Notes on Econometrics

September 14, 2021

Simple linear model

- N observations of dependent variable $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix}$
- N observations of each explanatory variable $\boldsymbol{x_k} = \begin{pmatrix} x_{k1} \\ y_{k2} \\ \dots \\ y_{kN} \end{pmatrix}$
- Assume K explanatory variables
- Assume each y_n is related to x_{1n}, x_{2n}, ..., x_{Kn} linearly

$$y_n = \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_K x_{Kn} + \epsilon_n$$

where ϵ_n is an unobserved error

- **Our goal is to estimate the unknown parameters β_1, β_2, ..., β_k**

Simple linear model

The simple linear model is summarized as

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

$$\mathrm{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{pmatrix} \qquad (\textit{data matrix})$$

$$\mathrm{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_K \end{pmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_N \end{pmatrix}$$

Ordinary least squares

The method of **OLS (ordinary least squares)** estimates β by minimizing the **sum of squared residuals (SSR)**:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \arg\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} (y_n - x_{n1}\beta_1 - x_{n2}\beta_2 - \ldots - x_{nK}\beta_K)^2 \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\end{aligned}$$

Hypothesis testing

- We usually test **null hypotheses** of the form $\beta_k = 0$
- Under the null hypothesis, the **test statistic** $\hat{t} = \frac{\hat{\beta}_k}{\sqrt{\widehat{var}(\hat{\beta}_k)}}$ has a t-distribution with $N - K$ degrees of freedom
- The p-value is smallest **significant level** under which the null hypothesis $\beta_k = 0$ will be rejected