ANOMALIES OF THE MAGNITUDE OF THE BIAS OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF THE REGRESSION SLOPE

Diarmuid O'Driscoll [1], Donald E. Ramirez [2]

[1] Head of Department of Mathematics and Computer Studies
Mary Immaculate College
Ireland

[2] Department of Mathematics
University of Virginia
USA

Corresponding Author: diarmuid.odriscoll@mic.ul.ie

***Abstract*:**

The slope of the best-fit line $y = h(x) = \beta_0 + \beta_1 x$ from minimizing a function of the squared vertical and horizontal errors is the root of a polynomial of degree four which has exactly two real roots, one positive and one negative, with the global minimum being the root corresponding to the sign of the correlation coefficient. We solve second order and fourth order moment equations to estimate the variances of the errors in the measurement error model. Using these solutions as an estimate of the error ratio κ in the maximum likelihood estimator, we introduce a new estimator β_1^{kap}. We create a function ψ which relates κ to the oblique parameter λ, used in the parameterization of the line from $(x, h(x))$ to $(h^{-1}(y), y)$, to introduce an oblique estimator β_1^{lam}. A Monte Carlo simulation study shows improvement in bias and mean squared error of each of these two new estimators over the ordinary least squares estimator. In O'Driscoll and Ramirez (2011), it was noted that the bias of the MLE estimator of the slope is monotone decreasing as the estimated variances error ratio $\tilde{\kappa}$ approaches the true variances error ratio $\kappa = \sigma_\tau^2/\sigma_\delta^2$. However for a fixed estimated variances error ratio $\tilde{\kappa}$, it was noted that the bias is not monotone decreasing as the true error ratio κ approaches $\tilde{\kappa}$. This paper explains this anomaly by showing that as κ approaches a fixed $\tilde{\kappa}$, the bias of the MLE estimator of the slope is also dependent on the magnitude of σ_δ^2. Other anomalies with the MLE estimator of the slope in the presence of errors in both *x* and *y* are discussed.

***Keywords*:** Maximum likelihood estimation, Measurement errors, Moment estimating equations, Oblique estimators

1 Introduction

With ordinary least squares $OLS(y|x)$ regression we have data $\{(x_1, Y_1|X = x_1),...,(x_n, Y_n|X_n = x_n)\}$ and we minimize the sum of the squared vertical errors to find the best-fit line $y = h(x) = \beta_0 + \beta_1 x$ where it is assumed that the independent or causal variable X is measured without error. The measurement error model does not assume that X is measured without error, has wide interest with many applications and has been studied in depth by many, for example, Carroll et al. (2006) and Fuller (1987). As in the regression procedure of Deming (1943) to account for both sets of errors σ_X^2 and σ_Y^2, we determine a fit so that a function of both the squared vertical and the squared horizontal errors will be minimized. In Section 2, we outline the Oblique Error Method and the measurement error model and introduce second order and fourth order equations to estimate $\kappa = \sigma_\tau^2/\sigma_\delta^2$ in the maximum likelihood estimator. We also introduce two new estimators β_1^{kap} and β_1^{lam} and describe our Monte Carlo simulations. We report on our findings in Section 3 and conclude that that our estimators β_1^{kap} and β_1^{lam} greatly reduce the Bias and MSE associated with the ordinary least squares estimator β_1^{ver} .

2 Methodology

2.1 Minimizing Squared Oblique Errors

From the data point (x_i, y_i) to the fitted line $y = h(x) = \beta_0 + \beta_1 x$, define the vertical length $v_i = |y_i - \beta_0 - \beta_1 x_i|$ from which it follows that the sum of the squares of the oblique lengths from (x_i, y_i) to $(h^{-1}(y_i) + \lambda(x_i - h^{-1}(y_i)), y_i + \lambda(h(x_i) - y_i))$ is

$$SSE_o(\beta_0, \beta_1, \lambda) = (1-\lambda)^2 \sum v_i^2 / \beta_1^2 + \lambda^2 \sum v_i^2 . \qquad (1)$$

In a comprehensive paper by Riggs et al. (1978), the authors state that: "It is a poor method indeed whose results depend upon the particular units chosen for measuring the variables." As in O'Driscoll et al. (2011), so that our equation is dimensionally correct we consider a standardized weighted model

$$SSE_o(\beta_0, \beta_1, \lambda) = (1-\lambda)^2 s_{yy} \sum v_i^2 / \beta_1^2 + \lambda^2 s_{xx} \sum v_i^2$$

where

$s_{xx} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$, $s_{yy} = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n}$ and $s_{xy} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$.

The solution of $\frac{d(SSE_o)}{d\beta_0} = 0$ is given by $\beta_0 = \bar{y} - \beta_1 \bar{x}$ and the solutions of $\frac{d(SSE_o)}{d\beta_1} = 0$ are the roots of the fourth degree polynomial, $P_4(\beta_1)$,

$$\lambda^2 (s_{xx}/s_{yy})^{1.5} \beta_1^4 - \lambda^2 \rho\, s_{xx}/s_{yy}\, \beta_1^3 + (1-\lambda)^2 \rho \beta_1 - (1-\lambda)^2 (s_{yy}/s_{xx})^{0.5}. \qquad (2)$$

From O'Driscoll et al. (2008), the Complete Discrimination System

$\{D_1,...,D_n\}$ of Yang (1999) is a set of explicit expressions that determine the number (and multiplicity) of roots of a polynomial. This system is used to show that the fourth order polynomial $P_4(\beta_1)$ has exactly two real roots, one positive and one negative with the global minimum being the positive (respectively negative) root corresponding to the sign of s_{xy}.

With λ = 1 we recover the minimum squared vertical errors with estimated slope β_1^{ver} and with λ = 0 we recover the minimum squared horizontal errors with estimated slope β_1^{hor}. The geometric mean estimator $\beta_1^{gm} = \sqrt{s_{yy} / s_{xx}}$ has the fixed oblique parameter λ = 0.5 and for the measurement error model, when both the vertical and horizontal models are reasonable, a compromise estimator such as β_1^{gm} is widely used and is hoped to have improved efficiency. However, Lindley and El-Sayyad (1968) proved that the expected value of β_1^{gm} is biased unless $\kappa = \sigma_Y^2/\sigma_X^2$ where, as before, $\kappa = \sigma_\tau^2/\sigma_\delta^2$ is the ratio of the errors in *y* and *x* respectively.

2.2 Measurement Error Model; Second and Fourth Moment Estimation

We now consider the measurement error model as follows. In this paper it is assumed that *X* and *Y* are random variables with respective finite variances σ_X^2 and σ_Y^2, finite fourth moments and have the linear functional relationship $Y = \beta_0 + \beta_1 X$. The observed data $\{(x_i, y_i), 1 \le i \le n\}$ are subject to error by $x_i = X_i + \delta_i$ and $y_i = Y_i + \tau_i$ where it is also assumed that δ is $N(0,\sigma_\delta^2)$ and τ is $N(0,\sigma_\tau^2)$. It is well known, in a measurement error model, that the expected value for β_1^{ver} $(OLS(y\,|\,x))$ is attenuated to zero by the attenuating factor $\sigma_X^2/(\sigma_X^2 + \sigma_\delta^2)$ called the reliability ratio by Fuller (1987); and similarly the expected value for β_1^{hor} $(OLS(x\,|\,y))$ is amplified to infinity by the amplifying factor $(\sigma_Y^2 + \sigma_\tau^2)/\sigma_Y^2$. From Gillard and Iles (2009), second moment equations are

$$s_{xx} = \tilde{\sigma}_X^2 + \tilde{\sigma}_\delta^2;\ s_{yy} = \hat{\beta}_1^2\tilde{\sigma}_X^2 + \tilde{\sigma}_\tau^2\ ;\ s_{xy} = \hat{\beta}_1\tilde{\sigma}_X^2 \qquad (3)$$

and fourth moment equations are

$$s_{xxxy} = \hat{\beta}_1\tilde{\mu}_{X4} + 3\hat{\beta}_1\tilde{\sigma}_X^2\tilde{\sigma}_\delta^2;\ \ s_{xyyy} = \hat{\beta}_1^3\tilde{\mu}_{X4} + 3\hat{\beta}_1\tilde{\sigma}_X^2\tilde{\sigma}_\tau^2. \qquad (4)$$

These equations yield the estimators

$$\tilde{\sigma}_\delta^2 = s_{xx} - \frac{s_{xy}}{\hat{\beta}_1},\ \ \tilde{\sigma}_\tau^2 = s_{yy} - \hat{\beta}_1 s_{xy}, \qquad (5)$$

the Frisch hyperbola of Van Montfort (1987)

$$(s_{xx} - \tilde{\sigma}_\delta^2)(s_{yy} - \tilde{\sigma}_\tau^2) = s_{xy}^2 \qquad (6)$$

and the fourth order equation

$$(s_{xxxy} - 3s_{xy}\tilde{\sigma}_\delta^2)(s_{xy}^2) = (s_{xx} - \tilde{\sigma}_\delta^2)^2(s_{xyyy} - 3s_{xy}\tilde{\sigma}_\tau^2)\,. \qquad (7)$$

We use equations (6) and (7) to solve for $\tilde{\sigma}_{\delta}^{2}$ and $\tilde{\sigma}_{\tau}^{2}$ imposing suitable restrictions on the possible solutions; firstly the variances must be positive; secondly the kurtosis of the underlying distribution must be significantly different from the kurtosis of the normal distribution to assure the validity of Equation (4) and thirdly the sample sizes must be adequately large. We then use these solutions as estimates for the ratio κ in the maximum likelihood estimator as described in Section 2.3. An alternative procedure for computing κ will be shown in Section 5.

2.3 The Maximum Likelihood Estimator

If the ratio of the error variances $\kappa = \sigma_{\tau}^{2} / \sigma_{\delta}^{2}$ is assumed finite, then Madansky (1959), among others, showed that the maximum likelihood estimator for the slope is

$$\beta_1^{mle} = \hat{\beta}_1(\kappa) = \frac{(s_{yy} - \kappa s_{xx}) + \sqrt{(s_{yy} - \kappa s_{xx})^2 + 4\kappa\rho^2 s_{xx} s_{yy}}}{2\rho\sqrt{s_{xx} s_{yy}}}. \qquad (8)$$

For finite κ it also follows that the moment estimator agrees with the MLE. If $\kappa = 1$ in Equation (8) then the MLE (often called the Deming Regression estimator) is equivalent to the perpendicular estimator, β_1^{per}, first introduced by Adcock (1878). In the particular case where $\kappa = s_{yy}/s_{xx}$ then β_1^{mle} has a fixed λ value of 0.5.

If the researcher knows the true error ratio $\kappa = \sigma_{\tau}^{2}/\sigma_{\delta}^{2}$ then

$$E(\hat{\beta}_1(\kappa)) = 0.5\left((\beta_1 - \kappa/\beta_1) + \sqrt{(\beta_1 - \kappa/\beta_1)^2 + 4\kappa}\right) = \beta_1 \qquad (9)$$

and there are no bias problems. We will discuss the more realistic situation when κ is an unknown parameter and must be estimated by $\tilde{\kappa}$.

3. The Empirical Bias of the *MLE* for an incorrect choice of κ.

3.1 Empirical Bias

In practice, the researcher estimates κ by $\tilde{\kappa}$ with error $\epsilon = \kappa - \tilde{\kappa} \neq 0$. To develop an expression for the bias $E\left(\hat{\beta}_1(\tilde{\kappa}) - \beta_1\right)$, we recall equation (9) and write

$$E\left(\hat{\beta}_1(\tilde{\kappa}) - \beta_1\right) = E\left(\hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa)\right) + E\left(\hat{\beta}_1(\kappa) - \beta_1\right) = E\left(\hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa)\right).$$

We define the empirical bias, in using $\tilde{\kappa}$ to estimate κ, as $empbias(\tilde{\kappa}{:}\kappa) = \hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa)$, which in terms of $\{s_{xx}, s_{yy}, s_{xy}\}$ is

$$\frac{1}{2s_{xy}}\left(-s_{xx}(\tilde{\kappa} - \kappa) + \sqrt{(s_{yy} - \tilde{\kappa}s_{xx})^2 + 4\tilde{\kappa}s_{xy}^2} - \sqrt{\left(s_{yy} - \kappa s_{xx}\right)^2 + 4\kappa s_{xy}^2}\right) \qquad (10)$$

The empirical bias is an estimate of the error that occurs in $\widehat{\beta}_1(\kappa)$ as a result of using $\tilde{\kappa}$ for the unknown error ratio κ. In our simulation study we record $E\left(\hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa)\right)$ in Table 1.

Using the fact that the second sample moments converge in probability to their expectations, it follows from (3) that

$$\frac{s_{yy}-\kappa s_{xx}}{s_{xy}} = \frac{\widehat{\beta}_1^2\sigma_X^2+\sigma_\tau^2-\kappa(\sigma_X^2+\sigma_\delta^2)}{\widehat{\beta}_1\sigma_X^2} = \hat{\beta}_1 - \kappa/\hat{\beta}_1 \to \beta_1 - \kappa/\beta_1$$

and

$$\frac{s_{xx}}{s_{xy}} = \frac{\sigma_X^2+\sigma_\delta^2}{\widehat{\beta}_1\sigma_X^2} \to (1+\sigma_\delta^2/\sigma_X^2)/\beta_1$$

where $\sigma_\delta^2/\sigma_X^2$ is the noise to signal ratio of the model.

With $\epsilon = \kappa - \tilde{\kappa} \neq 0$ and $\theta = 1 + \sigma_\delta^2/\sigma_X^2$, the bias, $bias(\tilde{\kappa}{:}\kappa)$, in terms of $\{\beta_1, \varepsilon, \theta\}$ is then

$$-0.5\left(\beta_1 + \frac{\kappa}{\beta_1} + \frac{\theta\varepsilon}{\beta_1}\right) + 0.5\left(\beta_1 + \frac{\kappa}{\beta_1} - \frac{\theta\varepsilon}{\beta_1}\right)\sqrt{1 + \frac{4\varepsilon\left(1+\frac{\kappa\theta}{\beta_1^2}\right)}{\left(\beta_1+\frac{\kappa}{\beta_1}-\frac{\theta\varepsilon}{\beta_1}\right)^2}} \quad (11)$$

We note that $bias(\tilde{\kappa}{:}\kappa) = 0$ only when $\epsilon = \kappa - \tilde{\kappa} = 0$.

3.2 Series Expansion for the Bias

The series expansion, $bias(\tilde{\kappa}{:}\kappa)$, of the bias may be written in terms of ε as

$$-\frac{\varepsilon\beta_1(\theta-1)}{\beta_1^2+\kappa}\left(1 + \frac{\epsilon(\beta_1^2+\kappa\theta)}{(\beta_1^2+\kappa)^2} + \frac{\epsilon^2(-\kappa\theta+\theta\beta_1^2-2\beta_1^2)(\beta_1^2+\kappa\theta)}{(\beta_1^2+\kappa)^4}\right) + O(\epsilon^4) \quad (12)$$

Since

$$\frac{\varepsilon\beta_1(\theta-1)}{\beta_1^2+\kappa} = \frac{(\tilde{\kappa}-\kappa)\sigma_\delta^2}{\left(\beta_1+\frac{\kappa}{\beta_1}\right)\sigma_X^2},$$

Equation (12) shows that $bias(\tilde{\kappa}{:}\kappa)$ is not alone dependent on the magnitude of $\tilde{\kappa} - \kappa$ but is also dependent on the magnitude of σ_δ^2; that is, the magnitude of the bias is dependent on the magnitude of the difference $\tilde{\kappa}\sigma_\delta^2 - \sigma_\tau^2$ and our claim in the Abstract is justified.

3.2 Monte Carlo Simulation

We set $\beta_1 = 1$ and $\kappa = 1$. The X data was generated from a uniform distribution on $\left(0, \sqrt{12*100}\right)$ to set $\sigma_X^2 = 100$. The linear regression model had slope $\beta_1 = 1$, $\beta_0 = 0$ and sample size n = 50. For the measurement error model, we used normal errors with mean equal to zero and variances $\{\sigma_\tau^2, \sigma_\delta^2\}$ varying over {1,2,3,4,5,9}. Typical values for the bias, $bias(\tilde{\kappa}{:}\kappa)$, and the Third Order series approximation, $serbias(\tilde{\kappa}{:}\kappa)$, are shown in Table 1. We used Minitab for our simulation study setting the number of runs N=5000. The results for the bias $E\left(\hat{\beta}_1(\tilde{\kappa}) - \beta_1\right)$ and the estimated empirical bias $E\left(\hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa)\right)$ are recorded in Table 1.

(table 1 here)

The rows of Table 1 are sorted in ascending order of the theoretical bias, $bias(\tilde{\kappa}{:}\kappa)$ displayed in Column 7. Column 8 shows that our approximation, $serbias(\tilde{\kappa}{:}\kappa)$, is a good estimate for $bias(\tilde{\kappa}{:}\kappa)$. Columns 5 and 6 also show that our simulation study produced very good results for $E\left(\hat{\beta}_1(\tilde{\kappa}) - \beta_1\right)$ and $E(\hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa))$.

We make the following observations. Firstly, with $\tilde{\kappa} = 1$, the ranking for the bias concurs with the ranking of the differences in the error variances $\sigma_\tau^2 - \sigma_\delta^2$ but does not concur with the ranking for $\kappa = \sigma_\tau^2/\sigma_\delta^2$ in terms of its closeness to $\tilde{\kappa}$. The value for κ=0.555 in Row 2 is closer to the assumed value $\tilde{\kappa} = 1$ than the value for κ=0.500 in Row 3 is. However the absolute value for the bias 0.0198 in Row 2 is approximately double the absolute value for the bias 0.0100 in Row 3; that is, the magnitude of the bias for the *MLE* estimator $\hat{\beta}_1(\kappa)$ is not monotone in κ.

Secondly, for equal κ =3/1 in Row 7 and κ=9/3 in Row 10, the respective biases 0.0101 and 0.0304 are approximately proportional to the respective differences of the error variances 2 and 6.

4. The efficiency of different slope estimators

Using the solutions $\tilde{\sigma}_\tau^2$ and $\tilde{\sigma}_\delta^2$ from equations (6) and (7) as estimates for κ in β_1^{mle}, we introduce a new estimator β_1^{kap} which performs very well in our Monte Carlo simulation.

4.1 Relation between kappa and lambda

With κ estimated as in Section 2.2, the invertible function $\psi : [0,\infty] \to [0,1]$ defined by $\lambda = \psi(\kappa) = c\kappa/(c\kappa+1), c = s_{xx}/s_{yy}$, creates a new estimator β_1^{lam}. This proposed oblique estimator also performs very well in our Monte Carlo simulation. Since the range of κ includes infinity, we do not compute its average value in our simulation. Instead, we compute the average λ value for β_1^{lam}, and use $\psi^{-1}(\bar{\lambda})$ as the effective average $\tilde{\kappa}$ for κ. To determine the efficiency of the six estimators $\{\beta_1^{ver}, \beta_1^{gm}, \beta_1^{hor}, \beta_1^{per}, \beta_1^{kap}, \beta_1^{lam}\}$, we conducted a set of Monte Carlo simulations for varying values of the true slope β_1.

We report in Tables 2-5 the MSE, the Bias, the associated parameter λ and the associated oblique angle θ_λ for each of the six estimators above.

(tables 2 and 3 here)

In the cases represented by Tables 2 and 3 we can see that β_1^{kap} and β_1^{lam} make significant improvement in (MSE, Bias) over the estimator β_1^{ver} and each of the 'compromise' estimators β_1^{gm} and β_1^{per}. Of course β_1^{hor} performs well in each of these cases but its use would have been based on prior knowledge that $\sigma_\delta^2 >> \sigma_\tau^2$.

(tables 4 and 5 here)

In the cases represented by Tables 4 and 5 we again see that β_1^{kap} and β_1^{lam} make significant improvement in (MSE, Bias) over the estimators β_1^{ver} and β_1^{hor}. With $\kappa = 1$, β_1^{per} performed very well in each case as expected since $\sigma_\tau^2 = \sigma_\delta^2$. The condition of Lindley and El-Sayyad (1968) of $\kappa = \sigma_Y^2/\sigma_X^2$ is satisfied in the case represented by Table 4 but not by Table 5 and hence β_1^{gm} performed very well in Table 4 but not as well in Table 5. Riggs *et al.* (1978) state that "no one method of estimating the true slope is the best method under all circumstances". Tables 2-5 show that β_1^{kap} and β_1^{lam} perform well in all of the above four cases where no prior knowledge of the errors is assumed. Table 6 reports the effective average for $\tilde{\kappa}$, as described in Section 4.1, for $(\sigma_\delta^2, \sigma_\tau^2) \in \{1,4,9\} \times \{1,4,9\}$.

(table 6 here)

5. Anomalies with *MLE* Estimator for the Slope

The likelihood function for the measurement error model is proportional to

$$\left(\frac{1}{\sigma_\delta^2 \sigma_\tau^2}\right) exp\left(-\sum_{i=1}^{n} \frac{(x_i - X_i)^2}{2\sigma_\delta^2} - \sum_{i=1}^{n} \frac{(y_i - Y_i)^2}{2\sigma_\tau^2}\right)$$

with maximum likelihood equations for $\{X_i \mid_{i=1}^n, \beta_0, \beta_1\}$, with $\kappa = \sigma_\tau^2/\sigma_\delta^2$ as

$$\kappa(x_i - X_i) + (y_i - \beta_0 - \beta_1 X_i) = 0,$$

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_i) = 0\,, \quad \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_i)X_i = 0. \qquad (13)$$

With κ assumed to be known, the *n+2* Equations (13) yield *MLE* estimators for $\{X_i \mid_{i=1}^n, \beta_0, \beta_1\}$. The earliest reference we know of such equations is Dent ((1935), Equations 27–31) where she uses the additional equations

$$\sigma_\delta^2 = \sum_{i=1}^{n} \frac{(x_i - X_i)^2}{n} \text{ and } \sigma_\tau^2 = \sum_{i=1}^{n} \frac{(y_i - Y_i)^2}{n}$$

to give the *MLE* estimator for κ, and the functional relationship (Dent (1935), Equation 38)),

$$\hat{\beta}_1^2 = \frac{\sigma_\tau^2}{\sigma_\delta^2} = \hat{\kappa}.$$

Lindley ((1947), Equations 58–62) re-examined this derivation with respect to consistency. Lindley ((1947), Equation 73) showed that σ_δ^2 converges in probability to $\sigma_\delta^2/2$ and thus is not a consistent estimator. The information matrix for the *MLE* estimators has been given in Barnett (1970) where replicated observations are allowed. Lindley and El-Sayyad (1968) gave a simple explanation of the problem inherent in these *MLE* solutions, where they note that if there is a functional relationship among the estimators $\{\hat{\beta}_1, \tilde{\sigma}_\tau^2, \tilde{\sigma}_\delta^2\}$ and if all estimators are consistent, the identical functional relationship must also hold among the parameters $\{\beta_1, \sigma_\tau^2, \sigma_\delta^2\}$, which need not be the case.

Moment estimators for σ_δ^2 and σ_τ^2 using the geometric mean estimator $\beta_1^{gm} = s_{yy}/s_{xx}$ have a functional relationship among the estimators $\{\beta_1^{gm}, \tilde{\sigma}_\tau^2(\beta_1^{gm}), \tilde{\sigma}_\delta^2(\beta_1^{gm})\}$ as noted in O'Driscoll and Ramirez ((2011), Proposition 2)

$$(\beta_1^{gm})^2 = \tilde{\sigma}_\tau^2(\beta_1^{gm})\tilde{\sigma}_\delta^2(\beta_1^{gm}) = \tilde{\kappa}(\beta_1^{gm}). \quad (13)$$

The functional relationship in (13) shows that the estimators $\{\beta_1^{gm}, \tilde{\sigma}_\tau^2(\beta_1^{gm}), \tilde{\sigma}_\delta^2(\beta_1^{gm})\}$ cannot all be consistent estimators.

The *MLE* estimator $\hat{\beta}(\kappa)$ from Equation (8) requires users to estimate the unknown ratio $\kappa = \sigma_\tau^2/\sigma_\delta^2$. Table 2 in O'Driscoll and Ramirez (2011)) shows the error ratio $\kappa = \sigma_\tau^2/\sigma_\delta^2$ for varying estimators of the slope using standard methods. We can add to this table the fourth moment estimator β_1^{mom} which will be described shortly. This estimator will satisfy the functional relationships given in Equation (5).

With κ assumed to be an unknown parameter, Solari (1969) showed that the maximum likelihood estimator for the slope β_1 does not exist, as the maximum likelihood surface has a saddle point at the critical value. Sprent (1970) pointed out that the result of Solari does not imply the maximum likelihood principle has failed, but rather that the likelihood surface has no maximum at the critical value. Copas (1972) suggested a remedy. He assumed the data has rounding errors in the observations which allows for an approximate likelihood function to be used, and this approximated likelihood function, β_1^{copas} is bounded below by the standard *OLS* estimator, β_1^{ver} when $s_{yy}<s_{xx}$ or bounded above by β_1^{hor} when $s_{yy}>s_{xx}$. O'Driscoll and Ramirez (2011) used a simple data set {(1,1),(2,3),(3,2),(4,*x*)} to demonstrate the jump discontinuity inherent with β_1^{copas} . For this set, $\beta_1^{copas} = \beta_1^{ver} = 0.7970$ when *x*=3.99 and $\beta_1^{copas} = \beta_1^{hor} = 1.2528$ when *x*=4.01. Their remedy is to use the fourth moment slope estimator

$$\beta_1^{mom} = \sqrt{(s_{xyyy} - 3s_{xy}s_{yy})/(s_{xxxy} - 3s_{xy}s_{xx})}$$

described by Gillard and Iles (2005, Equation 26) to smooth out the jump discontinuity between β_1^{ver} and β_1^{hor}; that is use β_1^{mom} only when β_1^{ver} $<\beta_1^{mom}<\beta_1^{hor}$ and the *OLS* estimators otherwise. This procedure constrains the moment estimators so that the error variances are positive. We adopt this convention for β_1^{mom} in our simulation study below.

The estimator, β_1^{mom} can be shown to be in the family of *MLE* estimators with

$$\kappa = \frac{(s_{yy} - \beta_1^{mom}\, s_{xy})}{(s_{xx} - s_{xy}/\beta_1^{mom})}.$$

In studying the ratio of errors, it usually does not matter whether one uses the sum of squares such as $S_{xx} = \sum_{i=1}^n (x_i - X_i)^2$ or the normalized sum of squares $S_{xx} = \sum_{i=1}^n \frac{(x_i - X_i)^2}{n}$.

However, the formula for β_1^{mom} does require the normalized values, which is the notation that we have adopted throughout.

Solving Equations (6) and (7) yields a moment estimator, $\tilde{\kappa}(\beta_1)$, for $\kappa = \sigma_\tau^2/\sigma_\delta^2$. The slope estimator β_1^{mom} and the associated κ allows for a direct computation for the identical variance error ratio. Since the *MLE* estimator κ for the ratio of error variances has not been a useful estimator for κ, some authors such as Al-Nasser (2012) have suggested using the second order moments estimators in Equation (5) along with the fourth moment slope estimator β_1^{mom}. This was shown to have some promise in O'Driscoll and Ramirez (2011). We consider the moment estimator, $\tilde{\kappa}(\beta_1)$, as part of our simulation below.

Table 8 records the results for the moment estimators $\left\{\beta_1^{mom}, \tilde{\sigma}_\tau^2(\beta_1^{mom}), \tilde{\sigma}_\delta^2(\beta_1^{mom})\right\}$ for the expectation of the moment estimators for the error variances $E(\tilde{\sigma}_\tau^2)$ and $E(\tilde{\sigma}_\delta^2)$; the expected bias $E(\beta_1^{mom} - \beta_1)$ using β_1^{mom} and the expected standard deviation $E(std(\beta_1^{mom}))$.

(table 7 here)

6. Conclusion and References

Our simulations support the claim that our estimators β_1^{kap} and β_1^{lam}, under the conditions outlined above, greatly reduce the Bias and MSE associated with the ordinary least squares estimator β_1^{ver}. The expected values for the moment estimators for the error variances appear to suggest that the moment estimators are potentially useful for estimating the error ratios as suggested by Al-Nasser (2012).

Adcock, R. J. (1878). 'A problem in least-squares', *The Analys*t, 5: 53-54.

Al-Nasser A. (2012). 'On using the maximum entropy median for fitting the un-replicated functional model between the unemployment rate and the human development index in the Arab states', *J. Appl. Sci*., 12: 326–335.

Barnett V. (1970). 'Fitting straight lines—the linear functional relationship with replicated observations', *J. Roy. Statist. Soc. Ser. C Appl. Statist*., 19: 135–144.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models - A Modern Perspective,* Boca Raton: Chapman & Hall/CRC, Second Edition.

Copas J. (1972). 'The likelihood surface in the linear functional relationship problem', *J. Roy. Statist. Soc. Ser. B Methodological, 34:* 274–278.

Deming, W. E. (1943). *Statistical Adjustment of Data,* New York: Wiley.

Dent B. (1935), 'On observations of points connected by linear relation', *Proc. Phys.Soc.*, 47: 92–108.

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

Gillard J.,Iles T. (2005). 'Method of moments estimation in linear regression with errors in both variables'. *Cardiff University School of Mathematics Technical Report*, Cardiff, Wales, UK.

Gillard, J., Iles T. (2009). 'Methods of fitting straight lines where both variables are subject to measurement error', *Current Clinical Pharmacology*, 4: 164-171.

Lindley D. (1947). 'Regression lines and the linear functional relationship', *Supp. J.Roy. Stat. Soc.*, 9: 218–244.

Lindley, D., El-Sayyad, M. (1968). 'The Bayesian estimation of a linear functional relationship', *Journal of the Royal Statistical Society Series B (Methodological,)*, 30: 190-202.

Madansky, A. (1957). 'The fitting of straight lines when both variables are subject to error', *Journal of American Statistical Association,* 54: 173-205.

O'Driscoll, D., Ramirez, D., Schmitz, R. (2008). 'Minimizing oblique errors for robust estimation', *Irish. Math. Soc. Bulletin*, 62: 71-78.

O'Driscoll, D., Ramirez, D. (2011). 'Geometric View of Measurement Errors', *Communications in Statistics-Simulation and Computation,* 40: 1373-1382.

Riggs, D., Guarnieri, J., Addelman, S. (1978). 'Fitting straight lines when both variables are subject to error', *Life Sciences,* 22: 1305-1360.

Solari M. (1969). 'The "Maximum Likelihood Solution" of the problem of estimating a linear functional relationship', *J. Roy. Stat. Soc. Ser. B Methodological, 31:* 372–375.

Sprent P. (1970). 'The saddle point of the likelihood surface for a linear functional relationship', *J. Roy. Stat. Soc. Series B Methodological* , 32: 432–434.

Van Montfort, K., Mooijaart, A., and de Leeuw, J. (1987). 'Regression with errors in variables', *Statist Neerlandica,* 41: 223-239.

Yang L. (1999). 'Recent advances on determining the number of real roots of parametric polynomials', *J. Symbolic Computation,* 28: 225-242.

Table 1

$E_{bias} = E(\hat{\beta}_1(\tilde{\kappa}) - \beta_1)$; $E_{emp} = E(\hat{\beta}_1(\tilde{\kappa}) - \hat{\beta}_1(\kappa))$; $bias(\tilde{\kappa}{:}\kappa)$; $serbias(\tilde{\kappa}{:}\kappa)$ with $\{\beta_1 = 1,\ \beta_0 = 0,\ \sigma_X^2 = 100, \kappa = \sigma_\tau^2/\sigma_\delta^2,\ \tilde{\kappa} = 1, n = 50, N = 5000\}$.

σ_τ^2	σ_δ^2	κ	$\sigma_\tau^2 - \sigma_\delta^2$	E_{bias}	E_{emp}	$bias(\tilde{\kappa}{:}\kappa)$	$serbias(\tilde{\kappa}{:}\kappa)$
3	9	0.333	−6.0	−0.0298	−0.0286	−0.0296	−0.0281
5	9	0.555	−4.0	−0.0201	−0.0200	−0.0198	−0.0202
2	4	0.500	−2.0	−0.0100	−0.0095	−0.0100	−0.0103
1	3	0.333	−2.0	−0.0089	−0.0092	−0.0099	−0.0112
1	2	0.500	−1.0	−0.0048	−0.0047	−0.0050	−0.0052
2	1	2.000	1.0	0.0047	0.0051	0.0050	0.0048
3	1	3.000	2.0	0.0103	0.0104	0.0101	0.0088
4	2	2.000	2.0	0.0107	0.0118	0.0101	0.0097
9	5	1.800	4.0	0.0204	0.0201	0.0202	0.0197
9	3	3.000	6.0	0.0318	0.0306	0.0304	0.0286

Table 2

X is UD(0,20), β_1=1.0, β_0=0, R=1000, n=100, σ_τ=1, σ_δ=3

	MSE 10^{-3}	%Bias	λ	θ_λ
β_1^{ver}	46.569	-21.189	1	51.76
β_1^{gm}	11.897	-9.947	0.500	95.99
β_1^{hor}	4.402	2.957	0	134.17
β_1^{per}	15.130	-11.246	0.556	89.93
β_1^{kap}	4.625	-1.382	0.169	118.37
β_1^{lam}	4.442	-0.029	0.237	123.49

Table 3

X is UD(0,20), β_1=1.25, β_0=0, R =1000, n=100, σ_τ=1, σ_δ=3

	MSE 10^{-3}	%Bias	λ	θ_λ
β_1^{ver}	70.809	-20.929	1	45.33
β_1^{gm}	18.425	-10.036	0.500	83.29
β_1^{hor}	5.708	2.413	0	127.99
β_1^{per}	15.081	-8.546	0.434	89.90
β_1^{kap}	6.304	-1.180	0.171	114.70
β_1^{lam}	5.847	0.092	0.145	116.62

Table 4

X is UD(0,20), β_1=1.0, β_0=0, R =1000, n =100, σ_τ=2, σ_δ=2

	MSE 10^{-3}	%Bias	λ	θ_λ
β_1^{ver}	13.403	-10.688	1	48.23
β_1^{gm}	2.117	0.0989	0.500	89.94
β_1^{hor}	18.146	12.232	0	131.70
β_1^{per}	2.672	0.126	0.500	89.92
β_1^{kap}	4.432	0.295	0.495	90.38
β_1^{lam}	5.962	0.425	0.497	90.14

Table 5

X is UD(0,20), β_1 =0.75, β_0 =0, R =1000, n =100, σ_τ =2, σ_δ =2

	MSE 10^{-3}	%Bias	λ	θ_λ
β_1^{ver}	7.791	-10.518	1	56.13
β_1^{gm}	2.603	4.196	0.500	103.99
β_1^{hor}	28.487	21.417	0	137.68
β_1^{per}	2.041	0.169	0.640	89.96
β_1^{kap}	4.233	0.725	0.590	95.55
β_1^{lam}	5.402	-0.029	0.615	92.97

Table 6

Effective $\tilde{\kappa}$ average, X is UD(0,20), β_1 =1, β_0 = 0, R =1000, n =100

	σ_τ^2 =1	σ_τ^2 = 4	σ_τ^2 = 9
σ_δ^2 =1	1.1781	3.3975	6.1251
σ_δ^2 =4	0.3185	0.9169	1.9514
σ_δ^2 =9	0.1701	0.4090	1.1658

Table 7.

Simulation Study of $E(\tilde{\sigma}_\tau^2)$, $E(\tilde{\sigma}_\delta^2)$, $E(\beta_1^{mom} - \beta_1)$, $E(std(\beta_1^{mom})$ with

$$\{\beta_1 = 1, \beta_0 = 0, \sigma_X^2 = 100, \kappa = \sigma_\tau^2/\sigma_\delta^2,\ \ \tilde{\kappa} = 1, n = 50, N = 5000\}$$

σ_δ^2	σ_τ^2	$E(\tilde{\sigma}_\delta^2)$	$E(\tilde{\sigma}_\tau^2)$	$E(\beta_1^{mom} - \beta_1)$	$E(std(\beta_1^{mom}))$.
4	9	4.29	8.14	-0.00221	0.06372
1	1	0.95	0.96	0.00001	0.02131
1	2	1.16	1.71	-0.00213	0.02743
1	3	1.37	2.45	-0.00396	0.03238
1	4	1.47	3.30	-0.00500	0.03611
2	1	1.69	1.17	0.00247	0.02757
2	2	1.91	1.91	0.00123	0.03227
2	3	2.10	2.68	-0.00094	0.03666
2	4	2.23	3.51	-0.00257	0.03993
3	1	2.46	1.36	0.00456	0.03233
3	4	2.74	2.04	0.00212	0.03683
3	3	2.85	2.84	0.00098	0.04086
3	4	3.04	3.60	-0.00152	0.04419
4	1	3.25	1.52	0.00604	0.03663
4	2	3.50	2.20	0.00390	0.04107
4	3	3.68	2.30	0.00271	0.04107
4	4	3.83	3.80	0.00112	0.04784
9	4	8.01	4.29	0.00911	0.06466