Solution – Assignment 1

1) a) **Derivation for the weighted TLS Solution**

$u_1, u_2, u_3$ ….. $u_N$ and $y_1, y_2, y_3$ ….. $y_N$ be the set of independent and the corresponding dependent variables. Let $\sigma_\delta$ and $\sigma_\varepsilon$ be the standard deviation of errors in $u_i$ and $y_i$. Let $\rho = \sigma_\varepsilon^2/\sigma_\delta^2$ be the ratio of the error variances.

Setting up the WTLS objective function

$$J = Min \sum \frac{(y_i - \hat{y_i})^2}{\sigma_\delta^2} + \frac{(u_i - \hat{u_i})^2}{\sigma_\epsilon^2}$$

Assuming the regression model for the WTLS $\quad \hat{y_i} = a\,\hat{x_i} + b$

$$J = Min \sum \frac{(y_i - a\,\hat{u_i} - b)^2}{\sigma_\varepsilon^2} + \frac{(u_i - \hat{u_i})^2}{\sigma_\delta^2}$$

The decision variables for the objective function *J* are $\hat{x_i}, a \text{ and } b$. To find the basis for the minimum variance of the objective function

$$\frac{\partial J}{\partial a} = 0, \qquad \frac{\partial J}{\partial b} = 0 \quad and \quad \frac{\partial J}{\partial\,\hat{u_i}} = 0$$

$$\frac{\partial J}{\partial b} = 0 \quad \Leftrightarrow b = \bar{y} - a\,\bar{u}$$

$$\frac{\partial J}{\partial a} = 0 \quad \Leftrightarrow \sum(y_i - a\hat{u}_i - b)\hat{u}_i = 0 \qquad (1)$$

$$\frac{\partial J}{\partial\,\hat{u}_i} = 0 \quad \Leftrightarrow a y_i - a^2\hat{u}_i - ab - \rho(u_i - \hat{u_i}) = 0$$

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

Substituting for $\hat{u}_i$ in (1) and solving for *a*

$$a = \frac{(S_{yy} - \rho S_{uu}) + \sqrt{(S_{yy} - \rho S_{uu})^2 + 4\rho S_{uy}{}^2}}{2S_{uy}} \qquad \textbf{(2)}$$

b) $\rho \to \infty$ can occur if $\sigma_\varepsilon^2 \to \infty$ or $\sigma_\delta^2 \to 0$ . However, both the error variances are assumed to be non-zero (in setting up the objective function). Therefore, $\rho \to \infty$ is interpreted as $\sigma_\varepsilon^2 \to \infty$. This implies that $\boldsymbol{u_i}$ has negligible error as compared to $\boldsymbol{y_i}$. In order to evaluate the limit of (1) as $\rho \to \infty$, we can multiply the numerator and denominator by $\rho$, replace $1/\rho$ by $\rho'$ and find the limit as $\rho' \to 0$. Apply L'Hospital's rule and after simplification we get $a = \frac{S_{xy}}{S_{uu}}$ (the OLS solution). $\rho \to 0$ implies $\sigma_\epsilon^2 \to \infty$ , which implies that $y_i$ has negligible error as compared to $u_i$. From (2) we easily get the limiting value of $a = \frac{S_{yy}}{S_{uy}}$ (the inverse OLS solution).

c) When b = 0

The regression model for the WTLS $\quad \hat{y_i} = a\,\hat{u_i}$

$$J = Min\,\Sigma\frac{(y_i - a\widehat{u_\iota})^2}{\sigma_\delta^2} + \Sigma\frac{(u_i - \widehat{u_\iota})^2}{\sigma_\epsilon^2}$$

The decision variables for the objective function $J$ are $\widehat{u_\iota}, a$

To find the basis for the minimum variance of the objective function

$$\frac{\partial J}{\partial a} = 0\ and\ \frac{\partial J}{\partial\,\widehat{u_\iota}} = 0$$

$$a = \frac{\Sigma\, u_i y_i}{\Sigma\,\hat{u}_i^2}\ and\ \hat{x}_i = \frac{a y_i + \rho u_i}{a^2 + \rho}$$

$$a = \frac{(\Sigma y_i^2 - \rho\Sigma u_i^2) + \sqrt{(\Sigma y_i^2 - \rho\Sigma u_i^2)^2 + 4\rho(\Sigma u_i y_i)^2}}{2\Sigma u_i y_i}$$

(d) Solution for the estimates

OLS: $\hat{u}_i = u_i;\ \hat{y}_i = a\hat{u}_i + b$

IOLS: $\hat{y}_i = y_i;\ \hat{u}_i = \frac{\hat{y}_i - b}{a}$

TLS: $\hat{u}_i = \frac{a(y_i - b) + \rho u_i}{a^2 + \rho};\ \hat{y}_i = a\hat{u}_i + b$

(2) OLS estimate for slope is 0.0186 and intercept is -5.94. TLS estimate for slope is 0.0187 and intercept is -5.96 Estimate of maximum permissible CO2 level for a 3.6 deg F rise in temperature by OLS method is 512.1 ppm while TLS estimate is 511.7 ppm. Both methods yields similar estimates.

For fitting CO2 vs time we use OLS since t does not have any error. The slope is 1.782 and intercept is 340.91. The # years (starting from 1984) estimated to reach concentration of 512.7 ppm is 96 years which implies that the limit will be reached in year 2080.

3) We can use hypothesis testing to check whether the difference in the two measurements is equal to 0. This is called a pairwise t-test. This is performed using the following procedure.
    a. Compute the difference between the measurements
    b. Compute mean and standard deviation of the difference.
    c. Use student t- distribution with corresponding degrees of freedom (N-1) to test the hypothesis that the mean of the differences is zero
    d. Since the p-value is very small, the null hypotheses is not rejected. Alternatively, the 95% confidence interval contains 0, which implies that the null hypothesis is not rejected.

| EP | CF | Difference |
|---|---|---|
| 1.98 | 1.87 | 0.11 |
| 2.31 | 2.2 | 0.11 |
| 3.29 | 3.15 | 0.14 |
| 3.56 | 3.42 | 0.14 |
| 1.23 | 1.1 | 0.13 |
| 1.57 | 1.41 | 0.16 |
| 2.05 | 1.84 | 0.21 |
| 0.66 | 0.68 | -0.02 |
| 0.31 | 0.27 | 0.04 |
| 2.82 | 2.8 | 0.02 |
| 0.13 | 0.14 | -0.01 |
| 3.15 | 3.2 | -0.05 |
| 2.72 | 2.7 | 0.02 |
| 2.31 | 2.43 | -0.12 |
| 1.92 | 1.78 | 0.14 |
| 1.56 | 1.53 | 0.03 |
| 0.94 | 0.84 | 0.1 |
| 2.27 | 2.21 | 0.06 |
| 3.17 | 3.1 | 0.07 |
| 2.36 | 2.34 | 0.02 |
| | | |
| Mean | | 0.065 |
| Standard deviation | | 0.081013 |
| t-statistic | | 3.5882 |
| p-value of statistic | | 0.00098 |
| 95% Confidence interval | | [-2.028, 2.158] |

The other way of testing the same is by computing the regression coefficient between EP and CF. As they denote the same measurements the slope of the line should be close to 1. We fit a TLS model (since both measurements contain errors having same standard deviation) for $y = ax + b$ with y as CF and x as EP. We get the OLS slope estimate as $a = 0.9924$ and b = -0.0497. Because the slope is close to 1 we can consider the new technology as a replacement for old technology (Strictly we should perform a hypotheses test to verify that the coefficient is equal to 1). The two-sided 95% confidence interval (based on t-distribution) for the slope is given by $[a \pm 2.1\,\sigma^2/s_{xx}]$ where an estimate of $\sigma^2$ is obtained as

$$\hat{\sigma}^2 = (y_i - \hat{y}_i)^2/(n-2)$$

The confidence interval obtained is [0.9830, 1.0019] includes 1. We can also conduct a test to check whether $b$ is zero.

The IOLS estimate of the a = 0.999105 and b = -0.0632. The TLS estimate for a is 0.99574 (between the OLS and IOLS estimates) and b is -0.05641.

All approaches indicate that CF is a good alternative for EP.

(3b) Under OLS assumption EP is perfect and under IOLS assumption CF is perfect. This implies best estimate under OLS assumption is 2.31 mg/l and under IOLS assumption it is 2.2 mg/l. Under TLS assumption, the estimate (assuming a = 1 and b = 0) is the average of the two = 2.255 mg/l.

(4) The orientation of each defect (assuming them to be linear) can be obtained by fitting a line between y and x coordinates of the pixel annotations defining the boundary of the defect. Since there will be error in both the x and y coordinates of pixels, TLS is best option to find the slopes. Since the location of the defect is unknown, offset parameter has to be estimated. The slopes for the three defects are 0.0091, 0.928 and 28.67, respectively. The angles made with x –axis in degrees of the three defects are 0.5, 42.8 and 88 deg respectively, implying that defect 1 is aligned with a horizontal reinforcement steel bar, defect 3 is aligned with a vertical reinforcement steel bar, while the second is not aligned with any reinforcement bar.