Introductory Econometrics

Simple Linear Regression Model (II)

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2025

Recall: Bivariate Linear Regression Model

- Recall the bivariate linear regression model

$$Y = \beta_0 + \beta_1 X + u,$$

where β_0 and β_1 are parameters to be estimated, and
  - β_0 is referred to as the **intercept**.
  - β_1 is referred to as the **slope**.
- We observe Y and X as **sample**,

$$\{(X_i, Y_i) : i = 1, 2, \ldots, n\}$$

- **For each** i, bivariate linear regression model suggests

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

Recall: Ordinary Least Square Estimation

Normal equations

$$\begin{cases} \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \\ \sum X_i\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \end{cases}$$

Solution

$$\begin{cases} \hat{\beta}_0 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2} \\ \hat{\beta}_1 = \frac{n \sum Y_i X_i - \sum Y_i \sum X_i}{n \sum X_i^2 - (\sum X_i)^2} \end{cases}$$

Assumptions for Simple Linear Regression

Assumption 1: Model is correctly specified.

(1) Variables are correctly selected.

(2) Model is correctly specified.

Assumption 2: Explanatory variable converges in probability to constant.

$$P\lim_{n\to\infty}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2/n \to Q.$$

Assumption 3: Stochastic error has zero expectation.

$$\mathrm{E}\left(u_i \mid X\right) = 0.$$

Assumptions for Simple Linear Regression

$\mathrm{E}(u_i \mid X) = 0$ suggests that $\mathrm{Cov}(u_i, X) = 0$, i.e. u_i is uncorrelated with X. We need an important tool called **Law of Iterated Expectations, LIE**.

Assumption 4: Homeskedasticity and no correlation.

$$\mathrm{Var}(u_i \mid X) = \sigma^2 \quad i = 1, 2, \cdots, n$$

$$\mathrm{Cov}(u_i, u_j \mid X) = 0 \quad i \neq j$$

And similarly due to LIE, we have

$$\mathrm{Var}(u_i) = \sigma^2.$$

Assumptions for Simple Linear Regression

Assumption 5: Normality.

$$u_i \mid X \sim N\left(0, \sigma^2\right).$$

Assumption 1 to **Assumption 5** are usually referred to as the **classical assumption**, and the corresponding linear regression model is referred as to the **Classical Linear Regression Model, CLRM**.

- Specifically, **Assumption 5** can be relaxed in large sample, **Assumption 1** to **Assumption 4** refers to the **Gauss-Markov assumption**.

- For bivariate linear regression model

$$Y \mid X \sim N\left(\beta_0 + \beta_1 X, \sigma^2\right).$$

Statistical Properties

- Statistics and statistical properties of statistics.

  Small-sample-properties:

  - Unbiasedness
  - Efficiency

  Large-sample-properties:

  - Consistency

- $\hat{\theta}$ is unbiased if $\mathrm{E}\left(\hat{\theta}\right) = \theta$, that is the average value of $\hat{\theta}$ over all realizations is equal to the underlying population value.
- $\hat{\theta}$ is efficient if it has smaller variance than **any other unbiased estimator**.

Statistical Properties

- **Convergence in Probability**. *A sequence of random variables* $\{X_n\}$ *converges in probability to a random variable* X *if, for every* $\epsilon > 0$*,*

$$\lim_{n\to\infty} \mathrm{P}\left(\|X_n - X\| \geq \epsilon\right) = 0.$$

  Notation: $X_n \xrightarrow{p} X$*, or* $P\lim\left(X_n\right) = X$*.*

- An estimator $\hat{\theta}$ is consistent if $P\lim\left(\hat{\theta}\right) = \theta$.

Small-Sample-Properties of OLS Estimator

- OLS estimator is a linear combination of Y_i.

$$\hat{\beta}_1 = \sum k_i Y_i \quad \hat{\beta}_0 = \sum w_i Y_i$$

  where $k_i = \frac{x_i}{\sum x_i^2}$ and $w_i = \frac{1}{n} - \bar{X} k_i$.
- OLS estimator is unbiased.
- OLS estimator is efficient. More discussion
- OLS estimator is **best linear unbiased estimator (BLUE)**. This claim is usually referred to as the **Gauss-Markov theorem**.

Large-Sample Properties of OLS Estimator

- Large sample properties refer to the properties of statistics when sample size is "sufficiently large", i.e. $n \to \infty$.
- **Weak Law of Large Numbers**. *Suppose* $X_1, \ldots, X_n$ *and i.i.d. with* $\mathrm{E}\|X_1\| < \infty$, *then as* $n \to \infty$,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathrm{E}(X_1).$$

- OLS estimator is consistent.

$$\begin{aligned} P\lim\left(\hat{\beta}_1\right) &= P\lim\left(\beta_1 + \sum k_i u_i\right) = P\lim\left(\beta_1\right) + P\lim\left(\frac{\sum x_i u_i}{\sum x_i^2}\right) \\ &= \beta_1 + P\lim\left(\frac{\sum x_i u_i / n}{\sum x_i^2 / n}\right) = \beta_1. \end{aligned}$$

Distributions Associated with OLS Estimator

- If we stick to the Normality assumption (**Assumption 5**), it can be easily shown that for the finite sample

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right)$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2\right)$$

Distributions Associated with OLS Estimator

- By relaxing the normality assumption under the large sample, assymptotically we have

$$\hat{\beta}_1 \overset{a}{\sim} N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right)$$

$$\hat{\beta}_0 \overset{a}{\sim} N\left(\beta_0, \frac{\sum X_i^2}{n \sum x_i^2}\sigma^2\right)$$

Estimation of σ^2

$$\begin{aligned}\sum e_i^2 &= \sum (y_i - \hat{y}_i)^2 \\ &= \sum \left[\left(\beta_1 - \hat{\beta}_1\right) x_i + (u_i - \bar{u})\right]^2 \\ &= \sum \left(\beta_1 - \hat{\beta}_1\right)^2 x_i^2 + \sum (u_i - \bar{u})^2 - 2 \sum \left(\sum k_i u_i\right) x_i (u_i - \bar{u})\end{aligned}$$

- Note that

$$\begin{aligned}-2 \sum \left(\sum k_i u_i\right) x_i (u_i - \bar{u}) &= -2 \sum x_i u_i \sum k_i u_i + 2\bar{u} \sum x_i \sum k_i u_i \\ &= -2 \sum x_i u_i \sum k_i u_i\end{aligned}$$

- Hence

$$\sum e_i^2 = \sum \left(\beta_1 - \hat{\beta}_1\right)^2 x_i^2 + \sum (u_i - \bar{u})^2 - 2 \sum x_i u_i \frac{\sum x_i u_i}{\sum x_i^2}$$

Estimation of σ^2

$$\mathrm{E}\left[\sum\left(\beta_1-\hat{\beta}_1\right)^2 x_i^2 \mid X\right]=\sum x_i^2 \operatorname{Var}\left(\hat{\beta}_1 \mid X\right)=\frac{\sum x_i^2\sigma^2}{\sum x_i^2}=\sigma^2$$

$$\begin{aligned}\mathrm{E}\left[\sum\left(u_i-\bar{u}\right)^2 \mid X\right] &= \mathrm{E}\left[\left(\sum u_i^2-2\bar{u}\sum u_i+n\bar{u}^2\right) \mid X\right] \\ &= \mathrm{E}\left[\left(\sum u_i^2-n\bar{u}^2\right) \mid X\right]=(n-1)\sigma^2\end{aligned}$$

$$\mathrm{E}\left[\left.\frac{\left(\sum x_i u_i\right)^2}{\sum x_i^2}\right| X\right]=\mathrm{E}\left[\left.\frac{\sum x_i^2 u_i^2+2\sum_{i\neq j}\left(x_i x_j\right)\left(u_i u_j\right)}{\sum x_i^2}\right| X\right]=\sigma^2$$

- Finally

$$\mathrm{E}\left(\sum e_i^2 \mid X\right)=\sigma^2+(n-1)\sigma^2-2\sigma^2=(n-2)\sigma^2.$$

Sample Variance of OLS Estimator

- We denote the estimation of σ^2 by $\hat{\sigma}^2$ as follows

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}.$$

- With $\hat{\sigma}^2$, we obtain the sample variance of $\hat{\beta}_1$ and $\hat{\beta}_0$ respectively as follows

$$\begin{aligned} S_{\hat{\beta}_1}^2 &= \hat{\sigma}^2 / \sum x_i^2 \\ S_{\hat{\beta}_0}^2 &= \hat{\sigma}^2 \sum X_i^2 / n \sum x_i^2 \end{aligned}$$

Summary

- Classical assumptions for simple linear regression.
- Law of Iterated Expectations.
- Statistical properties of OLS estimator.
- Corresponding Distributions.

Variance of $\hat{\beta}_1$

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_1 \mid X\right) &= \operatorname{Var}\left(\sum k_i Y_i \mid X\right) \\
&= \sum k_i^2 \operatorname{Var}\left[\left(\beta_0 + \beta_1 X_i + u_i\right) \mid X\right] \\
&= \sum k_i^2 \operatorname{Var}\left(u_i \mid X\right) \\
&= \sum \left(\frac{x_i}{\sum x_i^2}\right)^2 \sigma^2 = \frac{\sigma^2}{\sum x_i^2}
\end{aligned}$$

Back