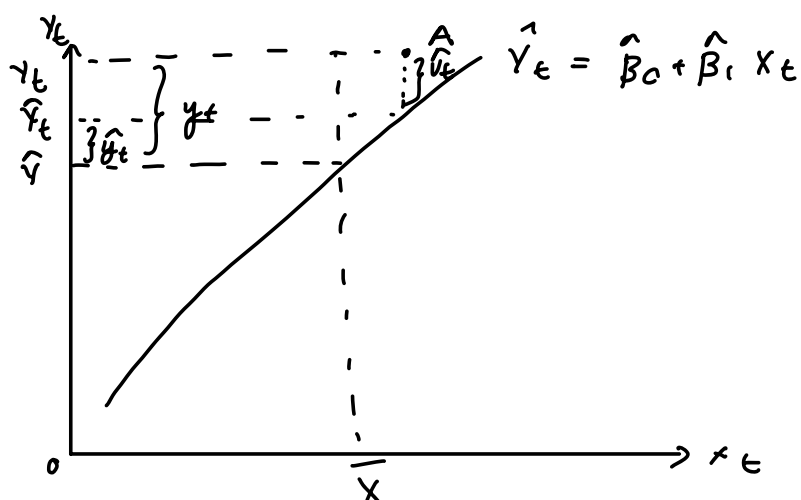


## Goodness of Fit

How good is the fit of  $\hat{Y}$  to the scatter of observations on  $Y$  vs  $X$



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

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

we have 3 different distances

1- Distance  $Y_t$  from origin (0)

$$Y_t - 0 = Y_t$$

2- Distance of  $Y_t$  from its sample mean

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

3- Distance of  $Y_t$  from fitted regression line

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

Distance to use to form a measure of goodness of fit is variation of  $Y_t$  around its sample mean

$$\sum_t (Y_t - \bar{Y})^2 = \sum_t y_t^2$$

part explained  
by  $\hat{Y}_t$   
( $Y_t - \bar{Y} = \hat{y}_t$ )

part explained  
by  $u_t$

$$\begin{aligned}\sum_t y_t^2 &= \sum_t (\hat{y}_t + \hat{u}_t)^2 = \sum_t \hat{y}_t^2 + \sum_t \hat{u}_t^2 + 2 \sum_t \hat{y}_t \hat{u}_t \\ &= \sum_t \hat{y}_t^2 + \sum_t \hat{u}_t^2 \quad (\text{divide both sides by } \sum_t y_t^2)\end{aligned}$$

coefficient of determination  $R^2 = \sum_t \hat{y}_t^2 / \sum_t y_t^2 =$   
 $1 - (\sum_t \hat{u}_t^2 / \sum_t y_t^2)$  ,  $0 \leq R^2 \leq 1$

$R^2 = 0 \Rightarrow$  observation points are scattered around regression line purely randomly

$R^2 = 1 \Rightarrow$  all observations are on regression line   
↳ perfect fit

low values of  $R^2 \rightarrow$  a poor fit of regression

high values of  $R^2 \rightarrow$  a good fit of regression

$R^2 = 0.99 \rightarrow$  maybe a good fit  
↳ spurious regression/multicollinearity

$$R^2 = \frac{[T \sum x_t y_t - (\sum x_t)(\sum y_t)]^2}{[T \sum x_t^2 - (\sum x_t)^2][T \sum y_t^2 - (\sum y_t)^2]}$$

$R^2 = r_{\hat{y}\hat{y}}^2$        $\sqrt{r_{\hat{y}\hat{y}}^2} \rightarrow$  does not lie in interval  $(-1, 1)$

$0 \leq \sqrt{r_{\hat{y}\hat{y}}^2} \leq 1$        $-1 \leq \sqrt{r_{xy}^2} \leq 1$

Finite Sample Properties of  $\hat{\beta}_0$  &  $\hat{\beta}_1$

Define a variable

$$a_{t1} = x_t / \sum_t x_t^2$$

$$\hat{\beta}_1 = \frac{\sum_t x_t y_t}{\sum_t x_t^2} = \frac{\sum_t a_{t1} y_t}{\sum_t x_t^2}$$

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

$$\hat{\beta}_1 = \sum_t a_{t1} (y_t - \bar{y}) = \sum_t a_{t1} y_t - \bar{y} \sum_t a_{t1} = \underbrace{\sum_t a_{t1} y_t}_{\text{linearity}}$$

Since  $\sum_t x_t = 0 \Rightarrow \sum_t a_{t1} = 0$

For  $\hat{\beta}_0$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \sum_t Y_t / T - \hat{\beta}_1 (\sum_t X_t / T) = \sum_t Y_t T^{-1} - \hat{\beta}_1 \sum_t X_t T^{-1}$$

$$= \sum_t y_t T^{-1} - \sum_t a_{t1} y_t \left( \sum_t x_t T^{-1} \right)$$

$$= \sum_t y_t T^{-1} - \sum_t a_{t1} y_t (\bar{x}) = \sum_t y_t T^{-1} - \bar{x} \sum_t a_{t1} y_t$$

$$= \sum_t \underbrace{(T^{-1} - \bar{X} a_{t1})}_{a_{t0}} y_t = \sum_t a_{t0} y_t \rightarrow \text{linearity}$$

we show that  $\hat{\beta}_0$  &  $\hat{\beta}_1$  belong to set of linear estimators

For unbiasedness of  $\hat{\beta}_0$  &  $\hat{\beta}_1$

Let  $\tilde{\beta}_1 = \sum_t \frac{c_{t1} Y_t}{c_{t0} c_{t1}}$  be any linear estimator of two regression coefficients

$$Y_t = \beta_0 + \beta_1 X_t + U_t \rightarrow \text{substitute this into } \tilde{\beta}_i$$

$$\tilde{\beta}_i = \sum_t c_{ti} (\beta_0 + \beta_1 x_t + v_t), \quad i=0,1$$

$$= \sum_t c_{t1} \beta_0 + \sum_t c_{t1} \beta_1 X_t + \sum_t c_{t1} u_t \quad E(\tilde{\beta}_i) = \tilde{\beta}_i$$

$$E(\tilde{\beta}_i) = \beta_0 \sum_t c_{ti} + \beta_1 \sum_t c_{ti} X_t$$

In order for  $\tilde{\beta}_1$  to be unbiased

$$\sum_t c_{t0} = 1 \quad \sum_t c_{t0} \bar{x}_t = 0$$

$$\sum_t c_{t1} = 0 \quad \sum_t c_{t1} x_t = 1$$

$$\sum_t a_{t1} = \sum_t x_t / \sum_t x_t^2 = 0$$

$$\sum_t a_{t1} x_t = \sum_t a_{t1} x_t = \sum_t x_t^2 / \sum_t x_t^2 = 1$$

$$\text{since } \sum_t a_{t1} = \sum_t x_t / \sum_t x_t^2 \Rightarrow \sum_t a_{t1} x_t = \sum_t x_t^2 / \sum_t x_t^2$$

$$\sum_t a_{t0} = 1 - \bar{x} \underbrace{\sum_t a_{t1}}_0 = 1$$

$$\text{since } a_{t0} = T^{-1} - \bar{x} a_{t1}$$

$$\sum_t a_{t0} x_t = \sum_t x_t (T^{-1} - \bar{x} a_{t1}) = \bar{x} - \bar{x} \underbrace{\sum_t a_{t1} x_t}_1 = 0$$

$\rightarrow a_{t1}$  satisfy these conditions of unbiasedness

the least square estimators  $\hat{\beta}_i$  are linear and unbiased estimators of  $\beta_i$

For efficiency

$$\tilde{\beta}_i = \beta_i + \sum_t c_{ti} u_t$$

$$\text{Var}(\tilde{\beta}_i) = \hat{\sigma}^2 \tilde{\beta}_i^2 = E[(\tilde{\beta}_i - \beta_i)]^2 = E[(\sum_t c_{ti} u_t)^2]$$

$$E(u_t^2) = \sigma^2$$

$$\text{Var}(\tilde{\beta}_i) = \sigma^2 \sum_t c_{ti}^2$$

$$\begin{aligned}\sigma^2 \hat{\beta}_1 &= \sigma^2 \sum_t a_{t1}^2 = \sigma^2 \cancel{\sum_t x_t^2} / (\sum_t x_t^2)^2 \\ &= \sigma^2 / \sum_t x_t^2 \\ \sigma^2 \hat{\beta}_0 &= \sigma^2 \sum_t a_{t0}^2 = \sigma^2 \sum_t (T^{-1} - \bar{X} a_{t1})^2 \\ &= \sigma^2 (T^{-1} + \bar{X}^2 \sum_t a_{t1}^2) = \sigma^2 (T^{-1} + (X^2 / \sum_t x_t^2)) \\ \text{Since } \sigma^2 \sum_t a_{t1}^2 &= \sigma^2 / \sum_t x_t^2\end{aligned}$$

$$\sigma_{\hat{\beta}_1}^2 = \sigma^2 \sum_t x_t^2 / T \sum_t x_t^2 \quad \text{since } \sum_t x_t^2 = \sum_t x_t^2 - T \bar{X}^2$$

HW: Show Efficiency of  $\hat{\beta}_0$  &  $\hat{\beta}_1$

consider difference b.w.  $\tilde{\beta}_i$  &  $\hat{\beta}_i$

$$\tilde{\beta}_i = \hat{\beta}_i = \sum_t \underbrace{(c_{ti} - a_{t1})}_{\parallel d_{ti}} y_t = \sum_t d_{ti} y_t$$

$$\sigma_{\tilde{\beta}_i}^2 = \underbrace{\sigma^2 \sum_t a_{ti}^2}_{\sigma_{\hat{\beta}_i}^2} + \sigma^2 \sum_t d_{ti}^2 + 2\sigma^2 \sum_t a_{ti} d_{ti} \quad i=0,1$$

$$\begin{aligned}\sum_t a_{t0} d_{t0} &= \sum_t (T^{-1} \bar{X} a_{t1}) d_{t0} - \\ &= T^{-1} \sum_t d_{t0} - \bar{X} \sum_t a_{t1} d_{t0}\end{aligned}$$

$$\sum_t d_{t0} = \sum_t a_{t0} - \sum_t c_{t0} = 1 - 1 = 0 \quad \text{for unbiasedness}$$

$$\sum_t a_{t1} d_{t0} = (\sum_t x_t c_{t0} - \sum_t x_t a_{t0}) / \sum_t x_t^2 = 0$$

$$\text{Since } \sum_t x_t c_{t0} = \sum_t x_t c_{t0} - \bar{X} \underbrace{\sum_t c_{t0}}_{=0} = -\bar{X} \sum_t x_t a_{t0} = -\bar{X}$$

$$\sum_t a_{t0} d_{t0} = 0$$

$$\sum_t = t, = \frac{\sum_t x_t}{\sum_t x_t^2} = 0$$

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

$$\text{Since } a_{t0} = T^{-1} - \bar{x} a_{t1}$$

$$\begin{aligned} \sum_t a_{t0} x_t &= \sum_t x_t (T^{-1} - \bar{x} a_{t1}) \\ &= \bar{x} - \bar{x} \sum_t \underbrace{a_{t1} x_t}_{1} = 0 \end{aligned}$$

Hence,  $a_{t1}$  satisfies conditions we have noted before

Thus, LS estimators are linear, unbiased & efficient estimators  
 $\hat{\beta}_i$  are BLUE (best, linear, unbiased estimators)

$$\sigma^2 \Rightarrow \hat{\sigma}^2 \quad u_t \Rightarrow \hat{u}_t$$

Estimating  $\sigma^2$

If disturbance  $u_t$  were observable a quadratic unbiased estimator of  $\sigma^2$ , it would be obtained as

$$S^2 = \sum_t (u_t - \bar{u})^2 / T-1$$

since  $u_t$ 's are unobservable, we need sample counter parts,  $\hat{u}_t$

$$\sum_t (\hat{u}_t - \bar{\hat{u}})^2 = \sum_t \hat{u}_t^2 \quad \text{since } \bar{\hat{u}} = T^{-1} \sum_t \hat{u}_t = 0$$

$$\hat{\sigma}^2 = \sum_t \hat{u}_t^2 / T-2 \quad \rightarrow y_t = \beta_0 + \beta_1 x_t - u_t$$

$$\sigma^2 = \sum_t a_{t1}^2 / (T-k-1) \quad \begin{matrix} \rightarrow \text{\# of explanatory} \\ \text{(independent variables)} \end{matrix}$$

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \beta_2 z_t + u_t \\ (T-2-1) &= T-3 \end{aligned}$$

HW: show that  $\hat{\sigma}^2$  is an unbiased estimator of  $\sigma^2$

$$\hat{\sigma}_{\beta_0}^2 = \hat{\sigma}^2 \sum_{t=0}^T x_t^2 / T \sum_{t=0}^T x_t^2 / \hat{\sigma}_{\beta_1}^2 = \hat{\sigma}^2 / \sum_{t=0}^T x_t^2$$

## Interval Estimation

for  $\beta_i$ ,  $i=0,1$

$$P[\hat{\beta}_i - \hat{\sigma}_{\hat{\beta}_i} t_{T-2, \alpha/2} \leq \beta_i \leq \hat{\beta}_i + \hat{\sigma}_{\hat{\beta}_i} t_{T-2, \alpha/2}] = 1-\alpha$$

for  $\sigma^2$ ,

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

$T-K-1$

$$P[(T-2) \hat{\sigma}^2 / \chi^2_{T-2, 1-\alpha/2} \leq \hat{\sigma}^2 \leq [(T-2) \hat{\sigma}^2 / \chi^2_{T-2, \alpha/2}] = 1-\alpha$$

## Test of Hypothesis

$$H_0: \beta_i = \beta_i^0$$

$$H_1: \beta_i \neq \beta_i^0 \Rightarrow \text{Reject } H_0 \text{ if } (\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} > t_{T-2, \alpha/2}$$

$$\text{or } \beta_1 < \beta_i^0 \Rightarrow \text{Reject } H_0 \text{ if } (\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} \leq -t_{T-2, \alpha}$$

$$\text{or } \beta_i > \beta_i^0 \Rightarrow \text{Reject } H_0 \text{ if } (\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} > t_{T-2, \alpha}$$

$$T = (\hat{\beta}_i - \beta_i^0) / \hat{\sigma}_{\hat{\beta}_i} \sim t_{T-2, \alpha}$$

$\sim t_{T-2, \alpha/2}$

## Test of Significance $\Rightarrow$ will be in the final

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

$$H_0: \beta_1 = 0$$

$$H_0: \beta_0 = 0 \rightarrow \text{Not reject } H_0$$

$$H_1: \beta_1 \neq 0$$

$$H_1: \beta_0 \neq 0 \quad Y_t = \beta_i X_t + U_t$$

$$Y_t = \beta_0$$

$$H_0: \beta_i = 0 \quad \left\{ \begin{array}{l} \text{reject } H_0 \text{ if} \\ t = \hat{\beta}_i / \hat{\sigma}_{\hat{\beta}_i} > t_{T-2, \alpha/2} \end{array} \right.$$

$$\text{or} \\ H_1: \beta_i \neq 0 \quad \left\{ \begin{array}{l} \text{reject } H_0 \text{ if} \\ t = \hat{\beta}_i / \hat{\sigma}_{\hat{\beta}_i} \geq -t_{T-2, \alpha} \end{array} \right.$$

$$\text{or} \\ H_1: \beta_i > 0 \quad \left\{ \begin{array}{l} \text{reject } H_0 \text{ if} \\ t = \hat{\beta}_i / \hat{\sigma}_{\hat{\beta}_i} > t_{T-2, \alpha} \end{array} \right.$$

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + U_t$$

$$Y_t = \beta_0 + \beta_1 k_t + \beta_2 C_t + U_t$$

$$H_0: \beta_1 + \beta_2 = 1 \rightarrow \text{constant elasticity of substitution}$$

let's say we have such a hypothesis

$$H_0: w_0 \beta_0 + w_1 \beta_1 = q \Rightarrow SSR_R \rightarrow S_R \rightarrow \text{restricted}$$

$$H_1: w_0 \beta_0 + w_1 \beta_1 \neq q \Rightarrow SSR_U \rightarrow S_U \rightarrow \text{unrestricted}$$

$$Q = [(S_R - S_U) / S_U] [(T-2) / p]$$

$$p \Rightarrow \# \text{ of restrictions}$$

we reject  $H_0$  if

$$Q > F_{p, T-2}^{\alpha}$$

$\Rightarrow$  for test of significance for slope of coeff ( $\hat{\beta}_1$ )

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$Q = [R^2 / (1-R^2)] (T-2) \sim F_{1, T-2}$$

this test is equivalent to t-test of significance

Joint Test of Significance