Sampling Distribution and Test of a Single Parameter in the Linear Regression Model

Econometrics

Péter Elek

Properties of the OLS Estimator

- The OLS estimator is:

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n} \hat{r}_{ji} y_i}{\sum_{i=1}^{n} \hat{r}_{ji}^2} = \beta_1 + \frac{\sum_{i=1}^{n} \hat{r}_{ji} u_i}{\sum_{i=1}^{n} \hat{r}_{ji}^2}.$$

- So far: $E(\hat{\beta}_j) = \beta_j$ and $Var(\beta_j \mid \mathbf{X}) = \frac{\sigma_u^2}{n * \hat{Var}(x_j) * (1 - R_j^2)}$
  - if heteroskedasticity, a robust variance formula applies in large samples
- Since OLS is unbiased and $Var(\hat{\beta}_j \mid \mathbf{X}) \sim K/n \to 0$ as $n \to \infty$, hence $\hat{\beta}_j$ is a consistent estimator of β_j.

Asymptotic Normality of the OLS Estimator

- $\hat{\beta}_j - \beta_j$ is a (weighted) sum of the u_i error terms. It follows from the Central Limit Theorem that as $n \to \infty$ $\hat{\beta}_j$ is approximately normally distributed:

$$\hat{\beta}_j \mid \mathbf{X} \overset{\text{A}}{\sim} N(\beta_j, Var(\hat{\beta}_j \mid \mathbf{X}))$$

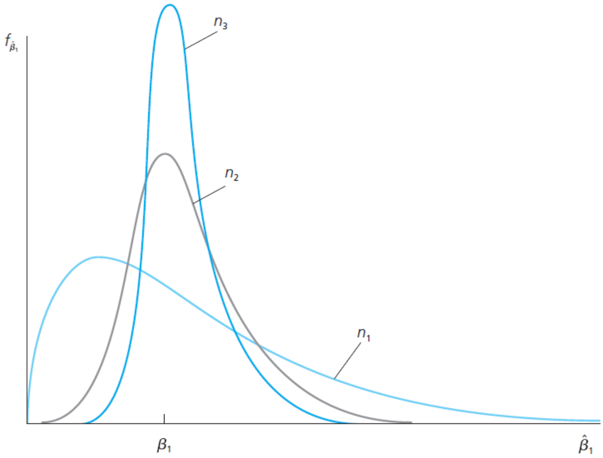
- Hence

$$\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \overset{\text{A}}{\sim} N(0, 1).$$

- Here we assume large samples and assumptions (1)-(4).
- Homoskedasticity is not assumed. Under heteroskedasticity, robust SE is used.

Simulation of the Sampling Distribution of $\hat{\beta}_j$ for Different Sample Sizes

FIGURE 5.1 Sampling distributions of $\hat{\beta}_1$ for sample sizes $n_1 < n_2 < n_3$.



Source: Wooldridge Fig. 5.1.

Large Sample Confidence Interval for β_j

- In large samples the $1-\alpha$ confidence interval is

$$\hat{\beta}_j \pm z_{1-\alpha/2} * SE(\hat{\beta}_j)$$

- where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $N(0,1)$.
- Example: the 95% CI is

$$\hat{\beta}_j \pm 1.96 * SE(\hat{\beta}_j).$$

- t-distribution appears in small samples, but we do not discuss it.

Test of $\beta_j = 0$

- $H_0 : \beta_j = 0$, $H_1 : \beta_j \neq 0$ (two-sided alternative)
  - Most often, a two-sided alternative is recommended if a single parameter is tested
- Under H_0, in large samples,

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \overset{\mathrm{A}}{\sim} N(0,1).$$

- Thus we reject H_0 at significance level α if

$$|t| > z_{1-\alpha/2}$$

- Example: we reject H_0 at the 5% level of significance if $|z| > 1.96$.
- Large-sample p-value of the test is $2 * (1 - \Phi(|z|))$ where $\Phi(.)$ is the cdf of $N(0,1)$.
- t-distribution appears in small samples, but we do not discuss it.

Test of $\beta_j = b$

- $H_0 : \beta_j = b$
- $H_1 : \beta_j \neq b$ (two-sided alternative)
- Under H_0, in large samples,

$$t = \frac{\hat{\beta}_j - b}{SE(\hat{\beta}_j)} \overset{A}{\sim} N(0, 1).$$

- Thus we reject H_0 at significance level α if

$$|t| > z_{1-\alpha/2}$$

- Example: we reject H_0 at the 5% level of significance if $|z| > 1.96$.
- p-value of the test is given as $2 * (1 - \Phi(|z|))$ where $\Phi(.)$ is the cdf of $N(0, 1)$.

Example

Dep. var: birth weight (gr)	**(1)** (bwghtgr $\sim$ cigs)	**(2)** (bwghtgr $\sim$ cigs + faminc)
Intercept	3395.53***	3316.22***
	(16.23)	(29.74)
cigarettes per day	-14.57***	-13.14***
	(2.57)	(2.60)
family income ($1000)	—	2.63***
		(0.83)
R^2	0.0227	0.0298
N	1388	1388

- Source: Wooldridge bwght data
- Standard errors in parentheses, $^{***}: p < 0.01$