CONSIDER $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$. $(k=3)$

OLS ESTIMATION: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$.

SOLUTION

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 .$$

$$\hat{\beta}_2 = \frac{\left(\sum_{i=1}^{n} y_i x_{2i}\right)\left(\sum_{i=1}^{n} x_{3i}^2\right) - \left(\sum_{i=1}^{n} y_i x_{3i}\right)\left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)}{\left(\sum_{i=1}^{n} x_{2i}^2\right)\left(\sum_{i=1}^{n} x_{3i}^2\right) - \left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)^2} .$$

$$\hat{\beta}_3 = \frac{\left(\sum_{i=1}^{n} y_i x_{3i}\right)\left(\sum_{i=1}^{n} x_{2i}^2\right) - \left(\sum_{i=1}^{n} y_i x_{2i}\right)\left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)}{\left(\sum_{i=1}^{n} x_{2i}^2\right)\left(\sum_{i=1}^{n} x_{3i}^2\right) - \left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)^2}$$

HOW ABOUT VARIANCE OF THE ESTIMATED PARAMETERS?

$$\mathrm{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2\left(\sum_{i=1}^{n} x_{3i}^2\right) + \bar{X}_3^2\left(\sum_{i=1}^{n} x_{2i}^2\right) - 2\bar{X}_2\bar{X}_3 \sum_{i=1}^{n} x_{2i}x_{3i}}{\left(\sum_{i=1}^{n} x_{2i}^2\right)\left(\sum_{i=1}^{n} x_{3i}^2\right) - \left(\sum_{i=1}^{n} x_{2i} x_{3i}\right)^2}\right]$$

$$\mathrm{var}(\hat{\beta}_2) = \left[\frac{\left(\sum x_{3i}^2\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2}\right]\sigma^2$$

$$\mathrm{var}(\hat{\beta}_3) = \left[\frac{\left(\sum x_{2i}^2\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2}\right]\sigma^2$$

ALTERNATIVE FORM:

$$\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 \left(1 - r_{23}^2\right)} \qquad \sigma^2 \rightarrow \text{UNKNOWN}$$

WHERE $r_{23}^2 = \dfrac{\left(\sum x_{2i} x_{3i}\right)^2}{\sum x_{2i}^2 \sum x_{3i}^2}$.

$$\mathrm{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 \left(1 - r_{23}^2\right)} \qquad \sigma^2 \text{ — UNKNOWN}$$

ESTIMATED BY $\hat{\sigma}^2$

NOTE: YOU CAN GET $se(\hat{\beta}_2) = \sqrt{\mathrm{var}(\hat{\beta}_2)}$ AND

$se(\hat{\beta}_3) = \sqrt{\mathrm{var}(\hat{\beta}_3)}$.

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}} \qquad \left(= \sum \hat{u}_i \hat{u}_i\right)$$

WHERE $\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$.

$$\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\,\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum x_{2i}^2}\cdot\sqrt{\sum x_{3i}^2}} .$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\,\sigma^2}{(1 - r_{23}^2)\sqrt{\sum x_{2i}^2}\cdot\sqrt{\sum x_{3i}^2}}.$$

PROPERTIES OF OLS ESTIMATORS IN MULTIPLE REGRESSION

① THE ESTIMATED REGRESSION LINE PASSES POINT

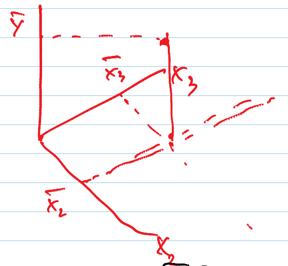
$(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k, \bar{Y})$ (PASSES THE MEAN)

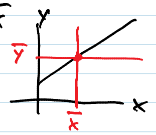
(WHY ?)

SINCE $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \ldots + \hat{\beta}_k \bar{X}_k$.

FOR $k = 3 \Rightarrow \bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$.

FOR $k = 2 \Rightarrow$ WE HAVE $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$





(2 DIMENSIONAL SPACES)

$(\bar{X}, \bar{Y})$ IS A FAMILY MEMBER OF THE REGRESSION LINE.

② $\bar{Y} = \bar{\hat{Y}}$

(AVERAGE VALUE) MEAN OF ACTUAL Y

(AVERAGE VALUE) MEAN OF ESTIMATED Y

③ THE SUM AND MEAN OF THE RESIDUALS ARE 0 :

$$\sum_{i=1}^{n} \hat{u}_i = 0 \quad \text{AND} \quad \bar{\hat{u}} = \frac{\sum_{i=1}^{n} \hat{u}_i}{n} = 0$$

MEAN OF $\hat{u}_i$

④ THE COVARIANCE BETWEEN EACH EXPLANATORY VARIABLE

AND THE OLS RESIDUAL IS 0 :

$$\sum_{i=1}^{n} \hat{u}_i X_{ij} = 0$$

i.e., $\sum_{i=1}^{n} \hat{u}_i X_{2i} = 0$ AND $\sum_{i=1}^{n} \hat{u}_i X_{3i} = 0$.

(THE RESIDUALS $\hat{u}_i$ ARE UNCORRELATED WITH X_{2i} AND X_{3i})

⑤ THE COVARIANCE BETWEEN ESTIMATED DEPENDENT VARIABLE AND THE OLS RESIDUAL IS 0 .

$$\sum_{i=1}^{n} \hat{u}_i \hat{Y}_i = 0 .$$

⑥ FROM THE VARIANCE EQUATIONS ABOVE, NAMELY

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

AND

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

YOU CAN SEE THAT, FOR A GIVEN σ^2, $\sum x_{2i}^2$, AND $\sum x_{3i}^2$, variances OF $\hat{\beta}_2$ AND OF $\hat{\beta}_3$ WILL RISE AS r_{23}^2 (= CORRELATION BETWEEN X_{2i} AND X_{3i}) INCREASES

WORSE CASE : IN LIMIT, WHEN $r_{23}^2 \to 1$ (PERFECT COLLINEARITY), THESE VARIANCES BECOME INDEFINITE !!!

$$r_{23}^2 \longrightarrow \begin{matrix} \operatorname{var}(\hat{\beta}_2) \to \operatorname{se}(\hat{\beta}_2) \to t_{\hat{\beta}_2} \\ \operatorname{var}(\hat{\beta}_3) \to \operatorname{se}(\hat{\beta}_3) \to t_{\hat{\beta}_3} \end{matrix}$$

HYPOTHESIS TESTING ~~WILL~~ BE MISLEADING (TYPE I

$$X_{3i} = 2X_{2i} + 3$$ EXAMPLE OF

$\hat{\beta}_3$ WILL BE MISLEADING (TYPE I ERROR)

ACCEPT H_0 WHEN IT IS ACTUALLY FALSE.

$$X_{3i} = 2X_{2i} + 3$$

OR

$$X_{3i} = 4X_{2i}$$

EXAMPLE OF PERFECT LINEAR RELATIONSHIP BETWEEN X_{2i} AND X_{3i}

(PROBLEM OF MULTICOLLINEARITY)

H - A - M

(7) GIVEN r_{23}^2, THE VARIANCE OF $\hat{\beta}_2$ AND $\hat{\beta}_3$ "VARIES" DIRECTLY PROPORTIONAL TO σ^2 AND INVERSELY PROPORTIONAL TO $\sum x_{2i}^2$ AND $\sum x_{3i}^2$.

(8) GIVEN THE ASSUMPTIONS OF THE CLRM, ONE CAN PROVE THAT

$\hat{\beta}_2$ AND $\hat{\beta}_3$ ARE BLUE.

i.e.,
- THEY ARE "UNBIASED" [$E(\hat{\beta}_2) = \beta_2$ AND $E(\hat{\beta}_3) = \beta_2$]
- THEY HAVE "MINIMUM VARIANCE" IN THE CLASS OF ALL LINEAR UNBIASED ESTIMATORS.

25.10.12 (MAKE UP)

<u>MULTIPLE COEFFICIENT OF DETERMINATION</u> (R^2)

WHEN $k = 3$. (INCLUDING INTERCEPT TERM)

$$Y_i = \underbrace{\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}}_{\hat{Y}_i} + \hat{u}_i \qquad (\text{OR } Y_i = \hat{Y}_i + \hat{u}_i)$$

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

IT CAN BE EXPRESSED AS

$$y_i = \underbrace{\hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}} + \hat{u}_i$$

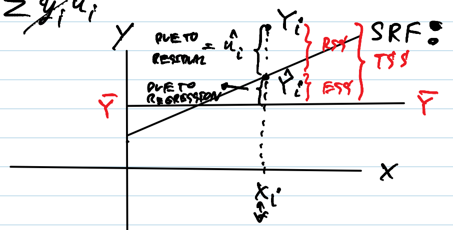
$$y_i = \hat{y}_i + \hat{u}_i$$ [IN DEVIATION FORM]

$$\sum y_i^2 = \sum (\hat{y}_i + \hat{u}_i)^2$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + \cancel{2 \sum \hat{y}_i \hat{u}_i} \rightarrow = 0$$

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$TSS = ESS + RSS$$



SRF: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$

I $\sum y_i^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$: TOTAL VARIATION OF THE ACTUAL Y_i FROM ITS AVERAGE VALUE (TOTAL SUM OF SQUARES: TSS)

II $\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2$: TOTAL VARIATION OF THE ESTIMATED Y VALUES FROM ITS AVERAGE VALUE (EXPLAINED SUM OF SQUARES: ESS)

III $\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$: UNEXPLAINED PART OF THE REGRESSION OR UNEXPECTED VARIATION OF THE Y VALUES FROM THE REGRESSION (RESIDUAL SUM OF SQUARES: RSS)

$$\text{I} = \text{II} + \text{III}$$

$$TSS = ESS + RSS$$

$$\boxed{R^2 = \frac{ESS}{TSS} = \frac{\text{EXPLAINED VARIATION}}{\text{TOTAL VARIATION}}}$$

FROM $TSS = ESS + RSS$

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

$$1 = R^2 + \frac{RSS}{TSS}$$

RSS → UNEXPLAINED VARIATION :(

TSS → TOTAL VARIATION

SO $R^2 = 1 - \frac{RSS}{TSS}$ OR $= 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n y_i^2}$

NOW, OTHER FORM OF RSS AND R^2...

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n \hat{u}_i \hat{u}_i = \sum_{i=1}^n \hat{u}_i (Y_i - \hat{Y}_i) = \sum_{i=1}^n \hat{u}_i (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})$$

$$= \sum_{i=1}^n \hat{u}_i (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})$$

$$= \sum_{i=1}^n \hat{u}_i y_i .$$

RSS

$$= \sum_{i=1}^n (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) y_i$$

THEREFORE $\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n y_i^2 - \hat{\beta}_2 \sum_{i=1}^n y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^n y_i x_{3i}$.

$$\underbrace{\sum_{i=1}^n y_i^2}_{TSS} = \underbrace{\hat{\beta}_2 \sum_{i=1}^n y_i x_{2i} + \hat{\beta}_3 \sum_{i=1}^n y_i x_{3i}}_{ESS} + \underbrace{\sum_{i=1}^n \hat{u}_i^2}_{RSS} .$$

CONSEQUENTLY, $R^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}_2 \sum_{i=1}^n y_i x_{2i} + \hat{\beta}_3 \sum_{i=1}^n y_i x_{3i}}{\sum_{i=1}^n y_i^2}$

NOTE: WHEN $k = 3$.

PROPERTIES OF R^2

① $0 < R^2 < 1$

② AS # OF EXPLANATORY VARIABLES INCREASES,

R^2 WILL "ALWAYS" INCREASE.

(SOME RESEARCHERS MIGHT HAVE A TEMPTATION TO IMPROVE R^2 BY ADDING MORE EXPLANATORY VARIABLES INTO THE MODEL. THIS IS UNETHICAL ACADEMICALLY.)

WAY OUT ?

USE ADJUSTED R^2 INSTEAD.

$$\text{ADJUSTED } R^2 \ (\bar{R}^2) = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2 / (n-k)}{\sum_{i=1}^{n} y_i^2 / (n-1)} .$$

OR $$\bar{R}^2 = 1 - (1-R^2)\frac{(n-1)}{(n-k)} .$$

EX: IF MORE X_{ki} ARE ADDED, FOR EXAMPLE

IF X_{4i} CANNOT EXPLAIN Y AT ALL $\rightarrow$ R^2 DO NOT $\downarrow$

BUT $\bar{R}^2$ WILL $\downarrow$.

SINCE $$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)}$$ # OF PARAMETERS

OF OBSERVATION

IT CAN BE COMPUTED BY

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} .$$

NOTE: TO COMPARE R^2 BETWEEN TWO MODEL, THE NUMBER OF OBSERVATION AND THE LEFT-HAND-SIDE VARIABLE MUST BE THE SAME !!!

VARIABLE MUST BE THE SAME !!!

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad n = 60 \qquad \Rightarrow R^2$$

$$\ln Y_i = \gamma_1 + \gamma_2 X_{2i} + \gamma_3 X_{3i} + v_i \qquad n = 60 \qquad \Rightarrow R^2$$

TABLE 7.2 Raw Data for Comparing Two R^2 Values

Year	Y_t (1)	$\hat{Y}_t$ (2)	$\widehat{\ln Y_t}$ (3)	Antilog of $\widehat{\ln Y_t}$ (4)	$\ln Y_t$ (5)	$\ln(\hat{Y}_t)$ (6)
1970	2.57	2.321887	0.843555	2.324616	0.943906	0.842380
1971	2.50	2.336272	0.853611	2.348111	0.916291	0.848557
1972	2.35	2.345863	0.860544	2.364447	0.854415	0.852653
1973	2.30	2.341068	0.857054	2.356209	0.832909	0.850607
1974	2.25	2.326682	0.846863	2.332318	0.810930	0.844443
1975	2.20	2.331477	0.850214	2.340149	0.788457	0.846502
1976	2.11	2.173233	0.757943	2.133882	0.746688	0.776216
1977	1.94	1.823176	0.627279	1.872508	0.662688	0.600580
1978	1.97	2.024579	0.694089	2.001884	0.678034	0.705362
1979	2.06	2.115689	0.731282	2.077742	0.722706	0.749381
1980	2.02	2.130075	0.737688	2.091096	0.703098	0.756157

Notes: Column (1): Actual Y values from Table 7.1.
Column (2): Estimated Y values from the linear model (7.8.8).
Column (3): Estimated log Y values from the double-log model (7.8.9).
Column (4): Antilog of values in column (3).
Column (5): Log values of Y in column (1).
Column (6): Log values of $\hat{Y}_t$ in column (2).

TO COMPARE R^2 OR ADJUSTED R^2 BETWEEN THE TWO MODELS ABOVE, WE HAVE TWO OPTIONS:

EX:

$$\hat{Y}_i = 2.6911 - 0.4795\, X_i \qquad R^2 = 0.6628. \quad \text{— MODEL 1}$$

$$\widehat{\ln Y_i} = 0.7774 - 0.253 \ln X_i \qquad R^2 = 0.7448 \quad \text{— MODEL 2}$$

IF YOU WANT TO COMPARE R^2 BETWEEN THE TWO MODELS, THERE ARE TWO OPTIONS:

OPTION 1:
1. TAKE ANTILOG OF $\widehat{\ln Y_t}$ (COLUMN 4)
2. CALCULATE A NEW R^2 BY USING INFORMATION IN COLUMN 4 AND COLUMN 1.

$$\text{A NEW } r^2 = \frac{\left(\sum y_i \hat{y}_i\right)^2}{\left(\sum y_i^2\right)\left(\sum \hat{y}_i^2\right)} = \ldots$$

$$(\Sigma y_i^2)(\Sigma \hat{y}_i^2)$$

BRING THIS TO COMPARE W/ R^2 IN THE FIRST MODEL.

OPTION 2 ⇒ 1 TAKE LOG OF ACTUAL Y VALUES, SO YOU WILL GET $\ln Y_i$ (COLUMN 5)

2 OBTAIN THE ESTIMATED Y_i VALUES: $\hat{Y}_i$ AND TAKE LOG OF $\hat{Y}_i$ → YOU WILL GET COLUMN 6 [$\ln(\hat{Y}_i)$]

3 CACULATE A NEW R^2 BETWEEN $\ln Y_i$ AND $\ln \hat{Y}_i$ BY USING THE FORMULA ABOVE:

$$r^2 = \frac{(\Sigma y_i \hat{y}_i)^2}{(\Sigma y_i^2)(\Sigma \hat{y}_i^2)} = \ldots\ldots$$

A NEW R^2

BRING TO COMPARE W/ R^2 FROM MODEL 2

EXAMPLE OF MULTIPLE REGRESSION MODEL.

Y = CHILD MORTALITY (CM)

X_2 = PER CAPITA GNP (PGNP)

X_3 = FEMALE LITERACY RATE (FLR)

$$CM_i = \beta_1 + \beta_2 PGNP_i + \beta_3 FLR_i + u_i.$$

(%)

$$\hat{CM}_i = 263.6416 - 0.0056\, PGNP_i - 2.2316\, FLR_i$$

$$se = (11.5932) \quad (0.0019) \quad (0.2099)$$

$R^2 = 0.7077$

$R^2 = 0.7077 \Rightarrow$ 70% OF THE VARIATION IN CM COULD BE EXPLAINED BY PGNP AND FLR

$\frac{ESS}{TSS}$

263.6416 $\Rightarrow$ IF VALUES OF PGNP AND FLR RATE WERE FIXED AT ZERO, ON AVERAGE, CM RATE WOULD BE ABOUT 263 DEATHS PER THOUSAND LIVE BIRTHS.

$-0.0056 \Rightarrow$ AS PGNP INCREASES BY ONE DOLLAR, ON AVERAGE, THE NUMBER OF DEATHS OF CHILDREN UNDER THE AGE OF 5 GOES DOWN BY ABOUT 0.0056 PER THOUSAND LIVE BIRTHS.

OR

AS PGNP ↑ BY 1,000 DOLLARS, CM RATE WILL GO DOWN BY ABOUT $0.0056 \times 1000 = 5.6$ PER THOUSAND LIVE BIRTHS, ON AVERAGE

$-2.2316 \Rightarrow$ HOLDING THE INFLUENCE OF PGNP CONSTANT, THE NUMBER OF DEATHS OF CHILDREN UNDER AGE 5 GOES DOWN BY ABOUT 2.23 PER 1000 LIVE BIRTHS IF FLR RISES BY 1 PERCENTAGE POINT.

THE COBB-DOUGLAS PRODUCTION FUNCTION

... BY 1 PERCENTAGE POINT.

THE COBB-DOUGLAS PRODUCTION FUNCTION

$Y = f(L, K)$

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \quad \text{--- (1)}$$

$(Y = A L^{\alpha} K^{\beta})$

$\alpha + \beta = 1 \rightarrow CRS$

$\alpha + \beta > 1 \rightarrow IRS$

$\alpha + \beta < 1 \rightarrow DRS$

Y = OUTPUT

X_2 = LABOUR INPUT

X_3 = CAPITAL INPUT

u_i = DISTURBANCE TERM

e = BASE OF NATURAL LOG

IF $\beta_2 + \beta_3 = 1$, THE PRODUCTION FUNCTION EXHIBITS CRS.

IF $\beta_2 + \beta_2 > 1$, " " IRS.

IF $\beta_2 + \beta_3 < 1$, " " DRS.

TAKE LOG ON EQUATION 1 THROUGHOUT:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$\ln Y_i = \beta_0 + \underbrace{\beta_2}_{\text{OUTPUT ELASTICITY OF LABOR}} \ln X_{2i} + \underbrace{\beta_3}_{\text{OUTPUT ELASTICITY OF CAPITAL}} \ln X_{3i} + u_i$$

$$\widehat{\ln OUTPUT_i} = 3.8876 + 0.4683 \ln LABOR_{2i} + 0.5213 \ln CAPITAL_{3i}$$

$t = (9.815) \quad (4.7342) \quad (5.3803)$

$t = \quad (9.815) \qquad (4.7342) \qquad (5.3803)$

$R^2 = 0.9642$

$\bar{R}^2 = 0.9627$

} $\bar{R}^2$ IS ALWAYS LESS THAN R^2.

$\hat{\beta}_2 + \hat{\beta}_3 = 0.4683 + 0.5213 = \boxed{\;\dots\;}$

$H_0: \beta_2 + \beta_3 = 1$ (CRS)

$H_0: \beta_2 + \beta_3 \neq 1$ (NOT CRS)

} LATER IN CHAPTER 8