TEST OF SEVERAL LINEAR RESTRICTIONS (CONTD)

CONSIDER A PARTITION OF X AS

$$X = [X_1 \; X_2]$$

& THE CORRESPONDING PARTITION OF β AS

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

THUS THE REGRESSION MODEL

$$y = X\beta + u$$ BECOMES

$$y = X_1\beta_1 + X_2\beta_2 + u$$

NOW CONSIDER THE RESTRICTIONS $\beta_2 = 0$.

THUS, $R = [0 \; I]$ & THE MATRIX FORM OF THE RESTRICTIONS IS

$$R\beta = [0 \; I]\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_2 = 0$$

SUPPOSE THAT THE RESTRICTED ESTIMATOR IS

$$\beta^* = \begin{bmatrix} \beta_1^* \\ 0 \end{bmatrix}$$

AND THE RESIDUALS FROM THE RESTRICTED MODEL IS

$$e^* = y - X\beta^*$$

$$= y - X\hat{\beta} - X\beta^* + X\hat{\beta} = (y - X\hat{\beta}) - X(\beta^* - \hat{\beta})$$

HENCE, $e^* = e - X(\beta^* - \hat{\beta})$.

FURTHER $e^{*\prime}e^* = e'e + (\beta^* - \hat{\beta})'X'X(\beta^* - \hat{\beta}) - 2(\beta^* - \hat{\beta})'X'e$

$$= e'e + (\beta^* - \hat{\beta})'X'X(\beta^* - \hat{\beta}) \quad [\text{BECAUSE } X'e = 0$$

HENCE, $e^{*\prime}e^* - e'e = (\beta^* - \hat{\beta})'X'X(\beta^* - \hat{\beta})$.

HOWEVER $\beta^* = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)$ IN THE GENERAL CASE.

HENCE $(\beta^* - \hat{\beta})'X'X(\beta^* - \hat{\beta})$

$$= (R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)$$

$$= (R\hat{\beta} - q)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q) \sim \chi^2_p$$

(AS SEEN BEFORE).

THUS
$$\frac{\left[e^{*\prime}e^{*} - e'e\right]/p}{\dfrac{e'e}{n-k}} \sim F_{p,\,n-k}$$

HERE p IS THE # OF COLUMNS OF X_2 (i.e. THE NUMBER OF VARIABLES EXCLUDED).

DEFINE $e'e = SSE_U$ (UNRESTRICTED)

& $e^{*\prime}e^{*} = SSE_R$ (RESTRICTED)

THEN
$$F = \frac{(SSE_R - SSE_U)/p}{\dfrac{SSE_U}{n-k}}$$

$$= \frac{\left(\dfrac{SSE_R}{SST} - \dfrac{SSE_U}{SST}\right)/p}{\dfrac{SSE_U/SST}{n-k}}$$

$$= \frac{\left[(1-R_R^2) - (1-R_U^2)\right]/p}{\dfrac{1-R_U^2}{n-k}}$$

$$= \frac{(R_U^2 - R_R^2)/p}{\dfrac{1-R_U^2}{n-k}} \sim F_{p,\,n-k}.$$

HERE R_U^2 = GOODNESS OF FIT OF UNRESTRICTED MODEL

R_R^2 = GOODNESS OF FIT OF RESTRICTED MODEL.

R^2 & $\bar{R}^2$:

GOODNESS OF FIT OF A MODEL IS MEASURED BY THE

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

FOR A GIVEN SAMPLE THE VALUES OF THE DEPENDENT VARIABLE ARE GIVEN & DONOT CHANGE WITH THE CHOICE OF EXPLANATORY VARIABLES. THUS SST (i.e. $\sum (Y_i - \bar{Y})^2$) IS CONSTANT. BUT AS THE NUMBER OF EXPLANATORY VARIABLES INCREASES SSE WILL EITHER DECLINE OR (AT WORST) REMAIN UNCHANGED. HENCE R^2 CAN ONLY INCREASE (STAY THE SAME) AS THE NUMBER OF X VARIABLES INCREASES. THERE IS, THUS, A BUILT IN BIAS IN FAVOR OF A MORE ELABORATE MODEL.

THEIL PROPOSED THE ADJUSTED R^2:

$$\bar{R}^2 = 1 - \frac{\sum e_i^2 / n-k}{\sum (Y_i - \bar{Y})^2 / n-1} = 1 - \frac{SSE / n-k}{SST / n-1}$$

HERE BOTH SSE & SST ARE SCALED BY THE RESPECTIVE DEGREES OF FREEDOM.

AS k INCREASES, $\bar{R}^2$ WILL INCREASE ONLY ~~OF~~ IF $SSE / n-k$ FALLS ($SST / n-1$ DOES NOT DEPEND ON k).

AS k INCREASES, SSE FALLS. BUT $n-k$ ALSO FALLS. THE RATIO FALLS ONLY IF SSE DECLINES FASTER THAN $n-k$.

NOTE $\frac{SSE}{n-k} = \hat{\sigma}^2$. $\bar{R}^2$ ~~FALL~~ INCREASES WHEN $\hat{\sigma}^2$ FALLS. IN GENERAL AS k INCREASES $\bar{R}^2$ MAY EITHER RISE OR FALL.

THE MODEL WITH A HIGHER $\bar{R}^2$ IS REGARDED

$$\bar{R}^2 = 1 - \left(\frac{SSE}{SST}\right)\left(\frac{n-1}{n-k}\right)$$

$$= 1 - (1-R^2)\left(\frac{n-1}{n-k}\right)$$

$$\Rightarrow \quad 1-\bar{R}^2 = (1-R^2)\left(\frac{n-1}{n-k}\right)$$

$$\Rightarrow \quad \frac{1-R^2}{1-\bar{R}^2} = \frac{n-k}{n-1} < 1 \qquad (\text{BECAUSE } k \geq 2).$$

$$\Rightarrow \quad 1-R^2 < 1-\bar{R}^2 \Rightarrow \boxed{\bar{R}^2 < R^2}$$

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

BUT $\bar{R}^2$ CAN BE NEGATIVE.

RELATION BETWEEN R^2, $\bar{R}^2$, & F:

RECALL THAT FOR THE SIGNIFICANCE OF OVERALL REGRESSION

$$F = \frac{R^2/k-1}{(1-R^2)/n-k} \Rightarrow \frac{1-R^2}{R^2} = \frac{n-k}{n-1}\cdot\frac{1}{F}$$

$$\frac{1}{R^2} = 1 + \left(\frac{n-k}{n-1}\right)\frac{1}{F}.$$

ASIDE: t, F & R^2 IN THE 2 VARIABLE REGRESSION

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X})$$

$$\sum(\hat{Y}_i - \bar{Y})^2 = (\hat{\beta}_2)^2 \sum(X_i - \bar{X})^2$$

FOR THE $H_0: \beta_2 = 0$

$$t = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} \Rightarrow t^2 = \frac{(\hat{\beta}_2)^2}{\widehat{Var}(\hat{\beta}_2)}$$

$$= \frac{(\hat{\beta}_2)^2}{\left[\frac{\sum e_i^2 / n-2}{\sum (x_i - \bar{x})^2}\right]} = \left[\frac{(\hat{\beta}_2)^2 \sum (x_i - \bar{x})^2}{(\sum e_i^2)}\right] n-2$$

$$t^2 = \left(\frac{SSR}{SSE}\right) n-2 = \left[\frac{SSR/SST}{SSE/SST}\right] n-2$$

$$= \left[\frac{R^2}{1-R^2}\right](n-2)$$

$$\frac{1-R^2}{R^2} = \frac{n-2}{t^2}$$

$$\frac{1}{R^2} - 1 = \frac{n-2}{t^2} \Rightarrow \frac{1}{R^2} = \frac{n-2}{t^2} + 1 = \frac{(n-2)+t^2}{t^2}$$

$$\Rightarrow \boxed{R^2 = \frac{t^2}{t^2 + (n-2)}}$$

UNRESTRICTED MODEL : $y = x_1\beta_1 + x_2\beta_2 + u_2$ (1)

RESTRICTED MODEL : $y = x_1\beta_1 + u$ (2)

SUPPOSE THAT (1) HAS k EXPLANATORY VARIABLES & (2) HAS k_1 EXPLANATORY VARIABLES. THUS, (2) EXCLUDES

$k_2 = k - k_1$ VARIABLES.

$H_0: \beta_2 = 0$

$$F = \frac{(SSE_R - SSE_U)/k_2}{SSE_U / n-k}$$

REJECT H_0 IF $F > F^*_{k_2, n-k}$.

NOW CONSIDER $\bar{R}^2$ CRITERION.

SUPPOSE $\bar{R}^2_U > \bar{R}^2_R$.

$$1 - \frac{SSE_U/n-k}{SST/n-1} > 1 - \frac{SSE_R/n-k_1}{SST/n-1}$$

$$\Rightarrow \frac{SSE_R/n-k_1}{SST/n-1} > \frac{SSE_U/n-k}{SST/n-1}$$

$$\Rightarrow \frac{SSE_R}{n-k_1} > \frac{SSE_U}{n-k}$$

$$\Rightarrow \frac{SSE_R}{SSE_U} > \frac{n-k_1}{n-k}$$

$$\Rightarrow \frac{SSE_R}{SSE_U} - 1 > \frac{n-k_1}{n-k} - 1$$

$$\Rightarrow \frac{SSE_R - SSE_U}{SSE_U} > \frac{k-k_1}{n-k} = \frac{k_2}{n-k}$$

$$\Rightarrow \frac{(SSE_R - SSE_U)/k_2}{SSE_U/n-k} > 1.$$

THUS $\bar{R}^2_U > \bar{R}^2_R \Rightarrow F > 1$.

HENCE THE $\bar{R}^2$ CRITERION SELECTS THE UNRESTRICTED MODEL WHENEVER $F > 1$ IRRESPECTIVE OF THE DEGREES OF FREEDOM.