Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 7 (1): 20-26, 2007
ISSN 1812-5654

An Investigation of Causation: The Unreplicated Linear Functional Relationship Model

[1]Chang Yun Fah, [2]Abdul Ghapor Hussin and [3]Omar Mohd Rijal
[1]Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia
[2]Centre for Foundation Studies in Science, [3]Institute of Mathematical Sciences,
University of Malaya, 50603 Kuala Lumpur, Malaysia

Abstract: This study investigates the Unreplicated Linear Functional Relationship (ULFR) model where the measurement error term, δ, is introduced. The coefficient of determination (COD) of ULFR, denoted by R_F^2 is proposed and it properties are investigated. When the introduction of δ increases significantly the COD, we say that the causation factor has been incorporated into the independent variable. Present result on the Malaysian road accident data illustrates the causal relationship between the socio-biological factors and road accident may be explained.

Key words: Causation, functional relationship, coefficient of determination, road accident

INTRODUCTION

The seriousness of the road accident problem warrants a deeper investigation into the possible causes. In particular the ULFR model is investigated. The studies of functional and structural relationships have been focused on estimating the parameters and the measurement error models since it was first introduced by Adcock in year 1877 (Sprent, 1990). Several methods were proposed, but the widely used estimator is the maximum likelihood method (Dolby, 1976; Fuller, 1987; Kendall and Stuart, 1979). An early study to check the adequacy of the ULFR model is given by Fuller (1987) who suggests using the residual plot as an indication of nonlinearity, lack of homogeneity of the error variances, nonnormality of the errors and outlier observations. In this study we propose the COD as a measure to check the model adequacy and the assumption of independent variable. Without loss of generality, the properties of this measurement for case $\lambda = 1$ is investigated. A large value of COD of the ULFR relative to the COD of ordinary Simple Linear (SL) regression model is indicative of causation.

As pointed out by Fuller (1987), the assumption that the independent variable can be measured exactly may not be realistic in many situations. The estimates of independent variable may contain measurement error arising from trying to quantify a variable that has no physical dimension. For example in the road accident problem, it has been shown that the male motorist is more prone to be involved in accident. One possible explanation is the performance of male drivers may vary with state of mind. Since this factor, the state of mind cannot be measured directly, it is more reasonable to consider them as a measurement error rather than introduce a new independent variable into the model.

UNREPLICATED LINEAR FUNCTIONAL RELATIONSHIP MODEL

Suppose that X and Y are two linearly related unobservable variables

$$Y_i = a + \beta X_i \qquad (1)$$

and the two corresponding random variables x and y are observed with error δ and ε, respectively as

$$\left.\begin{aligned} x_i &= X_i + d_i \\ y_i &= Y_i + \varepsilon_i \end{aligned}\right\} \quad i = 1, 2, ..., n \qquad (2)$$

The following conditions are assumed

$$\left.\begin{aligned} &E(\delta_i) = E(\varepsilon_i) = 0, \quad Var(\delta_i) = \sigma_\delta^2, \quad Var(\varepsilon_i) = \sigma_\varepsilon^2, \quad \forall i \\ &Cov(\delta_i, \delta_j) = Cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j \\ &Cov(\delta_i, \varepsilon_j) = 0, \quad \forall i, j \end{aligned}\right\} \qquad (3)$$

where, δ_i and ε_i are mutually independent and normally distributed random variables. Equation 1 and 2 are known as ULFR model when there is only one relationship between the two variables X and Y. When the ratio of the error variances is known, that is $\sigma_\varepsilon^2/\sigma_\delta^2 = \lambda$, then the maximum likelihood estimators of parameters α, β, σ_δ^2 and X_i are:

Corresponding Author: Dr. Abdul Ghapor Hussin, Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia Fax: 603 79576478

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (4)$$

$$\hat{\beta} = \frac{(S_{yy} - \lambda S_{xx}) + \left\{(S_{yy} - \lambda S_{xx})^2 + 4\lambda S_{xy}^2\right\}^{1/2}}{2S_{xy}} \quad (5)$$

$$\hat{\sigma}_\delta^2 = \frac{1}{n-2}\left[\sum\left(x_i - \hat{X}_i\right)^2 + \frac{1}{\lambda}\sum\left(y_i - \hat{\alpha} - \hat{\beta}\hat{X}_i\right)^2\right] \quad (6)$$

and
$$\hat{X}_i = \frac{\lambda x_i + \hat{\beta}\left(y_i - \hat{\alpha}\right)}{\lambda + \hat{\beta}^2} \quad (7)$$

where,
$$\bar{y} = \frac{\sum y_i}{n},\ \bar{x} = \frac{\sum x_i}{n}$$

$$S_{yy} = \sum\left(y_i - \bar{y}\right)^2,\ S_{xx} = \sum\left(x_i - \bar{x}\right)^2 \text{ and}$$
$$S_{xy} = \sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$$

DERIVATION OF THE COEFFICIENT OF DETERMINATION (COD)

In an ordinary SL regression analysis, we look at the COD as a measure of the variability in y explained by the regression model. The construction of the COD is a good practice when fitting the ULFR model.
The Eq. 1 and 2 can be rewritten as

$$y_i = \alpha + \beta X_i + \varepsilon_i \text{ for } i = 1,2,\ \ldots,\ n \quad (8)$$

If we substitute X_i by (x_i-δ_i) in (8), then we have the expression

$$y_i = \alpha + \beta x_i + (\varepsilon_i - \beta\delta_i)$$
$$= \alpha + \beta x_i + V_i$$

Where, the errors of the model $V_i = (\varepsilon_i - \beta\delta_i) = y_i - (\alpha + \beta x_i)$, for i = 1, 2, ..., n is a normally distributed random variable with zero mean and variance $\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2$. If $\hat{\alpha}$ and $\hat{\beta}$ are the estimates of α and β respectively, then

$$\hat{V}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta}x_i),\ \text{for } i = 1, 2, \ldots, n \quad (9)$$

will be the residual of the model. From Kendall and Stuart (1979) and Andrson (1984), that the sum of squared distances of the observed points from the fitted line or the residual sum of squares (SS_E) is given as:

$$SS_E = \frac{\sum\left\{y_i - \left(\hat{\alpha} + \hat{\beta}x_i\right)\right\}^2}{\left(\lambda + \hat{\beta}^2\right)} = \frac{S_{yy} - 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{xx}}{\left(\lambda + \hat{\beta}^2\right)} \quad (10)$$

We shall consider here that the ratio of the error variances is equal to one (λ = 1). For those cases when λ ≠ 1, we can always reduce this to the case of λ = 1 by dividing the observed values of y by $\lambda^{1/2}$ (Kendall and Stuart, 1979). Hence, we have:

$$SS_E = \frac{S_{yy} - 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{xx}}{1 + \hat{\beta}^2} \quad (11)$$

In the same way as ordinary linear regression, we can now define the COD of the ULFR (R_F^2) as the proportion of variation explained by the variable x, that is:

$$R_F^2 = \frac{SS_R}{S_{yy}} \quad (12)$$

where, SS_R is the regression sum of squares which can be derived as:

$$SS_R = S_{yy} - SS_E = S_{yy} - \frac{S_{yy} - 2\hat{\beta}S_{xy} + \hat{\beta}^2 S_{xx}}{1 + \hat{\beta}^2}$$
$$= \frac{\hat{\beta}^2 S_{yy} + 2\hat{\beta}S_{xy} - \hat{\beta}^2 S_{xx}}{1 + \hat{\beta}^2} = \frac{\hat{\beta}^2\left(S_{yy} - S_{xx}\right) + 2\hat{\beta}S_{xy}}{1 + \hat{\beta}^2} \quad (13)$$

We can summarise our proposed COD with the following results.

Result 1: Let the ratio of the error variances be known and is equal to one, $\sigma_e^2/\sigma_d^2 = \lambda = 1$, then the COD for the ULFR model is

$$R_F^2 = \frac{SS_R}{S_{yy}} = \frac{\hat{\beta}S_{xy}}{S_{yy}} \quad (14)$$

Proof: When λ = 1, then $\hat{\beta}$ in Eq. 5 becomes

$$\hat{\beta} = \frac{\left(S_{yy} - S_{xx}\right) + \left[\left(S_{yy} - S_{xx}\right)^2 + 4S_{xy}^2\right]^{1/2}}{2S_{xy}}$$

and $$\hat{\beta}^2 = \frac{(S_{yy} - S_{xx})\hat{\beta} + S_{xy}}{S_{xy}}$$

Notice that $\frac{SS_R}{S_{yy}} = \frac{\hat{\beta} S_{xy}}{S_{yy}} \Leftrightarrow SS_R = \hat{\beta} S_{xy}$ From Eq. 13, we have,

$$\frac{\hat{\beta}^2 (S_{yy} - S_{xx}) + 2\hat{\beta} S_{xy}}{1 + \hat{\beta}^2} = \hat{\beta} S_{xy}$$

$$\Leftrightarrow \hat{\beta}^2 (S_{yy} - S_{xx}) = \hat{\beta}^3 S_{xy} - \hat{\beta} S_{xy}$$

$$\Leftrightarrow \hat{\beta}(S_{yy} - S_{xx}) = \left(\hat{\beta}^2 - 1\right) S_{xy}$$

From the R.H.S.

$$\left(\hat{\beta}^2 - 1\right) S_{xy} = \left[\frac{(S_{yy} - S_{xx})\hat{\beta} + S_{xy}}{S_{xy}} - 1\right] S_{xy}$$

$$= \left[\frac{(S_{yy} - S_{xx})\hat{\beta} + S_{xy} - S_{xy}}{S_{xy}}\right] S_{xy}$$

$$= \hat{\beta}(S_{yy} - S_{xx})$$

Result 2: Let $\hat{\beta}_s$ and $\hat{\beta}$ be the slope estimator for the simple linear regression and the ULFR, respectively. The corresponding COD are:

$$R_S^2 = \frac{\hat{\beta}_S S_{xy}}{S_{yy}} \quad \text{and} \quad R_F^2 = \frac{\hat{\beta} S_{xy}}{S_{yy}} \quad (\text{when } l = 1)$$

Then $0 \le R_S^2 \le R_F^2 \le 1$.

Proof: From the regression sum of squares, we obtain

$$0 \le SS_R = S_{yy} - SS_E \le S_{yy}$$

$$0 \le R_F^2 = \frac{SS_R}{S_{yy}} \le 1$$

we also know that simple linear regression has $0 \le R_S^2 \le 1$. We now proof $R_S^2 \le R_F^2$ by contradiction. Assuming that

$$R_S^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} > \frac{\left\{(S_{yy} - S_{xx}) + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}\right\} S_{xy}}{2S_{xy} S_{yy}} = R_F^2$$

$$\frac{2S_{xy}^2}{S_{xx}} - (S_{yy} - S_{xx}) > \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}$$

$$\frac{4S_{xy}^4}{S_{xx}^2} + (S_{yy} - S_{xx})^2 - \frac{4S_{xy}^2 (S_{yy} - S_{xx})}{S_{xx}} > (S_{yy} - S_{xx})^2 + 4S_{xy}^2$$

$$S_{xy}^4 - S_{xy}^2 S_{xx} (S_{yy} - S_{xx}) > S_{xy}^2 S_{xx}^2$$

$$S_{xy}^4 > S_{xy}^2 S_{xx} (S_{xx} + S_{yy} - S_{xx})$$

$$S_{xy}^2 > S_{xx} S_{yy}$$

Also from $R_S^2 = S_{xy}^2 / S_{xx} S_{yy} < 1 \Rightarrow S_{xy}^2 < S_{xx} S_{yy}$, this is a contradiction. Hence, we have $R_S^2 \le R_F^2$.

PROPERTIES OF R_F^2

Since R_F^2 is computed following the same method as R_S^2, some of the properties of R_S^2 may also remain for R_F^2. It is known that R_S^2 does not measure the appropriateness of the linear model (Montgomery and Peck, 1992). This holds for R_F^2 too. As an example, when a nonlinear model has a large R_S^2 value, it is obvious that R_F^2 will also be large (Result 2).

Secondly, the magnitude of R_F^2 depends on the range of variability in the variables X and Y. From Eq. 14, we have

$$R_F^2 = \frac{\hat{\beta} S_{xy}}{S_{yy}}$$

$$= \frac{\left\{(S_{yy} - S_{xx}) + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}\right\} S_{xy}}{2S_{xy} S_{yy}}$$

$$= \frac{\left\{(S_{yy} - S_{xx}) + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}\right\}}{2S_{yy}}$$

Let var(X) and var(Y) be variances of X and Y, respectively then $\text{var}(X) = S_{xx} / n - 1$ and $\text{var}(Y) = S_{yy} / n - 1$. We consider three possible cases of relationship between var(X) and var(Y) when $S_{xx} > 0$ and $S_{yy} > 0$.

Case 1: If var(X) = var(Y) $\Leftrightarrow$ $S_{xx} = S_{yy}$, then

$$R_F^2 = \frac{\sqrt{4S_{xy}^2}}{2S_{yy}} = \frac{2S_{xy}}{2S_{yy}} = \frac{S_{xy}}{S_{yy}}$$

when compared with Eq. 14, this holds when $\hat{\beta} = 1$.

Case 2: If var(X) < var(Y) ⇔ $S_{yy}/k = S_{xx} < S_{yy}$ for k > 1, then

$$R_F^2 = \frac{\left(S_{yy} - \frac{S_{yy}}{k}\right) + \sqrt{\left(S_{yy} - \frac{S_{yy}}{k}\right)^2 + 4S_{xy}^2}}{2S_{yy}}$$

$$= \frac{\frac{(k-1)}{k}S_{yy} + \sqrt{\frac{(k-1)^2}{k^2}S_{yy}^2 + 4S_{xy}^2}}{2S_{yy}}$$

$$> \frac{\frac{(k-1)}{k}S_{yy} + \sqrt{4S_{xy}^2}}{2S_{yy}} = \frac{(k-1)}{2k} + \frac{S_{xy}}{S_{yy}} > \frac{S_{xy}}{S_{yy}}$$

when compared with Eq. 14, this holds when $\hat{\beta} > 1$.

Case 3: If var(X) > var(Y) ⇔ $kS_{yy} = S_{yy}$ for k>1, then

$$R_F^2 = \frac{-\left(kS_{yy} - S_{yy}\right) + \sqrt{\left(S_{yy} - kS_{yy}\right)^2 + 4S_{xy}^2}}{2S_{yy}}$$

$$= \frac{-(k-1)S_{yy} + \sqrt{(1-k)^2 S_{yy}^2 + 4S_{xy}^2}}{2S_{yy}}$$

$$< \sqrt{\frac{(1-k)^2 S_{yy}^2 + 4S_{xy}^2}{4S_{yy}^2}}$$

$$= \sqrt{\frac{(1-k)^2}{4} + \frac{S_{xy}^2}{S_{yy}^2}} < \frac{(1-k)}{2} + \frac{S_{xy}}{S_{yy}} < \frac{S_{xy}}{S_{yy}}$$

when compared with Eq. 14, this holds when $\hat{\beta} < 1$.

We can summarize the above properties with the following result.

Result 3: Let $S_{xx} > 0$, $S_{yy} > 0$ and $Sxx = kS_{yy}$ or var(x) = k var(y), then

a) $R_F^2 > \frac{S_{xy}}{S_{yy}} \Leftrightarrow \hat{\beta} > 1, \qquad 0 < k < 1$

b) $R_F^2 = \frac{S_{xy}}{S_{yy}} \Leftrightarrow \hat{\beta} = 1, \qquad k = 1$

c) $R_F^2 < \frac{S_{xy}}{S_{yy}} \Leftrightarrow \hat{\beta} < 1, \qquad k > 1.$

The numerical example of the above results is given in Fig. 1a-1f. To explain the relationship between R_S^2 and R_F^2, the value of CODs are compared for different sample sizes (n = 10, 50, 100, 1000, 5000 and 10000). For a given sample size, fifty samples were generated from a uniform distribution, each contains two random data sets (y_i, x_i). As an example, Fig. 1a shows fifty values of COD derived from SL and ULFR methods with sample size n = 10. Figure 1b-1f show the same plot with sample sizes n = 50, 100, 1000, 5000 and 10000, respectively. It is clearly shown that R_F^2 is always greater or equal to R_S^2. One interesting feature shown in Fig. 1 is that both R_F^2 and R_S^2 decreased when the sample size increased.

APPLICATION IN ROAD ACCIDENT PROBLEM

So far, various road accident models have been proposed to explain Malaysian road accident phenomena (Radin *et al*., 1995; Radin, 1996; Razali and Dahalan, 1992). However, PDRM road statistics show that the number of road accidents still remains high. One reason the Malaysian road accident problem has not been solved effectively is that the existing models are constructed for prediction purpose rather than causal inference. The crucial factor ignored in the existing models is the driver's state of mind that cannot be measured directly. The driver state of mind includes driver's attitude, ability to concentrate and tendency to take risks. It has been suggested that the driver's state of mind is an essential factor that causes road accident (Leeming, 1969; Forbes, 1972) .

In this example, we introduce the state of mind as a factor that causes variations in other physically measurable factors such as driver's sex, age and race. This is formally expressed using the ULFR model where the measurement error, **δ** is representing the state of mind. Difference between R_F^2 and R_S^2 is used to measure the effect of the state of mind to the road accident model. Since R_F^2 and R_S^2 have similar interpretations by construction, when R_F^2 is significantly larger than R_S^2 we say that the state of mind has been incorporated into the independent variable and the model now explains the causal relationship of road accident.

Table 1 shows the socio-biological variables of the Malaysian road accident data for 1996 obtained form Malaysian Royal Police (PDRM) headquarters. The Box-Cox power transformation (Box and Cox, 1964) is presented to access normality with $[\lambda_Y \;\; \lambda_{P_1} \;\; \cdots \;\; \lambda_{P_5}] = [0.25 \;\; 0.69 \;\; 1.25 \;\; 0.14 \;\; 0.23 \;\; 0.47]$. Without loss of generality, we assume that the ratio of variances, λ, equals one by dividing the number of state accidents by the total number of accidents in Malaysia.

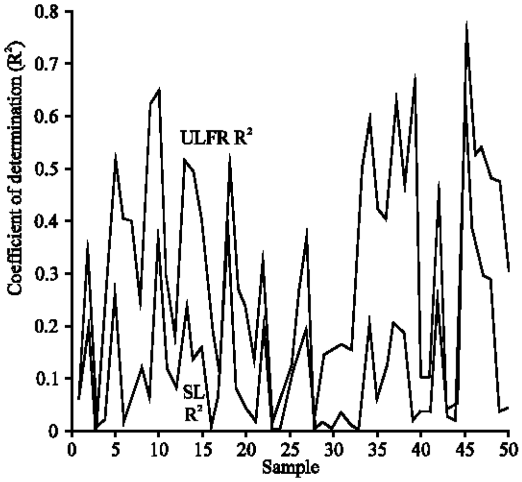

Fig. 1a: COD plots for SL and ULFR models with n = 10

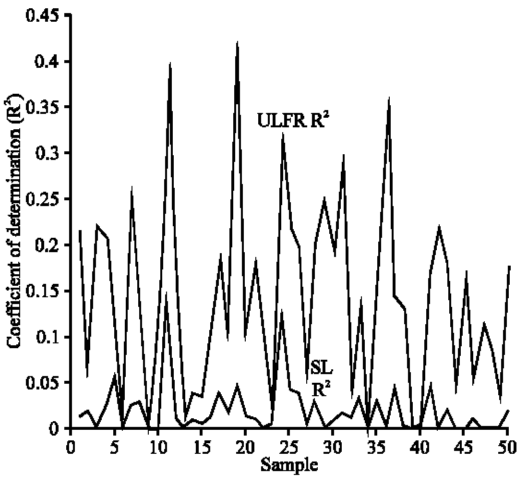

Fig. 1b: COD plot for SL and ULFR models with with n = 50

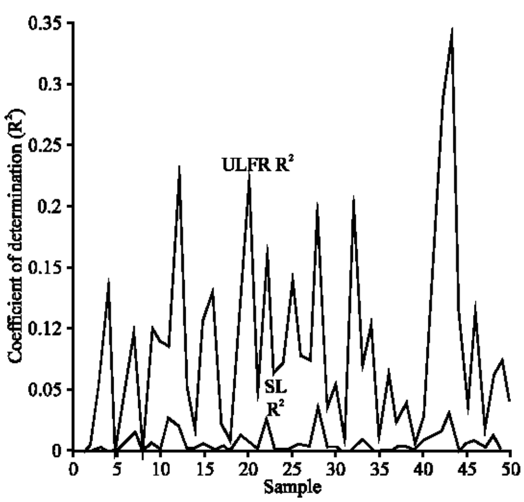

Fig. 1c: COD plots for SL and ULFR models with n = 100

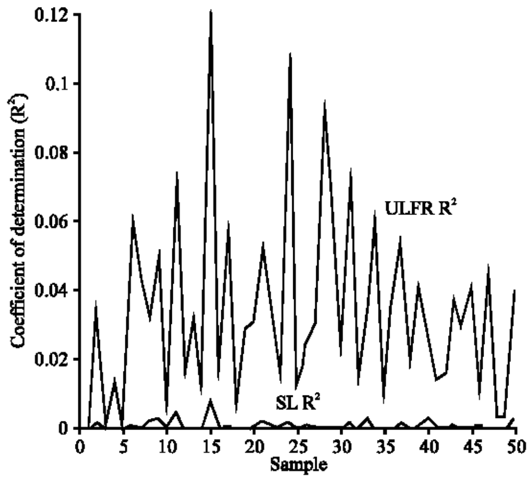

Fig. 1d: COD plot for SL and ULFR models with n = 1000

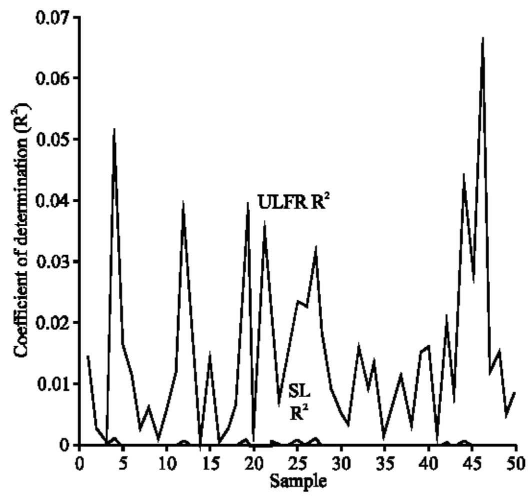

Fig. 1e: COD plots for SL and ULFR models with n = 5000

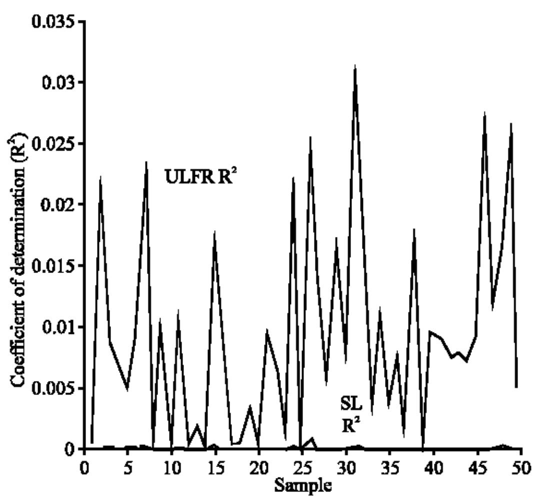

Fig. 1f: COD plot for SL and ULFR models with with n = 10000

Table 1: Socio-biological explanatory variables

State	Y	P_1	P_2	P_3	P_4	P_5
Perlis	0.208	0.4856	0.1491	0.02	0.4741	0.9051
Kedah	1.410	0.4686	0.2282	0.04	0.3965	0.9072
Perak	1.921	0.4670	0.3338	0.06	0.4440	0.9126
Penang	1.277	0.5600	0.3923	0.08	0.5542	0.9173
K.Lumpur	1.597	0.5401	0.3791	0.16	0.6119	0.9590
Selangor	2.208	0.5428	0.2800	0.09	0.4596	0.9401
N.Sembilan	1.121	0.5070	0.2638	0.04	0.5061	0.9385
Malacca	1.008	0.5342	0.3215	0.08	0.3386	0.9302
Johore	2.451	0.5173	0.3545	0.11	0.4363	0.9371
Pahang	0.971	0.5086	0.2196	0.02	0.3436	0.9496
Terengganu	0.716	0.4955	0.0554	0.01	0.4559	0.9278
Kelantan	0.934	0.4577	0.0880	0.07	0.3688	0.9203
Sabah	0.469	0.4479	0.2353	0.05	0.4128	0.9505
Sarawak	0.552	0.5122	0.4206	0.11	0.4452	0.9121

Source: PDRM (Malaysian Royal Police, 1995-1998)

Note:

- The number of vehicles involved in 1996 does not include accident of type 'slightly injured' and property damaged only
- Y = number of road accidents ('000)
  P_1 = Age (ratio of drivers aged 16-30 years to total number of drivers involved in accidents)
  P_2 = Race (ratio of Chinese drivers to total number of drivers involved in accidents)
  P_3 = Number of serious and fatal accidents due to influence of alcohol ('00)
  P_4 = Experience (ratio of vehicles with drivers of less than 5 years' experience to total number of vehicles involved in accidents)
  P_5 = Sex (ratio of male drivers to total number of drivers involved in accidents)
- The numbers given are ratios explained in (2). For example, 0.28 in column two row six means that for every 100 cases of road accident in Selangor during 1996 about 28 cases involved Chinese drivers.

First, we look at the estimation of parameters for the ULFR model containing the state of mind measurement error for drivers age (P_1) is given as

$$S_{yy} = 0.0199 \qquad S_{P_1P_1} = 0.0154 \qquad S_{P_1y} = 0.0062$$

$$\hat{\beta} = \frac{(0.0199 - 0.0154) + \left[\begin{matrix}(0.0199-0.0154)^2 \\ +4(0.0062)^2\end{matrix}\right]^{1/2}}{2(0.0062)} = 1.4264$$

$$\hat{\alpha} = 0.0714 - 1.4264(0.5032) = -0.6463$$

therefore, the ULFR model is

Table 2: Estimated parameters, coefficient of determination for the ULFR and SL models and the effect of measurement error

Variables	$\hat{\alpha}$	$\hat{\beta}$	R_F^2	$\hat{\alpha}_S$	$\hat{\beta}_S$	R_S^2	E_e
P_1	-0.6463	1.4264	0.4408	-0.1290	0.3983	0.1231	0.3177
P_2	0.0275	0.1652	0.2010	0.0318	0.1490	0.1814	0.0196
P_3	0.0164	0.7698	0.4215	0.0411	0.4245	0.2324	0.1891
P_4	0.0171	0.1218	0.0421	0.0308	0.0911	0.0315	0.0106
P_5	-7.8582	8.5348	0.8162	-0.3842	0.4904	0.0469	0.7693

$$\hat{Y} = -0.6463 + 1.4264\hat{P}_1$$

with the coefficient of determination, $R_F^2 = 0.4408$ $\left(> \frac{S_{xy}}{S_{yy}} = 0.3116 \text{ and } \hat{\beta} > 1 \right)$ (see Result 3). Comparing with the SL model

$$\hat{Y} = -0.1290 + 0.3983P_1 \qquad \text{and} \qquad R_S^2 = 0.1231$$

the result shows that the ULFR model explains the variability of Y better than the SL model $(R_F^2 > R^2)$. The SL model explains 12.31% of the variance of the number of road accidents but it increases to 44.08% when conditioned to 'state of mind' factor. This suggests that the state of mind explains 31.77% ($E_e = R_F^2 - R_S^2$) of the variance of road accident number.

It is clearly show (Table 2) that the ULFR model gives a better fit than the SL model with higher COD. The SL model indicates that variables P_1, P_3 and P_5 are not significantly contribute to the road accident rates. The first interpretation opposes other studies (Arthur and Little, 1970; IATSS, 1988; OECD Road Research Group, 1978; OECD Scientific Export Group, 1986) that age, gentle and alcoholic are important factors. However, when the state of mind is introduced, the three variables, especially P_5 become important and it may use to explain causal relationship of the road accident. For example, we say the younger driver may be the cause of higher accident rates because of their driving attitude or tendency to take risks.

The introduction of state of mind does not increase the COD of variables P_2 and P_4 significantly. This suggests that the driver's race and experience are two exact predictor of road accident. On the other hand, both R_S^2 and R_F^2 indicate race and driving experience are not significantly explained the road accident model. Henceforth, only variables P_1, P_3 and P_5 are recommended for further analysis such as multiple regression.

CONCLUSION AND FURTHER WORK

In this study, the COD of ULFR is proposed and it properties are investigated. The proposed COD and the

COD of ordinary linear regression model are used to investigate the effect of measurement error and hence try to explain the causal relationship of the model. It is a useful way to check whether a variable is reasonably assumed to be independent. This study on the Malaysian road accident problem demonstrates that the effect of the state of mind, indicated by the measurement error is clear but different in all factors.

However, confirmation of the causal properties in the ULFR model may be further strengthening by studying further the distribution of R_F^2. The multivariate version of ULFR may also be considered in further studies.

REFERENCES

Anderson, T.W., 1984. The 1992 Wald memorial lectures: Estimating linear statistical relationships. Ann. Stat., 12: 1-18.

Arthur, D. and I.N.C. Little, 1970. The State of the Art of Traffic Safety: A Comprehensive Review of Existing Information, New York: Praeger Publishers, pp: 38-42.

Box, G.E.P. and D.R. Cox, 1964. An analysis of transformation. J. Royal Stat. Soc., B, 26: 211-252.

Dolby, G.R., 1976. The ultrastructural relation: A synthesis of the functional and structural relations. Biometrika, 63: 30-50.

Forbes, T.W., 1972. Human Factors in Highway Traffic Safety Research, Canada: John Wiley and Sons.

Fuller, W.A., 1987. Measurement Error Models, New York: Wiley.

IATSS., 1988. Statistics ‘87/’88: Road Accidents Japan, Tokyo: Traffic Bureau, National Police Agency.

Kendall, M.G. and A. Stuart, 1979. The advanced theory of statistics, London: Griffin, Vol. 2.

Leeming, J.J., 1969. Road accident: Prevent or punish?, London: Cassell and Co. Ltd.

Montgomery, D.C. and E.A. Peck, 1992. Introduction to Linear Regression Analysis. 2nd Edn., New York: John Wiley and Sons.

OECD Road Research Group, 1978. New research on the role of alcohol and drugs in road accident: A Report, Paris: OECD.

OECD Scientific Expert Group, 1986. Guidelines for improving the safety of elderly road users, Paris: OECD.

Polis Diraja Malaysia (PDRM), 1995-1998. Laporan Tahunan Cawangan Traffik Bukit Aman, Kuala Lumpur: PDRM.

Radin, U.R.S., G.M. Mackay and B.L. Hills, 1995. Preliminary analysis of exclusive motorcycle lanes along the federal highway FO2, Shah Alam, Malaysia. J. IATSS Res., 19: 93-98.

Radin, U.R.S., 1996. Model kematian dan kecederaan di malaysia unjuran tahun 2000. Kertas Dasar Keselamatan Jalan Raya, Serdang: Kementerian Pengangkutan dan Universiti Putra Malaysia.

Razali, A.M. and B.H. Dahalan, 1992. Analysis Data Kamalangan Jalan Raya, Bangi: Jabatan Statistik Universiti Kebangsaan Malaysia.

Sprent, P., 1990. Some history of functional and structural relationships, Contemporary Math., 112: 3-15.