

Simple Linear Regression Model

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

↓ ↓ ↓ ↓
dependent variable coefficients independent variable disturbance term

β_0 & β_1 are linear + Y_t & X_t do not have to be linear

examples:

$$\begin{aligned} \ln Y_t &= \beta_0 + \beta_1 \ln X_t + u_t \\ \sqrt{Y_t} &= \beta_0 + \beta_1 X_t^2 + u_t \end{aligned} \quad \left. \vphantom{\begin{aligned} \ln Y_t &= \beta_0 + \beta_1 \ln X_t + u_t \\ \sqrt{Y_t} &= \beta_0 + \beta_1 X_t^2 + u_t \end{aligned}} \right\} \text{does not violate linearity}$$
$$Y_t = \alpha_0 + \alpha_1^2 X_t \rightarrow \text{violates}$$

disturbance/error term (u_t)

why we have u_t ?

- 1- unpredictable effects. No single pattern for every observation
- 2- Omitted relevant variables

ex:

$$\text{Exports}_t = \beta_0 + \beta_1 \text{Imports} + \beta_2 \text{RGDP} + u_t$$

if this is the correct model, than an equation omitting either Imports & RGDP will be misspecified due to the omission of relevant variable

if a researcher estimates

$$\text{Exports}_t = \beta_0 + \beta_1 \text{Imports}_t + u_t$$

u_t will contain also RGDP

Gauss - Markov Assumptions

1- $E(u_t) = 0$

$Y_t = \beta_0 + \beta_1 X_t$ then $E(u_t) = 0$

the sum of deviations from the mean is always zero

2- $Var(u_t) = \sigma^2$ homoscedasticity or equal variance assumption

$Var(u_t) = E(u_t - E(u_t))^2 = E(u_t)^2$, since $E(u_t) = 0$

the smaller the variance, the more variable the estimator

Homoscedasticity: All observations are equally reliable / important

3- $Cov(u_t, u_s) = 0$ (no autocorrelation, no serial correlation)

$Cov(u_t, u_s) = E[(u_t - E(u_t))(u_s - E(u_s))] = E(u_t u_s)$

since $E(u_t) = E(u_s) = 0$
disturbance terms are independent of each other

Assumptions Concerning X

4- X is non stochastic

the values of X are fixed in repeated samples
this assumption corresponds to the case that the value of X is not affected by Y

$X \rightarrow Y$ but not the reverse ($Y \rightarrow X$)

explanatory variables are exogenous

Y is the endogenous variable

this assumption is violated if there is a simultaneous relationship between X & Y

5- $\text{Cov}(X_t, U_t) = 0$

X is non-stochastic and u is stochastic/random
 ↳ by definition they are not correlated

If X is not exogenous → this assumption is violated

6- $\text{Var}(X_t) \neq 0$

Some of the observations for X are different from others

7- $T > k+1$

positive degrees of freedom

Ordinary Least Squares (OLS)

$$Y_t = \beta_0 + \beta_1 X_t + U_t \rightarrow \text{statistical/economical model}$$

Data: X_t, Y_t

β_0 & $\beta_1 \rightarrow$ unknown parameters to be estimated

observations:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

$$y = X\beta + u \rightarrow \text{matrix notation}$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

Find estimators for $\beta_0, \beta_1 \Rightarrow$

Estimates: $\hat{\beta}_0, \hat{\beta}_1$

to obtain

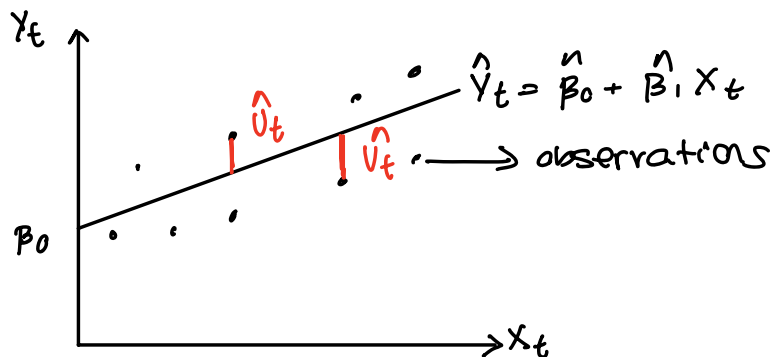
$$\hat{Y}_t = \underbrace{\hat{\beta}_0}_{\text{intercept}} + \underbrace{\hat{\beta}_1 X_t}_{\text{slope}}$$

$$Y_t - \hat{Y}_t = \underbrace{\hat{u}_t}_{\text{estimated error term}} \quad \text{residual}$$

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$$

$$Y_t - \hat{Y}_t = \hat{u}_t$$



Least Squares Rule

We get a line so that the sum of the squares of the vertical distances from each actual observation is minimum (SSR is minimum)

SSR (sum of squares due to regression)

$$SSR = \sum (Y_t - \hat{Y}_t)^2 = \sum \hat{u}_t^2$$

$$\begin{aligned} \text{minimize } SSR(\hat{\beta}_0, \hat{\beta}_1) &= \sum \hat{u}_t^2 \\ &= \sum (Y_t - \hat{Y}_t)^2 \\ &= \sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t)^2 \end{aligned}$$

First order conditions

$$\begin{aligned} (1) \quad \frac{\partial SSR}{\partial \hat{\beta}_0} &= -2 \sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t) = 0 \\ (2) \quad \frac{\partial SSR}{\partial \hat{\beta}_1} &= -2 \sum X_t (Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t) = 0 \end{aligned} \quad \left. \vphantom{\begin{aligned} (1) \\ (2) \end{aligned}} \right\} \text{normal equations}$$

$$\text{from (1)} = \sum \hat{u}_t = 0$$

$$(2) = \sum \hat{u}_t X_t = 0$$

$$\text{from (1)} \Rightarrow \sum Y_t - T \hat{\beta}_0 - \hat{\beta}_1 \sum X_t = 0$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\text{where, } \bar{Y} = \frac{\sum Y_t}{T}, \quad \bar{X} = \frac{\sum X_t}{T}$$

insert it into (2)

$$\begin{aligned} \sum X_t Y_t - \hat{\beta}_0 \sum X_t - \hat{\beta}_1 \sum X_t^2 &= 0 \\ = \sum X_t Y_t - (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum X_t - \hat{\beta}_1 \sum X_t^2 &= 0 \\ = \sum X_t Y_t - T \bar{Y} \bar{X} + \hat{\beta}_1 T \bar{X}^2 &= 0 \end{aligned}$$

$$\hat{\beta}_1 = \frac{\sum X_t Y_t - T \bar{Y} \bar{X}}{\sum X_t^2 - T \bar{X}^2}$$

note that

$$\begin{aligned}\sum (x_t - \bar{x})^2 &= \sum x_t^2 - 2\bar{x} \sum x_t + T\bar{x}^2 \\ &= \sum x_t^2 - 2T\bar{x}^2 + T\bar{x}^2 \\ &= \sum x_t^2 - T\bar{x}^2\end{aligned}$$

$$\begin{aligned}\sum (x_t - \bar{x})(y_t - \bar{y}) &= \sum x_t y_t - \bar{y} \sum x_t - \bar{x} \sum y_t + T \bar{y} \bar{x} \\ &= \sum x_t y_t - T \bar{y} \bar{x}\end{aligned}$$

$$\hat{\beta}_1 = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} = \frac{\sum x_t y_t}{\sum x_t^2}$$

$$\hat{\beta}_1 = \frac{\text{Cov}(X_t, Y_t)}{\text{Var}(X_t)} \quad (T-1)'s \text{ cancel})$$

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

define

$$a_{t_1} = \frac{X_t}{\sum X_t^2}$$

So that

$$\hat{\beta}_r = \sum a_{t,r} y_t = \sum a_{t,r} y_t$$

$$\hat{\beta}_0 = \sum a_{10} \gamma_t$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= \frac{\sum Y_t}{T} - \bar{X} \sum a_t, Y_t = \sum \underbrace{\left(\frac{1}{T} - \bar{X} a_t \right)}_{a_{t0}} Y_t$$

note that

$$\sum a_{t1} = \frac{\sum x_t}{\sum x_t^2} = 0 \quad \text{since } \sum x_t = 0$$

$$\begin{aligned} \sum a_{t1} y_t &= \sum a_{t1} (y_t - \bar{y}) = \sum a_{t1} y_t - \underbrace{\bar{y} \sum a_{t1}}_{0''} \\ &= \sum a_{t1} y_t \end{aligned}$$

$$\sum a_{t1} x_t = \frac{\sum x_t^2}{\sum x_t^2} = 1$$

$$\begin{aligned} \sum a_{t1} x_t &= \sum a_{t1} (x_t + \bar{x}) = \sum a_{t1} x_t + \bar{x} \sum a_{t1} \\ &= \sum a_{t1} x_t \end{aligned}$$

$$\sum a_{t0} = \sum \left(\frac{1}{T} - \bar{x} a_{t1} \right) = 1 - \bar{x} \sum a_{t1} = 1$$

$$\begin{aligned} \sum a_{t0} x_t &= \sum \left(\frac{1}{T} - \bar{x} a_{t1} \right) x_t \\ &= \bar{x} - \bar{x} \underbrace{\sum a_{t1} x_t}_{=1} = \bar{x} - \bar{x} = 0 \end{aligned}$$

Linearity

$$\hat{\beta}_i = \sum a_{ti} y_t, \quad i=0,1$$

$$\hat{\beta}_1 = a_{11} y_1 + a_{21} y_2 + \dots + a_{T1} y_T$$

$\hat{\beta}_i$ is a linear function of y_t

$\hat{\beta}_i$ is a linear estimator

Unbiasedness

definition: An estimator $\hat{\theta}$ is an unbiased estimator if $E(\hat{\theta}) = \theta$

$\hat{\theta}$ converge to the true population value θ on average

$$E(\hat{\theta}) - \theta = \text{biasedness}$$

consistency: bias $\rightarrow 0$ as $T \rightarrow \infty$

$$\hat{\beta}_1 = \sum a_t, Y_t$$

$$= \sum a_t, (\beta_0 + \beta_1 X_t + u_t)$$

$$= \beta_0 \underbrace{\sum a_t}_0 + \beta_1 \underbrace{\sum a_t X_t}_1 + \sum a_t u_t$$

$$\hat{\beta}_1 = \beta_1 + \sum a_t u_t$$

$$E(\hat{\beta}_1) = \beta_1 + \sum a_t \underbrace{E(u_t)}_{=0}$$

$$E(\hat{\beta}_1) = \beta_1 : \hat{\beta}_1 \Rightarrow \text{unbiased}$$

Variance

* hitting the target on average is a desired property, but we may also will not to deviate from the target each time



E.A.



E.T.

An estimator may be unbiased, but still may be unreliable

⇒ Variance of the estimator?

$\text{Var}(\hat{\beta}_i)$

$$\begin{aligned}\text{Var}(\hat{\beta}_i) &= E(\hat{\beta}_i - E(\hat{\beta}_i))^2 = E(\hat{\beta}_i - \beta_i)^2 \\ &= E\left[\sum a_{ti} u_t\right]^2 \\ &= \sigma^2 \sum a_{ti}^2\end{aligned}$$

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \sum a_{t1}^2 = \frac{\sigma^2}{\sum x_t^2} = \sigma_{\hat{\beta}_1}^2$$

$$\text{Var}(\hat{\beta}_0) = \sigma_{\hat{\beta}_0}^2 = \frac{\sigma^2 \sum x_t^2}{T \sum x_t^2}$$

Estimated Variances

$$\text{Est. Var}(\hat{\beta}_1) = \hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{\sum x_t^2}$$

$$\text{Est. Var}(\hat{\beta}_0) = \hat{\sigma}_{\hat{\beta}_0}^2 = \frac{\hat{\sigma}^2 \sum x_t^2}{T \sum x_t^2}$$

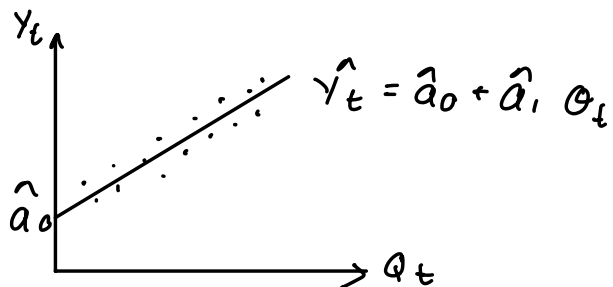
where $\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{T-k-1}$

The Coefficient of Determination (R^2)

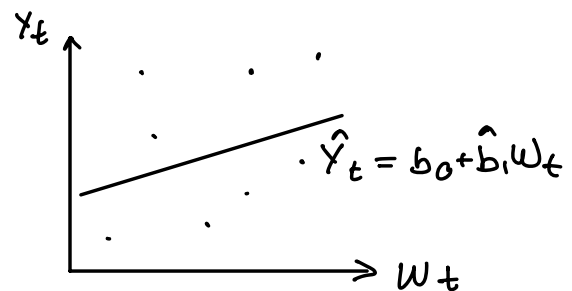
two alternative models to explain y

$$\text{Model A: } Y_t = a_0 + a_1 Q_t + U_{t1}$$

$$\text{Model B: } Y_t = b_0 + b_1 W_t + U_{t2}$$



Model A



Model B

Let,

$$Y_t - \bar{Y} = y_t$$

$$Y_t - \hat{Y}_t = \hat{U}_t$$

$$\hat{Y}_t - \bar{Y} = \hat{y}_t$$

Then,

the deviation of Y_t from the mean $\bar{Y}$

$$y_t = \hat{U}_t + \hat{y}_t \rightarrow \text{explained}$$

$\downarrow \qquad \downarrow$
total unexplained

$$y_t = \hat{u}_t + \hat{y}_t$$

$$\sum y_t^2 = \sum \hat{u}_t^2 + \sum \hat{y}_t^2 + 2 \sum \hat{u}_t \hat{y}_t$$

$$\begin{aligned} \sum \hat{u}_t \hat{y}_t &= \sum \hat{u}_t (\hat{y}_t - \bar{y}) \\ &= \underbrace{\sum \hat{u}_t \hat{y}_t}_{=0} - \bar{y} \underbrace{\sum \hat{u}_t}_{=0} \end{aligned}$$

└ unexplained ($\hat{u}_t$) & explained ($\hat{y}_t$) are not correlated

then

$$\underbrace{\sum y_t^2}_{\text{TSS}} = \underbrace{\sum \hat{u}_t^2}_{\substack{\text{SSR} \\ \text{RSS}}} + \underbrace{\sum \hat{y}_t^2}_{\text{ESS}}$$

TSS: Total Sum of Squares

(The variation of y around $\bar{y}$)

SSR: Sum of the Squared Residuals

Part of the model that is NOT explained by model

RSS: Residual Sum of Squares

ESS: Explained Sum of Squares

(the variation of the systematic part around $\bar{y}$)

$$TSS = SSR + ESS$$

divide both sides by TSS

$$\frac{TSS}{TSS} = \frac{SSR}{TSS} + \frac{ESS}{TSS}$$

$$\frac{\sum y_t^2}{\sum y_t^2} = \frac{\sum \hat{u}_t^2}{\sum y_t^2} + \boxed{\frac{\sum \hat{y}_t^2}{\sum y_t^2}} \rightarrow R^2$$

$\underbrace{\frac{\sum y_t^2}{\sum y_t^2}}_{=1}$

$$R^2 = \frac{\sum \hat{y}_t^2}{\sum y_t^2} \quad \text{or} \quad 1 - \frac{\sum \hat{u}_t^2}{\sum y_t^2}$$

$$0 \leq R^2 \leq 1$$

$$R^2 = 1 - \frac{\sum \hat{u}_t^2}{\sum y_t^2} = 1 - \frac{\sum \hat{u}_t^2}{\sum \hat{y}_t^2 + \sum \hat{u}_t^2}$$

$$\text{If } \sum \hat{u}_t^2 = 0 \rightarrow R^2 = 1$$

$$\text{If } \sum \hat{y}_t^2 = 0 \rightarrow R^2 = 1$$

$$R^2 = 1 \rightarrow \text{Perfect fit,}$$

all observations fit regression line