BARC0130 : Advanced Mathematical Modelling & Analysis
2021-2022
Topic : Applications of Linear Algebra

CANDIDATE CODE : RWVW7

COURSEWORK 4

Part A:

On analyzing the data provided it can be observed that it forms a linear model of the form:

$$AX = B + v$$

It forms an overdetermined system wherein we have 3 parameters to be observed with 25 different observations, so:

$$m = 25$$
$$n = 3$$

Here, the quantity matrix A that can be defined as:

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ \dots & \dots & \dots \\ a_{25} & b_{25} & c_{25} \end{pmatrix}$$

wherein a_1-a_{25} , b_1-b_{25} , c_1-c_{25} correspond to quantities of glass (m^2) , concrete (m^3) and steel (tons) used in each of the 25 skyscrapers built by a single construction company.

We have a quantity vector X that can be defined as

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

wherein x, y, z corresponds to the least squares estimate of the unit costs of glass, concrete, and steel.

We have a quantity vector B that can be defined as

$$B = \begin{pmatrix} d_1 \\ \dots \\ d_{25} \end{pmatrix}$$

wherein d_1-d_{25} correspond to the total costs of materials used in each of the 25 skyscrapers built by a single construction company.

$$v = \text{residuals}$$

For an unweighted system W = I (identity matrix)

We can calculate least squares estimate of X using the expression

$$X = (A^T W A)^{-1} A^T W B$$

$$\Rightarrow X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 37.8061 \\ 75.1257 \\ 556.4197 \end{pmatrix}$$

Hence, we have:

Cost of 1 m^2 of glass = £ 37.8061= **£ 37.81**

Cost of 1 m^3 of concrete = £ 75.1257= **£ 75.13**

Cost of 1 ton of steel = £ 556.4197= **£ 556.42**

We can calculate the residuals v using

$$AX = B + v$$
$$\Rightarrow v = AX - B$$

We calculate the unit variance using

$$\sigma_o^2 = \frac{\nu^T W \nu}{m - n}$$

$$\Rightarrow \quad \sigma_o^2 = £^2\, 5.1042 \times 10^8 = \mathbf{£^2\, 5.10 \times 10^8}$$

The covariance matrix was calculated using

$$c_x = \sigma_o^2 (A^T W A)^{-1}$$

$$\Rightarrow \quad c_x = \begin{pmatrix} \boldsymbol{\sigma_a} & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \boldsymbol{\sigma_b} & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \boldsymbol{\sigma_c} \end{pmatrix} = £^2 \begin{pmatrix} \mathbf{62.8781} & -4.9200 & -6.7190 \\ -4.9200 & \mathbf{1.2489} & -0.7034 \\ -6.7190 & -0.7034 & \mathbf{2.4707} \end{pmatrix}$$

To statistically assess the residuals, it is useful to calculate the sum of the squares of the residuals and the RMS error of the residuals. It is not useful to calculate the mean residuals as the residuals have significant positive and negative values. The linear regression ensures that sum of the square of the residuals is minimized.

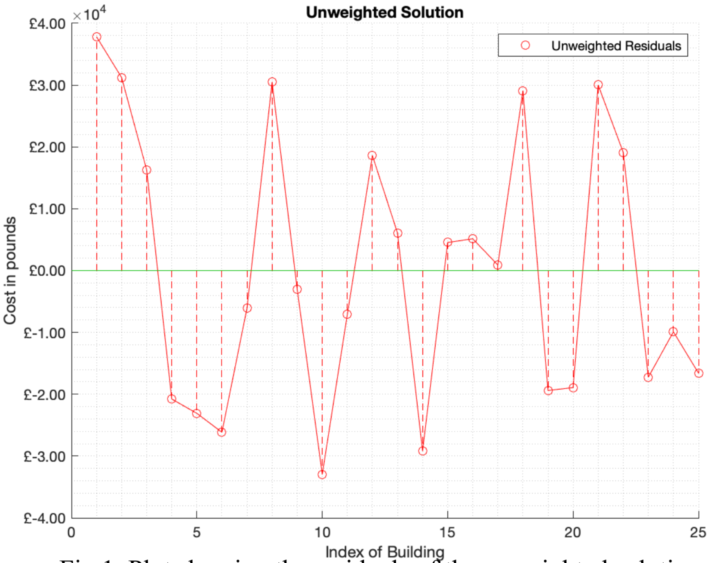

Fig 1. Plot showing the residuals of the unweighted solution.

$$v_{squared} = \Sigma(\nu)^2$$

$$\Rightarrow \quad v_{squared} = £^2\, 1.1229 \times 10^{10}$$

$$\Rightarrow \quad \mathbf{v_{squared} = £^2\, 1.12 \times 10^{10}}$$

$$v_{rms\text{-}error} = \sqrt{\frac{\Sigma(v)^2}{m}}$$

$$\Rightarrow \quad v_{rms\text{-}error} = £\, 2.1194 \times 10^4$$

$$\Rightarrow \quad \mathbf{v_{rms\text{-}error} = £\, 2.12 \times 10^4}$$

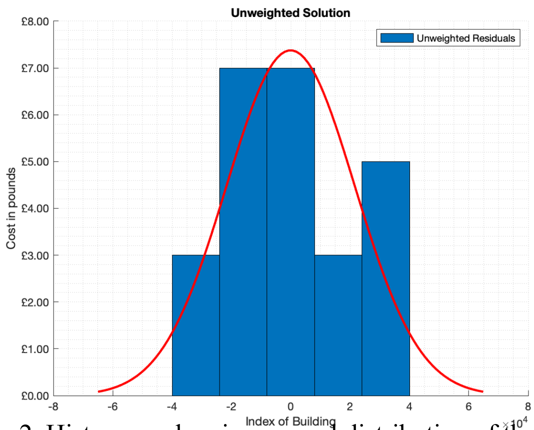

Fig 2. Histogram showing normal distribution of the residuals of the unweighted solution.

The calculated unit costs show large variances. It can be implied that the least squares analysis did not fulfill its purpose of minimizing the sum of the squared residuals as the values are extremely large. This implies that the unit costs calculated are not the best estimation.

Further, on plotting the residuals it is evident that the residuals are evenly scattered and follow no general trend on pattern. This implies that the linear regression is an accurate method for these calculations. To further confirm, a chi square test for goodness of fit conducted at 5% significance, confirmed the normal distribution of the residuals which can be visualized in the plot shown alongside.

Part B:

We repeat the same calculations as shown in part one for a weighted solution.

$$W = \begin{pmatrix} w_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & w_{25} \end{pmatrix}$$

where w is a 25x25 diagonal matrix wherein $w_1 \dots w_{25}$ correspond to the weights of each of the 25 observation equations provided in the excel sheet.

$$\Rightarrow \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 43.9687 \\ 74.3534 \\ 556.1765 \end{pmatrix}$$

Hence, we have

Cost of 1 m^2 of glass = £ 43.9687= **£ 43.97**

Cost of 1 m^3 of concrete = £ 74.3534= **£ 74.35**

Cost of 1 ton of steel = £ 556.1765= **£ 556.18**

Unit Variance,

$$\sigma_o^2 = £^2\ 0.9340$$

$$\sigma_o^2 = \mathbf{£^2\ 0.93}$$

Covariance Matrix,

$$C_x = \begin{pmatrix} \boldsymbol{\sigma_a} & \sigma_{ab} & \sigma_{ac} \\ \sigma_{ab} & \boldsymbol{\sigma_b} & \sigma_{bc} \\ \sigma_{ac} & \sigma_{bc} & \boldsymbol{\sigma_c} \end{pmatrix} = £^2 \begin{pmatrix} \mathbf{1.6721} & -0.3005 & 0.0605 \\ -0.3005 & \mathbf{0.2222} & -0.2566 \\ 0.0605 & -0.2566 & \mathbf{0.3613} \end{pmatrix}$$

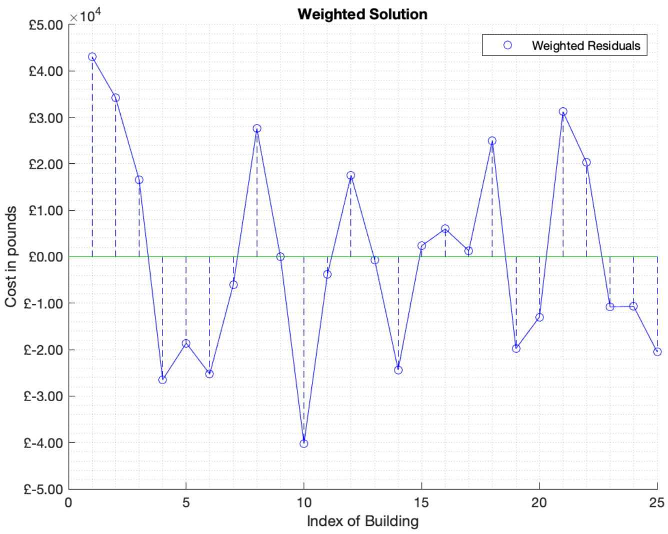

Fig 3. Plot showing the residuals of the weighted solution.

$$v_{squared} = \Sigma(\ v^T W v\ )^2$$

$$v_{squared} = £^2\ 20.5479$$

$$\Rightarrow \quad \mathbf{v_{squared} = £^2\ 20.55}$$

$$v_{rms\text{-}error} = \sqrt{\frac{\Sigma(\ v^T W v\ )^2}{m}}$$

$$\Rightarrow \quad v_{rms\text{-}error} = £\ 0.9066$$

$$\mathbf{v_{rms\text{-}error} = £\ 0.91}$$

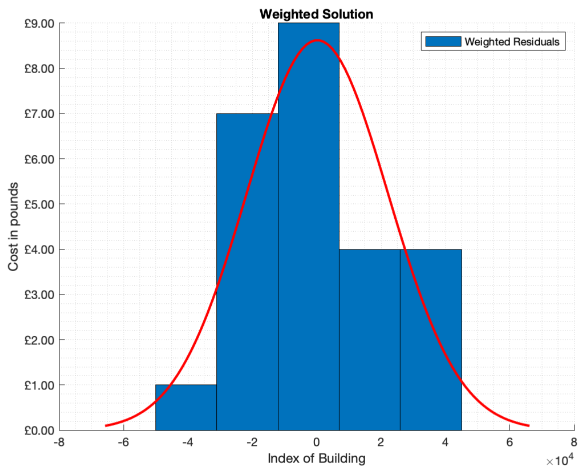

Fig 4. Histogram showing normal distribution of the residuals of the weighted solution.

The calculated unit costs show low variances. It can be implied that the least squares analysis successfully calculated the unit costs with minimized sum of the squared residuals. The RMS error of the residuals calculated is extremely low (approaches 0). It can hence be deduced that the unit costs calculated are a fairly accurate estimation.

Further, on plotting the residuals it is evident that the residuals are evenly scattered and follow no general trend on pattern. This implies that the linear regression is an accurate method for these calculations. To further confirm, a chi square test for goodness of fit conducted at 5% significance, confirmed the normal distribution of the residuals which can be visualized in the plot shown alongside.

Material		Unweighted Solution	Weighted Solution
Glass	Value	£ 37.81	£ 43.97
	Variance	$£^2$62.88	$£^2$ 1.67
Concrete	Value	£ 75.13	£ 74.35
	Variance	$£^2$ 1.25	$£^2$0.22
Steel	Value	£ 556.42	£ 556.18
	Variance	$£^2$ 2.47	$£^2$0.36
Total cost unit variance		£ 5.10 x 10^8	£ 0.93
Sum of Squared Residuals		$£^2$ 1.12 x 10^{10}	$£^2$ 20.55
Residual RMS Error		£ 2.12 x 10^4	£ 0.91

Table 1. Comparison of the unweighted and weighted solution.

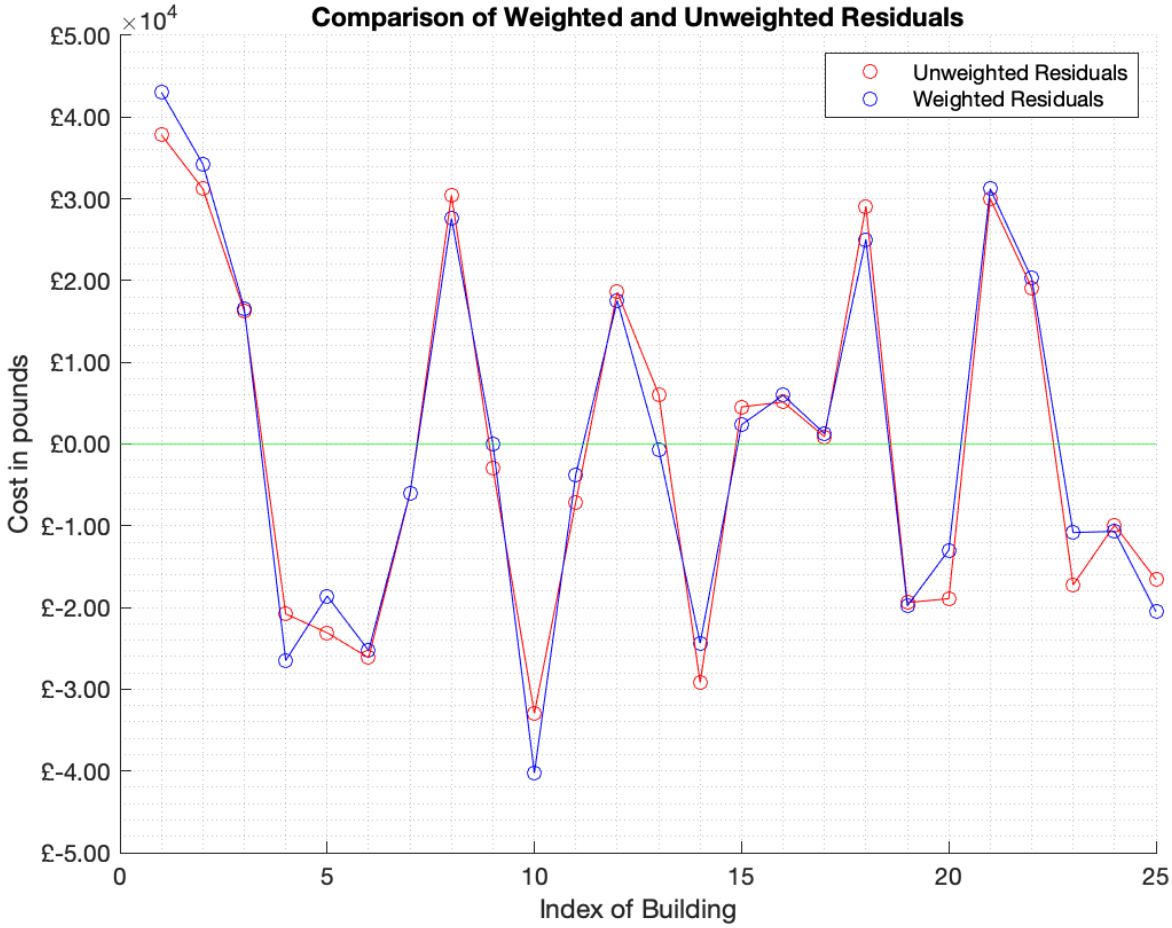

Fig 5. Plot comparing the residuals of the unweighted and weighted solution.

On comparing the parameters calculated for the unweighted and weighted analysis, it can be observed that the variances of the estimated unit costs of materials are clearly lower for the weighted analysis. Further, the total cost unit variance for the weighted solution is about 10^8 times smaller than that calculated for the unweighted solution. It can hence be deduced that due to lower variances in the unit cost estimation in the weighted analysis, the weighted analysis provides more accurate unit costs than the unweighted analysis. This can be explained by the simple fact that the weights add value to certain datasets over others that could relate to their observational accuracy and hence help provide a more accurate solution.

Further, it can be concluded that the weighted analysis was much more successful least square analysis as it lowered the sum of the squares of the residuals (i.e. brought it closest to zero) whereas the unweighted analysis gave a fairly large value for the sum of the squares of the residuals (of the order 10^{10}). Consequently, the RMS errors for the residuals were also much lower in the case of the weighted solution compared to the unweighted solution.

Overall, the weighted analysis was more successful and hence would be used for further calculations.

Part C:

Final Cost = Sum of ((Material Quantities) x X)

= £ 67077095.10420680

= £ 6.7077 x 10^7

= **£ 6.71 x 10^7**

This, however, is the estimated cost of the materials. This would vary slightly from the actual cost as the unit cost of the material is calculated using linear regressions and is an approximated value.