15

Anomalies of the Magnitude of the Bias of the Maximum Likelihood Estimator of the Regression Slope

Diarmuid O'Driscoll, Mary Immaculate College, Ireland
Donald Ramirez, University of Virginia, USA

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

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}.

Methodology

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

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

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 and Ramirez (2011), so that our equation is dimensionally correct we consider a standardized weighted model

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

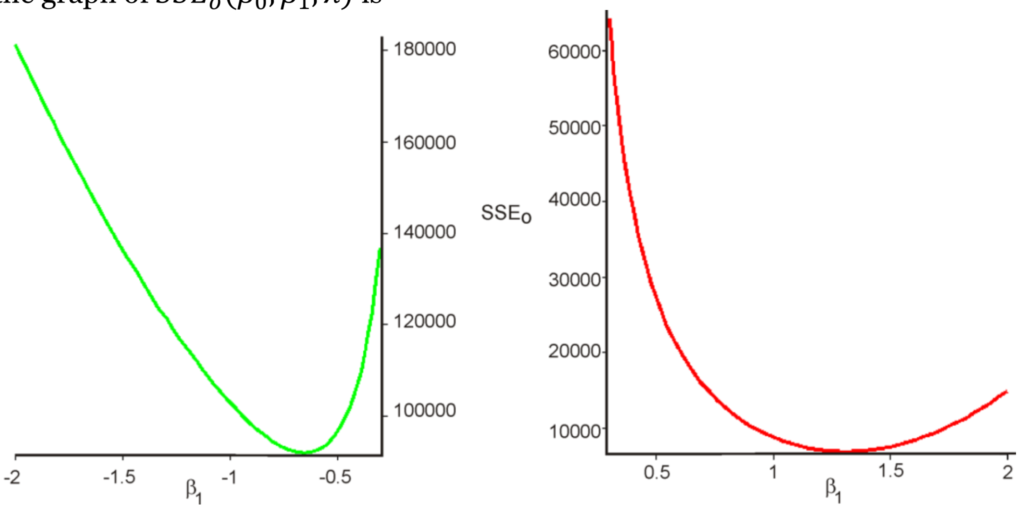
where

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

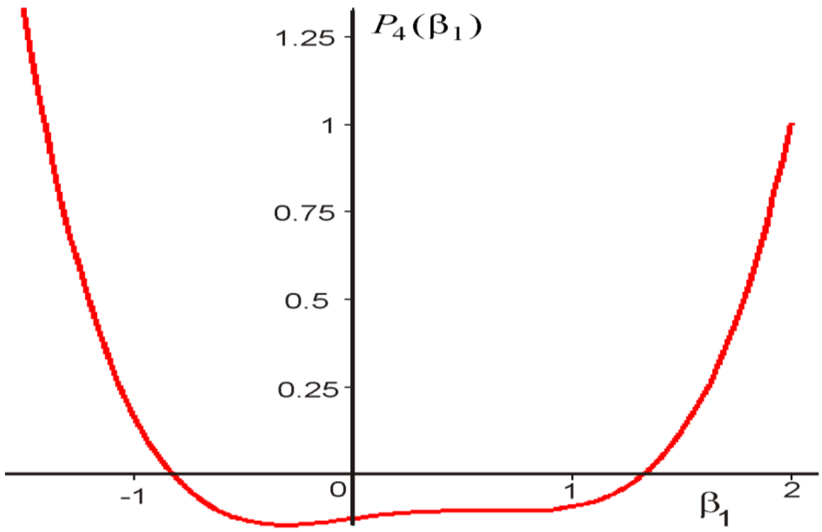
The solution of $\frac{d(SSE_O)}{d\beta_0} = 0$ is given by $\beta_0 = \bar{y} - \beta_1 \bar{x}$ and the solutions of $\frac{d(SSE_O)}{d\beta_1} = 0$ are the roots of the fourth degree polynomial, $P_4(\beta_1)$,

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

For example, if $s_{xx} = 31690, s_{yy} = 63610, s_{xy} = 40611$ and $\lambda = 0.7$, the graph of $SSE_o(\beta_0, \beta_1, \lambda)$ is



From O'Driscoll et al. (2008), the Complete Discrimination System $\{D_1,\ldots,D_n\}$ 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}. For $s_{xx} = 31690, s_{yy} = 63610, s_{xy} = 40611$ and $\lambda = 0.7$ the graph of $P_4(\beta_1)$ is



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 $\kappa = \sigma_Y^2/\sigma_X^2$.

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

$$\delta_i \sim N(0, \sigma_\delta^2);\ \tau_i \sim N(0, \sigma_\tau^2)\ ;$$

$$E(\delta_i, \delta_j) = 0, i \neq j; E(\tau_i, \tau_j) = 0, i \neq j;\ E(\delta_i, \tau_j) = 0, \forall\ i \text{ and } j.$$

It is well known, in a measurement error model, that the expected value for β_1^{ver} ($OLS(y\,|\,x)$) is attenuated towards 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 towards infinity by the amplifying factor $(\sigma_Y^2 + \sigma_\tau^2)/\sigma_Y^2$.
From Gillard and Iles (2009), second moment equations are

$$s_{xx} = \sigma_X^2 + \sigma_\delta^2;\ s_{yy} = \beta_1^2\sigma_X^2 + \sigma_\tau^2\ ;\ s_{xy} = \beta_1\sigma_X^2 \quad (3)$$

and fourth moment equations are

$$s_{xxxy} = \beta_1\mu_{X4} + 3\beta_1\sigma_X^2\sigma_\delta^2;\ \ s_{xyyy} = \beta_1^3\mu_{X4} + 3\beta_1\sigma_X^2\sigma_\tau^2. \quad (4)$$

These equations yield the estimators

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

the Frisch hyperbola of Van Montfort (1987)

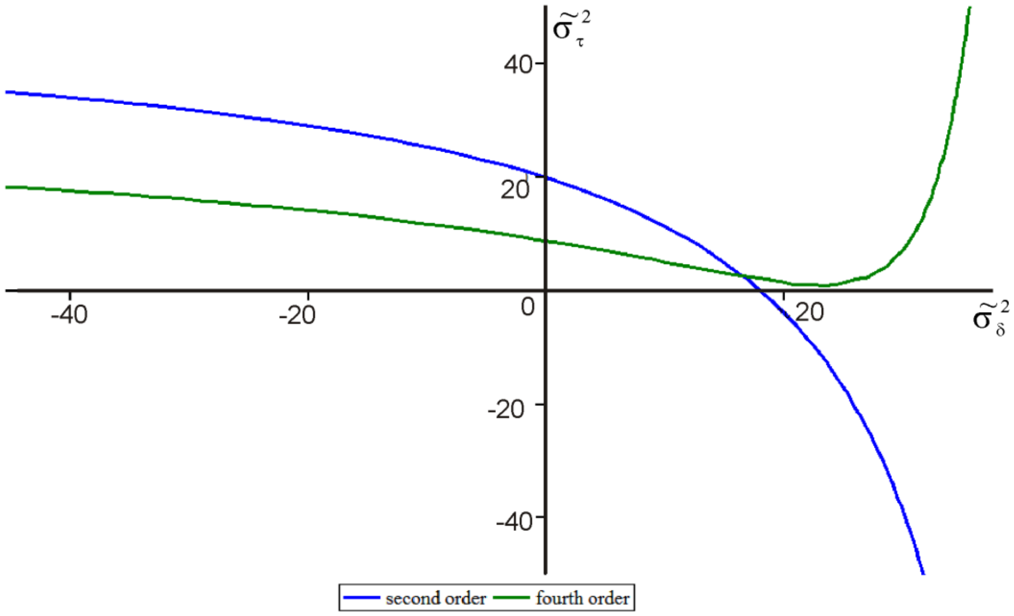
$$(s_{xx} - \sigma_\delta^2)(s_{yy} - \sigma_\tau^2) = s_{xy}^2 \qquad (6)$$

and the fourth order equation

$$(s_{xxxy} - 3s_{xy}\sigma_\delta^2)(s_{xy}^2) = (s_{xx} - \sigma_\delta^2)^2(s_{xyyy} - 3s_{xy}\sigma_\tau^2) \qquad (7)$$

We use equations (6) and (7) to find estimators for σ_δ^2 and σ_τ^2 , namely $\tilde{\sigma}_\delta^2$ 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. A typical graph of equations (6) and (7) is

True Ratio $\kappa = 1/9$: Solution $\tilde{\kappa} = 0.14962$



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 $\tilde{\kappa}$.

Monte Carlo Simulation

The Bias of the MLE for an Incorrect Choice of κ

We set the estimated ratio of the error variances $\tilde{\kappa} = 1$. 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}. We used Minitab for our simulation study setting the number of runs N = 5000. The results for the bias $E\left(\hat{\beta}_1(\tilde{\kappa}\,) - \beta_1\,\right)$ are recorded in Table 1.

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

σ_τ^2	σ_δ^2	κ	$\sigma_\tau^2 - \sigma_\delta^2$	E_{bias}
3	9	0.333	−6.0	−0.0298
5	9	0.555	−4.0	−0.0201
2	4	0.500	−2.0	−0.0100
1	3	0.333	−2.0	−0.0089
1	2	0.500	−1.0	−0.0048
2	1	2.000	1.0	0.0047
3	1	3.000	2.0	0.0103
4	2	2.000	2.0	0.0107
9	5	1.800	4.0	0.0204
9	3	3.000	6.0	0.0318

The rows of Table 1 are sorted in ascending order of the theoretical bias, E_{bias}, displayed in Column 5. 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 0.0201 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 0.0103 and 0.0318 are approximately proportional to the respective differences of the error variances $3 - 1 = 2$ and $9 - 3 = 6$.

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.

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. The orientation for θ_λ is chosen such that for β_1^{ver}, $0 < \theta_\lambda < 90^0$ and for β_1^{hor}, $90^0 < \theta_\lambda < 180^0$.

Table 2. X *is UD(0,20),* $\beta_1 = 1.0, \beta_0 = 0, N = 1000, n = 100, \sigma_\tau = 1, \sigma_\delta = 3$

	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),* $\beta_1 = 1.25, \beta_0 = 0, N = 1000, n = 100, \sigma_\tau = 1, \sigma_\delta = 3$

	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

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 $\sigma_\delta^2 >> \sigma_\tau^2$.

Table 4. *X is UD(0,20),* $\beta_1 = 1.0, \beta_0 = 0, N = 1000, n = 100, \sigma_\tau = 2, \sigma_\delta = 2$

	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),* $\beta_1 = 0.75$*,* $\beta_0 = 0$*, N =1000, n =100,* $\sigma_\tau = 2$*,* $\sigma_\delta = 2$

	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

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 3.3, for $(\sigma_\delta^2, \sigma_\tau^2) \in \{1,4,9\} \times \{1,4,9\}$.

Table 6. *Effective* $\tilde{\kappa}$ *average; X is UD(0,20),* $\beta_1 = 1$*,* $\beta_0 = 0$*, N =1000, n = 100*

	$\sigma_\tau^2 = 1$	$\sigma_\tau^2 = 4$	$\sigma_\tau^2 = 9$
$\sigma_\delta^2 = 1$	1.1781	3.3975	6.1251
$\sigma_\delta^2 = 4$	0.3185	0.9169	1.9514
$\sigma_\delta^2 = 9$	0.1701	0.4090	1.1658

Conclusion

Our simulation study in *3.1* illustrates that the bias of the MLE estimator of the regression slope is dependent on the magnitude of σ_δ^2, the variance of the errors in *x*.

Our simulation studies in *3.3* support the claim that our estimators β_1^{kap} and β_1^{lam}, under the conditions outlined in *2.2*, greatly reduce the Bias and MSE associated with the ordinary least squares estimator β_1^{ver}.

References

Adcock, R.J. 1878. A problem in least-squares, *The Analys*t, 5: 53-54.
Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. 2006. *Measurement Error in Nonlinear Models - A Modern Perspective,* Boca Raton: Chapman & Hall/CRC, Second Edition.
Deming, W.E. 1943. *Statistical Adjustment of Data,* New York: Wiley.
Fuller, W.A. 1987. *Measurement Error Models*, New York: Wiley.
Gillard, J. and Iles T. 2009. Methods of fitting straight lines where both variables are subject to measurement error, *Current Clinical Pharmacology*, 4, 164-171.
Lindley, D. and 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. and Schmitz, R. 2008. Minimizing oblique errors for robust estimation, *Irish. Math. Soc. Bulletin*, 62, 71-78.
O'Driscoll, D. and Ramirez, D. 2011. Geometric View of Measurement Errors, *Communications in Statistics-Simulation and Computation,* 40, 1373-1382.
Riggs, D., Guarnieri, J. and Addelman, S. (1978). Fitting straight lines when both variables are subject to error, *Life Sciences,* 22, 1305-1360.
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.