Erasmus School of Economics

MOOC Econometrics

Lecture 2.3 on Multiple Regression: Estimation

Christiaan Heij

OLS criterion

- Model: $y = X\beta + \varepsilon$
- Dimensions: y $(n \times 1)$, X $(n \times k)$: observed data
  β $(k \times 1)$, ε $(n \times 1)$: unobserved
- Objective:
  $\rightarrow$ Estimate β by $(k \times 1)$ vector b so that Xb is 'close' to y.

OLS criterion

- We assume that $(n \times k)$ matrix X has $\text{rank}(X) = k$.

Test

Prove that $\#(\text{parameters}) = k \leq n = \#(\text{observations})$.

- Answer: X is $(n \times k)$ matrix, hence $k = \text{rank}(X) \leq n$.
- Wish: small vector of residuals $y - Xb = e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$.
- Least squares criterion ('ordinary least squares', OLS):
  $\rightarrow$ minimize $S(b) = e'e = \sum_{i=1}^{n} e_i^2$.

OLS estimation

- $S(b) = e'e = (y - Xb)'(y - Xb)$
  $= y'y - y'Xb - b'X'y + b'X'Xb$
  $= y'y - 2b'X'y + b'X'Xb$

Test

We used $y'Xb = b'X'y$. Prove this result.

- Answer: $y'Xb$ is (1×1), so $y'Xb = (y'Xb)' = b'X'y$.
- Facts of matrix derivatives (see Building Blocks):
  $\frac{\partial b'a}{\partial b} = a$
  $\frac{\partial b'Ab}{\partial b} = (A + A')b$

OLS estimation

- First order conditions for $S(b) = y'y - 2b'X'y + b'X'Xb$:

  $\frac{\partial S}{\partial b} = -2X'y + (X'X + X'X)b = -2X'y + 2X'Xb = 0.$

- So: $X'Xb = X'y$.

Test

Prove that $\text{rank}(X) = k$ implies that $X'X$ is invertible.

- Answer: $X'X$ is $(k \times k)$ matrix, and

  $X'Xa = 0 \Rightarrow a'X'Xa = (Xa)'Xa = 0 \Rightarrow Xa = 0 \Rightarrow a = 0.$
  Last step follows from $\text{rank}(X) = k$.

- So: $\boxed{b = (X'X)^{-1}X'y}$

Geometric aspects

- y is $(n \times 1)$, X is $(n \times k)$

- Define $H = X(X'X)^{-1}X'$

  $M = I - H = I - X(X'X)^{-1}X'$

Test

Show that $M' = M$, $M^2 = M$, $MX = 0$, $MH = 0$.

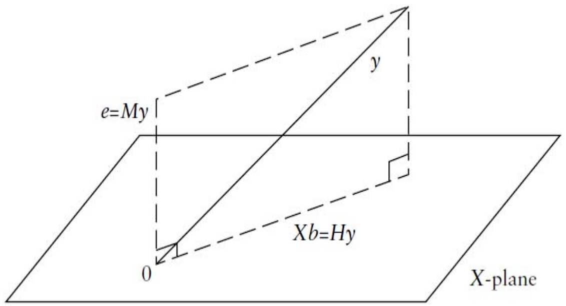
- Answer: Direct calculations.
  Use $(X'X)^{-1}$ symmetric and $(X'X)^{-1}X'X = I$.

- Fitted values: $\hat{y} = Xb = X(X'X)^{-1}X'y = Hy$.

  Residuals: $e = y - Xb = y - Hy = My$.

- e and $\hat{y}$ orthogonal: $e'\hat{y} = (My)'Hy = y'M'Hy = 0$.

Relation between y, X, b, and e



The 'X-plane' is k-dimensional subspace spanned by columns of X, that is, set of vectors Xa with a arbitrary $(k \times 1)$ vector.

Estimation of error variance σ^2

- $\sigma^2 = E(\varepsilon_i^2)$

  Estimate unknown $\varepsilon = y - X\beta$ by residuals $e = y - Xb$.

- Sample variance of residuals: $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(e_i - \bar{e})^2$.

Test

Check that the $(n \times 1)$ vector of residuals e satisfies k linear restrictions, so that e has $(n - k)$ 'degrees of freedom'.

- Answer: $\text{rank}(X) = k$, and $X'e = X'(y - Xb) = X'y - X'Xb = 0$.

- OLS estimator: $s^2 = \frac{1}{n-k}e'e = \frac{1}{n-k}\sum_{i=1}^{n} e_i^2$

- Unbiased under standard assumptions (see next lecture).

R-squared (R^2)

- Definition: $R^2 = \Big(\text{cor}(y, \hat{y})\Big)^2 = \frac{\left(\sum(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})\right)^2}{\sum(y_i - \overline{y})^2 \sum(\hat{y}_i - \overline{\hat{y}})^2}$,

  where 'cor' is correlation coefficient and $\hat{y} = Xb$.

- Higher R^2 means better fit of Xb to observed y.

- If model contains constant term ($x_{1i} = 1$ for all $i = 1, \dots n$):

  $R^2 = 1 - \frac{e'e}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$.

TRAINING EXERCISE 2.3

- Train yourself by making the training exercise (see the website).

- After making this exercise, check your answers by studying the webcast solution (also available on the website).