BULLETIN
of the
MALAYSIAN MATHEMATICAL SOCIETY

Bull. Malaysian Math. Soc. (Second Series) **21** (1998) 79-86

The Unreplicated Complex Linear Functional Relationship Model and Its Application

ABDUL GHAPOR HUSSIN
Institute of Mathematical Sciences, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
e-mail: j2ghapor@umcsd.um.edu.my

Abstract. A model is developed by fitting a straight line when both the variables are circular, subject to measurement and other errors. This is done by extending the complex linear regression model, called the unreplicated complex linear functional relationship model. This paper deals with the maximum likelihood estimation of this model assuming the ratio between the two error variances is known. We also show that the closed-form expression for the maximum likelihood estimators are not available and the estimates may be obtained iteratively by choosing a suitable initial value. The model is illustrated with an application to the analysis of the wind direction data recorded by two different techniques: the HF radar system and the anchored wave buoy.

1. Introduction

The fitting of a linear relationship with errors in the continuous linear variables or error-in-variables model (EIVM) had been explored since the latter part of the 19th century when Adcock [1, 2] investigated estimation properties under somewhat restrictive but realistic assumptions in ordinary linear regression models. Both variables are subject to errors. In essence, Adcock obtained the least squares solution for the slope parameters by assuming that both variables have equal error variances. Since then, several authors have worked on the problem of estimating the parameters: in particular when Kendall [6, 7] formally made a distinction between functional and structural relationship between the two variables.

Fitting a straight line when both variables are circular and subject to errors has not received any attention. What we mean by a circular variable is one which takes values on the circumference of a circle, i.e., they are angles in the range $(0, 2\pi)$ radians or $(0^\circ, 360^\circ)$. This variable must be analysed by techniques different from those appropriate for the usual Euclidean type variable because the circumference is a bounded closed space where the concept of origin is arbitrary or undefined. A continuous linear variable is a variable with realisations on the straight line which may be analysed straightforwardly with the usual techniques. Laycock [8] proposed a complex linear regression model for the circular variables. The maximum likelihood estimators was given by Hussin [5]. This sort of model was used in the analysis of wind direction data measured by two different methods, the anchored wave buoy and HF radar system, Hussin [5]. By looking at the nature of the experiment and our interest in calibrating one measurement technique against the other, the functional relationship model is more appropriate than the regression model. Since the variables are circular, the model is termed the *errors-in-circular-variables model* as an analogy to the error-in-variables model (EIVM) for the continuous linear variables.

For this, we extended the complex linear regression model as proposed by Laycock [8] to the case when both circular variables are subject to errors. Thus the *unreplicated complex linear functional relationship model* fits a straight line for the unreplicated data obtained from circular variables. Here, "linear" refers to the fact that the relationship itself is linear but the variables involved are circular.

In the following section we present the unreplicated complex linear functional relationship model. Section 3 deals with the maximum likelihood estimation and asymptotic properties of parameters for such model. The application of the model is given in Section 4. It analyses the wind direction data measured by two different methods: the anchored wave buoy and HF radar system. Some comparison is also given regarding the use of complex linear regression when both variables are subject to errors. All the calculations and estimations have been done by a specially written FORTRAN programme and can be obtained on request.

2. The complex functional model for circular variables

In this section we extend the complex linear regression model to the case when both variables *X* and *Y* are subject to errors. This model is also known as the unreplicated complex linear functional relationship model, where only unreplicated data were considered. We may assume *X* and *Y* are two circular variables which are linearly related with observed errors. As an analogy to the linear functional relationship model or EIVM, we will propose the complex linear functional relationship model, which is a functional relationship on vectors of direction cosines. We also assumed that *X* is a mathematical or fixed variable (without specific distribution). If *X* is a random variable which has a specified distribution then this is termed a structural relationship model between *X* and *Y*.

Let the direction cosines of n directions in two dimensions of, i.e., (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n), where $0 \le x_j,\ y_j < 2\pi$, for $j = 1, 2, \cdots, n$ be denoted by a series of a complex number (real and imaginary part) in the form of $(\cos x_1 + i \sin x_1, \cos y_1 + i \sin y_1), \cdots, (\cos x_n + i \sin x_n, \cos y_n + i \sin y_n)$. For any fixed X_j, we assume x_j and y_j have been measured with errors δ_j and ε_j respectively. The complex linear functional relationship model may be written as

$$\left(\cos x_j + i \sin x_j\right) = \left(\cos X_j + i \sin X_j\right) + \delta_j$$

and

$$\left(\cos y_j + i \sin y_j\right) = \left(\cos Y_j + i \sin Y_j\right) + \varepsilon_j,$$

where

$$\left(\cos Y + i \sin Y\right) = \alpha + \beta\left(\cos X + i \sin X\right).$$

We also assume δ_j and ε_j are independently distributed with the bivariate complex Gaussian distribution, (Goodman [4]), with zero mean and variances σ_1^2 and σ_2^2 respectively.

3. The maximum likelihood estimation of parameters

Parameters to be estimated are α, β, X_j and σ_1^2, and the estimation will be obtained by the maximum likelihood method. Suppose we assume $\frac{\sigma_2^2}{\sigma_1^2} = \lambda$, the ratio of two error variances is known, the same assumption being used for the unreplicated linear functional relationship model (EIVM) for the continuous linear variable. This constraint is important in order to overcome the problem of unbounded log likelihood function.

Then the log likelihood function is given by

$$\begin{aligned}
&\log L\left(\alpha, \beta, \sigma_1^2, X_1, \cdots, X_n;\ x_1, \cdots, x_n, y_1, \cdots, y_n\right) \\
&= -2n\log(\pi) - n\log(\lambda\sigma_1^2) - \frac{1}{\sigma_1^2}\sum_j \left|\delta_j\right|^2 - \frac{1}{\lambda\sigma_1^2}\sum_j \left|\varepsilon_j\right|^2 \\
&= -2n\log(\pi) - n\log(\lambda\sigma_1^2) - \frac{1}{\sigma_1^2}\sum\left\{2 - 2\cos x_j \cos X_j - 2\sin x_j \sin X_j\right\} \\
&\quad - \frac{1}{\lambda\sigma_1^2}\sum\left\{1 + \alpha^2 + \beta^2 + 2\alpha\left(\beta\cos X_j - \cos y_j\right)\right\} \\
&\quad + \frac{2\beta}{\lambda\sigma_1^2}\sum\left(\cos y_j \cos X_j + \sin y_j \sin X_j\right)
\end{aligned}$$

The function log L is differentiated with respect to parameters α and then equated to zero. This is followed by other parameters β, X_j and σ_1^2. Hence $\hat{\alpha}, \hat{\beta}, \hat{X}_j$ and $\hat{\sigma}_1^2$ are obtained and are given by

$$\hat{\alpha} = \frac{1}{n}\sum\left(\cos y_j - \hat{\beta}\cos\hat{X}_j\right),$$

$$\hat{\beta} = \frac{1}{n}\sum\left(\cos y_j \cos\hat{X}_j + \sin y_j \sin\hat{X}_j - \hat{\alpha}\cos\hat{X}_j\right),$$

$$\hat{X}_j = \tan^{-1}\left\{\frac{\lambda\sin x_j + \hat{\beta}\sin y_j}{\lambda\cos x_j + \hat{\beta}\cos y_j - \hat{\alpha}\hat{\beta}}\right\},\ \text{ for } j = 1, \cdots, n$$

and

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum\left\{2 - 2\cos x_j \cos \hat{X}_j - 2\sin x_j \sin \hat{X}_j\right\} + \frac{1}{\lambda n}\sum\left\{1 + \hat{\alpha}^2 + \hat{\beta}^2 + 2\hat{\alpha}\left(\hat{\beta}\cos \hat{X}_j - \cos y_j\right)\right\} - \frac{2\hat{\beta}}{\lambda n}\sum\left(\cos y_j \cos \hat{X}_j + \sin y_j \sin \hat{X}_j\right).$$

The above equations for $\hat{\alpha}, \hat{\beta}, \hat{X}_j$ and $\hat{\sigma}_1^2$ suggest that there is no closed-form available for $\hat{\alpha}, \hat{\beta}, \hat{X}_j$ and $\hat{\sigma}_1^2$, thus the estimates may be obtained iteratively and starting values for the iteration can be chosen from the complex regression model given by

$$\hat{\alpha}_0 = \frac{n\sum\cos y_j - \Psi\left(\sum\cos x_j\right)}{n^2 - \left(\sum\cos x_j\right)^2}$$

and

$$\hat{\beta}_0 = \frac{1}{n}\left\{\Psi - \hat{\alpha}\sum\cos x_j\right\},$$

where

$$\Psi = \sum\left(\cos y_j \cos x_j + \sin y_j \sin x_j\right).$$

By using these initial values we can then update $\hat{\alpha}, \hat{\beta}, \hat{X}_j$ and $\hat{\sigma}_1^2$, and this iteration procedure will continue until the convergence criterion is satisfied. It was found that stopping the iteration when both $\hat{\alpha}, \hat{\beta}, \hat{X}_j$ and $\hat{\sigma}_1^2$ change by no more than 0.00001 gives an acceptable accuracy.

Further, the asymptotic properties of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_1^2$ can be obtained from Fisher's information matrix. As usual the appropriate regularity conditions have been taken into consideration (see as an example, Cox & Hinkley [3]). It can be shown that the asymptotic properties for $\hat{\alpha}$ and $\hat{\beta}$ is given by

$$Var(\hat{\alpha}) = \frac{a_1 - b_4}{(a_1 - b_1)(a_1 - b_4) - (a_2 - b_2)^2},$$

$$Var(\hat{\beta}) = \frac{a_1 - b_1}{(a_1 - b_1)(a_1 - b_4) - (a_2 - b_2)^2}$$

and

$$Cov\left(\hat{\alpha}, \hat{\beta}\right) = \frac{b_2 - a_2}{\left(a_1 - b_1\right)\left(a_1 - b_4\right) - \left(a_2 - b_2\right)^2},$$

where

$$a_1 = \frac{2n}{\lambda\hat{\sigma}_1^2}, \qquad a_2 = \frac{2\sum \cos \hat{X}_j}{\lambda\hat{\sigma}_1^2},$$

$$b_1 = \frac{\sum W_{j1}^2 R_{jj}}{\sum R_{jj}}, \qquad b_2 = \frac{\sum W_{j2} R_{jj} W_{j1}}{\sum R_{jj}}, \qquad b_4 = \frac{\sum W_{j2}^2 R_{jj}}{\sum R_{jj}},$$

and also

$$W_{j1} = \frac{2\hat{\beta}\sin \hat{X}_j}{\lambda\hat{\sigma}_1^2},$$

$$W_{j2} = \frac{2}{\lambda\hat{\sigma}_1^2}\left(\cos y_j \sin \hat{X}_j - \sin y_j \cos \hat{X}_j - \hat{\alpha}\sin \hat{X}_j\right),$$

$$R_{jj} = \frac{2}{\hat{\sigma}_1^2}\left(\cos x_j \cos \hat{X}_j + \sin x_j \sin \hat{X}_j\right) + \frac{2\hat{\beta}}{\lambda\hat{\sigma}_1^2}\left(\sin y_j \sin \hat{X}_j + \cos y_j \sin \hat{X}_j - \hat{\alpha}\cos \hat{X}_j\right).$$

The asymptotic properties for $\hat{\sigma}_1^2$ is given by

$$Var\left(\hat{\sigma}_1^2\right) = S^{-1},$$

where

$$S = \frac{2}{\lambda\hat{\sigma}_1^6}\sum\left\{1 + \hat{\alpha}^2 + \hat{\beta}^2 + 2\hat{\alpha}\left(\hat{\beta}\cos \hat{X}_j - \cos y_j\right) - 2\hat{\beta}\left(\cos y_j \cos \hat{X}_j + \sin y_j \sin \hat{X}_j\right)\right\} + \frac{4}{\hat{\sigma}_1^6}\sum\left\{1 - \cos x_j \cos \hat{X}_j - \sin x_j \sin \hat{X}_j\right\} - \frac{n}{\hat{\sigma}_1^4}.$$

Thus for large values of *n*, the above expressions can be used to estimate the standard error of $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}_1^2$. These estimates are distributed normally, and further the estimate of error variances is uncorelated with the intercept and slope estimate.

4. Application of the model

In this section the wind direction data can be modelled by a complex linear functional relationship model by assuming the n observations as n direction cosines in two dimension. The full description of the wind direction data can be found in Hussin [5]. Figure 1 gives the scatter plot of such data. Note that the scale is artificially broken at 0 (or equivalently 2π) radians so the apparent outliers in the top left corner are in fact measurements which are close to those which appear in the bottom left corner or top right corner. This is due to wrap-around observations or measurements from 2π back to 0.

From the scatter plot of wind direction data (Figure 1), it is reasonable to assume a linear relationship between the direction measured by an anchored wave buoy and by HF radar. Suppose we assume that for observation j, x_j measures the underlying direction measured by HF radar, X_j, with some random error, δ_j. Similarly, y_j is assumed to be the real direction measured by an anchored wave buoy, Y_j, recorded with some random error, ε_j. Thus we have

$$\cos x_j + i \sin x_j = \cos X_j + i \sin X_j + \delta_j$$

and

$$\cos Y_j + i \sin Y_j = \alpha + \beta\left(\cos X_j + i \sin X_j\right) + \varepsilon_j .$$

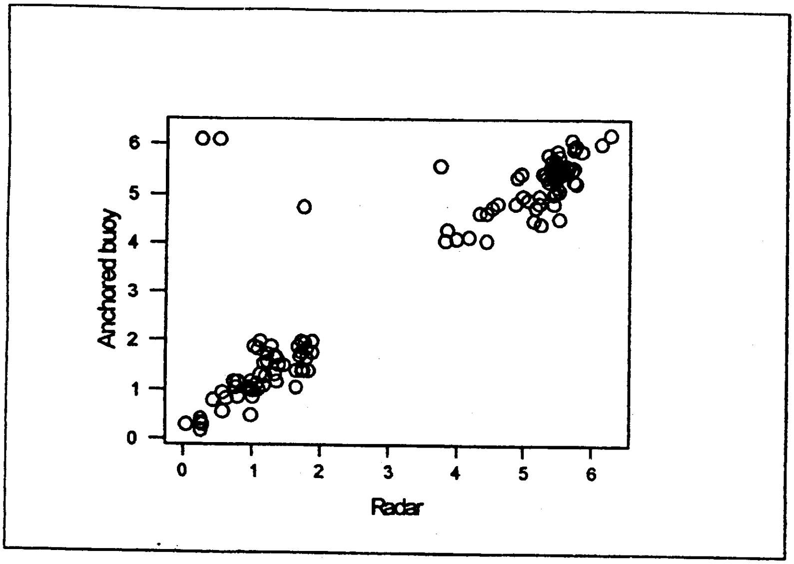

Figure 1. A scatter plot of wind direction data measured by HF radar system and anchored wave buoy.

This model allows relative calibration of observation measured by anchored wave buoy and by HF radar and vice-versa since the model is entirely symmetrical in the two variables. This distinguishes it from the complex linear regression model. The prime interest is how the measurement of the direction made by anchored wave buoy would be related to the measurement made by HF radar.

The parameters $\alpha, \beta,$ and σ_1^2 involved can be interpreted as

α : the additive correction factor to be applied when converting response measured by HF radar to the response measured by anchored wave buoy,

β : the multiplicative factor involved and

σ_1^2 : the error variance (representing the precision of the measurement) for the HF radar data.

The main interest is in detecting any departure of α from 0 and β from 1, since $\alpha = 0$ and $\beta = 1$ implies no difference in calibration between the two techniques. In addition we can estimate $X_1, \cdots, X_n,$ the underlying directions measured by the HF radar.

By assuming that the ratio of error variances is known and equal to 1, estimates for the wind direction data is given in Table 1, together with their estimated standard error.

Hence the relationship between variable X (HF radar) and Y (an anchored wave buoy) for the wind direction data is given by

$$\left(\cos Y + i \sin Y\right) = -5.330 \times 10^{-3} + 0.997\left(\cos X + i \sin X\right).$$

Table 1. Estimates for wind direction data for the unreplicated complex linear functional relationship model.

Parameter	**Estimate**	**St. error**
α	-5.330×10^{-3}	1.998×10^{-2}
β	0.977	1.886×10^{-2}
σ_1^2	9.130×10^{-2}	8.038×10^{-3}

Table 2. Parameter estimates for wind direction data for the complex linear regression model.

Parameter	**Estimate**	**St. error**
α	-2.592×10^{-3}	2.619×10^{-2}
β	0.927	2.619×10^{-2}
σ^2	0.143	1.259×10^{-2}

However, the 95% confidence interval for $\hat{\alpha}$ is $(-4.449 \times 10^{-2},\ 3.383 \times 10^{-2})$ and for $\hat{\beta}$ is (0.940, 1.014), which suggests no reason to doubt that $\alpha = 0.0$ and $\beta = 1.0$.

Table 2 gives the estimates of the wind direction data analysed by using the complex linear regression model (i.e., assuming only *Y* variable is subject to error). Comparison of the estimates of parameters given in Tables 1 and 2 shows that ignoring the error in the *X* variable will give a biased estimate of parameters.

5. Conclusion

In this paper we present the unreplicated complex linear functional relationship model, which is a statistical method of fitting a straight-line relationship when both circular variables are subject to errors. It is also better to compare or calibrate the two circular variables which are subject to measurements error, than to predict one variable from the other. The model is analogous to the unreplicated linear functional relationship model for continuous linear variable and is also the extension of the complex linear regression model. Hence we assume that the circular random errors are distributed according to a complex Gaussian distribution.

Estimation of parameters have been obtained by maximum likelihood estimation assuming a known ratio of error variances. The same assumption is being imposed for the unreplicated linear functional relationship model for continuous linear variable (EIVM). Estimates have been obtained numerically by an iterative method, since the closed-form expression for each estimate is not available and the starting values for iteration have been chosen from the complex linear regression model.

It is found that, by assuming a known ratio of error variances, the unreplicated complex linear functional relationship model reasonably represents the relationship between the two circular variables. This is observed with error by considering their vector of direction cosines as shown in the analyses of the wind direction data.

References

1. R.J. Adcock, Note on the method of least squares, *Analyst* **4** (1877), 183-184.
2. R.J. Adcock, A problem in least squares, *Analyst* **5** (1878), 53-54.
3. D.R. Cox, and D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London, 1974.
4. N.R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction), *Ann. Math. Stat.* **34** (1963), 152-165.
5. A.G. Hussin, The maximum likelihood estimation for complex linear regression model, *J. Sains* **5:2** (1997)**,** 705-711.
6. M.G. Kendall, Regression, structural and functional relationship I, *Biometrika* **38** (1951), 11-25.
7. M.G. Kendall, Regression, structural and functional relationship II, *Biometrika* **39** (1952), 96-108.
8. P.J. Laycock, Optimal design: regression model for directions, *Biometrika* **62** (1975), 305-311.