sigma^2)$$

- Normal regression is a parametric model (i.e. we assume a particular distribution for the random variables).
- Estimation of $\boldsymbol{\beta}$ and σ^2: Maximum likelihood

Maximum Likelihood

- Suppose y_i has a pdf $f(y_i; \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is $m \times 1$ unknown parameter.
- Random sample: $(y_1, ... y_n)$ are iid.
- Joint density: $f_n(y_1, ..., y_n; \boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta})$
- Likelihood $L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\beta})$
- Log-likelihood: $\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\beta})$
- MLE $\widehat{\boldsymbol{\beta}}$ maximizes the (log-)likelihood.

MLE in Normal Regression Model

- Assume $y_i | \mathbf{x}_i \sim N(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2)$.
- Normal pdf:

$$f(y_i | \mathbf{x}_i; \boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2}{2\sigma^2}}$$

- Log-likelihood:

$$\log L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2$$

MLE and Normal Regression Model (cont.)

- MLE $(\widehat{\boldsymbol{\beta}}_{\boldsymbol{mle}}, \widehat{\sigma}^2_{mle})$ maximizes $\log L(\boldsymbol{\beta}, \sigma^2)$.
- Maximizing $\log L(\boldsymbol{\beta}, \sigma^2)$ with respect to $\boldsymbol{\beta}$ is equivalent to the OLS estimator:

$$\widehat{\boldsymbol{\beta}}_{mle} = \widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- First order conditions with respect to σ^2:

$$\widehat{\sigma}^2_{mle} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mathbf{x}_i' \widehat{\boldsymbol{\beta}}_{mle})^2 = \frac{1}{n} \sum_{i=1}^{n} \widehat{e}_i^2 = \widehat{\sigma}^2.$$

- The log likelihood function evaluated at MLE $\log L(\widehat{\beta}_{mle}, \widehat{\sigma}^2_{mle})$ typically reported as a measure of fit.