On Errors-in-Variables in Binary Regression—Berkson Case

by

Deborah Burr*

Department of Statistics
Florida State University
Tallahassee, Florida 32306-3033

FSU Technical Report No. M-775

November, 1987

* Deborah Burr is Assistant Professor, Department of Statistics and Statistical Consulting Center, Florida State University, Tallahassee, Florida 32306. This paper is based on the author's Ph.D. dissertation, written under the supervision of the late Professor Rupert G. Miller, Jr., who suggested the topic, guided the work, and contributed valuable ideas. The author thanks two referees for very thorough and helpful reports, which led to substantial improvements in content and clarity of this paper.
ABSTRACT

In binary regression, the predictor variables may be measured with error. The Berkson case of the errors-in-variables problem is considered, under which the values of the predictor variables are set by the experimenter, but are not achieved exactly. A particular model for this case is considered, with probit regression and normally distributed errors of observation. The regression parameters for intercept and slope are to be estimated. Two estimators are studied, the maximum likelihood estimator (MLE) and a modification of it. The modified MLE is shown in a simulation study to improve on the MLE in a situation with substantial measurement error.

Quantal bioassay is an important field of application of the Berkson case of the binary regression model. In quantal bioassay, there is a stimulus, perhaps a carcinogen or poison, with doses $X$ to be determined by the experimenter. In order to address questions about carcinogenicity or toxicity of the stimulus, each experimental animal is assigned a dose of the stimulus, and a binary response such as death or survival is observed. A need for errors-in-variables models could arise in quantal bioassay if there is difficulty in achieving the desired dose. Or, the amount of the injected substance which remains in the bloodstream may be the variable which most directly affects the outcome. This amount could be modelled by a formula of the form $\log(X) +$ error, say.

When analyzing a data set, it is important to remember that even if there is error in $X$, an errors-in-variables model may not be needed. Consider the situation in which one can never hope to observe $X$ without error and is only interested in estimating the probability of success for a new observed $X$ which is measured with error. In this situation, under the particular model considered in this paper, the ordinary binary regression model, assuming no error in $X$, is all that is required; see Section 2 for details. On the other hand, if one is interested in modelling the relationship between the true dose and the outcome or in predicting the outcome given a true dose, then the errors-in-variables model is needed. Madansky (1959) discusses these points for errors-in-variables in linear regression.

In this paper it is shown that when an errors-in-variables model is appropriate and measurement error is large, the usual probit regression MLE is a poor estimator of the intercept and slope parameters $\alpha$ and $\beta$. In particular, this naive MLE is inconsistent. It is also shown that to improve upon the naive MLE, one must know or estimate in a separate experiment the measurement error variance, for otherwise the regression parameters are unidentifiable.

KEY WORDS: Logistic regression; Maximum likelihood estimation; Probit regression; Quantal bioassay.
1. INTRODUCTION

In binary regression, one observes a 0 – 1 outcome $Y$, along with an explanatory or predictor variable $X$. In this paper $X$ is taken to be univariate. The results generalize to the multivariate case. Suppose that the probability of success increases with $X$. Then it is often natural to assume $P(Y = 1 | X = x) = G(\alpha + \beta x)$, $G$ a cumulative distribution function. Two common models are the logistic regression model, with $G(x) = (1 + e^{-x})^{-1}$, and the probit regression model, with $G = \Phi$, the standard normal cumulative distribution function. These two models agree very closely over almost the whole range of values of the predictor variable; they can be used interchangeably unless questions about extreme quantiles are being addressed (Cox 1970, p. 27).

The term “errors-in-variables” in the context of binary regression refers to error in $X$. That is, the observable $X$ is the sum of the “true” explanatory variable and an error of measurement.

The linear regression errors-in-variables problem has been extensively studied. Not nearly as much has been done on errors-in-variables in binary regression, but some models for this problem have recently been suggested and studied. All of these models were chosen to be appropriate for the situation of the Framingham Heart Study (Gordon and Kannel 1968). In this prospective study of the development of heart disease, factors thought to influence heart disease, such as systolic blood pressure and serum cholesterol, were measured. The presence or absence of coronary heart disease was recorded. The need for errors-in-variables models arose because measurements of variables such as blood pressure are highly variable, with error arising from actual measurement error, plus time-of-day and day-of-week variability. Two approaches were taken to cope with the unknown true predictor variable. These approaches are the classical ones for the error-in-$X$ problem in linear regression (Kendall and Stuart 1979, chap. 29). Carroll et al. (1984) and Clark (1982) get a structural model by assuming the true predictor variable is a random variable. Stefanski and Carroll (1985) get a functional model by assuming the true predictors are fixed, unknown parameters. In all these papers, the MLEs assuming no error in $X$ are found to be inconsistent under the models studied, and alternative estimators are suggested and studied.

The starting point for the present work comes from the work of Carroll et al. on the structural case. These authors consider a particular case of the structural model, in which the regression is probit, the predictor variables are normally distributed, and the errors of observation are normally distributed. The advantage of the normality assumptions is that the likelihood can be expressed in a convenient form. Carroll et al. suggest an estimator of $\alpha$ and $\beta$ which is a pseudo MLE. They do Monte Carlo studies comparing the bias and mean squared error of their estimator to the bias and
mean squared error of the MLE for ordinary probit regression, that is, the MLE assuming no error in $X$. Their studies demonstrate inconsistency of the ordinary MLE yet indicate that it is worthwhile to use the error-in-$X$ MLE only when the measurement error is substantial and the sample size large.

Recall that in quantal bioassay the $X$-values are fixed by the experimenter. In the linear regression errors-in-variables problem, Berkson first noticed that the situation in which $X$ is fixed by the experimenter differs fundamentally from the situation in which $X$ is not fixed by the experimenter, and the former case has been named after him (Berkson 1950). He showed that in linear regression with error in $X$ the ordinary least-squares estimator is consistent when $X$ is fixed. Under the particular model for binary regression with error in $X$ considered in this paper, the ordinary MLE is not consistent whether or not $X$ is fixed; a heuristic argument for inconsistency of the ordinary MLE is given in Section 2.

Section 2 begins by defining the model for the Berkson case (Model 1) which is analogous to Carroll et al.'s for the structural case. It is noted that this errors-in-variables model is a reparameterization of the ordinary probit regression model. Several useful facts are made clear by the reparameterization, such as inconsistency of the ordinary MLE, $\hat{\beta}_0 = (\hat{\alpha}_0, \hat{\beta}_0)$, and unidentifiability of the parameters if the error variance is unknown. The inconsistency of $\hat{\beta}_0$ motivates consideration of an estimator which has better asymptotic behavior; the MLE for Model 1, $\hat{\beta}_1$, is discussed in Section 3. There is noticeable skewness in the distribution of $\hat{\beta}_1$ for sample sizes up to 300; hence its mean squared error is disappointingly large. In Section 4, a new estimator, $\hat{\beta}_2$, is defined in the hope of doing better for moderate sample sizes by reducing the skewness of $\hat{\beta}_1$. Results of a Monte Carlo study comparing $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are given in Section 5, and $\hat{\beta}_2$ is seen to be a substantial improvement over $\hat{\beta}_0$ and $\hat{\beta}_1$ in a situation with large measurement error variance.

2. THE BERKSON MODEL REPARAMETERIZED

The experimenter chooses $k$ fixed dose levels $x_j, j = 1, \ldots, k$, and observes $n$ independent binary outcomes at each dose level, $Y_{ij}, i = 1, \ldots, n$. Henceforth subscripts are omitted except where needed for clarity. The usual probit regression model is

\begin{align*}
\text{Model 0.} \quad & P(Y = 1|x) = \Phi(\alpha_0 + \beta_0 x), \\
& -\infty < \alpha_0 < \infty, \beta_0 \geq 0.
\end{align*}
The model assumed here is:

Model 1. \[ U = x + \epsilon, \]
\[ P(Y = 1|U = u) = \Phi(\alpha_1 + \beta_1 u), \]
\[ -\infty < \alpha_1 < \infty, \beta_1 \geq 0. \]

The errors \( \epsilon_{ij} \) are independent and normally distributed with mean zero and known variance \( \sigma^2_\epsilon \). Call the MLE under Model 0 \( \hat{\theta}_0 \) and the MLE under Model 1 \( \hat{\theta}_1 \).

It is not necessary in the development of statistical methods for Models 0 and 1 to assume \( \beta \geq 0 \). The results given here can be stated for the case \( -\infty < \beta < \infty \). The restriction \( \beta \geq 0 \) is imposed here to allow some simplification of the arguments and formulas; also, this assumption is usually implicit in applications of the Berkson model such as quantal bioassay.

In the structural case of the errors-in-variables model, it is usually assumed that the true predictor, \( U \), and error in observing it, \( \epsilon \), are independent. Note that these two random variables are not independent in the Berkson model; in fact, \( \text{cov}(U_{ij}, \epsilon_{ij}) = \sigma^2_\epsilon \).

To find the likelihood under Model 1, write

\[
P(Y = 1|X = x) = \int \Phi(\alpha_1 + \beta_1 u)f_{U|x}(u|x)du, \tag{2.1}
\]

where \( f_{U|x} \) is the normal density with mean \( x \) and variance \( \sigma^2_\epsilon \). The right side of equation (2.1) simplifies to

\[
P(Y = 1|X = x) = \Phi((\alpha_1 + \beta_1 x)(1 + \beta_1^2 \sigma^2_\epsilon)^{-\frac{1}{2}})
\]
[= \Phi(\alpha_R + \beta_R x),
\]

with \( \alpha_R = \alpha_1(1 + \beta_1^2 \sigma^2_\epsilon)^{-\frac{1}{2}} \), and \( \beta_R = \beta_1(1 + \beta_1^2 \sigma^2_\epsilon)^{-\frac{1}{2}} \). \tag{2.2}

Also note that as \( \beta_1 \uparrow \infty, \beta_R \uparrow 1/\sigma_\epsilon \). Therefore Model 1 is equivalent to

Model 1R. \[ P(Y = 1|x) = \Phi(\alpha_R + \beta_R x), \]
\[ -\infty < \alpha_R < \infty, 0 \leq \beta_R < 1/\sigma_\epsilon. \]

That is, Model 1 is just an ordinary probit regression model with slope parameter constrained to be less than \( 1/\sigma_\epsilon \). The MLE of \( (\alpha_R, \beta_R) \), \( \hat{\theta}_R \), is a constrained MLE.
The reparameterization makes clear that it is impossible to estimate the parameters if \( \sigma^2 \) is unknown. For any positive value of \( \sigma^2 \), Model 1 is indistinguishable from an ordinary probit regression model, in which \( \sigma^2 = 0 \). More formally, for any set of parameter values \( \nu = (\alpha, \beta, \sigma^2) \), there exist infinitely many other sets of parameter values \( \nu' \neq \nu \) such that for all \( x \), \( P_\nu(Y = 1|x) = P_{\nu'}(Y = 1|x) \).

One does not expect \( \sigma^2 \) to be known in practice, so it is unfortunate that if the parameters \( \alpha, \beta, \sigma^2 \) are all unknown, they are unidentifiable. Even if we know the intercept \( \alpha \), the slope \( \beta \) and error variance \( \sigma^2 \) are unidentifiable. We must know either \( \beta \) or \( \sigma^2 \) to get started. The usual situation is that we want to estimate \( \beta \), and \( \sigma^2 \) is a nuisance parameter which must be estimated.

What can be said about convergence in probability or almost surely of the MLE for Model 0 when the assumptions of Model 1 hold? The reparameterization makes the answer obvious; since Model 1 with parameters \( \alpha_1 \) and \( \beta_1 \) "looks like" Model 0 with parameters \( \alpha_R \) and \( \beta_R \), the MLE for Model 0 will be a good estimator – of \( \alpha_R \) and \( \beta_R \). Fahrmeir and Kaufmann (1985, p. 362; also see Correction 1986) give conditions under which the probit regression MLE is weakly consistent and asymptotically normal, as a special case of results for generalized linear models. Since \( \alpha_R \neq \alpha_1 \) and \( \beta_R \neq \beta_1 \) except in the case \( \beta_1 = 0 \), \( \hat{\theta}_0 \) is inconsistent for \( \theta_j \) unless \( \beta_1 = 0 \). In fact, we know more; we know to what value \( \hat{\theta}_0 \) converges. Note in particular, from Equation (2.2), that the usual probit regression slope estimator is asymptotically biased downward from \( \beta_1 \). This is in contrast with the Berkson case in errors-in-variables in linear regression, in which the least-squares estimate is exactly unbiased.

In quantal bioassay, the binary regression model is often expressed in terms of a tolerance distribution. The main interest is estimation of the median of the tolerance distribution, the ED50 (LD50), or perhaps some other quantiles of the tolerance distribution such as the ED75 (LD75) [see Finney (1964)]. The effect of measurement error on quantile estimation is thus of interest. A little notation is needed. For \( 0 < \tau < 1 \), let \( z_\tau \) be the \( \tau \)th quantile of the standard normal distribution. That is, for \( Z \sim N(0,1) \), define \( z_\tau \) by \( P(Z \leq z_\tau) = \tau \). Let \( x_\tau \) be defined so that under Model 1R \( P(Y = 1|x_\tau) = \tau \); let \( u_\tau \) be defined so that under Model 1 \( P(Y = 1|U = u_\tau) = \tau \). Then \( x_\tau \) and \( u_\tau \) are the observed and actual dose levels respectively which produce a response in \( 100\tau \% \) of the animals.

By Equations (2.2) and the expressions for \( P(Y = 1|U = u) \) and \( P(Y = 1|x) \) in Model 1 and Model 1R, we see that

\[
x_\tau - u_\tau = \frac{z_\tau}{\beta_1} \left[ (1 + \beta_1^2 \sigma^2)^{\frac{1}{2}} - 1 \right].
\]
Thus, (i) if \( r = \frac{1}{2} \) then \( z_r = 0 \) and \( x_r = u_r \), (ii) if \( r > \frac{1}{2} \) then \( z_r > 0 \) and \( x_r > u_r \), and (iii) if \( r < \frac{1}{2} \) then \( z_r < 0 \) and \( x_r < u_r \). The relationship between \( x_r \) and \( u_r \) reflects the attenuation in the observable response function due to the measurement error. It is easy to see the effect of this attenuation on the naive estimates of quantiles of the tolerance distribution for the true dose \( U \). From (i) the estimate of the ED50 will be consistent for the true ED50; from (ii) and (iii) the estimates of quantiles above (below) the median will be asymptotically positively (negatively) biased.

3. THE MAXIMUM LIKELIHOOD ESTIMATOR

The relationship between the two versions of the error-in-\( X \) model, Model 1 and Model 1R, can be written

\[
\beta_1 = \beta_R (1 - \beta_R^2 \sigma_e^2)^{-\frac{1}{2}}, \\
\alpha_1 = \alpha_R (1 - \beta_R^2 \sigma_e^2)^{-\frac{1}{2}}.
\]  

(3.1)

Say \( \hat{\theta}_1 = f(\hat{\theta}_R) \).

By the invariance property of maximum likelihood estimators, \( \hat{\theta}_1 = f(\hat{\theta}_R) \). Now \( \hat{\theta}_R \) is the same as the MLE for Model 0, \( \hat{\theta}_0 \), except for the constraint \( \beta_R < 1/\sigma_e \). So if \( \hat{\beta}_0 < 1/\sigma_e \), it is clear that \( \hat{\theta}_R = \hat{\theta}_0 \), and \( \hat{\theta}_1 = f(\hat{\theta}_0) \).

What if \( \hat{\beta} \geq 1/\sigma_e \)? Now in the ordinary probit regression model, either the MLE exists and is unique, or it is infinite. (Silvapulle (1981), gives conditions on the predictor variables which are necessary and sufficient for existence and uniqueness of the MLE). In addition, the negative of the log likelihood function is convex and has continuous first and second derivatives. Therefore, when the MLE exists, it is the unique zero of the likelihood equations. Thus if \( \hat{\beta}_0 \geq 1/\sigma_e \), the likelihood for Model 1R increases as \( \beta_1 \uparrow 1/\sigma_e \), and the maximum likelihood occurs on the boundary of the set \( \{ (\alpha_R, \beta_R) : |\alpha_R| < \infty, 0 \leq \beta_R < 1/\sigma_e \} \). Then it is clear from Equations (3.1) that if \( \hat{\beta}_0 \geq 1/\sigma_e \), the MLE for Model 1, \( \hat{\theta}_1 \), diverges. To summarize, a computational formula for \( \hat{\theta}_1 \) is:

\[
\hat{\theta}_1 = \frac{\hat{\theta}_0}{\sqrt{1 - \hat{\beta}_0^2 \sigma_e^2}} \text{ if } \hat{\beta}_0 < 1/\sigma_e \\
\hat{\theta}_1 \text{ diverges} \text{ if } \hat{\beta}_0 \geq 1/\sigma_e.
\]  

(3.2)

Note that by Equations (3.2), \( -\hat{\alpha}_1/\hat{\beta}_1 = -\hat{\alpha}_0/\hat{\beta}_0 \) when \( \hat{\beta}_0 < 1/\sigma_e \). That is, when \( \hat{\beta}_0 < 1/\sigma_e \), the MLE of the ED50 under Model 1 is the same as the MLE of the ED50 under Model 0. To see what happens when \( \hat{\beta}_0 \geq 1/\sigma_e \), we first establish some notation. Let \( \beta_R(\delta) = (\sigma_e - \delta)^{-1} \), and let \( \alpha_R(\delta) \) be
the value of $\alpha_R$ which maximizes the Model 1R likelihood when $\beta_R = \beta_R(\delta)$. Let $\ell(\cdot, \cdot)$ denote the likelihood function. If $\ell(\alpha(\delta), \beta(\delta)) \uparrow$ as $\delta \downarrow 0$, then $\hat{\alpha}_R = \alpha(0)$ and $\hat{\beta}_R = \beta(0) = \sigma_e^{-1}$. Now fixing $\delta > 0$ and solving Equations (3.1) for $\beta_1(\delta)$ and $\alpha_1(\delta)$ gives

$$
\beta_1(\delta) = (\sigma_e - \delta)^{-1}(1 - \sigma_e^2/(\sigma_e - \delta)^2)^{-\frac{1}{2}}
$$

$$
\alpha_1(\delta) = \alpha(\delta)(1 - \sigma_e^2/(\sigma_e - \delta)^2)^{-\frac{1}{2}}.
$$

As $\delta \to 0$ it follows that $|\beta_1(\delta)|$ and $|\alpha_1(\delta)|$ both diverge, but the ratio $\alpha_1(\delta)/\beta_1(\delta) \to \sigma_e\alpha(0) = \hat{\alpha}_R/\hat{\beta}_R$. Thus the Model 1 MLE of the ED50 equals the Model 1R MLE of the ED50.

Since $\hat{\theta}_1$ is the MLE for Model 1, we expect it to be consistent and asymptotically normal under Model 1. Verification proceeds as follows. The naive MLE $\hat{\theta}_0$ is a consistent estimator of $\theta_R$; see Section 2. Weak consistency of $\hat{\theta}_0$ for $\theta_R$ establishes weak consistency of $\hat{\theta}_1 = f(\hat{\theta}_0)$ for $\theta_1 = f(\theta_R)$, since $f(x, y)$ is a continuous function on $(-\infty, \infty) \times (0, 1/\sigma_e)$ and $\beta_R \in (0, 1/\sigma_e)$. Asymptotic normality of $\hat{\theta}_1$ also follows from asymptotic normality of $\hat{\theta}_0$. If the error variance $\sigma_e^2$ is not known but is estimated from a previous data set with $m$ observations, then in order for these asymptotic results to hold we must have both $m \to \infty$ and $n \to \infty$.

The asymptotic normal distribution of $\hat{\theta}_1$ will not be used to form confidence intervals for $\theta_1$, because Monte Carlo studies indicated that this normal distribution is reached only for very large $n$. Consider instead how to form confidence intervals for $\theta_1$ using the transformation given in Equations (3.1). First the standard MLE-theory confidence intervals for $\alpha_0$ and $\beta_0$ are formed. These can be modified in the obvious way if necessary to satisfy the constraint in Model 1R, so that they are reasonable confidence intervals for $\alpha_R$ and $\beta_R$. By Equations (3.1), $\beta_1$ is an increasing function of $\beta_R$. Therefore, a $100(1 - \alpha)$% confidence interval for $\beta_1$ can be formed by transforming the lower and upper limits of a $100(1 - \alpha)$% interval for $\beta_R$ according to this increasing function. Say $\alpha_1 = g(\alpha_R, \beta_R)$. To get a $100(1 - \alpha)$% interval for $\alpha_1$ requires simultaneous $100(1 - \alpha)$% intervals for $\alpha_R$ and $\beta_R$. Simultaneous coverage is obtained here by the Bonferroni method. That is, non-simultaneous $100(1 - \alpha/2)$% intervals are obtained; these have simultaneous coverage probability greater than or equal to $1 - \alpha$. Since the function $g$ is monotone in each argument, it is straightforward to transform these intervals and obtain a conservative $100(1 - \alpha)$% interval for $\alpha_1$, taking care to observe that the direction of the monotonicity in $\beta_R$ changes according to the sign of $\alpha_R$. These intervals will be referred to as Method 1 intervals for $\theta_1$. They will be compared to the naive intervals, that is the standard MLE-theory confidence intervals for $\theta_0$, which will be called Method 0 intervals for $\theta_1$. 

6
4. A MODIFICATION OF THE MLE

Monte Carlo studies of \( \hat{\beta}_1 \) showed that its distribution is skewed, even for sample size as large as 300. At the smaller sample sizes there is substantial probability that \( \hat{\beta}_1 \) will be infinite. This behavior is due to the denominator in Equation (3.2), which may be very small or zero.

Call the square of the denominator of the right side of Equation (3.2), \( \text{den}_1^2 \). Then the graph of \( \text{den}_1^2 \) as a function of \( \hat{\beta}_2^2 \) is shown in Figure 1. Also write \( \hat{\theta}_0 = \hat{\beta}_0 / \text{den}_0 \), where \( \text{den}_0 = 1 = \text{den}_0^2 \).

When the sample size is moderate, the skewness of the distribution of \( \hat{\beta}_1 \) causes it to have larger mean squared error than \( \hat{\theta}_0 \) despite the bias of \( \hat{\theta}_0 \). The gap between the constant function and the linear function in Figure 1 suggests that using some intermediate function as \( \text{den}_1^2 \) might yield a better estimator than either \( \hat{\theta}_0 \) or \( \hat{\beta}_1 \) by reducing the bias of \( \hat{\theta}_0 \) while avoiding the skewness in the distribution of \( \hat{\beta}_1 \). A common smooth function which falls between the two lines in Figure 1 is the exponential function. This is the motivation for considering

\[
\hat{\theta}_2 = \frac{\hat{\theta}_0}{e^{-\gamma \hat{\beta}_2^2 \sigma^2_2}}. \tag{4.1}
\]

There is no claim of optimality for this particular choice of denominator; the aim here is simply to show improvement is possible at moderate sample sizes.

Burr (1985) derives formulas for the moments of \( \hat{\beta}_1 \) and \( \hat{\theta}_2 \). These formulas are based on the assumption that \( \hat{\theta}_0 \) has reached its asymptotic normal distribution, and they are accurate only for large sample sizes. The true MLE will improve on the modified MLE eventually, for large enough sample size. Computations of the moments showed that \( \hat{\beta}_1 \) has smaller mean squared error than \( \hat{\theta}_2 \) when the sample size is very large (800 - 1600) and \( \sigma^2_2 / \sigma^2_1 \) and \( \beta_1 \) are large.

5. A MONTE CARLO STUDY

Tables 1 and 2 summarize the results of a Monte Carlo study comparing the three estimators \( \hat{\theta}_0 \), \( \hat{\beta}_1 \) and \( \hat{\theta}_2 \) and the three methods of forming confidence intervals.

Doses had to be selected for the study. Here, the aim in dose selection was to compare the estimators under the most favorable, even if not realistic, conditions. Therefore, doses were selected according to the criterion of D-optimality (Abdelbasit and Plackett 1984). Two dose levels were used. They were selected to be symmetric about the mean of the tolerance distribution, and such as to maximize the determinant of the information matrix for \( \theta_k \) under Model 1R. Of course, this choice of doses would be impossible in practice since it depends on knowing the parameters which are
being estimated. In other studies not reported here, similar results were obtained when doses were selected to be more like those that occur in practice.

The model parameters in this study are \( \alpha_1 = -1.4, \beta_1 = 1.4, \) and \( \sigma_\varepsilon^2 = .26. \) The number of observations at each of the \( k = 2 \) doses is \( n = 80. \) Table 1 gives the following information about the three estimators of \( \alpha_1 \) and \( \beta_1: \) the number of finite estimates (\# finite); the bias and mean squared error (mse) of the finite estimates; the median of all the estimates minus the true parameter value (median bias); and the median absolute deviation from the true parameter value (madt) of all estimates.

Because the distribution of \( \hat{\alpha}_1 \) is not normal, bias and mean squared error are not appropriate measures of its behavior. Therefore the median bias and median absolute deviation from the true value are given here as alternate, more robust measures of the estimators' behavior. The choice of median absolute deviation from the true value seems arbitrary, and this measure is not as well established as mean squared error. The goal is to give as complete as possible a comparison of the two methods of estimation. One might, for instance, give 25\% and 75\% points of the distribution of the absolute deviation of the estimate from the true value. A study of behavior of confidence intervals for \( \alpha_1 \) and \( \beta_1 \) based on the three methods would give an even more illuminating comparison.

In Table 2, Method 0 uses the maximum likelihood estimator normal-theory intervals, assuming no error in \( x. \) The expected Fisher information is used to estimate the variance. Method 1 gives confidence intervals for \( \alpha_1 \) and \( \beta_1 \) by transforming the intervals for \( \alpha_R \) and \( \beta_R \) as described in Section 3. Method 2 intervals are formed by transforming the intervals for \( \alpha_R \) and \( \beta_R \) using the relationship in Equation (4.1), in the same way Method 1 intervals are formed. The claimed coverage probability for all the intervals is \( .95. \)

Table 2 gives the following information about the methods of forming confidence intervals: the number of confidence intervals of the form \( (a, b), \) with \( a \) and \( b \) finite (\# finite); the percentage of confidence intervals, either half-lines or finite, containing the true value (\% covering); the mean length of finite confidence intervals; and the median length of all confidence intervals.

This is a study of a situation with large measurement error variance, where the ratio of error variance to variance of the distribution of tolerances for the true dose is \( \sigma_\varepsilon^2/\sigma_1^2 = .5. \)

In this study, the large variance of \( \hat{\beta}_1 \) causes it to have greater mean squared error than \( \hat{\beta}_0, \) despite its smaller bias. It is interesting that \( \hat{\beta}_1 \) does better than \( \hat{\beta}_0 \) in terms of median absolute deviation from the true value. Method 0 confidence intervals have much lower coverage probability
than claimed. Method 1 intervals have a coverage probability which is better than claimed, but they are unsatisfactorily wide. The modified MLE $\hat{\theta}_2$ is seen to improve on the other two estimators in this situation. Not only does $\hat{\theta}_2$ have smaller mean squared error and median absolute deviation from the true value than the other two estimators, but Method 2 confidence intervals provide a useful middle ground between Method 0 and Method 1 intervals. Method 2 intervals have shorter expected length than Method 1 intervals and much greater coverage probability than Method 0 intervals.

Other studies not reported here indicated that it is probably not worthwhile to attempt to correct for measurement error when $\sigma^2_x/\sigma^2_1 < .33$, when $\beta_1 < .5$, or when the total sample size is less than 60. In such situations the naive MLE $\hat{\theta}_0$ did better than both the true MLE $\hat{\theta}_1$ and the modified MLE $\hat{\theta}_2$ in terms of mean squared error and median absolute deviation from the true value. When measurement error variance and slope parameter are large ($\sigma^2_x/\sigma^2_1 \geq .33, \beta_1 \geq 1.4$) and the total sample size is between 80 and 600, it is reasonable to estimate $\hat{\theta}_1$ using the modified MLE. In Monte Carlo studies of such situations, $\hat{\theta}_2$ did as well or better than $\hat{\theta}_0$, and Method 2 confidence intervals has much closer to the assumed coverage probability than Method 0 intervals.

A FORTRAN program to carry out these studies was developed and run on a VAX 11/780; the program was compiled using the f77 compiler. The data were simulated using the IMSL linear congruential random number generator GGUBS and normal deviate generator GGNML. The number of simulations in the study reported here was 1000. All pairwise differences between mean-squared errors of the estimators of $\alpha$ and between mean-squared errors of the estimators of $\beta$ in Table 1 are significantly different from zero with $p < .001$ by the signed rank test.
REFERENCES


Table 1. Distribution characteristics of the three estimators, from a Monte Carlo study with 1000 simulations, and $\alpha_1 = -1.4, \beta_1 = 1.4, \sigma^2 = .26, \sigma^2/\sigma^2 = .50, k = 2, x_1 = 0.0$ and $x_2 = 2.0, n = 80.$

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\alpha}_0$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_2$</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td># finite</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>bias</td>
<td>.23(.006)</td>
<td>-.08(.011)</td>
<td>-.00(.009)</td>
<td>-.24(.004)</td>
<td>.07(.009)</td>
<td>-.00(.007)</td>
</tr>
<tr>
<td>mse</td>
<td>.09</td>
<td>.12</td>
<td>.08</td>
<td>.08</td>
<td>.09</td>
<td>.05</td>
</tr>
<tr>
<td>median bias</td>
<td>.25</td>
<td>-.04</td>
<td>.02</td>
<td>-.24</td>
<td>.04</td>
<td>-.02</td>
</tr>
<tr>
<td>madt</td>
<td>.25</td>
<td>.20</td>
<td>.18</td>
<td>.24</td>
<td>.17</td>
<td>.14</td>
</tr>
</tbody>
</table>

Note: The values in parentheses are standard errors.

Table 2. 95% confidence intervals by the three methods, from a Monte Carlo study with 1000 simulations, and $\alpha_1 = -1.4, \beta_1 = 1.4, \sigma^2 = .26, \sigma^2/\sigma^2 = .50, k = 2, x_1 = 0.0$ and $x_2 = 2.0, n = 80.$

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Method O</td>
<td>Method 1</td>
</tr>
<tr>
<td># finite</td>
<td>1000</td>
<td>998</td>
</tr>
<tr>
<td>% covering</td>
<td>69</td>
<td>97</td>
</tr>
<tr>
<td>mean length</td>
<td>.71</td>
<td>1.52</td>
</tr>
<tr>
<td>median length</td>
<td>.70</td>
<td>1.32</td>
</tr>
</tbody>
</table>
Figure 1. Graphs of $\text{den}_i^2$ as functions of $\hat{\beta}_0^2$ for $i = 0, 1, 2$, when $\sigma_i^2 = 1.0$. 