A Least Deviation Approach of Fitting Regression Line: An Alternative Approach to Least Square Estimates

Silvey Shamsi, Mian Arif Shams Adnan, Rahmatullah Imon
Ball State University, 2000 W University Avenue, Muncie, IN 47306

Abstract
All the existing regression estimates suffer from scale problem that exaggerate the contribution of extreme observation(s) as well as outlier(s). Unlike the traditional fitting procedure of a regression line based on the least squares estimates; it uses one dimensional transformed deviation for minimizing the total sum of errors. It also suggests some new measures of Coefficient of Determination.

Key Words: Absolute Deviation, Fit by Regression, Intercept, Outlier, Relative Coefficient of Determination, Slope.

1. Introduction

All the existing regression including Simple Linear Regression, Multiple Regression, etc use Least Square Method for estimating regression parameters. Unfortunately, the least square estimates of regression parameters leave the presence of extreme observations and/or outliers exaggerated that mislead a researcher or analyst with significant (or insignificant) value of the parameters with insignificant (or significant) effects. So, one dimensional distances should be used instead of squared distances for estimating regression parameters. But, the one dimensional distances of the data from the fitted regression line makes the total sum of errors zero which does not help the mathematicians to differentiate with respect to the parameters to calculate least deviation estimate of regression parameters. This is due to the fact that sum of positive deviations (positive errors) of the dependent variable apart from the fitted regression line nullifies the negative deviations (negative errors). As a result, statisticians used least square deviations not only to make the deviations apart from the fitted regression line positive but also to make the sum of squares of errors differentiable with respect to parameters so that a class of normal equations are accessible that result least square estimates. So, there was no way of using the one dimensional naïve difference between observed values of the dependent variable and its expected or fitted values.

Fortunately, one dimensional transformed differences of the aforesaid values might be used for the sake of having the regression estimates free from exaggeration by the presence extreme observation and/or outlier(s). For estimating the regression parameters, if we retransform the normal equations for fitting the regression line, we should get a fitted regression line along with least deviation regression estimates that overcome the problem for the presence or extreme as well as outlier(s).

Attempt has been made here to find a proper transformation of the one dimensional concern difference so that we can smoothly estimate the regression parameters. It is beneficiary for us for considering one dimensional in case of regression estimates since the real observation are in one dimensional form. Using the proper transformation of the one dimensional distance from the fitted regression line we have estimated regression parameters and checked whether the estimators follow the BLUE properties. The performance and the limiting behavior of the parameters have also been observed under simulations.

2. Estimation Methods for Linear Regression

Let the simple linear regression model is

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{2.1}$$

where the intercept β_0 and the slope β_1 are unknown constant known as regression coefficients and ε is a random error component. The errors are assumed to have mean zero and unknown variance σ^2. Here the errors are uncorrelated. There is a Normal probability distribution for y at each possible value for x such that

$$E(y|x) = \beta_0 + \beta_1 x$$

and $$V(y|x) = V(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2.$$

Although the mean of y is a linear function of x that is the conditional mean of y depends on x, but the conditional variance of y does not depend on x.Moreover, the responses y are uncorrelated since the errors ε are uncorrelated. Since the parameters β_0 and β_1 are unknown, they should be estimated using sample data. Suppose that we have n pairs of data, say $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ obtained from a controlled experimental design or from an observational study or from existing historical records. Least Square method estimates β_0 and β_1 so that the sum of squares of differences between the observations $y_i$and the straight line is minimum. From equation 2.1 we can write

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i; i = 1, 2, \dots, n \tag{2.2}$$

Equation 2.1 presents the Population Regression Model and equation 2.2 expresses the Sample Regression Model.

2.1 Least Square Method for Simple Linear Regression

Now the sum of squares of deviations from the true line is

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{2.3}$$

Now the least square estimates of β_0 and $\beta_1$must satisfy

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{2.4}$$

and $$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i = 0. \tag{2.5}$$

After simplification the two normal equations are generally found such that

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{2.6}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \qquad (2.7)$$

The solution to the normal equations is

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (2.8)$$

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (2.9)$$

where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ and } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Therefore, $\hat{\beta}_0$ and $\hat{\beta}_1$ are the Least Square estimates of the intercept and slope respectively. The fitted Simple Linear Regression Model is

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x. \qquad (2.10)$$

2.2 Least Deviation Estimates for Simple Linear Regression

For the aforesaid regression model the sum of absolute deviations from the true line is

$$\sum_{i=1}^{n} |\varepsilon_i| = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|. \qquad (2.11)$$

Now the least deviation estimates of β_0 and β_1 will satisfy

$$\frac{\partial \sum_{i=1}^{n} |\varepsilon_i|}{\partial \beta_0} = 0 \qquad (2.12)$$

and

$$\frac{\partial \sum_{i=1}^{n} |\varepsilon_i|}{\partial \beta_1} = 0 \qquad (2.13)$$

which are equivalent to

$$\frac{\partial \sum_{i=1}^{n} ln|\varepsilon_i|}{\partial \beta_0} = -\infty \qquad (2.14)$$

and

$$\frac{\partial \sum_{i=1}^{n} ln|\varepsilon_i|}{\partial \beta_1} = -\infty. \qquad (2.15)$$

Therefore,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (2.16)$$

Moreover,

$$\hat{\beta}_1 = \overline{\left(\frac{y}{x}\right)} - \hat{\beta}_0 \overline{\left(\frac{1}{x}\right)} \qquad (2.17)$$

Now putting the value of $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$in equation 2.17 we get,

$$\hat{\beta}_1 = \overline{\left(\frac{y}{x}\right)} - (\bar{y} - \hat{\beta}_1 \bar{x})\overline{\left(\frac{1}{x}\right)} = \overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)} + \hat{\beta}_1 \bar{x}\overline{\left(\frac{1}{x}\right)}$$

$$\therefore \hat{\beta}_1 - \hat{\beta}_1 \bar{x}\overline{\left(\frac{1}{x}\right)} = \overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)}$$

$$\therefore \hat{\beta}_1 \left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right] = \overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)}$$

Therefore, $$\hat{\beta}_1 = \frac{\overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)}}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]} \qquad (2.18)$$

2.3 Least Square Method for Multiple Regression

Let the simple linear regression model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \qquad (2.19)$$

where the intercept β_0 and the slopes $\beta_1, \beta_2$are unknown constant known as regression coefficients and ε is a random error component. The errors are assumed to have mean zero and unknown variance σ^2. Here the errors are uncorrelated. There is a Normal probability distribution for y at each possible value for x such that

$$E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

and $$V(y|x_i) = V(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon) = \sigma^2.$$

Although the mean of y is a linear function of x that is the conditional mean of y depends on all x, but the conditional variance of y does not depend on any x. Responses y are uncorrelated since the errors ε are uncorrelated. Moreover, the independent variables are mutually independent.

Since the parameters β_i are unknown, they should be estimated using sample data. Suppose that we have $n$tuples of data, say $(y_1, x_{11}, x_{21}), (y_2, x_{12}, x_{22}), \ldots, (y_n, x_{1n}, x_{2n})$ obtained from a controlled experimental design or from an observational study or from existing historical records. Least Square method estimates β_i so that the sum of squares of differences between the observations y_i and the straight line is minimum. From equation 2.19 we can write

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i; i = 1, 2, \ldots, n \qquad (2.20)$$

Equation 2.19 presents the Population Multiple Regression Model and equation 2.20 expresses the Sample Multiple Regression Model. Now the sum of squares of deviations from the true line is

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2. \qquad (2.21)$$

Now the least square estimates of β_i must satisfy

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}) = 0, \qquad (2.22)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}) x_{1i} = 0, \qquad (2.23)$$

$$\frac{\partial S}{\partial \beta_2} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}) x_{2i} = 0. \qquad (2.24)$$

After simplification the normal equations are generally found such that

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{1i} + \hat{\beta}_2 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i \qquad (2.25)$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{1i} + \hat{\beta}_1 \sum_{i=1}^{n} x_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{1i}x_{2i} = \sum_{i=1}^{n} x_{1i}y_i \quad (2.26)$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{2i} + \hat{\beta}_1 \sum_{i=1}^{n} x_{2i}x_{1i} + \hat{\beta}_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i}y_i \quad (2.27)$$

The solution to the normal equations is

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2, \quad (2.28)$$

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} y_i x_{1i} - n\bar{y}\bar{x}_1\right)\left(\sum_{i=1}^{n} x_2{}^2 - n\bar{x}_2{}^2\right) - \left(\sum_{i=1}^{n} y_i x_{2i} - n\bar{y}\bar{x}_2\right)\left(\sum_{i=1}^{n} x_1x_2 - n\bar{x}_1\bar{x}_2\right)}{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}_1{}^2\right)\left(\sum_{i=1}^{n} x_2^2 - n\bar{x}_2{}^2\right) - \left(\sum_{i=1}^{n} x_1x_2 - n\bar{x}_1\bar{x}_2\right)^2}, \quad (2.29)$$

and $$\hat{\beta}_2 = \frac{\left(\sum_{i=1}^{n} y_i x_{2i} - n\bar{y}\bar{x}_2\right)\left(\sum_{i=1}^{n} x_1{}^2 - n\bar{x}_1{}^2\right) - \left(\sum_{i=1}^{n} y_i x_{1i} - n\bar{y}\bar{x}_1\right)\left(\sum_{i=1}^{n} x_1x_2 - n\bar{x}_1\bar{x}_2\right)}{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}_1{}^2\right)\left(\sum_{i=1}^{n} x_2^2 - n\bar{x}_2{}^2\right) - \left(\sum_{i=1}^{n} x_1x_2 - n\bar{x}_1\bar{x}_2\right)^2} \quad (2.30)$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i}$ and $\bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2i}$

Therefore, $\hat{\beta}_0$ and $\hat{\beta}_1$, $\hat{\beta}_2$ are the Least Square estimates of the intercept and slopes respectively. The fitted Simple Linear Regression Model is

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_1 x_2. \quad (2.31)$$

2.4 Least Deviation Estimates for Multiple Regression

For the aforesaid regression model the sum of absolute deviations from the true line is

$$\sum_{i=1}^{n} |\varepsilon_i| = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}|. \quad (2.32)$$

Now the least deviation estimates of β_0, $\beta_1$and β_2 will satisfy

$$\frac{\partial \sum_{i=1}^{n} |\varepsilon_i|}{\partial \beta_0} = 0, \quad (2.33)$$

$$\frac{\partial \sum_{i=1}^{n} |\varepsilon_i|}{\partial \beta_1} = 0, \quad (2.34)$$

and $$\frac{\partial \sum_{i=1}^{n} |\varepsilon_i|}{\partial \beta_2} = 0 \quad (2.35)$$

which are equivalent to

$$\frac{\partial \sum_{i=1}^{n} ln|\varepsilon_i|}{\partial \beta_0} = -\infty \quad (2.36)$$

$$\frac{\partial \sum_{i=1}^{n} ln|\varepsilon_i|}{\partial \beta_1} = -\infty, \quad (2.37)$$

and $$\frac{\partial \sum_{i=1}^{n} ln|\varepsilon_i|}{\partial \beta_2} = -\infty \quad (2.38)$$

Therefore, $$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 \quad (2.39)$$

Moreover, $$\hat{\beta}_1 = \overline{\left(\frac{y}{x_1}\right)} - \hat{\beta}_0\overline{\left(\frac{1}{x_1}\right)} - \hat{\beta}_2\overline{\left(\frac{x_2}{x_1}\right)} \qquad (2.40)$$

$$\therefore\ \hat{\beta}_1\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]-\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]\hat{\beta}_2-\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]=0 \qquad (2.41)$$

$$\therefore\ \hat{\beta}_1=\frac{\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]+\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]\hat{\beta}_2}{\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]} \qquad (2.42)$$

Again, $$\frac{\partial\sum_{i=1}^{n} ln|\varepsilon_i|}{\partial\beta_2}=\frac{\partial\sum_{i=1}^{n} ln|y_i-\beta_0-\beta_1x_{1i}-\beta_2x_{2i}|}{\partial\beta_2}=-\infty$$

$$\therefore\ \hat{\beta}_2=\overline{\left(\frac{y}{x_2}\right)}-\hat{\beta}_0\overline{\left(\frac{1}{x_2}\right)}-\hat{\beta}_1\overline{\left(\frac{x_1}{x_2}\right)} \qquad (2.43)$$

$$\therefore\ \hat{\beta}_1\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]-\hat{\beta}_2\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]+\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]=0, \qquad (2.44)$$

$$\therefore\ \hat{\beta}_2=\frac{\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]+\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]\hat{\beta}_1}{\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]}, \qquad (2.45)$$

Therefore, from (2.40) and (2.43) we get the following equations of $\hat{\beta}_1$, $\hat{\beta}_2$such that

$$\hat{\beta}_1\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]-\hat{\beta}_2\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]-\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]=0$$

$$\hat{\beta}_1\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]-\hat{\beta}_2\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]+\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]=0$$

After the cross multiplication we get the following equations

$$\frac{\hat{\beta}_1}{-\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]-\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]}$$

$$=\frac{\hat{\beta}_2}{-\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]-\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]}$$

$$=\frac{1}{-\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]+\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]}$$

Therefore,

$$\therefore\ \hat{\beta}_1=\frac{\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]+\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]}{\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]-\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]} \qquad (2.46)$$

$$\therefore\ \hat{\beta}_2=\frac{\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]\left[\overline{\left(\frac{y}{x_1}\right)}-\bar{y}\overline{\left(\frac{1}{x_1}\right)}\right]+\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]\left[\overline{\left(\frac{y}{x_2}\right)}-\bar{y}\overline{\left(\frac{1}{x_2}\right)}\right]}{\left[1-\bar{x}_1\overline{\left(\frac{1}{x_1}\right)}\right]\left[1-\bar{x}_2\overline{\left(\frac{1}{x_2}\right)}\right]-\left[\bar{x}_1\overline{\left(\frac{1}{x_2}\right)}-\overline{\left(\frac{x_1}{x_2}\right)}\right]\left[\bar{x}_2\overline{\left(\frac{1}{x_1}\right)}-\overline{\left(\frac{x_2}{x_1}\right)}\right]} \qquad (2.47)$$

3. Properties of the Least Deviation Regression Estimators

Unlike the least square estimates, least deviations estimates may follow other properties which might not be BLUE.

Theorem 3.1: *If for the simple linear regression model* $y = \beta_0 + \beta_1 x + \varepsilon$, β_0 *and* β_1 *are the unknown the intercept and the slope constant known as regression coefficients and* ε *is a random error component, and if we have* n *pairs of data, say* $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, *then R be the percentage of observations that explains the extent fit of the model such that*

$$0 < \frac{\sum_{i=1}^{n}|\hat{y}-\bar{y}|}{\sum_{i=1}^{n}|y_i-\bar{y}|} = R < 1.$$

Here R presents the percentage of observations being explained by the fitted model. The proof has been shown in Appendix A1.

Theorem 3.2: *If for the simple linear regression model* $y = \beta_0 + \beta_1 x + \varepsilon$, β_0 *and* β_1 *are the unknown the intercept and the slope constant known as regression coefficients and* ε *is a random error component, and if we have* n *pairs of data, say* $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ *obtained from a controlled experimental design or from an observational study or from existing historical records, then the estimator of* β_1 *will be*

$$\hat{\beta}_1 = \frac{\overline{\left(\frac{y}{x}\right)} - \bar{y}\,\overline{\left(\frac{1}{x}\right)}}{\left[1 - \bar{x}\,\overline{\left(\frac{1}{x}\right)}\right]}$$

such that $\hat{\beta}_1$ *is an unbiased estimator i.e.* $E(\hat{\beta}_1) = \beta_1$.

Therefore, $\hat{\beta}_1$ is an unbiased estimator for the Least Deviation Method. The proof has been placed in Appendix A2.

Theorem 3.3: *If for the simple linear regression model* $y = \beta_0 + \beta_1 x + \varepsilon$, β_0 *and* β_1 *are the unknown the intercept and the slope constant known as regression coefficients and* ε *is a random error component, and if we have* n *pairs of data, say* $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, *then the estimator of* β_1 *will be*

$$\hat{\beta}_1 = \frac{\overline{\left(\frac{y}{x}\right)} - \bar{y}\,\overline{\left(\frac{1}{x}\right)}}{\left[1 - \bar{x}\,\overline{\left(\frac{1}{x}\right)}\right]}$$

such that the variance of the estimator $\hat{\beta}_1$ *will be*

$$V(\hat{\beta}_1) = \frac{\sigma^2}{n}\frac{1}{\left[1-\bar{x}\overline{\left(\frac{1}{x}\right)}\right]^2}\left[\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\right].$$

The proof has been addressed in Appendix A3.

Proposition 3.3.1: The variance of the Least Square estimator and the Least Deviation estimators are $V(\hat{\beta}_1) = \left\{\frac{1}{\left[1-\bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\}^2 \frac{\sigma^2}{n}\left[\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\right]$ and $V(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ respectively.

$$\therefore \left\{\frac{1}{\left[1-\bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\}^2 \frac{\sigma^2}{n}\left[\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\right] : \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \text{ or, } \frac{1}{n}\left[\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\right] : \frac{\left[1-\bar{x}\overline{\left(\frac{1}{x}\right)}\right]^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]}$$

$$\therefore \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]}{n}\left[\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\right] : \left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]^2$$

4. Real Life Examples and Simulation

Various real life examples along with data will be cited to explain the credibility of the Least Deviation Estimates over Least Square Estimates. Multiple simulations will be backed to simplify and amplify the limiting behaviors of the Least Deviation Estimators and Estimates.

Example 4.1 1: Shelf-Stocking Data

For a given shelf stocking data of size 15 of paired observations of time (minutes), y, and cases stocked, x, as referred to the page 50 in the text of Montogomery on Introduction to Linear Regression Analysis (fifth edition), we created a scatter plot where we observe that almost all paired observations form a linear trend. Two linear regressions lines having following equations have been fitted according to the Least Square Method and Least Deviation Method respectively as

$$\hat{y}_{OLS} = -0.09 + 0.41x,$$

$$\hat{y}_{OLD} = -0.24 + 0.42x.$$

We also observe the following dispersion measures.

$$\sum(y-\bar{y})^2 = \sum(y-\hat{y}_{OLS})^2 + \sum(\hat{y}_{OLS}-\bar{y})^2 < \sum(y-\hat{y}_{OLD})^2 + \sum(\hat{y}_{OLD}-\bar{y})^2$$

$$229.53 = 1.21 + 228.32 < 1.36 + 240.06$$

$$\sum|y-\bar{y}| < \sum|y-\hat{y}_{OLS}| + \sum|\hat{y}_{OLS}-\bar{y}| < \sum|y-\hat{y}_{OLD}| + \sum|\hat{y}_{OLD}-\bar{y}|$$

$$51.82 < 3.51 + 51.95 < 3.46 + 53.27$$

$$\sum(y-\hat{y}_{OLS})^2 < \sum(y-\hat{y}_{OLD})^2 \text{ but } \sum|y-\hat{y}_{OLS}| > \sum|y-\hat{y}_{OLD}|$$

$$R^2{}_{OLS} = \frac{\sum(\hat{y}_{OLS}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{228.32}{229.53} = 0.995 < R^2{}_{OLD} = \frac{\sum(\hat{y}_{OLD}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{240.06}{229.53} = 1.05$$

$$R_{OLS} = \frac{\sum|\hat{y}_{OLS}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{51.95}{51.82} = 1.00 < R_{OLD} = \frac{\sum|\hat{y}_{OLD}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{53.27}{51.82} = 1.03$$

It is evident from the aforesaid dispersion measures that $\sum e^2$ is less for OLS method but $\sum|e|$ is less for OLD method. Moreover, total variation of the unidimensional data (y), being explained by two dimensional co-efficient of determination, is less for OLS fit, and also being explained by one dimensional co-efficient of determination, is also less for OLS fit. So, one dimensional scaled dispersion not only commit less error but also provide better quality of estimation for OLD method compared to OLS method. Therefore, through the current example we come to know that Ordinary Least Dispersion Method performs better than Ordinary Least Square Method at least for simple linear regression.

Example 4.2: Voltage Drop Data having extreme observation(s) and/or outlier(s)

Referring to a given voltage-drop-data of page 232 in the same book a set of 41 paired observations of voltage drop, y, and the time (seconds), x. From the scatter plot we observe that almost all paired observations form a curvy trend except one/two outliers. Two linear regressions lines having following equations have been fitted according to the Least Square Method and Least Deviation Method respectively as

$$\hat{y}_{OLS} = 39.02 - 0.64x,$$

$$\hat{y}_{OLD} = 9.25 + 1.80x.$$

We also observe the following dispersion measures.

$$\sum(y-\bar{y})^2 = \sum(y-\hat{y}_{OLS})^2 + \sum(\hat{y}_{OLS}-\bar{y})^2 < \sum(y-\hat{y}_{OLD})^2 + \sum(\hat{y}_{OLD}-\bar{y})^2$$

$$682461 = 639411 + 3449 < 689474 + 27230$$

$$\sum|y-\bar{y}| < \sum|y-\hat{y}_{OLS}| + \sum|\hat{y}_{OLS}-\bar{y}| < \sum|y-\hat{y}_{OLD}| + \sum|\hat{y}_{OLD}-\bar{y}|$$

$$1584 < 1646 + 189 < 1609 + 532$$

$$\sum(y-\hat{y}_{OLS})^2 < \sum(y-\hat{y}_{OLD})^2 \text{ but } \sum|y-\hat{y}_{OLS}| > \sum|y-\hat{y}_{OLD}|$$

$$R^2{}_{OLS} = \frac{\sum(\hat{y}_{OLS}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{3449}{682461} = 0.005 < R^2{}_{OLD} = \frac{\sum(\hat{y}_{OLD}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{27230}{682461} = 0.04$$

$$R_{OLS} = \frac{\sum|\hat{y}_{OLS}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{189}{1584} = 0.12 < R_{OLD} = \frac{\sum|\hat{y}_{OLD}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{532}{1584} = 0.34$$

It is evident from the aforesaid dispersion measures that $\sum e^2$ is less (639411<689474) for OLS method but $\sum|e|$ is less (1646>1609) for OLD method. Moreover, total variation of the unidimensional data (y), being explained by two dimensional co-efficient of determination, is less(0.005<0.04) for OLS fit, and also being explained by one dimensional co-efficient of determination, is alsoless (0.12<0.34) for OLS fit. So, one dimensional scaled dispersion not only commit less error but also provide acutely better quality of estimation for OLD method compared to OLS method. So, the Ordinary Least Dispersion Method is behaving sharply better than Ordinary Least Square Method for the current example.

Example 4.3: Voltage Drop Data without extreme observation(s) and/or outlier(s)

The set of data has 3 pairs of extreme observations and/or outliers which are (8.33, 0.01), (823, 5), (14, 95). After dropping these 3 extreme observations, we have found the following features. Two linear regressions lines having following equations have been fitted according to the Least Square Method and Least Deviation Method respectively as

$$\hat{y}_{OLS} = 9.69 - 0.16x,$$

$$\hat{y}_{OLD} = 7.41 + 0.38x.$$

We also observe the following dispersion measures.

$$\sum(y-\bar{y})^2 = \sum(y-\hat{y}_{OLS})^2 + \sum(\hat{y}_{OLS}-\bar{y})^2 < \sum(y-\hat{y}_{OLD})^2 + \sum(\hat{y}_{OLD}-\bar{y})^2$$

$$236 = 203 + 33 < 261 + 179$$

$$\sum|y-\bar{y}| < \sum|y-\hat{y}_{OLS}| + \sum|\hat{y}_{OLS}-\bar{y}| < \sum|y-\hat{y}_{OLD}| + \sum|\hat{y}_{OLD}-\bar{y}|$$

$$83 < 79 + 31 < 81 + 72$$

$\sum(y-\hat{y}_{OLS})^2 = 203 \ll \sum(y-\hat{y}_{OLD})^2$=261and $\sum|y-\hat{y}_{OLS}|$=79 < $\sum|y-\hat{y}_{OLD}|$ = 81

$$R^2{}_{OLS} = \frac{\sum(\hat{y}_{OLS}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{33}{236} = 0.14 < R^2{}_{OLD} = \frac{\sum(\hat{y}_{OLD}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{179}{236} = 0.76$$

$$R_{OLS} = \frac{\sum|\hat{y}_{OLS}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{31}{83} = 0.37 < R_{OLD} = \frac{\sum|\hat{y}_{OLD}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{72}{83} = 0.87$$

$\sum e^2$ is more less (203<<261) but $\sum|e|$ is slight less (79<81) for OLS method. Moreover, total variation of the unidimensional data (y), being explained by two dimensional co-efficient of determination, is less (0.14<0.76) for OLS fit, and also being explained by one dimensional co-efficient of determination, is also less (0.37<0.87) for OLS fit. So, one dimensional scaled dispersion not only conductalmost same error but also provide better quality of estimation for OLD method compared to OLS method. So, the Ordinary Least Dispersion Method worth better estimation compared to Ordinary Least Square Method after eliminating the outliers.

Example 4.4: Bacteria in Canned Food: Exponential Data

In page 203of the same text, a set of 12 paired observations of average number of surviving bacteria in a canned food product, y, and the times of exposure to 300 degree of Fahrenheit heat (minutes), x, are available. From the scatter plot we observe that the paired observations form an exponential trend.

Two linear regressions lines having following equations have been fitted according to the Least Square Method and Least Deviation Method respectively as

$$\hat{y}_{OLS} = 142.20 - 12.50x,$$

$$\hat{y}_{OLD} = 166.25 - 16.18x.$$

We also observe the following dispersion measures.

$$\sum(y-\bar{y})^2 = \sum(y-\hat{y}_{OLS})^2 + \sum(\hat{y}_{OLS}-\bar{y})^2 < \sum(y-\hat{y}_{OLD})^2 + \sum(\hat{y}_{OLD}-\bar{y})^2$$

$$22269 = 3348 + 18921 < 5307 + 582$$

$$\sum|y-\bar{y}| < \sum|y-\hat{y}_{OLS}| + \sum|\hat{y}_{OLS}-\bar{y}| < \sum|y-\hat{y}_{OLD}| + \sum|\hat{y}_{OLD}-\bar{y}|$$

$$453 < 150 + 449 < 223 + 582$$

$\sum(y-\hat{y}_{OLS})^2 = 3348 \sum(y-\hat{y}_{OLD})^2$= 5307 and $\sum|y-\hat{y}_{OLS}|$= 150<$\sum|y-\hat{y}_{OLD}|$ = 223

$$R^2{}_{OLS} = \frac{\sum(\hat{y}_{OLS}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{18921}{22269} = 0.85 > R^2{}_{OLD} = \frac{\sum(\hat{y}_{OLD}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{582}{22269} = 0.03$$

$$R_{OLS} = \frac{\sum|\hat{y}_{OLS}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{449}{453} = 0.81 < R_{OLD} = \frac{\sum|\hat{y}_{OLD}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{582}{453} = 1.05$$

$\sum e^2$ is more less (3348<<5307) but $\sum|e|$ is not much more less (150<223) for OLS method. Moreover, total variation of the unidimensional data (y), being explained by two dimensional co-efficient of determination, is greater (0.85>0.03) for OLS fit, and being explained by one dimensional co-efficient of determination, is less (0.81<1.05) for OLS fit. Although, one dimensional scaled dispersion conduct more error for exponential data, but commit better quality of estimation for OLD method compared to OLS method since one dimensional coefficient of dispersion is greater for OLD (R_{OLS}= 0.81 $<$ R_{OLD} = 1.05). So, the Ordinary Least Dispersion Method worth better estimation compared to Ordinary Least Square Method even for the exponential data.

Example 4.5: Delivery time Data: Two independent variables

A set of 25 paired observations of average number of delivery time, y, and the number of cases, x_1, and the distances, x_1, are available in page 74 of the same text. From the scatter plot we observe that the 3-tupled observations form a 3D form.

Two linear regressions lines having following equations have been fitted according to the Least Square Method and Least Deviation Method respectively as

$$\hat{y}_{OLS} = 2.34 - 1.62x_1 + 0.01x_2,$$

$$\hat{y}_{OLD} = 3.94 + 1.68x_1 + 0.01x_2.$$

We also observe the following dispersion measures.

$$\sum(y-\bar{y})^2 = \sum(y-\hat{y}_{OLS})^2 + \sum(\hat{y}_{OLS}-\bar{y})^2 < \sum(y-\hat{y}_{OLD})^2 + \sum(\hat{y}_{OLD}-\bar{y})^2$$

5785 = 234 + 5551 > 280 + 4774

$$\sum|y-\bar{y}| < \sum|y-\hat{y}_{OLS}| + \sum|\hat{y}_{OLS}-\bar{y}| < \sum|y-\hat{y}_{OLD}| + \sum|\hat{y}_{OLD}-\bar{y}|$$

251 $<$ 57 + 257 > 58 + 235

$\sum(y-\hat{y}_{OLS})^2 = 234 < \sum(y-\hat{y}_{OLD})^2$= 280 and $\sum|y-\hat{y}_{OLS}|$ = 150 <$\sum|y-\hat{y}_{OLD}|$ = 223

$$R^2{}_{OLS} = \frac{\sum(\hat{y}_{OLS}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{5551}{5785} = 0.96 > R^2{}_{OLD} = \frac{\sum(\hat{y}_{OLD}-\bar{y})^2}{\sum(y-\bar{y})^2} = \frac{4774}{5785} = 0.83$$

$$R_{OLS} = \frac{\sum|\hat{y}_{OLS}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{257}{251} = 1.02 > R_{OLD} = \frac{\sum|\hat{y}_{OLD}-\bar{y}|}{\sum|y-\bar{y}|} = \frac{235}{251} = 0.94$$

$\sum e^2$ is less (234<280) for OLS but $\sum|e|$ is not much more (150<223) for OLD method. Moreover, total variation of the unidimensional data (y), being explained by two dimensional co-efficient of determination, is greater (0.96>0.83) for OLS fit, and also being explained by one dimensional co-efficient of determination, is also greater (1.02>0.94) for OLS fit. One dimensional scaled dispersion conduct little bit more error and commit less good quality of estimation for OLD method compared to OLS method. So, the Ordinary Least Dispersion Method does not worth better estimation compared to Ordinary Least Square Method for multiple regression.

5. Comparison Between Least Deviation Estimates and Least Square Estimates

Performances and potentialities will be better for the Least Deviations Estimates compared to those of Least Square Method due to the reason is that the estimators of the regression coefficients are maintaining the similar dimension of the true data since they are expressed in terms of the mathematical operations with unidimensional scaling. Besides, for the presence of extreme values or outliers, the two dimensional scaling of the dispersions of the fitted line from the data are being exaggerated. So, the malicious influences of the extreme value(s) and/or outlier(s) drastically affects the overall estimation of parameters in regression based on Least Square Method. On the other hand, the absolute deviations for the dispersion of the observations apart from fitted regression line in regression estimators are apathetic to the extreme observation(s) and/or outlier(s) and equi-sensitive to all observations. In all of the aforesaid examples (1 to 3), OLD method gives batter estimation than the OLS method since one dimensional co-efficient of determination is greater for OLD rather than OLS for each case.

Moreover, error and quality estimation are two complementary factors. If error increases and quality estimation decreases or vice versa, then it is better to assess the relative performance like estimation by error. If for one dimensional data, one dimensional relative variation due to regression with respect to total error encountered for fitting the model by one method is greater than that of the other method, then it is better to use the first method for fitting regression line. Interestingly enough it is observed that the relative fit to the error as

$$FBR = \frac{Fit}{Error} = \frac{\sum|\hat{y}_{Method} - \bar{y}|}{\sum|y - \hat{y}_{Method}|} \tag{5.1}$$

is greater for Ordinary Least Deviation Method compared to Ordinary Least Square Method for all four cases. That is

$$FBR_{OLD} = \frac{Fit_{OLD}}{Error_{OLD}} = \frac{\sum|\hat{y}_{OLD} - \bar{y}|}{\sum|y - \hat{y}_{OLD}|} > FBR_{OLS} = \frac{Fit_{OLS}}{Error_{OLS}} = \frac{\sum|\hat{y}_{OLS} - \bar{y}|}{\sum|y - \hat{y}_{OLS}|}. \tag{5.2}$$

In the afore-described three examples, the FBR for OLD is always greater than OLS (15.4>14.9, 0.33>0.11, 0.9>0.16).

In example 4 and 5, one dimensional FBR for OLS is greater than one dimensional FBR for OLD (2.99>2.61, 4.50>4.04). Two dimensional FBR for OLS is lower than two dimensional FBR for OLD (6.65<7.05) for example 4 but higher for example 5 (23.74>17.08). In example 4, the data follow exponential pattern. As a result, one dimensional and two dimensional relative co-efficient of determination (Fit by Error [FBR]) are almost equivalent (4.50$\approx$ 4.04, 6.65$\approx$ 7.05). But in example 5, two dimensional relative co-efficient of determination for OLS is greater than that for OLD, because in the current multiple regression there are two independent variables and the regression model is a plane rather that a line in 3D space. The two dimensional co-efficient of determination can be described as below

$$FBR = \frac{Fit}{Error} = \frac{\sum(\hat{y}_{Method} - \bar{y})^2}{\sum(y - \hat{y}_{Method})^2}. \tag{5.3}$$

Here we also observe the following inequalities.

$$FBR_{OLS} = \frac{Fit_{OLS}}{Error_{OLS}} = \frac{\sum(\hat{y}_{OLS} - \bar{y})^2}{\sum(y - \hat{y}_{OLS})^2} > FBR_{OLD} = \frac{Fit_{OLD}}{Error_{OLD}} = \frac{\sum(\hat{y}_{OLD} - \bar{y})^2}{\sum(y - \hat{y}_{OLD})^2} \tag{5.4}$$

As a result, the distance between any point/data and the fitted plane is a perpendicular plane rather than a perpendicular line. So, two dimensional relative co-efficient of determination is greater and better for measuring the quality of fit of the model based on OLS estimation in example 5.

Conclusion

Various Random Effect Models of the Analysis of Variance and Design of Experiment and several Time Series Models and multiple Time Series Regression Models and multiple Non-parametric Regression Models can be modified using the same concept of Least Deviation Method instead of Least Square Method and Weighted Least Square Method or Reweighted Least Square Method. The form of Coefficient of Determination, Outlier Detection and Cut-off Bandwidth will also be modified for the same method.

References

Adnan, M. A. S, (2015). An Alternative Approach of Fitting Regression Line not based on least Square Estimates. Unpublished manuscript for the presentation in the Joint Statistical Meetings 2015, American Statistical Association.

Cook R. D. and Weisberg S. (1999). Applied Regression including Computing and Graphics. Wiley Publishers.

Dick R. Wittink (1988). The Application of Regression Analysis. Allyn and Bacon Publishers.

Draper N. R., Smith H. (1981). Applied Regression Analysis. Second Edition. Wiley Publisher.

Netter J., Wasserman W. and Kutner M. H. (1983). Applied Linear Regression Models. Irwin Publisher.

Montogomery D. C., Peck E. A. and Vining G. G. (2012). Introduction to Linear Regression Analysis. Fifth Edition. Wiley Publisher.

Myers R. H. (1986). Classical and Modern Regression with Applications. PWS Publishers.

Robinson E. A. (1981). Least Square Regression Analysis in terms of Linear Algebra. Goose Pond Press.

Rousseeuw P. J. and Leroy A. M. (1987). Robust Regression and Outlier Detection. Wiley Publisher.

Seber, G. A. F. (1977). Linear Regression Analysis. Wiley Publishers.

Appendix A

A1: Proof of Theorem 3.1

$$\sum_{i=1}^{n}|y_i - \bar{y}| = \sum_{i=1}^{n}|(y_i - \hat{y}) + (\hat{y} - \bar{y})| \leq \sum_{i=1}^{n}|y_i - \hat{y}| + \sum_{i=1}^{n}|\hat{y} - \bar{y}|$$

$$\therefore 1 \leq \frac{\sum_{i=1}^{n}|y_i - \hat{y}|}{\sum_{i=1}^{n}|y_i - \bar{y}|} + \frac{\sum_{i=1}^{n}|\hat{y} - \bar{y}|}{\sum_{i=1}^{n}|y_i - \bar{y}|}$$

$$\therefore 1 - \frac{\sum_{i=1}^{n}|y_i - \hat{y}|}{\sum_{i=1}^{n}|y_i - \bar{y}|} \leq \frac{\sum_{i=1}^{n}|\hat{y} - \bar{y}|}{\sum_{i=1}^{n}|y_i - \bar{y}|} = \mathrm{R}$$

$$\therefore 0 \leq R \leq 1.$$

A2: Proof of Theorem 3.2

$$E(\hat{\beta}_1) = E\left\{\frac{\overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)}}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\} = \frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\left[E\left\{\overline{\left(\frac{y}{x}\right)}\right\} - \overline{\left(\frac{1}{x}\right)}E(\bar{y})\right].$$

Now,

$$E\left[\overline{\left(\frac{y}{x}\right)}\right] = E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}\right) = \frac{1}{n}\mathrm{E}\left(\frac{y_1}{x_1} + \frac{y_2}{x_2} + \cdots + \frac{y_n}{x_n}\right) = \frac{1}{n}\left(\frac{\beta_0}{x_1} + \beta_1 + \frac{\beta_0}{x_2} + \beta_1 + \cdots + \frac{\beta_0}{x_n} + \beta_1\right)$$

$$= \beta_0 \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} + \beta_1 = \beta_0\overline{\left(\frac{1}{x}\right)} + \beta_1$$

Since, $y \sim N(\beta_0 + \beta_1 x, \sigma^2), \therefore \frac{y}{x} \sim N(\frac{\beta_0}{x} + \beta_1, \frac{\sigma^2}{x^2})$

$$\therefore E(\hat{\beta}_1) = \frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\left[\beta_0\overline{\left(\frac{1}{x}\right)} + \beta_1 - \overline{\left(\frac{1}{x}\right)}(\beta_0 + \beta_1\bar{x})\right] = \frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\left[\beta_1 - \beta_1\bar{x}\overline{\left(\frac{1}{x}\right)}\right] = \beta_1 .$$

A3: Proof of Theorem 3.3

$$V(\hat{\beta}_1) = V\left\{\frac{\overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)}}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\} = \left\{\frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\}^2\left[V\left\{\overline{\left(\frac{y}{x}\right)}\right\} + \left\{\overline{\left(\frac{1}{x}\right)}\right\}^2 V(\bar{y})\right].$$

Now,

$$V\left[\overline{\left(\frac{y}{x}\right)}\right] = V\left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}\right) = \frac{1}{n^2}V\left(\frac{y_1}{x_1} + \frac{y_2}{x_2} + \cdots + \frac{y_n}{x_n}\right) = \frac{1}{n^2}\left(\frac{\sigma^2}{{x_1}^2} + \frac{\sigma^2}{{x_2}^2} + \cdots + \frac{\sigma^2}{{x_n}^2}\right)$$

$$= \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\frac{1}{{x_i}^2} = \frac{\sigma^2}{n}\overline{\left(\frac{1}{x^2}\right)}$$

$$V\left[\bar{y}\overline{\left(\frac{1}{x}\right)}\right] = \overline{\left(\frac{1}{x}\right)}^2 V(\bar{y}) = \overline{\left(\frac{1}{x}\right)}^2\frac{\sigma^2}{n}$$

$$V(\hat{\beta}_1) = V\left\{\frac{\overline{\left(\frac{y}{x}\right)} - \bar{y}\overline{\left(\frac{1}{x}\right)}}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\} = \left\{\frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\}^2\left[V\left\{\overline{\left(\frac{y}{x}\right)}\right\} + V\left\{\bar{y}\overline{\left(\frac{1}{x}\right)}\right\}\right]$$

$$= \left\{\frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\}^2\left[\frac{\sigma^2}{n}\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\frac{\sigma^2}{n}\right]$$

$$\therefore V(\hat{\beta}_1) = \left\{\frac{1}{\left[1 - \bar{x}\overline{\left(\frac{1}{x}\right)}\right]}\right\}^2\frac{\sigma^2}{n}\left[\overline{\left(\frac{1}{x^2}\right)} + \overline{\left(\frac{1}{x}\right)}^2\right].$$