Solution – Assignment 1

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

x_1, x_2, x_3 ….. x_N and y_1, y_2, y_3 ….. y_N be the set of independent and the corresponding dependent variables. Let σ_δ and σ_ε be the standard deviation of errors in x_i and y_i. Let $\alpha = \sigma_\varepsilon^2/\sigma_\delta^2$ be the ratio of the error variances.

Setting up the WTLS objective function

$$\boldsymbol{J = Min \sum \frac{(y_i - \hat{y}_i)^2}{\sigma_\delta^2} + \frac{(x_i - \hat{x}_i)^2}{\sigma_\epsilon^2}}$$

Assuming the regression model for the WTLS $\boldsymbol{\hat{y}_i = a\,\hat{x}_i + b}$

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

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

$$\boldsymbol{\frac{\partial J}{\partial a} = 0, \quad \frac{\partial J}{\partial b} = 0 \quad and \quad \frac{\partial J}{\partial\,\hat{x}_i} = 0}$$

$$\frac{\partial J}{\partial b} = 0 \iff b = \bar{y} - a\,\bar{x}$$

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

$$\frac{\partial J}{\partial\,\hat{x}_i} = 0 \iff a y_i - a^2\hat{x}_i - ab - \alpha(x_i - \hat{x}_i) = 0$$

$$\hat{x}_i = \frac{a(y_i - b) + \alpha x_i}{a^2 + \alpha}$$

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

$$\boldsymbol{a = \frac{(S_{yy} - \alpha S_{xx}) + \sqrt{(S_{yy} - \alpha S_{xx})^2 + 4\alpha S_{xy}{}^2}}{2S_{xy}}} \qquad \textbf{(2)}$$

b) $\alpha \to \infty$ 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, $\alpha \to \infty$ is interpreted as $\sigma_\varepsilon^2 \to \infty$. This implies that $\boldsymbol{x_i}$ has negligible error as compared to $\boldsymbol{y_i}$**.** In order to evaluate the limit of (1) as $\alpha \to \infty$, we can multiply the numerator and denominator by α, replace 1/α by α' and find the limit as $\alpha' \to 0$. Apply L'Hospital's rule and after simplification we get $\boldsymbol{a = \frac{S_{xy}}{S_{xx}}}$ (the OLS solution). $\alpha \to 0$ implies $\sigma_\epsilon^2 \to \infty$, which implies that $\boldsymbol{y_i}$ has negligible error as compared to $\boldsymbol{x_i}$**.** From (2) we easily get the limiting value of $\boldsymbol{a = \frac{S_{yy}}{S_{xy}}}$ (the inverse OLS solution).

c) When b = 0

The regression model for the WTLS $\boldsymbol{\hat{y}_i = a\,\hat{x}_i}$

$$J = Min\, \Sigma \frac{(y_i - a\hat{x}_i)^2}{\sigma_\delta^2} + \Sigma \frac{(x_i - \hat{x}_i)^2}{\sigma_\epsilon^2}$$

The decision variables for the objective function J are $\hat{x}_i, a$

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

$$\frac{\partial J}{\partial a} = 0 \; and \; \frac{\partial J}{\partial \hat{x}_i} = 0$$

$$a = \frac{\Sigma\, \hat{x}_i y_i}{\Sigma\, \hat{x}_i^2} and\; \hat{x}_i = \frac{a y_i + \alpha x_i}{a^2 + \alpha}$$

$$a = \frac{\left(\Sigma y_i^2 - \alpha \Sigma x_i^2\right) + \sqrt{\left(\Sigma y_i^2 - \alpha \Sigma x_i^2\right)^2 + 4\alpha(\Sigma x_i y_i)^2}}{2\Sigma x_i y_i}$$

(d) Solution for the estimates

OLS: $\hat{x}_i = x_i$; $\hat{y}_i = m\hat{x}_i + c$

IOLS: $\hat{y}_i = y_i$; $\hat{x}_i = \frac{\hat{y}_i - c}{m}$

TLS: $\hat{x}_i = \frac{m(y_i - c) + \lambda x_i}{m^2 + \lambda}$; $\hat{y}_i = m\hat{x}_i + c$

2) **Problem 2**

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 σ^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.

(2b) 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.

3) OLS estimate for slope is 0.018906 and intercept is -6.04232. IOLS estimate for slope is 0.153947 and intercept is -55.929. Estimate of maximum permissible CO_2 level for a 2.7 deg F rise in temperature by OLS method is 462.4 ppm while IOLS estimate is 380.84 ppm. IOLS gives more conservative estimate.