$H_0: \beta_{(1)} = \beta_{(2)}$ vs $H_1: \beta_{(1)} \neq \beta_{(2)}$

$H_0: \gamma = 0$ vs $H_1: \gamma \neq 0$.

Chow Test of Predictive Stability:

Suppose we have n_1 obs for the model: $Y_{(1)} = X_{(1)}\beta_{(1)} + U_{(1)}$
where $\beta_{(1)} = (k \times 1)$ vector for beta coeffs.

Consider $n_2 < k$ "out of sample" obs & we want to test whether the parameters are consistent across the two datasets of size $n_1 > k$ & $n_2 < k$ respectively.

RM:

$$Y_{(1)} = X_{(1)}\beta + U_{(1)}$$
$$Y_{(2)} = X_{(2)}\beta + U_{(2)}$$
$$\Rightarrow \begin{bmatrix} \overset{n_1 \times 1}{Y_{(1)}} \\ \underset{n_2 \times 1}{Y_{(2)}} \end{bmatrix}_{(n_1+n_2)\times 1} = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix} [\beta] + \begin{bmatrix} U_{(1)} \\ U_{(2)} \end{bmatrix}$$

On fitting RM, using ANOVA, we will obtain $\boxed{RRSS}$.

UM:

$$Y_{(1)} = X_{(1)}\beta_{(1)} + U_{(1)} \quad \text{----- (i)}$$
$$Y_{(2)} = X_{(2)}\beta_{(2)} + U_{(2)} \quad \text{--- (ii)}$$

where $\beta_{(1)} = (k \times 1)$ coeff vector for pre-covid.
$\beta_{(2)} = (k \times 1)$ coeff vector for post-covid.

$\therefore$ Here, we want to test $H_0: \beta_{(1)} = \beta_{(2)}$ vs $H_1: \beta_{(1)} \neq \beta_{(2)}$

Note: Here the issue is that as $n_2 < k$, $\hat{\beta}_{(2)}$ cannot be estimated directly.

(i) can be estimated and $\hat{\beta}_{(1)}$ can be obtained. & using ANOVA we can obtain RSS_1.

From (ii)

$$Y_{(2)} = X_{(2)}\beta_{(2)} + U_{(2)}$$
$$Y_{(2)} = X_{(2)}\beta_{(2)} + \left\{X_{(2)}\beta_{(1)} - X_{(2)}\beta_{(1)}\right\} + U_{(2)}$$
$$Y_{(2)} = X_{(2)}\beta_{(1)} + \left(X_{(2)}\left[\beta_{(2)} - \beta_{(1)}\right]\right) + U_{(2)}$$
$$Y_{(2)} = X_{(2)}\beta_{(1)} + \gamma + U_{(2)}$$

where $\underset{(n_2 \times 1)}{\gamma} = \underset{n_2 \times k}{X_{(2)}} \left[\overset{k\times 1}{\beta_{(2)}} - \overset{k \times 1}{\beta_{(1)}}\right]_{k \times 1}$

If $\gamma = 0 \Rightarrow \beta_{(2)} = \beta_{(1)}$ i.e the coeff vector is consistent across both the datasets.

$\therefore$ Our testing becomes: $H_0: \gamma = 0$ vs $H_1: \gamma \neq 0$.

. Summarizing:

$$Y_{(1)} = X_{(1)}\beta_{(1)} + U_{(1)} \quad \text{RSS}_1$$

$$Y_{(2)} = X_{(2)}\beta_{(1)} + \gamma + U_{(2)} \quad \text{RSS}_2$$

$\rightarrow$ Final set of eqns to estimate $\beta_{(1)}$ & γ

$\boxed{URSS = RSS_1 + RSS_2}$

$$\begin{bmatrix} Y_{(1)} \\ Y_{(2)} \end{bmatrix} = \underbrace{\begin{bmatrix} X_{(1)} & 0 \\ X_{(2)} & I_{n_2} \end{bmatrix}}_{\text{combined coeff matrix } (X)} \begin{bmatrix} \beta_{(1)} \\ \gamma \end{bmatrix} + \begin{bmatrix} U_{(1)} \\ U_{(2)} \end{bmatrix}$$

We will estimate $\hat{\beta}_{(1)}$ and $\hat{\gamma}$:

$$\underbrace{\begin{bmatrix} X_{(1)}' & X_{(2)}' \\ 0 & I_{n_2} \end{bmatrix}\begin{bmatrix} X_{(1)} & 0 \\ X_{(2)} & I_{n_2} \end{bmatrix}}\begin{bmatrix} \hat{\beta}_{(1)} \\ \hat{\gamma} \end{bmatrix} = \underbrace{\begin{bmatrix} X_{(1)}' & X_{(2)}' \\ 0 & I_{n_2} \end{bmatrix}\begin{bmatrix} Y_{(1)} \\ Y_{(2)} \end{bmatrix}}$$

$$\begin{bmatrix} X_{(1)}'X_{(1)} + X_{(2)}'X_{(2)} & X_{(1)}' \\ X_{(2)} & I_{n_2} \end{bmatrix}\begin{bmatrix} \hat{\beta}_{(1)} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X_{(1)}'Y_{(1)} + X_{(2)}'Y_{(2)} \\ Y_{(2)} \end{bmatrix}$$

PRF: $Y = \tilde{X}\beta + U$

$\hat{\beta} = (X'X)^{-1}X'Y$

$(X'X)\hat{\beta} = X'Y$.

SRF: $\hat{Y} = X\hat{\beta}$

$Y = \hat{Y} + e$.

$Y = X\hat{\beta} + e$.

$\therefore$ Now, $\hat{\beta}_{(1)}$ is estimated.

Writing the 2nd eqn from above: $X_{(2)}\hat{\beta}_{(1)} + \hat{\gamma} = Y_{(2)}$.

$$\boxed{\hat{\gamma} = Y_{(2)} - X_{(2)}\hat{\beta}_{(1)}} \quad \text{---- (ii)}$$

$\therefore$ Now, writing the results in the $Y = \hat{Y} + e = X\hat{\beta} + e$ format:-

$$\begin{bmatrix} Y_{(1)} \\ Y_{(2)} \end{bmatrix} = \begin{bmatrix} X_{(1)} & 0 \\ X_{(2)} & I_{n_2} \end{bmatrix}\begin{bmatrix} \hat{\beta}_{(1)} \\ \hat{\gamma} \end{bmatrix} + \begin{bmatrix} e_{(1)} \\ e_{(2)} \end{bmatrix}$$

Writing the 2nd eqn from above:

$$Y_{(2)} = X_{(2)}\hat{\beta}_{(1)} + \hat{\gamma} + e_{(2)}$$

From (ii) $\cancel{Y_{(2)}} = \cancel{X_{(2)}\hat{\beta}_{(1)}} + \cancel{Y_{(2)}} - \cancel{X_{(2)}\hat{\beta}_{(1)}} + e_{(2)} \Rightarrow e_{(2)} = 0$.

$RSS_2 = e_{(2)}'e_{(2)} = 0$.

$\Rightarrow e_{(2)} = 0.$

$$RSS_2 = e_{(2)}' e_{(2)} = 0.$$

$$\therefore \quad URSS = RSS_1 + RSS_2 = RSS_1.$$

$$\boxed{\text{Our Test statistic is, } F = \frac{(RRSS - URSS)/q}{URSS/(n-k)}.}$$

$\rightarrow$ standard test-statistic to be used for RM, UM Framework.

Here, $URSS = RSS_1$.

$q = n_2$ [$\because$ we are testing $H_0: \gamma = 0 \Rightarrow n_2 \times 1$ vector]

i) the no. of parameters estimated in UM. $= k + n_2$.

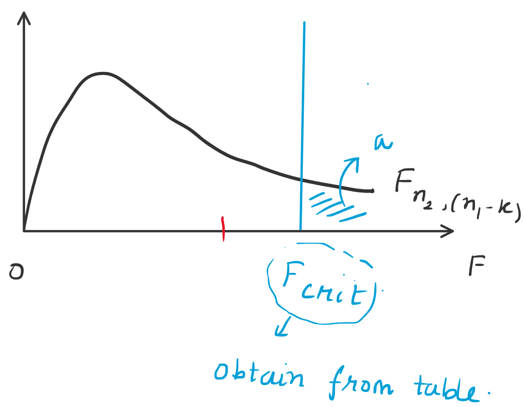
ii) No. of restrictions on the parameters $= n_2$.

iii) d.f of URSS $= (n_1 + n_2) - (k + n_2) = (n_1 - k)$

$$\therefore \text{ Test-statistic } \quad F = \frac{(RRSS - RSS_1)/n_2}{RSS_1/(n_1 - k)} \sim F_{n_2, (n_1 - k)}$$

We reject H_0 at $a\%$ l.o.s if

$F > F_{crit}$.



Q. Suppose an analyst collects data on 3 variables: Y, X_2, X_3 and creates a random sample of size n.

The analyst proposes 2 models for modelling Y:

Y	X_2	X_3
Y_1	X_{21}	X_{31}
Y_2	X_{22}	X_{32}
$\vdots$	$\vdots$	$\vdots$

$M_1: \quad Y_i = \beta_1 + \beta_2 X_{2i} + u_i \quad \text{---- [SLRM]}$

$M_2: \quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad [\dots]$

⋮	⋮	⋮
Y_n	X_{2n}	X_{3n}

$M_2: Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \cdots$ [CLRM]

How to determine which model is a better fit to the given data.

Goodness of Fit Measure $R^2 = \dfrac{ESS}{TSS}$.

Fit $M_1 \rightarrow$ Obtain R_1^2

Fit $M_2 \rightarrow$ Obtain R_2^2

$\rightarrow$ Compare b/w them, the model with the higher R^2 will be a fit to the data.

However, as the no. of explanatory variables increase, R^2 increases as ESS increases (without any consideration to whether the added explanatory variables are significant or not).

Hence, in the above case as M_2 has 2 explanatory variables, & M_1 has 1, $R_2^2 \geq R_1^2$. Hence R^2 cannot be used to choose the better model that fits the model.

$\therefore$ (i) If 2 or more competing models have the same no. of explanatory variables $\Rightarrow$ Goodness of Fit can be assessed using R^2 measure

(ii) If 2 or more competing models have different no. of explanatory variables $\Rightarrow$ Goodness of Fit can be assessed using "Adjusted R^2" measure.

In General, Adjusted R^2 $(\bar{R}^2) = 1 - \dfrac{RSS/(n-k)}{TSS/(n-1)}$.