Econometrics

Preliminaries

by Vanessa Berenguer-Rico

University of Oxford

Michaelmas Term 2016

Outline

The Linear Model

- The Simple Regression Model
- Multiple Regression Model

The Simple Regression Model

"The simple regression model can be used to study the relationship between two variables [...] It has limitations as a general tool for empirical analysis. Nevertheless it is sometimes appropriate as an empirical tool. (Moreover) Learning how to interpret the simple regression model is good practice for studying multiple regression," Wooldridge (2013) p. 20

The Simple Regression Model

Examples (Stock and Watson, 2012, p. 149)

- A state implements tough new penalties on drunk drivers: What is the effect on highway fatalities?
- A school district cuts the size of its elementary school classes: What is the effect on its students' standardized test scores?
- You successfully complete one more year of college classes: What is the effect on your future earnings?

The Simple Regression Model

$$y_i = \beta_1 + \beta_2 x_i + u_i,$$

- y_i and x_i are observable random scalars
- u_i is the unobservable random disturbance or error
- β_1 and β_2 are the parameters (constants) we would like to estimate

The Simple Regression Model: OLS

- The OLS objective function

$$\min_{b \in \mathbb{R}^2} \sum_{i=1}^{n} u_i^2 = \min_{b \in \mathbb{R}^2} \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2 = L$$

- System of Normal Equations: First Order Conditions

$$\begin{aligned} \frac{\partial L}{\partial b_1} &= -2 \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i) = 0 \\ \frac{\partial L}{\partial b_2} &= -2 \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)\, x_i = 0 \end{aligned}$$

The Simple Regression Model: OLS

- The OLS solution

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n^{-1}\sum_{i=1}^{n} y_i x_i - \bar{y}\bar{x}}{n^{-1}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

The Simple Regression Model: OLS

- Example: Wage and Education (Wooldridge, 2013, p. 31)

$$\widehat{wage}_i = -0.90 + 0.54educ_i$$

- Interpreting estimates (caution!)

The Multiple Regression Model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + ... + \beta_K x_{Ki} + u_i$$

Example:

$$wage_i = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + u_i$$

Multiple Regression: OLS

- The OLS objective function:

$$L = \min_{b \in \mathbb{R}^K} \sum_{i=1}^{n} u_i^2$$

- In this case:

$$L = \min_{b \in \mathbb{R}^K} \sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i} - b_3 x_{3i} - ... - b_K x_{Ki})^2$$

Multiple Regression: OLS

System of Normal Equations: First Order Conditions

$$
\begin{aligned}
\frac{\partial L}{\partial b_1} &= -2\sum_{i=1}^{n}\left(y_i - b_1x_{1i} - b_2x_{2i} - b_3x_{3i} - \dots - b_Kx_{Ki}\right)x_{1i} = 0 \\
\frac{\partial L}{\partial b_2} &= -2\sum_{i=1}^{n}\left(y_i - b_1x_{1i} - b_2x_{2i} - b_3x_{3i} - \dots - b_Kx_{Ki}\right)x_{2i} = 0 \\
&\vdots \\
\frac{\partial L}{\partial b_K} &= -2\sum_{i=1}^{n}\left(y_i - b_1x_{1i} - b_2x_{2i} - b_3x_{3i} - \dots - b_Kx_{Ki}\right)x_{Ki} = 0
\end{aligned}
$$

Multiple Regression: OLS

Too Long: Matrix Notation!

$$Y = X\beta + U$$

where

$$\begin{aligned} Y &= \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & x_{22} & \cdots & x_{K2} \\ x_{13} & x_{23} & \cdots & x_{K3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{Kn} \end{pmatrix}, U = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix} \\ \beta &= \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_K \end{pmatrix}' \end{aligned}$$