BY: S. VISHAL
CH18B020

CH5440 : MULTIVARIATE DATA ANALYSIS ASSIGNMENT-1

1 a) i) ρ known, b unknown

objective : $$\min_{a,b,\hat{u}_i} \sum \frac{(y_i - a\hat{u}_i - b)^2}{\sigma_\varepsilon^2} + \frac{(u_i - \hat{u}_i)^2}{\sigma_\delta^2}$$

multiply & divide by σ_ε^2 : $$\min_{a,b,\hat{u}_i} \frac{1}{\sigma_\varepsilon^2} \sum_{\forall i} (y_i - a\hat{u}_i - b)^2 + \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}(u_i - \hat{u}_i)^2$$

this is equivalent to : $$\min_{a,b,\hat{u}_i} \sum_{i=1}^{N} (y_i - a\hat{u}_i - b)^2 + \rho(u_i - \hat{u}_i)^2$$
$= E$ (say)

$$\frac{\partial E}{\partial a} = 0 \Rightarrow 2\sum_{\forall i} \hat{u}_i (y_i - a\hat{u}_i - b) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \hat{u}_i (y_i - a\hat{u}_i - b) = 0 \quad \text{--- (1)}$$

$$\frac{\partial E}{\partial b} = 0 \Rightarrow 2\sum_{\forall i} (y_i - a\hat{u}_i - b) = 0$$

$$\Rightarrow a\sum_{\forall i} \hat{u}_i + b\sum_{\forall i} 1 = \sum_{\forall i} y_i$$

$$\Rightarrow a \cdot \sum_{\forall i} \hat{u}_i + Nb = \sum_{\forall i} y_i \quad \text{--- (2)}$$

$$\frac{\partial E}{\partial \hat{u}_i} = 0 \;\forall i \Rightarrow \quad a(y_i - a\hat{u}_i - b) + \rho(u_i - \hat{u}_i) = 0$$

$$\Rightarrow \boxed{\hat{u}_i = \frac{a y_i - ab + \rho u_i}{a^2 + \rho}} \quad \text{③}$$

Substitute ③ in ②,

$$N b = \sum y_i - a\sum\left(\frac{a y_i - ab + \rho u_i}{a^2+\rho}\right)$$

$$\Rightarrow N(a^2+\rho)b = (a^2+\rho)\sum y_i + \left(-a^2\sum y_i + a^2 b - a\rho u_i\right)$$

$$\Rightarrow N\rho b = \rho\sum_{\forall i} y_i - \rho a\sum_{\forall i} u_i$$

$$\Rightarrow b = \frac{\sum y_i}{N} - \frac{a\sum u_i}{N}$$

$$\Rightarrow \boxed{b = \bar{y} - a\bar{u}} \quad \text{④}$$

Substitute ③ in ①

$$\sum \hat{u}_i\left(y_i - a\left(\frac{a y_i - ab + \rho u_i}{a^2+\rho}\right) - b\right) = 0$$

$$\Rightarrow \sum \hat{u}_i\left(\cancel{a^2 y_i} + \rho y_i - \cancel{a^2 y_i} + \cancel{a^2 b} - a\rho u_i - \cancel{a^2 b} - b\rho\right) = 0$$

$$\Rightarrow \sum \hat{u}_i\left(\rho y_i - (a\rho u_i) - \rho\bar{y} - a\rho\bar{u}\right) = 0$$

(Substituting b value from ④)

$\cancel{\Rightarrow \sum \hat{u}_i (\rho y_i - \alpha \rho u_i}$

Cancelling ρ,

$$\Rightarrow \sum_{\forall i} \hat{u}_i \left((y_i - \bar{y}) - \alpha (u_i - \bar{u}) \right) = 0$$

Substituting value of $\hat{u}_i$ from (3),

$$\sum \left(\frac{a y_i - ab + \rho u_i}{a^2 + \rho} \right) \left((y_i - \bar{y}) - a(u_i - \bar{u}) \right) = 0$$

Substituting b from (4)

$$\Rightarrow \sum (a y_i - a(\bar{y} - a\bar{u}) + \rho u_i) ((y_i - \bar{y}) - a(u_i - \bar{u})) = 0$$

$$\Rightarrow \sum \left(a(y_i - \bar{y}) + a^2 \bar{u} + \rho u_i \right) \left((y_i - \bar{y}) - a(u_i - \bar{u}) \right) = 0$$

$$\Rightarrow \sum_{\forall i} a(y_i - \bar{y})^2 - a^2 (y_i - \bar{y})(u_i - \bar{u}) + a^2 \bar{u}(y_i - \bar{y}) - a^3 \bar{u}(u_i - \bar{u}) + \rho u_i (y_i - \bar{y}) - \rho a u_i (u_i - \bar{u}) = 0 \quad (5)$$

a^3 term: $-\sum \bar{u}(u_i - \bar{u}) = \bar{u}\left(\sum u_i - \bar{u}\sum 1\right)$

$= \bar{u}\left(\sum u_i - N\bar{u}\right)$

$= \bar{u}(N\bar{u} - N\bar{u}) = 0$

a^2 term: $\sum \bar{u}(y_i - \bar{y}) - (y_i - \bar{y})(u_i - \bar{u})$

$= \bar{u}\left(\sum y_i - N\bar{y}\right) - \left(\sum y_i u_i + N\bar{y}\bar{u} - \bar{y}\sum u_i - \bar{u}\sum y_i\right)$

$$= -\left(\sum_i y_i u_i + N\bar{y}\bar{u} - N\bar{y}\bar{u} - N\bar{y}\bar{u}\right)$$

$$= -\left(\sum_i y_i u_i - N\bar{y}\bar{u}\right) \qquad \left(\because \frac{\sum a_i}{N} = \bar{a}\right)$$

$$= -N S_{yu} \qquad (\text{definition of } S_{yu}) \rightarrow E(YU) - E(Y)E(U)$$

a term: $\sum_i (y_i - \bar{y})^2 - \rho u_i (u_i - \bar{u})$

$$= N S_{yy} - \left[\rho \sum_{\forall i} (u_i^2) - \rho \bar{u} \sum_i u_i\right]$$

$$= N S_{yy} - \rho\left[\sum_i (u_i^2) - N(\bar{u})^2\right]$$

$$= N S_{yy} - \rho N S_{uu} = N(S_{yy} - \rho S_{yu})$$

constant: $\sum_i \rho u_i (y_i - \bar{y}) = \rho\left(\sum_i u_i y_i - \sum_i u_i \bar{y}\right)$

$$= N\rho\left(\frac{\sum (u_i y_i)}{N} - \bar{u}\bar{y}\right)$$

$$= N\rho S_{yu} \qquad (\text{definition of } S_{yu}) \downarrow E(YU) - E(Y)E(U)$$

So, (5) becomes:

$$-N S_{yu} a^2 + (N S_{yy} - \rho N S_{uu}) a + N\rho S_{yu} = 0$$

$$\Rightarrow S_{yu} a^2 + (\rho S_{uu} - S_{yy}) a - \rho S_{yu} = 0$$

~~$\Rightarrow a = \frac{S_{yy} - \rho S_{uu} + \sqrt{\ldots}}{}$~~

Applying Quadratic Formula,

$$a = \frac{S_{yy} - \rho S_{uu} \pm \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S_{yu}^2}}{2 S_{yu}} \quad \text{—} (6)$$

For minimum 2nd order condition : $\frac{d^2E}{da^2} > 0$

$$\frac{d^2E}{da^2} = 2 S_{yu}\, a + (\rho S_{uu} - S_{yy})$$

Subst + root

$$2 S_{yu} \left(\frac{S_{yy} - \rho S_{uu} + \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S_{yu}^2}}{2 S_{yu}} \right) + (\rho S_{uu} - S_{yy})$$

$$= \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S_{yu}^2} > 0 \quad \text{—} (7)$$

other root will give $-\sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S_{yu}^2}$

that is not a minimum

$$\therefore \boxed{\hat{a} = \frac{S_{yy} - \rho S_{uu} + \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho \rho^2_{yu}}}{2 S_{yu}} \quad \text{—} (8)}$$

ii) If b is known to be 0.

objective becomes $\min\limits_{a, b, \tilde{u}_i} \sum \frac{(y_i - a\tilde{u}_i)^2}{\sigma_\varepsilon^2} + \frac{(u_i - \tilde{u}_i)^2}{\sigma_\delta^2}$

$$\frac{\partial E}{\partial \hat{u}_i} = 0 \Rightarrow a(y_i - a\hat{u}_i) + \rho(u_i - \hat{u}_i) = 0$$

$$\Rightarrow \boxed{\hat{u}_i = \frac{a y_i + \rho u_i}{a^2 + \rho}} \quad \text{— (9)}$$

$$\frac{\partial E}{\partial a} = 0 \Rightarrow \sum \hat{u}_i (\cancel{a u_i} - a\hat{u}_i + y_i) = 0 \quad \text{— (10)}$$

Subst (9) in (10),

$$\sum \hat{u}_i \left(-a \left(\frac{a y_i + \rho u_i}{a^2 + \rho} \right) + y_i \right) = 0$$

$$\Rightarrow \sum \hat{u}_i (\rho)(y_i - a u_i) = 0$$

Subst from (9) again,

$$\sum \frac{a y_i + \rho u_i}{a^2 + \rho} (y_i - a u_i) = 0$$

$$\Rightarrow \sum a y_i^2 - a y_i u_i + \rho y_i u_i - a^2 y_i u_i = 0$$

$$\Rightarrow a^2 \left(\sum y_i u_i \right) + a \left(\rho \sum u_i^2 - \sum y_i^2 \right) - \rho \sum y_i u_i = 0$$

these coeff can be termed as respective variances around zero !

Solving for a & ensuring $\frac{d^2 E}{da^2} > 0$ as in previous case,

$$\hat{a} = \frac{\left(\sum y_i^2 - \rho \sum x_i^2 \right) + \sqrt{\left(\sum y_i^2 - a \sum x_i^2 \right)^2 + 4\rho \left(\sum x_i y_i \right)^2}}{2 \sum x_i y_j}$$

b) $$\hat{a} = \frac{S_{yy} - \delta S_{uu} + \sqrt{(S_{yy} - \delta S_{uu})^2 + 4\delta S_{yu}^2}}{2 S_{yu}}$$

$$\lim_{\delta \to 0} \hat{a} = \frac{S_{yy} + \sqrt{(S_{yy} - 0)^2 + 0}}{2 S_{yu}}$$

$$\Rightarrow \hat{a}_{IOLS} = \frac{S_{yy}}{S_{yu}}$$

$$\hat{b}_{IOLS} = \bar{y} - \hat{a}_{IOLS}\bar{u} = \bar{y} - \hat{a}_{IOLS}\bar{u}$$

$$\hat{u}_{IOLS,i} = \frac{\hat{a} y_i - \hat{a}\hat{b}_{IOLS}}{\hat{a}^2} = \frac{y_i - \hat{b}_{OLS}}{\hat{a}_{IOLS}}$$

$$\hat{y}_{i,IOLS} = y_i$$

(Subst $\delta = 0$ in ③)

~~OLS~~

$$\hat{a}_{OLS} = \lim_{\delta \to \infty} \frac{S_{yy} - \delta S_{uu} + \sqrt{(S_{yy} - \delta S_{uu})^2 + 4\delta S_{yu}^2}}{2 S_{yu}}$$

$$= \lim_{\delta \to \infty} \frac{\frac{S_{yy}}{\delta} - S_{uu} + \sqrt{\left(\frac{S_{yy}}{\delta} - S_{uu}\right)^2 + \frac{4 S_{yu}^2}{\delta}}}{\frac{2 S_{yu}}{\delta}}$$

Let $\xi = \frac{1}{\delta}$

$$\Rightarrow \hat{a}_{OLS} = \lim_{\xi \to 0} \frac{\xi S_{yy} - S_{uu} + \sqrt{(\xi S_{yy} - S_{uu})^2 + 4\xi S_{yu}^2}}{2 \xi S_{yu}}$$

OLS :

Recall the quadratic eqn derived in (a),

$$a^2 S_{yu} + (\delta S_{uu} - S_{yy}) a - \delta S_{yu} = 0$$

Divide by δ,

$$a^2 \left(\frac{S_{yu}}{\delta}\right) + \left(S_{uu} - \frac{S_{yy}}{\delta}\right) a - S_{yu} = 0$$

Let $\xi = \frac{1}{\delta}$, $a^2 (\xi S_{yu})$ $\cancel{a^2(\xi \delta)}$ $+ (S_{uu} - \xi S_{yy}) a - S_{uu} = 0$

$\lim_{\delta \to \infty} \Rightarrow \lim_{\xi \to 0} \Rightarrow \lim_{\xi \to 0} (\xi S_{yu}) a^2 + (S_{uu} - \xi S_{yy}) a - S_{yu} = 0$

$$\Rightarrow a(S_{uu}) - S_{yu} = 0$$

$$\Rightarrow \hat{a}_{OLS} = \frac{S_{yu}}{S_{uu}}$$

$$\hat{b}_{OLS} = \bar{y} - \hat{a}_{OLS} \bar{u}$$

$\hat{u}_{i,OLS} = u_i \Rightarrow \hat{u}_{i,OLS} = \hat{u}_i$

$\hat{y}_{i,OLS} = \hat{a}_{OLS} u_i + \hat{b}_{OLS}$

WTLS solutions (derived in (a)) :

$$\hat{a}_{WTLS} = \frac{S_{yy} - \rho S_{uu} + \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S^2_{yu}}}{2 S_{yu}}$$

$$\hat{b}_{WTLS} = \bar{y} - \hat{a}_{WTLS}\,\bar{u}$$

$$\cancel{\hat{u}_{WTLS} = \hat{a}}$$

$$\hat{u}_{i,WTLS} = \frac{\cancel{\alpha}\ \cdot \left(\hat{a}_{WTLS}(y_i - \hat{b}_{WTLS}) + \rho u_i\right)}{(\alpha^2 + \rho)}$$

$$\hat{y}_{i,WTLS} = \hat{a}_{WTLS}\,\hat{x}_{WTLS} + \hat{b}_{WTLS}$$

(2) a) Here, I consider CO_2 concentration as input variable and temperature deviation as output variable

This is because CO_2 concentration measurements are likely ~~to~~ have more errors. And, from an engineering viewpoint, CO_2 emission causes temperature deviation

$N = 31$ samples.

~~Perform~~ Computing sample statistics using MATLAB we get,

$$\bar{y} = \frac{\sum_i y_i}{N} = 0.8512$$

$$\bar{u} = \frac{\sum_i u_i}{N} = \cancel{0.08875}\ 369.4197$$

$$S_{yy} = \frac{1}{N} \sum_i (y_i - \bar{y})^2 = 0.0875$$

$$S_{yu} = \frac{1}{N} \sum_i (y_i - \bar{y})(u_i - \bar{u}) = 4.192$$

$$S_{uu} = \frac{1}{N} \sum_i (u_i - \bar{u})^2 = 255.74$$

$$\hat{\alpha}_{OLS} = \frac{S_{yu}}{S_{uu}} = 0.01639$$

$$\hat{\beta}_{OLS} = \bar{y} - \hat{\alpha}\bar{u} = -5.2044$$

OLS estimate of CO_2 concentration reqd. for $3.6°F$ raise $= \dfrac{3.6 - \hat{\beta}_{OLS}}{\hat{\alpha}_{OLS}}$

$$= \boxed{537.1075 \text{ ppm}}$$

$$\hat{\alpha}_{TLS} = \frac{S_{yy} - S_{uu} + \sqrt{(S_{yy} - S_{uu})^2 + 4S_{yu}^2}}{2 S_{yu}} = 0.016393$$

$$\hat{\beta}_{TLS} = \bar{y} - \hat{\alpha}_{TLS}\bar{u} = -5.20482$$

TLS estimate of CO_2 concentration reqd. for $3.6°F$ raise $= \dfrac{3.6 - \hat{\beta}_{TLS}}{\hat{\alpha}_{TLS}}$

$$= \boxed{537.0952 \text{ ppm}}$$

b) Now, we keep time (year) as input variable u

and CO_2 as output variable y.

We know that there is no error in the input, i.e., year. So we choose to go with an OLS estimator

Computing sample statistics in MATLAB,

$\bar{y} = 369.4197$ ppm $\quad \bar{u} = 1999$

$s_{yy} = 255.7407 \quad s_{yu} = 142.5342 \quad s_{uu} = 80$

$$\hat{\alpha}_{OLS} = \frac{s_{yu}}{s_{uu}} = 1.7817$$

$$\hat{\beta}_{OLS} = \bar{y} - \alpha\bar{u} = -3192.153.$$

i) Finding time taken to reach the OLS estimate of concentration

$$t = \frac{537.1075 - \hat{\beta}}{\hat{\alpha}} = \boxed{2093.1179}$$

ii) Finding time taken to reach the TLS estimate of concentration

$$t = \frac{537.0952 - \hat{\beta}}{\hat{\alpha}} = \boxed{2093.111}$$

Both the estimates are similar. Conservatively, we can say by year $\boxed{2093}$ & (72 more years) we will reach the maximum level of CO_2 to cause 3.6°F rise

③ a) Since we are performing OLS as well as EOLS: choice of input-output doesnt matter for this analysis

Set input, u as ~~CF~~ EP method

output, y as CF method.

Sample statistics. [All computations done in MATLAB]

$S_{yy} = 0.9272$ $\quad S_{yu} = 0.9241$ $\quad S_{uu} = 0.9311$

$\bar{y} = 1.9505$ $\quad \bar{u} = 2.0155$

Our aim to see whether the 95% confidence interval of $\hat{\alpha}$ contains 1 (so that $y = u$ this model comes) and if $\hat{\beta}$ is close to zero

95% CI : $\hat{\alpha} \pm$ ~~t_{α}~~ $2.16\,\sigma_\varepsilon^2$

$\because \sigma_\varepsilon^2$ is unknown, $\hat{\sigma}_\varepsilon^2 = \sum \dfrac{(y_i - \hat{\alpha} u_i - \hat{\beta})^2}{N-2}$

OLS (Assumption: EP method is free of error, CF method has error)

$\hat{\alpha}_{OLS} = \dfrac{S_{yu}}{S_{uu}} = \boxed{0.9924}$ $\quad \hat{\beta}_{OLS} = \boxed{-0.0497}$

95% confidence interval of $\hat{\alpha}$: $[0.9776, 1.0073]$

C.I. contains 1! $\quad \hat{\beta}$ close to 0!

IOLS

Assumption : EP has error, CF method has no error.

$$\hat{\alpha}_{\cancel{6}IOLS} = \frac{S_{yy}}{S_{yu}} = \boxed{0.9991}$$

$$\hat{\beta}_{IOLS} = \bar{y} \cancel{+} - \hat{\alpha}_{IOLS}\,\bar{u} = \boxed{-0.0564}$$

95% ~~CI~~ confidence interval for $\hat{\alpha}_{IOLS}$: ~~0.9809~~ [0.9842, 1.0074]

C.I. contains 1! $\hat{\beta}$ close to 0!
for $\hat{\alpha}$

TLS

Assumption : both EP & CF method has error.

$$\hat{\alpha}_{TLS} = \frac{S_{yy} - S_{uu} + \sqrt{(S_{yy} - S_{uu})^2 + 4S^2_{yu}}}{2S_{yu}} = \cancel{0.016893} \boxed{0.9957}$$

$$\hat{\beta}_{TLS} = \bar{y} - \hat{\alpha}_{TLS}\,\bar{u} = \cancel{8.2048}\ \boxed{-0.0564}$$

95% confidence interval for $\hat{\alpha}_{TLS}$ = [0.9809, 1.0106]

C.I. contains $\hat{\alpha} = 1$! $\hat{\beta}$ close to 0

All 3 approaches have $\hat{\alpha}$ close to 1 and $\hat{\beta}$ close to 0

So we can conclude that the new method (CF) is a good ~~sg~~ substitute for the established method (EP)

b) OLS : Input is error free → EP measmt is error free

~~⇒ $\hat{u}_i = u_i =$~~

⇒ Estimate $= u_i = \boxed{2.31 \text{ mg/L}}$

IOLS

Output is error free → CF is error free

⇒ Estimate $= y_i = \boxed{2.20 \text{ mg/L}}$

TLS

Approximating $\alpha = 1$, $\beta = 0$,

~~Assuming~~ $\alpha = 1$,

estimate $\hat{y}_i = \hat{u}_i = \dfrac{\alpha(y_i - \beta) + u_i}{\alpha^2 + 1}$

$= \dfrac{y_i + u_i}{2} = \dfrac{2.31 + 2.2}{2}$

⇒ estimate of phytic acid $= \boxed{2.255 \text{ mg/L}}$

④ The objective is to find if the defect is aligned across vertical or horizontal axis.

~~&~~ So by performing Linear regression, ~~if~~

→ If we get equation as $y = \beta$ it is parallel to x-axis

→ if we get equation as $x = \beta$ it is parallel to y-axis

Essentially we can do 2 regression for each data:

$$y = \alpha_1 x + \beta_1 + \epsilon \qquad x = \alpha_2 y + \beta_2$$

and check if α_1 or α_2 ~~are~~ is small enough / close to zero

Since swapping x & y is eqnt to EOLS we need ~~does~~ not do that separately. Just need to do OLS ~~and~~ to cover EOLS also.

In addition, we can check using TLS. Also, value of ~~β~~ doesn't matter in terms of deciding defect orientation

Defect - ①

~~$y = \alpha_1 x + \beta_1$ for.~~

$\bar{y} = 824.044 \qquad \bar{x} = 3774.6$

$S_{yy} = 267.0202 \qquad S_{yx} = 171.9279 \qquad S_{xx} = 1876.5$

Model 1: $y = \alpha_1 x + \beta_1$

$$\alpha_{1,OLS} = \frac{S_{yy}}{S_{yx}} = \underline{0.0092} \quad \beta_{1,0}$$

$$\alpha_{1,TLS} = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{yx}^2}}{2 S_{yx}}$$

$$= \underline{0.0093}$$

Both OLS & TLS suggest, the line is of the form

$$\underline{y = \beta}$$

~~Model 2: $x = \alpha_2 y + \beta_2$~~ (~~Note that actual slope is inverse of~~ this

So we have successfully identified the defect 1 along x-axis

$\therefore$ $\boxed{\text{Defect 1 is due to corrosion of steel reinforcement bars}}$

<u>Defect 2</u>

$\bar{x} = 2143.9 \quad \bar{y} = 3190.2$

$S_{yy} = 1904.0 \quad S_{yx} = 1708.0$

$S_{xx} = 2153.9$; Model ①: $y = \hat{\alpha} x + \hat{\beta}$

$$\hat{\alpha}_{1,OLS} = \frac{S_{yu}}{S_{yu}} = \cancel{0.0092} \quad \underline{0.793}$$

$$\hat{\alpha}_{1,TLS} = \frac{S_{uy} - S_{uu} + \sqrt{(S_{yy}-S_{uu})^2 + 4S^2_{yu}}}{2S_{yu}} = \underline{0.9295}$$

Both $\hat{\alpha}$ estimates are close to 1. Trying with

model 2.

Model ②: $x = \hat{\alpha} y + \hat{\beta}$

$$\hat{\alpha}_{2,OLS} = \frac{S_{yu}}{S_{yy}} = \underline{0.8971}$$

(using u, y nomenclature as defined earlier)

$$\hat{\alpha}_{2,TLS} = \frac{S_{yu} - S_{yy} + \sqrt{(S_{uu}-S_{yy})^2 + 4S^2_{uy}}}{2S_{uy}}$$

$$= \underline{1.0258}$$

Once again both are close to 1. and not the equation is of the form $y = x$

along vertical or horizontal axis (instead oriented at 45° — along $y = x$)

$\Rightarrow$ Defect 2 is not due to corrosion of steel reinforcement bars

Defect 3

$\bar{y} = 3522.3$ $\quad \bar{x} = 2832.7$ $\quad S_{yy} = 3099.5$

$S_{yx} = 910.12$ $\quad S_{xx} = 5152.5$

Model ①: $y = \hat{\alpha} x + \hat{\beta}$

$\hat{\alpha}_{1,OLS} = \cancel{28.4298}\ 0.1766$ $\quad \hat{\alpha}_{1,TLS} = \cancel{0.1766}\ 28.4298$

Model ②: $x = \hat{\alpha} y + \hat{\beta}$

$\hat{\alpha}_{2,OLS} = 0.0294$ $\quad \hat{\alpha}_{2,TLS} = 0.0352$

We see that $\hat{\alpha}_{2,OLS}$ & TLS are close to zero.

$\Rightarrow$ the line is of $x = \hat{\beta}$. the defect is along

the y-axis

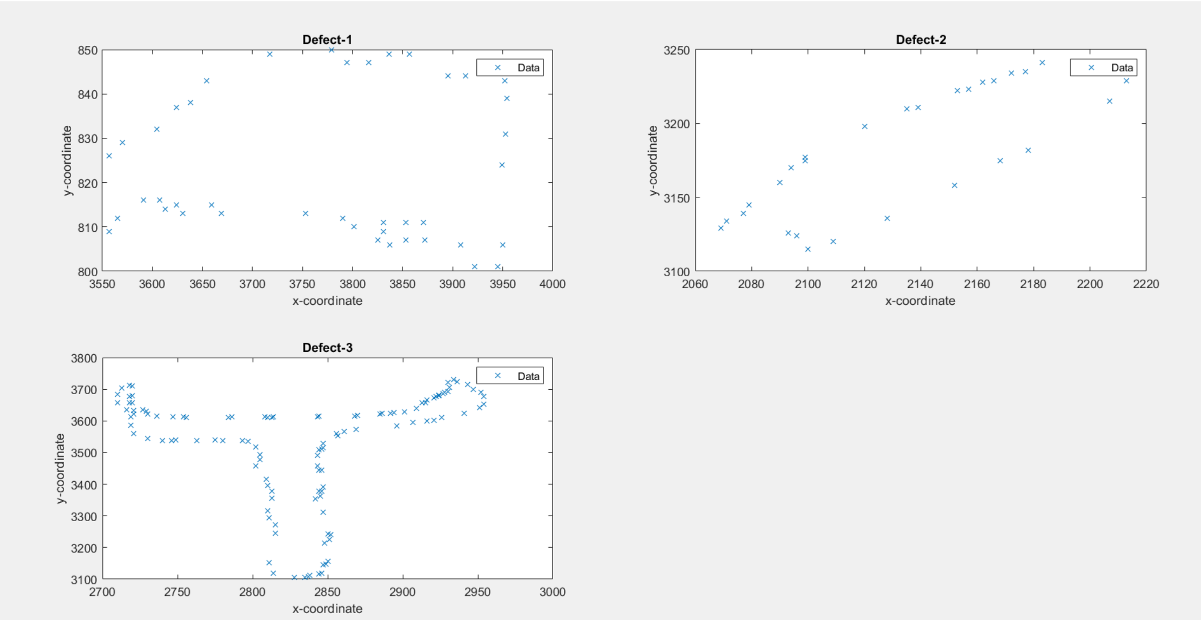
$\therefore$ Defect 3 is due to corrossion of steel reinforcement bars.

conclusion: Defects 1 and 2 are due to corrossion of steel reinforcement bars

BY S.VISHAL
CH18B020

Plots

Question-4 plots



The above plots agree with the inferences concluded using OLS/TLS.

Defect-1: along horizontal axis

Defect-2: along y = x

Defect-3: along vertical axis

Code

OLS.m (function to perform OLS)

```
function [alpha, beta, uhat, yhat, s] = OLS(u,y)
    N = length(u);
    ybar = mean(y);
    ubar = mean(u);
    syy = var(y,1);
    suu = var(u,1);
    syu = 1/N*sum((y-ybar).*(u-ubar));
    uhat = u;
    alpha = syu/suu;
    beta = ybar - alpha*ubar;
    yhat = alpha*uhat + beta;
    s = 
struct('alpha',alpha,'beta',beta,'ybar',ybar,'ubar',ubar,
'syy',syy,'syu',syu,'suu',suu);
end
```

IOLS.m (function to perform IOLS)

```
function [alpha, beta, uhat, yhat, s] = IOLS(u,y)
    N = length(u);
    ybar = mean(y);
    ubar = mean(u);
    syy = var(y,1);
    suu = var(u,1);
    syu = 1/N*sum((y-ybar).*(u-ubar));
    yhat = y;
    alpha = syy/syu;
    beta = ybar - alpha*ubar;
    uhat = (yhat - beta)/alpha;
    s = 
struct('alpha',alpha,'beta',beta,'ybar',ybar,'ubar',ubar,
'syy',syy,'syu',syu,'suu',suu);
end
```

TLS.m (function to perform TLS)

```
function [alpha, beta, uhat, yhat, s] = TLS(u,y)
    N = length(u);
    ybar = mean(y);
    ubar = mean(u);
    syy = var(y,1);
    suu = var(u,1);
    syu = 1/N*sum((y-ybar).*(u-ubar));
    alpha = (syy - suu + sqrt((syy-suu)^2 + 
4*syu^2))/2/syu;
    beta = ybar - alpha*ubar;
    uhat = (alpha*(y-beta)+u)/(alpha^2+1);
    yhat = alpha*uhat + beta;
    s = 
struct('alpha',alpha,'beta',beta,'ybar',ybar,'ubar',ubar,
'syy',syy,'syu',syu,'suu',suu);
end
```

Question-2

```
clear; close all;
%% Open data
A = readmatrix('CO2.csv');
time = A(:,1);
CO2 = A(:,2);
temp = A(:,3);
temp_cut = 3.6;
%% Predict maximum permissible level of CO2 - OLS
% y -> temperature deviation
% u -> CO2
```

```
u = CO2;
y = temp;
[alpha, beta, uhat, yhat, s] = OLS(u,y);
disp(s);
CO2_cut_OLS = (temp_cut-beta)/alpha;
%% Predict maximum permissible level of CO2 - TLS
[alpha_TLS, beta_TLS, uhat_TLS, yhat_TLS,s_TLS] =
TLS(u,y);
CO2_cut_TLS = (temp_cut-beta_TLS)/alpha_TLS;
%% Predict year
ut = time;
yt = CO2;
[alpha_t, beta_t, uhat_t, yhat_t, s_t] = OLS(ut,yt);
t_OLS = (CO2_cut_OLS-beta_t)/alpha_t;
tpred_OLS = ceil(t_OLS);
t_TLS = (CO2_cut_TLS-beta_t)/alpha_t;
tpred_TLS = ceil(t_TLS);
```

Question-3

```
close all; clear;
%% Data
EP =
[1.98,2.31,3.29,3.56,1.23,1.57,2.05,0.66,0.31,2.82,0.13,3
.15,2.72,2.31,1.92,1.56,0.94,2.27,3.17,2.36]';
CF =
[1.87,2.2,3.15,3.42,1.1,1.41,1.84,0.68,0.27,2.8,0.14,3.2,
2.7,2.43,1.78,1.53,0.84,2.21,3.10,2.34]';
% u - EP, y - CF
u = EP;
y = CF;
N = length(u);
%% Part a)
% OLS
[alpha_OLS, beta_OLS, uhat_OLS, yhat_OLS,s_OLS] =
OLS(u,y);
sigma_e_OLS = 1/(N-2)*sum((y-alpha_OLS*u-beta_OLS).^2);
CI_OLS = [alpha_OLS-
2.16*sigma_e_OLS,alpha_OLS+2.16*sigma_e_OLS];
% IOLS
[alpha_IOLS, beta_IOLS, uhat_IOLS, yhat_IOLS,s_IOLS] =
IOLS(u,y);
sigma_e_IOLS = 1/(N-2)*sum((y-alpha_IOLS*u-
beta_IOLS).^2);
CI_IOLS = [alpha_IOLS-
2.16*sigma_e_IOLS,alpha_OLS+2.16*sigma_e_IOLS];
% TLS
```

```
[alpha_TLS, beta_TLS, uhat_TLS, yhat_TLS,s_TLS] =
TLS(u,y);
sigma_e_TLS = 1/(N-2)*sum((y-alpha_TLS*u-beta_TLS).^2);
CI_TLS = [alpha_TLS-
2.16*sigma_e_TLS,alpha_TLS+2.16*sigma_e_TLS];
%% Part b)
ui = 2.31;
yi = 2.20;
% OLS: u is perfect
OLS_pred = ui;
% IOLS: y is perfect
IOLS_pred = yi;
% TLS: both imperfect. doing a perpendicular projection
will give us an
% estimate for EP and CF. Since we need a single estimate
we will assume
% alpha = 1 and beta = 0. So yhat = uhat = (u+y)/2
TLS_pred = (ui+yi)/2;
```

Question-4

```
clear; close all;
%% Open data
A = readmatrix('defects_annotation_data.csv');
x1 = rem_NaN(A(:,1)); y1 = rem_NaN(A(:,2));
x2 = rem_NaN(A(:,4)); y2 = rem_NaN(A(:,5));
x3 = rem_NaN(A(:,7)); y3 = rem_NaN(A(:,8));
%% Defect-1
N = length(x1);
% TLS
[alpha_TLS, beta_TLS, uhat_TLS, yhat_TLS,s_TLS] =
TLS(x1,y1);
sigma_e_TLS = 1/(N-2)*sum((y1-alpha_TLS*x1-beta_TLS).^2);
CI_TLS = [alpha_TLS-
2.16*sigma_e_TLS,alpha_TLS+2.16*sigma_e_TLS];
% TLS - inverted
[alpha_TLS2, beta_TLS2, uhat_TLS2, yhat_TLS2,s_TLS2] =
TLS(y1,x1);
sigma_e_TLS2 = 1/(N-2)*sum((y1-alpha_TLS2*x1-
beta_TLS2).^2);
CI_TLS2 = [alpha_TLS2-
2.16*sigma_e_TLS2,alpha_TLS2+2.16*sigma_e_TLS2];
% OLS
[alpha_OLS, beta_OLS, uhat_OLS, yhat_OLS,s_OLS] =
OLS(x1,y1);
sigma_e_OLS = 1/(N-2)*sum((y1-alpha_OLS*x1-beta_OLS).^2);
```

```
CI_OLS = [alpha_OLS-
2.16*sigma_e_OLS,alpha_OLS+2.16*sigma_e_OLS];
% OLS - inverted
[alpha_OLS2, beta_OLS2, uhat_OLS2, yhat_OLS2,s_OLS2] =
OLS(y1,x1);
sigma_e_OLS2 = 1/(N-2)*sum((y1-alpha_OLS2*x1-
beta_OLS2).^2);
CI_OLS2 = [alpha_OLS2-
2.16*sigma_e_OLS2,alpha_OLS2+2.16*sigma_e_OLS2];
% Plot
subplot(2,2,1);
plot(x1,y1,'x');
title('Defect-1'); xlabel('x-coordinate'); ylabel('y-
coordinate');
legend('Data');
%% Defect-2
% TLS
[alpha_TLS_def2, beta_TLS_def2, uhat_TLS_def2,
yhat_TLS_def2,s_TLS_def2] = TLS(x2,y2);
sigma_e_TLS_def2 = 1/(N-2)*sum((y2-alpha_TLS_def2*x2-
beta_TLS_def2).^2);
CI_TLS_def2 = [alpha_TLS_def2-
2.16*sigma_e_TLS_def2,alpha_TLS_def2+2.16*sigma_e_TLS_def
2];
% TLS - inverted
[alpha_TLS2_def2, beta_TLS2_def2, uhat_TLS2_def2,
yhat_TLS2_def2,s_TLS2_def2] = TLS(y2,x2);
sigma_e_TLS2_def2 = 1/(N-2)*sum((y2-alpha_TLS2_def2*x2-
beta_TLS2_def2).^2);
CI_TLS2_def2 = [alpha_TLS2_def2-
2.16*sigma_e_TLS2_def2,alpha_TLS2_def2+2.16*sigma_e_TLS2_
def2];
% OLS
[alpha_OLS_def2, beta_OLS_def2, uhat_OLS_def2,
yhat_OLS_def2,s_OLS_def2] = OLS(x2,y2);
sigma_e_OLS_def2 = 1/(N-2)*sum((y2-alpha_OLS_def2*x2-
beta_OLS_def2).^2);
CI_OLS_def2 = [alpha_OLS_def2-
2.16*sigma_e_OLS_def2,alpha_OLS_def2+2.16*sigma_e_OLS_def
2];
% OLS - inverted
[alpha_OLS2_def2, beta_OLS2_def2, uhat_OLS2_def2,
yhat_OLS2_def2,s_OLS2_def2] = OLS(y2,x2);
sigma_e_OLS2_def2 = 1/(N-2)*sum((y2-alpha_OLS2_def2*x2-
beta_OLS2_def2).^2);
```

```
CI_OLS2_def2 = [alpha_OLS2_def2-
2.16*sigma_e_OLS2_def2,alpha_OLS2_def2+2.16*sigma_e_OLS2_
def2];
% Plot
subplot(2,2,2);
plot(x2,y2,'x');
title('Defect-2'); xlabel('x-coordinate'); ylabel('y-
coordinate');
legend('Data');
%% Defect-3
% TLS
[alpha_TLS_def3, beta_TLS_def3, uhat_TLS_def3,
yhat_TLS_def3,s_TLS_def3] = TLS(x3,y3);
sigma_e_TLS_def3 = 1/(N-2)*sum((y3-alpha_TLS_def3*x3-
beta_TLS_def3).^2);
CI_TLS_def3 = [alpha_TLS_def3-
2.16*sigma_e_TLS_def3,alpha_TLS_def3+2.16*sigma_e_TLS_def
3];
% TLS - inverted
[alpha_TLS2_def3, beta_TLS2_def3, uhat_TLS2_def3,
yhat_TLS2_def3,s_TLS2_def3] = TLS(y3,x3);
sigma_e_TLS2_def3 = 1/(N-2)*sum((y3-alpha_TLS2_def3*x3-
beta_TLS2_def3).^2);
CI_TLS2_def3 = [alpha_TLS2_def3-
2.16*sigma_e_TLS2_def3,alpha_TLS2_def3+2.16*sigma_e_TLS2_
def3];
% OLS
[alpha_OLS_def3, beta_OLS_def3, uhat_OLS_def3,
yhat_OLS_def3,s_OLS_def3] = OLS(x3,y3);
sigma_e_OLS_def3 = 1/(N-2)*sum((y3-alpha_OLS_def3*x3-
beta_OLS_def3).^2);
CI_OLS_def3 = [alpha_OLS_def3-
2.16*sigma_e_OLS_def3,alpha_OLS_def3+2.16*sigma_e_OLS_def
3];
% OLS - inverted
[alpha_OLS2_def3, beta_OLS2_def3, uhat_OLS2_def3,
yhat_OLS2_def3,s_OLS2_def3] = OLS(y3,x3);
sigma_e_OLS2_def3 = 1/(N-2)*sum((y2-alpha_OLS2_def3*x2-
beta_OLS2_def3).^2);
CI_OLS2_def3 = [alpha_OLS2_def3-
2.16*sigma_e_OLS2_def3,alpha_OLS2_def3+2.16*sigma_e_OLS2_
def3];
% Plot
subplot(2,2,3);
plot(x3,y3,'x');
title('Defect-3'); xlabel('x-coordinate'); ylabel('y-
coordinate');
```

```
legend('Data');
%% function to remove NaN values
function vnew = rem_NaN(v)
    vnew = v(~isnan(v));
end
```