Lecture notes IV

The least squares assumption:

Under certain key assumptions the OLS estimator is:

i) unbiased: $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_0) = \beta_0$

ii) consistent: $\hat{\beta}_1 \xrightarrow{p} \beta_1$ and $\hat{\beta}_0 \xrightarrow{p} \beta_0$

iii) $$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim N(0,1)$$ when n is large, i.e. asymptotic normality

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \sim N(0,1)$$

Key OLS assumptions:

A.1 $E(u_i | X_i) = 0$ ($\Rightarrow \text{Cov}(u_i, X_i) = 0$)

A.2 (X_i, Y_i) for $i = 1, \dots, n$ are i.i.d vectors of random variables

[Alternatively: (X_i, u_i) are i.i.d]

A.3 $E(X_i^4) < \infty$ and $E(u_i^4) < \infty$

$\Rightarrow$ the probability of ~~the~~ extreme data is negligible

Why is $\hat{\beta}_k$ consistent for β_k $(k=0,1)$?

Note that
$$\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X}) Y_i}{\sum_i (X_i - \bar{X}) X_i} = \frac{\sum_{i=1}^n (X_i - \bar{X})(\beta_0 + \beta_1 X_i + u_i)}{\sum_i (X_i - \bar{X}) X_i}$$

$$= \frac{\beta_0 \sum_{i=1}^n (X_i - \bar{X})}{\sum_i (X_i - \bar{X}) X_i} + \beta_1 \frac{\sum_i (X_i - \bar{X}) X_i}{\sum_i (X_i - \bar{X}) X_i} + \frac{\sum_{i=1}^n u_i (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X}) X_i}$$

$$= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^n u_i (X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}) X_i} \xrightarrow{P} \beta_1 + \frac{Cov(u_i, X_i)}{Var(X_i)}$$

$$= \beta_1$$

Furthermore
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \xrightarrow{P} \mu_Y - \beta_1 \mu_X = \beta_0$$

What is variance of $\hat{\beta}_1$?

Since
$$\hat{\beta}_1 - \beta_1 = \frac{\sum_i (X_i - \bar{X}) u_i}{\sum_i (X_i - \bar{X})^2}$$

$$Var(\hat{\beta}_1 | X) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 \, Var(u_i | X)}{\left[\sum_i (X_i - \bar{X})^2\right]^2}$$

$X = (X_1, \ldots, X_n)$

Assume that $Var(u_i | X_i) = \sigma_u^2$ for all X, i.e. u_i is independent of X_i. Then

$$Var(\hat{\beta}_1 | X) = \frac{\sigma_u^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$ where

$X = (X_1, \ldots, X_n)$

Note that

$$Var(\hat{\beta}_1) = E\left(Var(\hat{\beta}_1 | X)\right) + Var\left(E(\hat{\beta}_1 | X)\right)$$

$$= E\left(\frac{\sigma_u^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}\right) + \underbrace{Var(\beta_1)}_{=0}$$

$$= E\left(\frac{\sigma_u^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}\right) \to 0 \text{ when } n \to \infty$$

if $\sum_{i=1}^{n} (X_i - \bar{X})^2 \to \infty$

Measures of fit:

$$Y_i = \hat{Y}_i + \hat{u}_i$$ where

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ is <u>predicted Y_i</u>

and $\hat{u}_i = Y_i - \hat{Y}_i$ is the <u>residual</u>

We have $(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + \hat{u}_i$ and

$$\underbrace{\sum_i (Y_i - \bar{Y})^2}_{TSS} = \underbrace{\sum_i (\hat{Y}_i - \bar{Y})^2}_{ESS} + \underbrace{\sum_i \hat{u}_i^2}_{SSR} \qquad (1)$$

(1) holds since $\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}) \hat{u}_i = 0$, i.e.

the residuals ($\hat{u}_i$) are uncorrelated with $\hat{Y}_i$ (the predicted values).

Goodness of fit measure R^2:

$R^2 = \dfrac{ESS}{TSS}$ Since $TSS = ESS + SSR$ we also have that

$$R^2 = \cancel{\frac{ESS}{ESS + SSR}} \; \frac{TSS - SSR}{TSS} = 1 - \frac{SSR}{TSS}$$

Standard error of regression (SER)

SER is the estimate of σ_u

$$SER = s_{\hat{u}} \quad \text{where} \quad s_{\hat{u}}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$$

$$= \frac{SSR}{n-2}$$

Therefore Standard error (SE) of $\hat{\beta}_1$ is $\dfrac{\sigma_u}{\sqrt{\sum_i (X_i-\bar{X})^2}}$, which can be estimated by $\dfrac{SER}{\sqrt{\sum_i (X_i-\bar{X})^2}}$ and hence

$$\frac{\hat{\beta}_1 - \beta_1}{\dfrac{SER}{\sqrt{\sum_i (X_i-\bar{X})^2}}} \sim N(0,1) \quad \text{asymptotically}$$

Sampling distribution of OLS estimator: Why is is normally distributed as $n \to \infty$?

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \xrightarrow{P} 0$$

To obtain non-degenerate limit distribution, necessary to scale $\hat{\beta}_1 - \beta_1$ so that $Var(\hat{\beta})$ stabilizes.

Look at $\sqrt{n}(\hat{\beta}_1 - \beta_1) = \dfrac{\frac{1}{\sqrt{n}}\sum(X_i - \bar{X})u_i}{\frac{1}{n}\sum(X_i - \bar{X})^2}$

As $n \to \infty$, $\bar{X} \to \mu_X$ an $\frac{1}{n}\sum(X_i - \bar{X})^2 \to Var(X_i)$ by Law of Large Numbers

Hence $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ has same limit distribution as

$$\frac{\frac{1}{\sqrt{n}}\sum(X_i - \mu_X)u_i}{Var(X_i)}$$

The latter expression is a sum of i.i.d random variables multiplied by a scaling factor $\frac{1}{\sqrt{n}\,Var(X_i)}$ and hence is asymptotically normal because of CLT.

Moreover $E(\sqrt{n}(\hat{\beta}_1 - \beta_1)) = 0$

$$Var(\sqrt{n}(\hat{\beta} - \beta_1)) = n \cdot Var(\hat{\beta} - \beta_1)$$

$$\xrightarrow{P} \frac{n \cdot \sigma_u^2}{\sum_i (X_i - \bar{X})^2} \xrightarrow{P} \frac{\sigma_u^2}{Var(X_i)}$$

Hence

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{D} N\left(0, \frac{\sigma_u^2}{Var(X_i)}\right)$$

SW Chapter 5: Hypothesis tests and confidence intervals

1. Two sided ~~hypothesis~~ test

$H_0: \beta_1 = \beta_1^* \quad \text{vs} \quad H_1: \beta_1 \neq \beta_1^*$

Test-statistic: $t = \dfrac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \overset{a}{\sim} N(0,1)$

when H_0 is true

p-value: The probability under H_0 that t is larger in absolute value than $|t^{obs}|$, where t^{obs} is the observed value of t in the actual data

For ~~example~~ example, if $\hat{\beta}_1 = 4$, $\beta_1^* = 0$ and $SE(\hat{\beta}_1) = 2$, then $t^{obs} = \frac{4-0}{2} = 2$

$$p\text{-value} = Pr(|t| \geq |t^{obs}|)$$

$$= Pr(t \geq |t^{obs}| \cup t \leq -|t^{obs}|)$$

$$= Pr(t \geq |t^{obs}|) + Pr(t \leq -|t^{obs}|)$$

$$= 1 - \Phi(|t^{obs}|) + \Phi(-|t^{obs}|)$$

(= c.d.f. of standard normal distr.)

$= 2\Phi(-|t^{obs}|)$ — because of symmetry of normal distribution

One-sided hypotheis 1

$H_0: \beta_1 = \beta_1^*$ vs $H_1: \beta_1 < \beta_1^*$

negative values of $t = \dfrac{\hat{\beta}_1 - \beta_1^*}{SE(\hat{\beta}_1)}$ is

adverse towards H_0 and in favour of H_1. $\Rightarrow$ p-value $= Pr(t \le t^{obs})$

$= \Phi(t^{obs})$

One sided hypothesis 2

$H_0: \beta_1 = \beta_1^*$ vs $H_1: \beta_1 > \beta_1^*$

p-value $= Pr(t > t^{obs}) = 1 - \Phi(t^{obs})$

$= \Phi(-t^{obs})$

Confidence intervals (CI)

95% CI: $\left[\hat{\beta}_1 - 1.96\, SE(\hat{\beta}_1),\ \hat{\beta}_1 + 1.96\, SE(\hat{\beta}_1)\right]$

$1-\alpha$ CI: $\left[\hat{\beta}_1 - Z_{\frac{\alpha}{2}}\, SE(\hat{\beta}_1),\ \hat{\beta}_1 + Z_{\frac{\alpha}{2}}\, SE(\hat{\beta}_1)\right]$

where $Z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ fractile in the $N(0,1)$ distribution:

$$Pr(Z \geq Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}, \quad Z \sim N(0,1)$$

E.g. $\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025,\ Z_{\frac{\alpha}{2}} = 1.96$

$\alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.5 \Rightarrow Z_{\frac{\alpha}{2}} = 1.64$