STRUCTURAL CHANGE & DUMMY VARIABLES

SUPPOSE THAT THE OBSERVATIONS IN THE SAMPLE CAN BE GROUPED INTO 2 CATEGORIES: GROUP 1 (e.g. MALE) & GROUP 2 (e.g. FEMALE).

SUPPOSE

$$E(Y|X=X_i) = \beta_1 + \beta_2 X_i \quad \text{(FOR GROUP 1)}$$

$$E(Y|X=X_i) = \alpha_1 + \beta_2 X_i \quad \text{(FOR GROUP 2)}$$

THE REGRESSIONS ~~FUNCTIONS~~ CAN BE WRITTEN AS:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{(GROUP 1)}$$

$$Y_i = \alpha_1 + \beta_2 X_i + u_i \quad \text{(GROUP 2)}$$

NOTE: (A) THE POPULATION REGRESSION FUNCTIONS ARE PARALLEL LINES

(B) DISTRIBUTION OF u_i IS THE SAME FOR BOTH GROUPS.

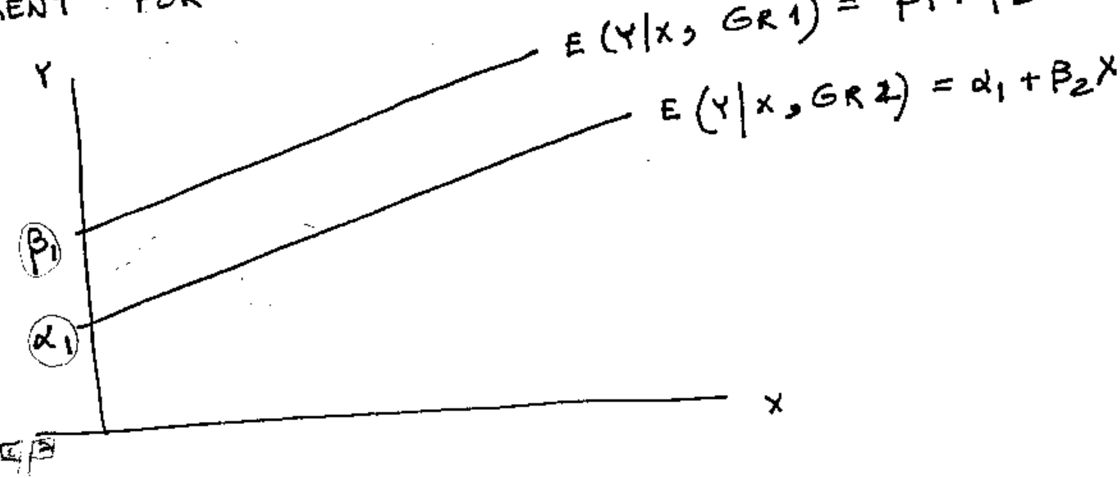
DEFINE $\delta = \alpha_1 - \beta_1 \Rightarrow \alpha_1 = \beta_1 + \delta$

THEN

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{(GR1)}$$

$$Y_i = \beta_1 + \delta + \beta_2 X_i + u_i \quad \text{(GR2)}$$

IF $\delta \neq 0$ FOR THE SAME LEVEL OF $X (= X_i)$, $E(Y|X=X_i)$ IS DIFFERENT FOR OBSERVATIONS FROM THE 2 DIFFERENT GROUPS.



DEFINE A BINARY VARIABLE D.

$D_i = 0$ FOR ANY OBSERVATION FROM GROUP 1

$D_i = 1$ FOR ANY OBSERVATION FROM GROUP 2.

NOW CONSIDER THE REGRESSION MODEL

$$Y_i = \beta_1 + \delta D_i + \beta_2 x_i + u_i$$

THAT APPLIES TO <u>BOTH</u> GROUPS.

<u>GROUP 1</u>

$$E(Y \mid x = x_i, \text{GROUP } 1) = E(Y \mid x = x_i, D_i = 0)$$
$$= \beta_1 + \delta . 0 + \beta_2 x_i$$
$$= \beta_1 + \beta_2 x_i$$

<u>GROUP 2:</u>

$$E(Y \mid x = x_i, \text{GROUP } 2) = E(Y \mid x = x_i, D_i = 1)$$
$$= \beta_1 + \delta . 1 + \beta_2 x_i$$
$$= (\beta_1 + \delta) + \beta_2 x_i = \alpha_1 + \beta_2 x_i$$

THIS IS A STANDARD 3-VARIABLE REGRESSION & A 't' TEST OF SIGNIFICANCE OF THE COEFFICIENT δ CONSTITUTES A TEST OF DIFFERENCE IN THE INTERCEPT BETWEEN THE 2 GROUPS [<u>SLOPES ARE IDENTICAL BY ASSUMPTION</u>]

A SPECIAL CASE

$Y_i = \beta_1 + u_i$ (GROUP 1)

$Y_i = \beta_1 + \delta + u_i$ (GROUP 2)

$Y_i = \beta_1 + \delta D_i + u_i$

$D_i = 0$ FOR GROUP 1

$D_i = 1$ FOR GROUP 2

ASSUME THAT n_1 OBSERVATIONS ARE FROM GROUP 1. $n_2 = n - n_1$ OBSERVATION ARE FROM GROUP 2.

THE OLS NORMAL EQNS ARE

$$\boxed{\begin{aligned} n\hat{\beta}_1 + \left(\sum D_i\right)\hat{\beta}_2 &= \sum Y_i \\ \left(\sum D_i\right)\hat{\beta}_1 + \left(\sum D_i^2\right)\hat{\beta}_2 &= \sum D_i Y_i \end{aligned}}$$

$$\sum D_i = \sum_{i \in \text{GROUP 1}} (D_i = 0) + \sum_{i \in \text{GROUP 2}} (D_i = 1) = n_2$$

$$\sum D_i^2 = \sum_{i \in GR1} (D_i = 0)^2 + \sum_{i \in GR2} (D_i = 1)^2 = n_2$$

$$n\hat{\beta}_1 + n_2\hat{\beta}_2 = \sum Y_i = n_1\overline{Y}_1 + n_2\overline{Y}_2$$

$$n_2\hat{\beta}_1 + n_2\hat{\beta}_2 = n_2\overline{Y}_2$$

$$(n - n_2)\hat{\beta}_1 = n_1\hat{\beta}_1 = n_1\overline{Y}_1$$

$$\Rightarrow \hat{\beta}_1 = \overline{Y}_1$$

$$n_2\hat{\beta}_2 = n_2\left(\overline{Y}_2 - \overline{Y}_1\right) \Rightarrow \boxed{\hat{\beta}_2 = \overline{Y}_2 - \overline{Y}_1}$$

BOTH PARAMETERS (INTERCEPT & SLOPE) DIFFERENT

GROUP 1 : $E(Y|X=x_i) = \beta_1 + \beta_2 x_i$

GROUP 2 : $E(Y|X=x_i) = \alpha_1 + \alpha_2 x_i$

DEFINE $\delta_1 = \alpha_1 - \beta_1 \Leftrightarrow \alpha_1 = \beta_1 + \delta_1$

$\delta_2 = \alpha_2 - \beta_2 \Leftrightarrow \alpha_2 = \beta_2 + \delta_2$

THEN

GROUP 2 : $E(Y|X=x_i) = (\beta_1 + \delta_1) + (\beta_2 + \delta_2) x_i$

AS BEFORE DEFINE THE DUMMY VARIABLE D :

$D_i = 0$ FOR $i \in$ GROUP 1

$D_i = 1$ FOR $i \in$ GROUP 2 .

NOW CONSIDER THE MODEL :

$$Y_i = \beta_1 + \delta_1 D_i + \beta_2 x_i + \delta_2 (D_i x_i) + u_i$$

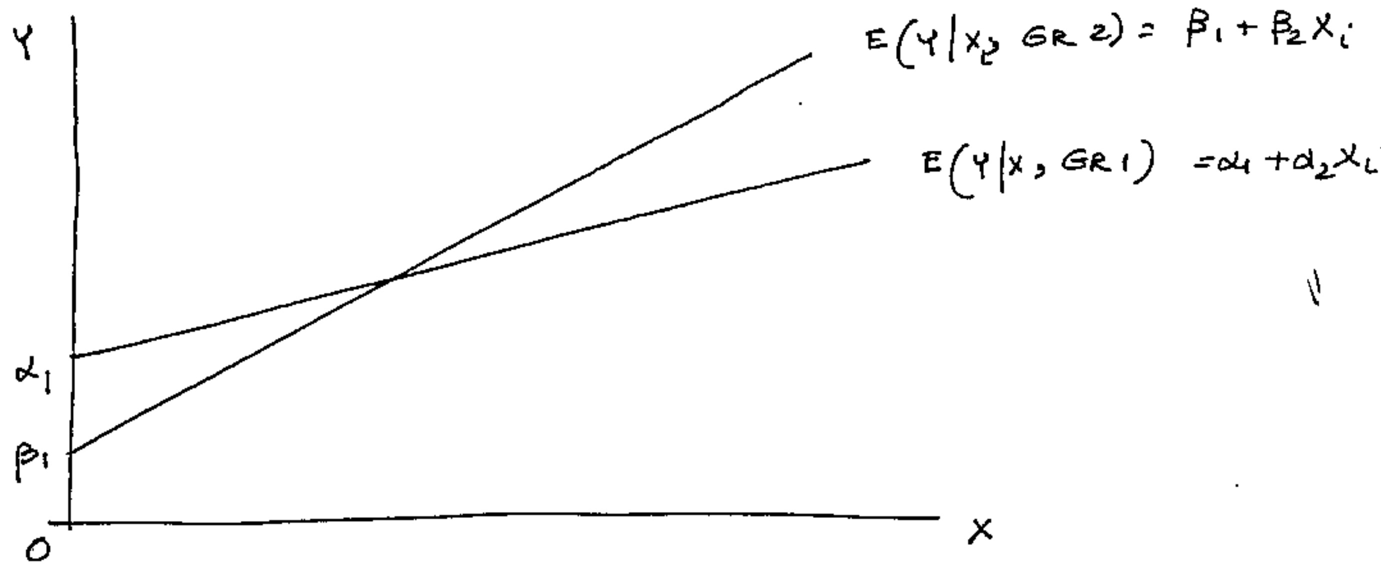
$$D_i x_i = (D_i)(x_i)$$

GROUP 1 : $D_i = 0$, $D_i x_i = 0$

GROUP 2 : $D_i = 1$; $D_i \cdot x_i = x_i$

$$E(Y|X=x_i, \text{GROUP 1}) = E(Y|X=x_i, D_i=0, D_i x_i = 0)$$
$$= \beta_1 + \delta_1 (0) + \beta_2 x_i + \delta_2 (0) =$$
$$\beta_1 + \beta_2 x_i$$

$$E(Y|X=x_i; \text{GROUP 2}) = E(Y|X=x_i, D_i=1, D_i x_i = x_i)$$
$$= \beta_1 + \delta_1 (1) + \beta_2 x_i + \delta_2 (x_i)$$
$$= (\beta_1 + \delta_1) + (\beta_2 + \delta_2) x_i$$
$$= \alpha_1 + \alpha_2 x_i$$



TEST OF NO STRUCTURAL DIFFERENCE

H_0: $\boxed{\begin{array}{c} \alpha_1 = \beta_1 \\ \& \ \alpha_2 = \beta_2 \end{array}} \iff \boxed{\begin{array}{c} \delta_1 = 0 \ \ \& \\ \delta_2 = 0 \end{array}}$

UNERSTRICTED MODEL:

$$U: \quad Y_i = \beta_1 + \delta_1 D_i + \beta_2 X_i + \delta_2 (D_i X_i) + u_i$$

$$R: \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$

TEST STATISTIC

$$F = \frac{(SSE_R - SSE_U)/2}{SSE_U / n-4} \sim F_{2,\, n-4}$$

n = # OF OBSERVATIONS.

1 CATEGORICAL VARIABLE WITH MULTIPLE CATEGORIES

SUPPOSE THAT INDIVIDUALS ARE GROUPED BY RACE:

GR 1 : WHITE
GR 2 : BLACK
GR 3 : HISPANIC & OTHERS.

3 PARALLEL REGRESSIONS:

$$Y_i = \beta_1 + \beta_2 x_i + u_i \quad \text{(WHITE)}$$
$$Y_i = \alpha_1 + \beta_2 x_i + u_i \quad \text{(BLACK)}$$
$$Y_i = \gamma_1 + \beta_2 x_i + u_i \quad \text{(HISPANIC)}$$

DEFINE

$$\delta_1 = \alpha_1 - \beta_1 \quad \text{(BLACK-WHITE DIFFERENCE)}$$
$$\delta_2 = \gamma_1 - \beta_1 \quad \text{(HISPANIC-WHITE DIFFERENCE)}$$

THEN

$$Y_i = \beta_1 + \beta_2 x_i + u_i \quad \text{(GR 1)}$$
$$Y_i = (\beta_1 + \delta_1) + \beta_2 x_i + u_i \quad \text{(GR 2)}$$
$$Y_i = (\beta_1 + \delta_2) + \beta_2 x_i + u_i \quad \text{(GR 3)}.$$

NOW CONSIDER 2 DUMMY VARIABLES

$$D_{1i} = 1 \quad \text{IF } i \in \text{GROUP 2}$$
$$= 0 \quad \text{IF } i \notin \text{GROUP 2}$$

$$D_{2i} = 1 \quad \text{IF } i \in \text{GROUP 3}$$
$$= 0 \quad \text{IF } i \notin \text{GROUP 3}.$$

	D_1	D_2
GROUP 1	0	0
GROUP 2	1	0
GROUP 3	0	1

NOW CONSIDER THE REGRESSION

$$Y_i = \beta_1 + \delta_1 D_{1i} + \delta_2 D_{2i} + \beta_2 x_i + u_i .$$

$$\boxed{\begin{aligned} E(Y_i \mid x = x_i, \text{ GROUP 1}) &= \beta_1 + \delta_1(0) + \delta_2(0) + \beta_2 x_i \\ &= \beta_1 + \beta_2 x_i . \end{aligned}}$$

$$\boxed{\begin{aligned} E(Y_i \mid x = x_i, \text{ GROUP 2}) &= \beta_1 + \delta_1(1) + \delta_2 .(0) + \beta_2 x_i \\ &= (\beta_1 + \delta_1) + \beta_2 x_i \\ &= \alpha_1 + \beta_2 x_i \end{aligned}}$$

$$\boxed{\begin{aligned} E(Y_i \mid x = x_i, \text{ GROUP 3}) &= \beta_1 + \delta_1(0) + \delta_2(1) + \beta_2 x_i \\ &= (\beta_1 + \delta_2) + \beta_2 x_i \\ &= \gamma_1 + \beta_2 x_i \end{aligned}}$$

H_0: $\delta_1 = \delta_2 = 0$ $\Rightarrow$ NO DIFFERENCE ACROSS GROUPS

H_0: $\delta_1 = 0$ $\Rightarrow$ BLACK-WHITE DIFFERENCE NOT 0 BUT WHITE-HISPANIC DIFFERENCE IS 0.

SEVERAL CATEGORICAL VARIABLES

SUPPOSE THAT INDIVIDUALS ARE CROSS-CLASSIFIED AS MALE/FEMALE & WHITE/NON-WHITE.

MALE - FEMALE DIFFERENCE $= \delta_1$

WHITE - NON-WHITE DIFFERENCE $= \delta_2$

$$E(Y \mid X = X_i, \text{ MALE, WHITE}) = \beta_1 + \beta_2 X_i$$

$$E(Y \mid X = X_i, \text{ FEMALE, WHITE}) = (\beta_1 + \delta_1) + \beta_2 X_i$$

$$E(Y \mid X = X_i, \text{ MALE, NON-WHITE}) = (\beta_1 + \delta_2) + \beta_2 X_i$$

$$E(Y \mid X = X_i, \text{ FEMALE, NON-WHITE}) = (\beta_1 + \delta_1 + \delta_2) + \beta_2 X_i$$

DEFINE THE DUMMY VARIABLES:

$$D_{1i} = 0 \quad \text{FOR MALES}$$
$$= 1 \quad \text{FOR FEMALES}$$

$$D_{2i} = 0 \quad \text{FOR WHITES}$$
$$= 1 \quad \text{FOR NON-WHITES}$$

	WHITE	NON-WHITE
MALE	$D_1 = 0$, $D_2 = 0$	$D_1 = 0$, $D_2 = 1$
FEMALE	$D_1 = 1$, $D_2 = 0$	$D_1 = 1$, $D_2 = 1$

NOW CONSIDER THE REGRESSION MODEL

$$Y_i = \beta_1 + \delta_1 D_{1i} + \delta_2 D_{2i} + \beta_2 X_i + u_i$$

$$E(Y \mid X = X_i, \text{ MALE}, \text{ WHITE})$$
$$= \beta_1 + \delta_1(0) + \delta_2(0) + \beta_2 X_i$$
$$= \beta_1 + \beta_2 X_i$$

$$E(Y \mid X = X_i, \text{ FEMALE}, \text{ WHITE})$$
$$= \beta_1 + \delta_1(1) + \delta_2(0) + \beta_2 X_i$$
$$= (\beta_1 + \delta_1) + \beta_2 X_i$$

$$E(Y \mid X = X_i, \text{ MALE}, \text{ NON-WHITE})$$
$$= \beta_1 + \delta_1(0) + \delta_2(1) + \beta_2 X_i$$
$$= (\beta_1 + \delta_2) + \beta_2 X_i$$

$$E(Y \mid X = X_i, \text{ FEMALE}, \text{ NON-WHITE})$$
$$= \beta_1 + \delta_1(1) + \delta_2(1) + \beta_2 X_i$$
$$= (\beta_1 + \delta_1 + \delta_2) + \beta_2 X_i$$

PIECEWISE LINEAR REGRESSION & SPLINE FUNCTIONS

CONSIDER A 2-VARIABLE REGRESSION MODEL:

$$Y_i = \alpha + \beta x_i + u_i$$

ASSUME THAT THERE IS A THRESHOLD LEVEL $x = x_0$ SUCH THAT

$$Y_i = \alpha_1 + \beta_1 x_i + u_i \quad \text{FOR} \quad x \le x_0$$

$$Y_i = \alpha_2 + \beta_2 x_i + u_i \quad \text{FOR} \quad x > x_0 .$$

THE STANDARD PROCEDURE IS TO DEFINE A DUMMY VARIABLE D:

$$D_i = 0 \quad \text{FOR} \quad x_i \le x_0$$

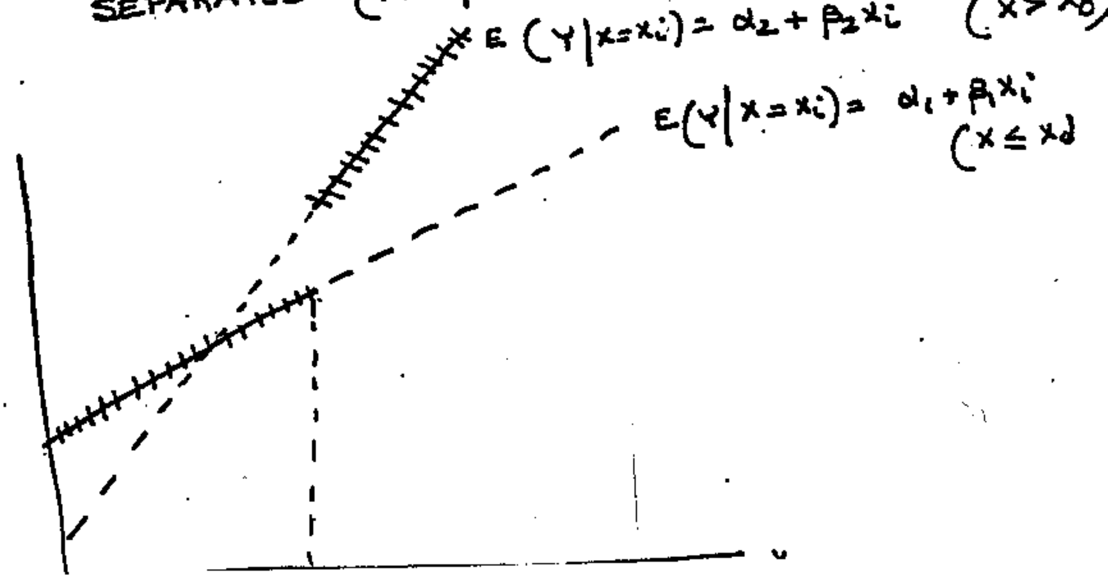
$$D_i = 1 \quad \text{FOR} \quad x_i > x_0$$

THE 2 SEPARATE REGRESSIONS CAN BE COMBINED AS:

$$Y_i = \alpha_1 + \delta_1 D_i + \beta_1 x_i + \delta_2 (D_i x_i) + u_i$$

(AS DESCRIBED BEFORE).

THE PROBLEM IS THAT THE 2 REGRESSION LINES ARE SEPARATED (RATHER THAN CONNECTED) AT $x = x_0$.



SUPPOSE THAT WE REQUIRE THE 2 REGRESSION LINES TO BE CONNECTED AT $X = X_0$. THEN

$$\alpha_1 + \beta_1 X_0 = \alpha_2 + \beta_2 X_0$$

$$\Rightarrow \quad \alpha_2 - \alpha_1 = -(\beta_2 - \beta_1) X_0$$

$$\Rightarrow \quad \boxed{\delta_1 = -\delta_2 X_0}$$

THE DUMMY VARIABLE REGRESSION BECOMES

$$Y_i = \alpha_1 - \delta_2 X_0 D_i + \beta_1 X_i + \delta_2 D_i X_i + u_i$$

$$\Rightarrow \quad \boxed{Y_i = \alpha_1 + \beta_1 X_i + \delta_2 D_i (X_i - X_0) + u_i}$$

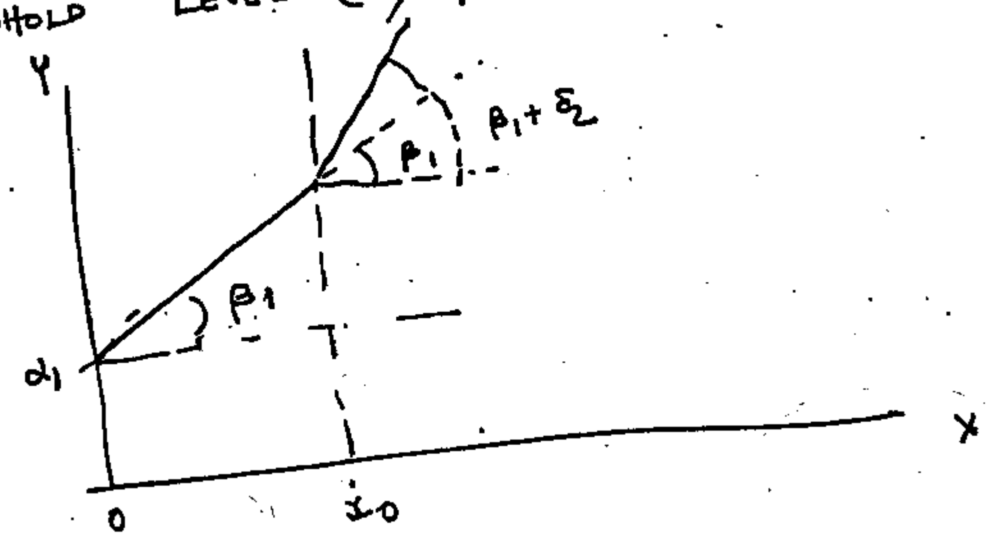
FOR $X_i \le X_0$, $D_i = 0 \Rightarrow$ (AS SPECIFIED)

$$Y_i = \alpha_1 + \beta_1 X_i + u_i$$

FOR $X_i > X_0$,

$$Y_i = \alpha_1 + \beta_1 X_i + \delta_2 (X_i - X_0) + u_i$$

HERE δ_2 IS THE INCREMENT IN THE MARGINAL EFFECT OF A CHANGE IN X ON Y THAT APPLIES ONLY TO THE PART OF X THAT EXCEEDS THE THRESHOLD LEVEL (X_0).



CHOW TEST OF STRUCTURAL CHANGE

CONSIDER THE 2 SEPARATE REGRESSIONS

$$Y_i = \alpha_1 + \beta_1 x_i + u_i \quad \text{FOR GROUP 1}$$
$$Y_i = \alpha_2 + \beta_2 x_i + u_i \quad \text{FOR GROUP 2}$$

NO STRUCTURAL BREAK IMPLIES

$$\alpha_1 = \alpha_2 \quad \& \quad \beta_1 = \beta_2$$

THAT IS, THE SAME REGRESSION APPLIES TO BOTH GROUP

IN THAT CASE,

$$Y_i = \alpha + \beta x_i + u_i$$

IS THE COMMON REGRESSION.

SEPARATE REGRESSIONS

GROUP 1 : n_1 OBSERVATIONS

$$y_1 = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_1} \end{pmatrix} \qquad X_1 = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_{n_1} \end{pmatrix}$$

PARAMETERS : $\delta_1 = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}$

GROUP-1 REGRESSION : $y_1 = X_1 \delta_1 + u_1$

GROUP 2 : n_2 OBSERVATIONS

$$y_2 = \begin{pmatrix} y_{n_1+1} \\ \vdots \\ y_n \end{pmatrix} \qquad X_2 = \begin{pmatrix} 1 & x_{n_1+1} \\ 1 & \\ 1 & x_n \end{pmatrix}$$

PARAMETERS $\delta_2 = \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}$

GROUP-2 REGRESSION : $y_2 = X_2 \delta_2 + u_2$

POOL THE REGRESSIONS:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

DEFINE $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$; $X = \begin{pmatrix} x_1 & 0 \\ 0 & X_2 \end{pmatrix}$; $\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$; $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$

THEN, THE MODEL IS

$$y = X\delta + u$$

$$X'X = \begin{pmatrix} x_1'x_1 & 0 \\ 0 & x_2'x_2 \end{pmatrix} \qquad X'y = \begin{pmatrix} x_1' & 0 \\ 0 & x_2' \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1'y_1 \\ x_2'y_2 \end{pmatrix}$$

$$\hat{\delta} = (X'X)^{-1}X'y$$

$$\Rightarrow \begin{pmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \end{pmatrix} = \begin{pmatrix} x_1'x_1 & 0 \\ 0 & x_2'x_2 \end{pmatrix}^{-1} \begin{pmatrix} x_1'y_1 \\ x_2'y_2 \end{pmatrix} = \begin{pmatrix} (x_1'x_1)^{-1} & 0 \\ 0 & (x_2'x_2)^{-1} \end{pmatrix} \begin{pmatrix} x_1'y_1 \\ x_2'y_2 \end{pmatrix}$$

THUS $$\hat{\delta} = \begin{pmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \end{pmatrix} = \begin{pmatrix} (x_1'x_1)^{-1}x_1'y_1 \\ (x_2'x_2)^{-1}x_2'y_2 \end{pmatrix}$$

THE POOLED REGRESSION YIELDS THE SAME ESTIMATORS AS THE SEPARATE REGRESSIONS.

$$e = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} y_1 - \hat{\delta}_1 x_1 \\ y_2 - \hat{\delta}_2 x_2 \end{pmatrix}$$

$$\boxed{e'e = e_1'e_1 + e_2'e_2}$$

IS THE UNRESTRICTED ERROR SUM OF SQUARES.

DEFINE

$$\left.\begin{aligned} SSE_1 &= e_1'e_1 \\ SSE_2 &= e_2'e_2 \end{aligned}\right\}$$ THESE ARE THE ERROR SUM OF SQUARES FROM THE 2 SEPARATE REGRESSIONS.

HENCE $SSE_U = SSE_1 + SSE_2$

IF WE RESTRICTED $\alpha_1 = \alpha_2$ &
$\beta_1 = \beta_2$

THERE WOULD BE A SINGLE REGRESSION
$Y_i = \alpha + \beta x_i + u_i$ USING ALL THE DATA

DEFINE $e_i^* = y_i - (\hat{\alpha} + \hat{\beta} x_i)$

THEN $e^{*\prime}e^* = SSE_3$

THIS IS THE RESTRICTED ERROR SUM OF SQUARES
$e^{*\prime}e^* = SSE_R$.

HENCE THE TEST STATISTIC IS

$$F = \frac{\left[SSE_3 - (SSE_1 + SSE_2)\right] / k}{\dfrac{(SSE_1 + SSE_2)}{n - 2k}} \sim F_{k,\, n-2k}.$$

HERE, k = # OF COEFFICIENTS IN EACH SEPARATE MODEL

n = # OF OBSERVATIONS (GROUP 1 & GROUP 2 COMBINED).

THIS TEST IS KNOWN AS THE **CHOW TEST**.

PERFORMING THE CHOW TEST
WITH TOO FEW OBSERVATIONS FOR ONE GROUP

CONSIDER THE REGRESSIONS

$$y_1 = X_1\beta_1 + u_1 \qquad \text{FOR GROUP 1}$$
$$y_2 = X_2\beta_2 + u_2 \qquad \text{FOR GROUP 2}$$

WE ASSUME THAT U_1 & U_2 ARE IDENTICALLY DISTRIBUTED. THE OBJECTIVE IS TO TEST

$$H_0: \beta_1 = \beta_2 = \beta \qquad \text{(NO STRUCTURAL DIFFERENCE}$$

SUPPOSE THAT β_1 & β_2 ARE k-ELEMENT VECTORS

WE HAVE n_1 OBSERVATIONS FOR GROUP 1
n_2 OBSERVATIONS FOR GROUP 2

$$n_1 > k \qquad \text{BUT} \qquad n_2 < k$$

IN THIS CASE WE CAN RUN A REGRESSION FOR GROUP 1 ONLY TO GET SSE_1 BUT CANNOT GET SSE_2 BECAUSE WE HAVE TOO FEW OBSERVATIONS IN GROUP 2.

IN THE PRESENT CONTEXT,

$$e_1'e_1 = (y_1 - X_1\hat{\beta}_1)'(y_1 - X_1\hat{\beta}_1)$$
IS ESS_1.

$$e_*'e_* = (y - X\hat{\beta})'(y - X\hat{\beta}) = ESS_3$$
(FROM A SINGLE REGRESSION WITH ALL OBSERVATIONS)