CORRECTING DATA FOR MEASUREMENT ERROR IN GENERALIZED LINEAR MODELS

Leonard A. Stefanski*

Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

Key Words and Phrases: errors-in-variables; functional models; generalized linear model; M-estimation; measurement error; structural models.

ABSTRACT

This paper discussed a general strategy for reducing measurement-error-induced bias in statistical models. It is assumed that the measurement error is unbiased with a known variance although no other distributional assumptions on the measurement-error distribution are employed.

Using a preliminary fit of the model to the observed data, a transformation of the variable measured with error is estimated. The transformation is constructed so that the estimates obtained by refitting the model to the "corrected" data yields estimates with smaller bias.

Whereas the general strategy can be applied in a wide variety of settings, this paper focuses on the problem of covariate measurement error in generalized linear models. Two estimators are derived and their effectiveness at reducing bias is demonstrated in a Monte Carlo study.

* Research supported by the National Science Foundation.

1. INTRODUCTION

In this paper a method of reducing measurement-error-induced bias in M-estimators is studied with emphasis on covariate measurement error in generalized linear models. It is assumed that given independent observations (Y_i, U_i) $(i=1,\ldots,n)$, the parameter of interest, Θ_0, is consistently estimated by the M-estimator Θ^* satisfying

$$\sum_{i=1}^{n} \psi(Y_i, U_i, \Theta^*) = 0, \tag{1.1}$$

where Θ_0 is uniquely determined by the requirement that

$$E\{\psi(Y_i, U_i, \Theta_0)\} = 0 , \quad (i=1,\ldots,n) .$$

Observable data consists of the pairs (Y_i, X_i) where

$$X_i = U_i + \sigma Z_i , \quad E(Z_i) = 0 , \quad E(Z_i Z_i^T) = \Omega ; \tag{1.2}$$

and conditional on U_i, Y_i and X_i are stochastically independent.

Unless $\psi(y,u,\Theta)$ is a linear function of u, the naive estimator of Θ_0, $\hat{\Theta}$, obtained by solving (1.1) with X_i in place of U_i, is generally inconsistent with an asymptotic bias which is $O(\sigma^2)$ under (1.2), (Stefanski, 1985). This paper outlines a strategy for reducing this asymptotic bias to $o(\sigma^2)$ under the assumption that σ^2 and Ω are known. Without additional assumptions on the measurement error distribution this is as much reduction in bias as can be expected except when $\psi(y,u,\Theta)$ is a quadratic function of u. Although an assumption of normality can always be defended on general principles, the two correlates of this assumption in linear measurement-error models, convenience and consistency (when the assumption of normality is correct and often when not), frequently are not obtained in nonlinear measurement-error models (Stefanski & Carroll, 1985; Carroll *et al*, 1984). Hence the motivation for (1.2) in place of the more restrictive and often unverifiable assumption of normality.

The strategy suggested in this paper can be described very simply. Let $C_i = C(Y_i, X_i, \Theta, \sigma)$ be a function of Y_i, X_i, Θ and σ. Then if Θ_C^* is a solution to the equations

$$\sum_{i=1}^{n} \Psi(Y_i, X_i + \sigma^2 C_i, \Theta) = 0$$

it would be expected that Θ_C^* converges in probability to some $\Theta = \Theta(\sigma)$ satisfying

$$\lim_n (1/n) \sum_{i=1}^{n} \Psi\{Y_i, X_i + \sigma^2 C_i, \Theta(\sigma)\} = 0 . \tag{1.3}$$

The idea is to determine choices of C_i which insure that $\Theta(\sigma) = \Theta_o + o(\sigma^2)$.

The formal implementation of this strategy does not depend on whether $U_1, \ldots, U_n$ is regarded as a fixed sequence of unknown constants, i.e., a functional model or as a sequence of independent and identically distributed random variables, i.e., a structural model, provided the Cesaro sum in (1.3) is sufficiently well behaved for large n. In the structural version of the model (1.3) would be replaced by $E\{\Psi(Y, X+\sigma^2 C, \Theta(\sigma))\}$ and the type of regularity conditions required are those that allow for the interchange of expectation and differentiation with respect to σ. However, justification for the proposed methods by way of rigorous mathematical analysis is not given; instead the strategy is defended by showing that the estimators so derived are reasonable, and in some cases identical to estimators proposed previously in the literature. Largely for convenience, attention is focused on generalized linear models.

In Section 2 a necessary condition for second-order unbiasedness ($\Theta(\sigma) = \Theta_o + o(\sigma^2)$) is derived and two estimators are proposed corresponding to two choices of C_i, satisfying this condition. In the course of deriving these estimators evidence is provided which indicates that functional maximum likelihood estimation generally fails to produce consistent estimators of regression parameters in nonlinear generalized linear model. In Section 3 Monte Carlo evidence is presented demonstrating the effectiveness of these estimators at reducing bias. For the simulation study a class of nonlinear models is identified which permit easy computation of bias terms.

The generalized linear models studied are specified by their mean function $E(Y|U) = \mu(U^T\Theta_o)$ and variance function $\mathrm{var}(Y|U) = \tau^2 v(U^T\Theta_o)$. The estimating equation $\psi(Y,U,\Theta)$ will be written

$$\psi(Y,U,\Theta) = Q_Y(U^T\Theta)U \text{ where } Q_Y(t) = \{Y-\mu(t)\}w(t) \qquad (1.4)$$

with $w(t) = \mu'(t)/v(t)$. With this notation the following identities are obtained for $t = U^T\Theta_o$

$$E\{Q_Y^2(t)\} = \tau^2\mu'w, \; E\{Q_Y'(t)\} = -\mu'w$$

$$E\{Q_Y''(t)\} = -\mu''w - 2\mu'w', \; E(Q_Y(t)Q_Y'(t)\}^2 = \tau\,\mu'w' \quad . \qquad (1.5)$$

Under appropriate regularity conditions the estimator Θ^* satisfying

$$\sum_{i=1}^{n} Q_{Y_i}(U_i^T\Theta^*)U_i = 0$$

is consistent for Θ_o and $n^{1/2}(\Theta^*-\Theta_o)$ has a limiting normal distribution with mean zero and covariance matrix $[\lim_n(\tau^2/n)\Sigma(\mu'(U_i^T\Theta_o)w(U_i^T\Theta_o)U_iU_i^T\}]^{-1}$.

2. DATA CORRECTIONS

2.1 Second-Order Unbiasedness

For ease of presentation, (1.3) will be written as

$$E[\psi\{Y,U + \sigma Z + \sigma^2 C, \Theta(\sigma)\}] = 0 \qquad (2.1)$$

for both the functional and structural versions of the model. It is assumed that $\psi(\cdot,\cdot,\cdot)$, C, the empirical distribution of the U_i's and the measurement error distribution are sufficiently well behaved so that (2.1) defines $\Theta(\sigma)$ implicitly as a three-times differentiable function of σ. Although for most models $\Theta(\sigma)$ cannot be obtained explicitly it is possible to develop $\Theta(\sigma)$ in a Maclaurin series; note that $\Theta(0) = \Theta_0$.

Now let $X_C = U + \sigma Z + \sigma^2 C$ and $\Theta = \Theta(\sigma)$. When $\psi(\cdot,\cdot,\cdot)$ has the form in (1.4) one differentiation with respect to σ results in the equation

$$0 = E\{Q_Y'(X_C^T\Theta)(X_C'^T\Theta + X_C^T\Theta')X_C + Q_Y(X_C^T\Theta)X_C'\} . \qquad (2.2)$$

Noting that $X_C(0) = U$ and $X_C'(0) = Z$, evaluating (2.2) at $\sigma = 0$ results in the equation

$$0 = E[Q_Y'(U^T\Theta_0)\{Z^T\Theta_0 + U^T\Theta'(0)\}U + Q_Y(U^T\Theta_0)Z] .$$

The unbiasedness condition (1.2) and the negative definiteness of $E\{Q_Y'(U^T\Theta_0)UU^T\}$ together imply that $\Theta'(0) = 0$. Upon taking two derivatives of (2.1) with respect to σ and evaluating the resulting expression at $\sigma = 0$ leads to the equation

$$-E\{Q_Y'(U^T\Theta_0)UU^T\}\Theta_0'' = E\{Q_Y''(U^T\Theta_0)U\Theta_0^T\Omega\Theta_0 + 2Q_Y'(U^T\Theta_0)\Omega\Theta_0$$
$$+ 2Q_Y'(U^T\Theta_0)U\,C^T\Theta_0 + 2Q_Y(U^T\Theta_0)C\} .$$

So $\Theta''(0) = 0$ whenever

$$0 = E\{Q_Y''(U^T\Theta_0)U\Theta_0^T\Omega\Theta_0 + 2Q_Y'(U^T\Theta_0)\Omega\Theta_0$$
$$+ 2Q_Y'(U^T\Theta_0)UC^T\Theta_0 + 2Q_Y(U^T\Theta_0)C\} . \qquad (2.3)$$

Thus (2.3) can be viewed as a necessary condition for second-order unbiasedness. In the next two sections (2.3) is used to construct two second-order unbiased estimators. This section closes by showing how it can be used to prove, at least formally, that functional maximum likelihood does not always yield consistent estimators in nonlinear generalized linear models.

Consider simple logistic regression through the origin with normal measurement error, i.e., $pr(Y_i = 1|U_i) = F(\Theta U_i)$, $F(t) = 1/\{1+\exp(-t)\}$, $X_i = U_i + \sigma Z_i$, $Z_i \sim N(0,1)$, σ known. Then functional maximum likelihood leads to the equations

$$\sum_{i=1}^{n} \{Y_i - F(\hat{U}_i\hat{\Theta})\}\hat{U}_i = 0 \ ;$$

$$\hat{U}_i = X_i + \sigma^2\hat{\Theta}\{Y_i - F(\hat{U}_i\hat{\Theta})\} \ , \qquad (i=1,\ldots,n) \ .$$

Now if $\hat{\Theta}$ converges at all it must converge to some $\Theta(\sigma)$ satisfying

$$E([Y - F\{X_C\Theta(\sigma)\}]X_C) = 0$$

where $X_C = X + \sigma^2\Theta(\sigma)[Y - F\{X_C\Theta(\sigma)\}]$. This has the form (2.1) with $C = \Theta(\sigma)[Y - F\{X_C\Theta(\sigma)\}]$ and at $\sigma = 0$, $C = C_o = \Theta_o\{Y - F(U\Theta_o)\}$. Now plugging C_o into the right hand side of (2.3), the expression reduces to $-E\{F''(U\Theta_o)U\Theta_o^2\}$ which is generally not equal to zero. Consequently the functional maximum likelihood estimator in a logistic regression errors-in-variables model has a bias which is $O(\sigma^2)$ under most conditions.

2.2 An Estimator With Reduced Bias

Consider (2.1) when $C = MX$ for some pxp matrix M. Then (2.3) holds provided

$$M^T = -(1/2)\left[E\{Q_Y'(U^T\Theta_o)UU^T\}\right]^{-1} E\{Q_Y''(U^T\Theta_o)U\Theta_o^T\Omega + 2Q_Y'(U^T\Theta_o)\Omega\} \ .$$

This is estimated to within $O(\sigma^2)$ by

$$\hat{M}^T = (-1/2)\left[\sum_1^n \{Q_{Y_i}'(X_i^T\hat{\Theta})X_iX_i^T\}\right]^{-1} \sum_1^n \{Q_{Y_i}''(X_i^T\hat{\Theta})\ X_i\hat{\Theta}^T\Omega + 2Q_{Y_i}'(X_i^T\hat{\Theta})\Omega\} \quad (2.4)$$

suggesting the following strategy. Ignoring the measurement error fit the model to the observed data to obtain $\hat{\theta}$. Compute $\hat{M}$ according to (2.4) and $\hat{X}_{i,C} = X_i + \sigma^2\hat{M}X_i$, $i=1,\ldots,n$. Refit the model using the corrected data set $(Y_i, \hat{X}_{i,C})$, $i=1,\ldots,n$, to obtain $\hat{\theta}_{C,1}$. Under reasonable regularity conditions $\hat{\theta}_{C,1}$ will have an asymptotic bias which is $o(\sigma^2)$. In a generalized linear model, it is not necessary to refit the model to obtain $\hat{\theta}_{C,1}$. Since $\hat{X}_{i,C} = (I + \sigma^2\hat{M})X_i$ invariance considerations show that

$$\hat{\theta}_{C,1} = (I + \sigma^2 \hat{M}^T)^{-1} \hat{\theta} , \tag{2.5}$$

where in (2.5) it is assumed that $I + \sigma^2\hat{M}^T$ is nonsingular.

In an earlier paper (Stefanski, 1985) the author proposed a second-order unbiased estimator obtained by subtracting an estimate of the bias from the naive estimator, a strategy which has since been refined as a recent paper by Whittemore & Keller (1987). Using current notation the estimator proposed in Stefanski (1985) is given by

$$\tilde{\theta} = (I - \sigma^2 \hat{M}^T)\hat{\theta} . \tag{2.6}$$

That both $\hat{\theta}_{C,1}$ and $\tilde{\theta}$ converge to limits which differ by no more than $o(\sigma^2)$ follows from the fact that $(I + \sigma^2\hat{M}^T)^{-1} = I - \sigma^2\hat{M}^T + o(\sigma^2)$.

When $\hat{M}^T$ is negative definite and $I + \sigma^2\hat{M}^T$ is positive definite $(I + \sigma^2\hat{M}^T)^{-1} > I - \sigma^2\hat{M}^T$ in the sense of positive definiteness. This means that the correction for attenuation in $\hat{\theta}_{C,1}$ is greater than that in $\tilde{\theta}$. Thus it would be expected that $\hat{\theta}_{C,1}$ will be less biased although more variable than $\tilde{\theta}$. This occurs, for example, in the linear model. In this case $\hat{\theta}$ converges to $(\Omega_U + \sigma^2\Omega)^{-1}\Omega_U\theta_o$ where $\Omega_U = E(UU^T)$. For this model $\hat{\theta}_{C,1} = (I - \sigma^2 S_{XX}^{-1}\Omega)^{-1}\hat{\theta}$ where $S_{XX} = n^{-1}\Sigma X_i X_i^T$. Since S_{XX}

converges to $\Omega_U + \sigma^2\Omega$, $\hat{\Theta}_{C,1}$ is consistent for Θ_o at the linear model. The estimator $\tilde{\Theta}$ is not consistent at the linear model. The simulations reported in Section 3 also support the claim that $\hat{\Theta}_{C,1}$ corrects for the effects of measurement error more so than $\tilde{\Theta}$.

2.3 A second estimator with reduced bias

Now consider (2.3) when C is given by $a_1\{Q_Y(U^T\Theta_o) - a_2\}\Omega\Theta_o$ for scalars a_1 and a_2 to be determined. Using the representation $Q_Y(t) = \{Y - \mu(t)\}w(t)$ and the relationships given in (1.5), (2.3) reduces to

$$0 = E\Big[\Theta_o^T\Omega\Theta_o\{2a_1(\tau^2\mu'w' + a_2\mu'w) - \mu''w - 2\mu'w'\}U + 2(\tau^2a_1\mu'w - \mu'w)\Omega\Theta_o\Big] \qquad (2.7)$$

where μ, w, μ', w', etc. are all evaluated at $U^T\Theta_o$. This equality holds true provided

$$\tau^2a_1 = 1,$$

and

$$2a_1(\tau^2\mu'w' + a_2\mu\, w) - \mu''w - 2\mu'w' = 0\ ,$$

which when solved yield $a_1 = 1/\tau^2$ and $a_2 = \tau^2\mu''/2\mu'$.

This suggests a second strategy for estimating Θ_o. From a preliminary fit to the observed data obtain $\hat{\Theta}$ and $\hat{\tau}^2$ and construct $\hat{X}_{i,C} = X_i + (\sigma^2/\hat{\tau}^2)\{Q_{Y_i}(X_i^T\hat{\Theta}) - \hat{\tau}^2\mu''(X_i^T\hat{\Theta})/2\mu'(X_i^T\hat{\Theta})\}\Omega\hat{\Theta}$.
Refit the model to the data $(Y_i,\hat{X}_{i,C})$, i=1,...,n, to obtain $\hat{\Theta}_{C,2}$. Under reasonable regularity conditions this estimator will possess a bias which is $o(\sigma^2)$.

Consider $\hat{\Theta}_{C,2}$ in the case of logistic regression. Here $Q_Y(t) = Y - F(t)$, $F(t) = 1/\{1 + \exp(-t)\}$, $\mu(t) = F(t)$ and $\tau^2 = 1$.

Since $F' = F - F^2$, $\mu' = F'$ and $\mu'' = F'(1-2F)$, $\mu''/2\mu' = 1/2 - F$ and X_C is given by

$$X_C = \sigma^2\{Y - 1/2\}\Omega\theta \ .$$

When θ is held fixed and Z is normally distributed it is shown in Stefanski & Carroll (1985, 1987) that $X_C = \sigma^2\{Y - 1/2\}\Omega\theta$ is a sufficient statistic for U and furthermore that

$$pr(Y=1|X_C) = F(X_C^T\theta) \ ,$$

i.e., conditioned on X_C, Y follows the logistic model with mean $F(X_C^T\theta)$. In Stefanski & Carroll (1987) it is shown that the score $\psi(Y,X,\theta) = \{Y - F(X_C^T\theta)\}X_C$ is efficient over a class of structural models for logistic regression with normally distributed measurement errors. The two-stage estimator for this model obtained by fitting the logistic model to $(Y_i, \hat{X}_{i,C} = X_i + \sigma^2(Y_i - 1/2)\Omega\hat{\theta})$ $i = 1,\ldots,n$, was investigated in Stefanski & Carroll (1985) and proved to significantly reduce bias in a simulation study.

Suppose now that Y|U has an exponential family density of the form

$$f_{Y|U}(y|u) = \exp\{ym(U^T\theta_o) - b\{m(U^T\theta_o)\} + d(y)\}$$

and that X|U is normally distributed with mean U and known covariance matrix $\sigma^2\Omega$. In this set up the functional maximum likelihood estimator of θ satisfies

$$\sum_{i=1}^{n} Q_{Y_i}(\hat{U}_i^T\hat{\theta}_m)\hat{U}_i = 0 \ ;$$

where

$$\hat{U}_i = X_i + \sigma^2 Q_{Y_i}(\hat{U}_i^T \hat{\theta}_m)\Omega\hat{\theta}_m \ , \qquad i=1,\ldots,n$$

and $Q_y(t) = [y - b\{m(t)\}]m'(t)$.

Apart from the two-stage feature of $\hat{\theta}_{C,2}$ the main difference between $\hat{\theta}_{C,2}$ and the functional maximum likelihood estimator lies the manner in which X_i is "corrected." Where $\hat{U}_i$ contains the term $Q_Y(\hat{U}_i^T\hat{\theta}_m)$, $\hat{X}_{i,C}$ contains the term

$$Q_Y(X_i^T\hat{\theta}) - \mu''(X_i^T\hat{\theta})/2\mu'(X_i^T\hat{\theta}).$$

With respect to bias this means that $\hat{\theta}_m$ and $\hat{\theta}_{C,2}$ will have the same second-order bias only if $\mu'' = 0$, or equivalently, when $\mu(t)$ is linear. Since $\hat{\theta}_{C,2}$ is constructed to have no second order bias, it follows that $\hat{\theta}_m$ is generally second-order biased except when $\mu(t)$ is linear.

A problem arises with $\hat{\theta}_{C,2}$ if the variation of Y around its mean is small, i.e., when τ^2 is small. Formally this can be seen by noting that it is not generally possible to solve (2.7) for a_1 and a_2 when $\tau^2 = 0$. The validity of the small-measurement error correction in $\hat{\theta}_{C,2}$ is more dependent on the ratio of the measurement error variance to the equation error variance whereas the validity of the correction in $\hat{\theta}_{C,1}$ depends more on the ratio of measurement-error variance to true predictor variation.

3. SIMULATION RESULTS

The study of linear measurement-error models is facilitated to a great extent by the ability to obtain closed-form expressions for bias terms. It is not generally possible to do the same in nonlinear measurement-error models and hence the motivation for the approximations derived in this paper and those employed by other investigators (Wolter & Fuller, 1982; Amemiya, 1982; Stefanski and Carroll, 1985; Armstrong, 1985 ; Stefanski, 1985 and Keller and Whittemore, 1987). In the simulations conducted for this paper three models were employed for which there exist simple and familiar expressions for the bias of the naive estimator. This provides a basis for comparing the corrected estimators and also allows them to be compared to a standard method-of-moments correction-for-attenuation estimator. The models are described in

Section 3.1 with derivations in the Appendix. Details of the simulations are given in Section 3.2 and the results are discussed in Section 3.3.

3.1 Some Nonlinear Measurement-Error Models with Tractable Bias Terms

Consider a regression model for which $E(Y|U) = \exp(\alpha_o + U^T\beta_o)$ and the conditional variance of $Y|U$ is $\mathrm{var}(Y|U) = \tau^2 \exp(\lambda\alpha_o + \lambda U^T\beta_o)$ for fixed constants τ^2 and λ. For this model the appropriate score function is

$$\psi(Y,U,\alpha,\beta) = (Y - e^{\alpha+U^T\beta})e^{(\alpha+U^T\beta)(1-\lambda)}\begin{bmatrix}1\\U\end{bmatrix} . \qquad (3.1)$$

Suppose that $X = U + Z$ with $E(Z) = 0$ and that conditioned on U, Y and X are independent. In the Appendix it is shown that when U and Z are independent normal vectors the solution, $(\alpha,\beta^T)^T$, to the system of equations

$$0 = E\{(Y - e^{\alpha+X^T\beta})e^{(\alpha+X^T\beta)(1-\lambda)}\begin{bmatrix}1\\X\end{bmatrix}\}$$

is given by

$$\beta = (\Omega_U + \Omega_Z)^{-1}\Omega_U\beta_o ,$$

$$\alpha = \alpha_o + \mu_U^T(\Omega_U + \Omega_Z)^{-1}\Omega_Z\beta_o + \beta_o^T\Omega_U(\Omega_U + \Omega_Z)^{-1}\Omega_Z\beta_o/2$$

where Ω_U and Ω_Z are the covariance matrices of U and Z respectively. Since α and β are the limits of the naive estimators $\hat{\alpha}$ and $\hat{\beta}$, this result indicates that the method-of-moments estimator

$$\hat{\beta}_{C,A} = (I - S_{XX}^{-1}\Omega_Z)^{-1}\hat{\beta}$$

is a consistent estimator of β_o under the assumed model with a corresponding correction for $\hat{\alpha}$ also possible.

For the simulations in this paper three univariate versions of this model were studied. In the first $Y|U$ is exponentially distributed with mean $\exp(\alpha+\beta U)$. In this model the likelihood score is given by (3.1) with $\lambda = 2$. In the second model $Y|U$ has a Poisson distribution with mean $\exp(\alpha + \beta U)$. The likelihood score for this model is given by (3.1) with $\lambda = 1$. In the third, $Y|U$ is normally distributed with mean $\exp(\alpha+\beta U)$ and constant variance τ^2.

3.2 Details of the Simulation

In the exponential and Poisson models the parameters were set at $\alpha = 0$ and $\beta = 1/2$. Five hundred simulated data sets were generated for each of three sample sizes n=50, 100, 200. In both models $U \sim N(0,1)$ and $Z \sim N(0,1/4)$. The normal random variates were generated using the normal random number generator supplied with GAUSS (1986). Exponential and Poisson variates were generated using GAUSS's uniform random number generator in conjunction with standard algorithms for producing exponential and Poisson variates. Further details are available from the author upon request.

Eight estimators were selected for study. These include the estimators derived previously, the second-order unbiased estimator proposed in Stefanski (1985) and certain modifications to these estimators as described below. The eight estimators are: 1) $\hat{\theta}$, the naive estimator; 2) $\hat{\theta}_{C,A}$, method-of-moments correction-for-attenuation estimator described in Section 3.1; 3-5) designated $\hat{\theta}_{C,1}(0)$, $\hat{\theta}_{C,1}(2)$ and $\hat{\theta}_{C,1}(6)$ respectively, $\hat{\theta}_{C,1}(0)$ is the estimator derived in Section 2.2, $\hat{\theta}_{C,1}(2)$ and $\hat{\theta}_{C,1}(6)$ are modifications obtained by replacing $\hat{M}^T$ in the definition of $\hat{\theta}_{C,1}(0)$ by $\hat{M}^T(n-\alpha)/n$, $\alpha=2,6$ respectively; 6) $\hat{\theta}_{C,2}$, the estimator derived in Section 2.3; 7-8) designated $\tilde{\theta}$ and $\tilde{\theta}_1$ respectively, $\tilde{\theta}$ is the estimator proposed in Stefanski (1985) and which appears in eq. (2.6), $\tilde{\theta}_1 = \tilde{\theta} + \sigma^4\hat{M}^T\hat{M}^T\hat{\theta}$.

The modifications to $\hat{\theta}_{C,1}(0)$ are suggested by the work of Fuller (1980). The modification to $\tilde{\theta}$ was suggested by comparing

(2.5) and (2.6); whereas $\tilde{\Theta}$ utilizes two terms in the expansion of $(I + \sigma^2\hat{M}^T)^{-1}$, $\tilde{\Theta}_1$ employs the first three terms in this expansion.

Tables 1 & 2 contain biases, standard deviations and mean-squared errors for the slope component of the various estimators in the exponential and Poisson models respectively. The fourth column in the tables report the frequency with which each estimator's absolute error was smaller than that of the naive estimator. For example, in Table 1 when n=50, $|\hat{\beta}_{C,2} - 1/2|$ was less than $|\hat{\beta} - 1/2|$ in 64.2% of the data sets generated. This frequency is (an estimator of) the relative performance measure $P(\hat{\beta}^*,\hat{\beta}) = pr\{|\hat{\beta}^* - 1/2| < |\hat{\beta} - 1/2|\}$ and is related to the concept of Pitman-closeness (Mood, Graybill & Boes, 1974, p. 290).

3.3 Discussion of the Simulation Results

Formal tests of significance were carried out comparing each of the estimators' mean squared errors to that of the naive estimator using paired difference tests. The tests were generally inconclusive for the smaller sample size, but without exception indicated a reduction in mean squared error for the two larger sample sizes. The number of simulations was not large enough to detect differences between the various corrected estimators on a case-by-case basis. However, the trends across sample sizes and models in Tables 1 and 2 and in additional simulations not reported here suggest some additional conclusions.

With respect to mean square error $\hat{\beta}_{C,2}$ performed best, followed closely by $\tilde{\beta}$. These estimators achieve this superiority at the expense of maintaining the largest negative bias, i.e., their corrections for attenuation are smallest. As suggested in Section 2.3 $\hat{\beta}_{C,1}(0)$ tends to overcompensate for the effects of measurement error and consistently has the largest positive bias. The modified versions of this estimator are more nearly unbiased as is the modified version of $\tilde{\beta}$. When an estimator is used as a pivotal quantity in a confidence interval the validity of the stated confidence levels depends critically on the bias/(standard deviation) ratio. Thus any of the modified estimators would be preferred in such an application although this recommendation assumes that reasonably good estimates of standard errors can be obtained, say by the delta method.

All of the nonzero proportions in column 4 of Tables 1 and 2 are significantly greater than 1/2. Thus each of the corrected estimators is Pitman-closer to its estimand than is the naive estimator.

Although the method-of-moments estimator $\hat{\beta}_{C,A}$, depends explicitly on the joint normality of the observed and true predictor, it did not seem to enjoy any consistent advantage over the other estimators whose derivations do not formally depend on normality. The mean squared error of $\hat{\beta}_{C,A}$ was consistently dominated by that of $\hat{\beta}_{C,2}$.

In both the exponential and Poisson models there is substantial variation in the conditional distribution of $Y|U$ at least for some U. Consequently the problems with $\hat{\theta}_{C,2}$ mentioned in Section 2.3 are not manifest in these models. To complement the Poisson and exponential simulations a third model was studied in which $(Y|U)$ is normally distributed with mean = $\alpha+\beta U$, $\alpha = 0$, $\beta = 1/2$ and variance τ^2. The measurement error was also normal with zero mean and variance = 1/16. For the three values of $\tau^2 = 1/4$, 1/16, 1/64 and sample size n = 25, five hundred simulated data sets were generated. The results from this experiment appear in Table 3.

The most interesting trend in this table is the increase in bias of $\hat{\beta}_{C,2}$ as τ^2 decreases while biases in the other estimators remain relatively constant, confirming the problems mentioned earlier. With samples this small the reduction in bias is generally offset by an increase in variability. Thus with respect to mean squared error there are no significant advantages to correcting for measurement error. However, with the exception of $\hat{\beta}_{C,2}(\tau = .125)$ all of the corrected estimators are Pitman-closer to β_0 than is the naive estimator, although in some cases not by much.

4. CONCLUDING REMARKS

Commonly in statistics the fitting of a complex model to data is avoided by transforming the data so they follow a simpler more familiar model. The benefits derived by doing so are well known, for example, the ability to use existing software for the fitting and diagnosing of the model under study. This paper is a first attempt at applying this principle to data analysis in the presence of measurement error. Its success at reducing measurement-error induced bias is apparent from the simulation study.

An issue not addressed in this paper is the estimation of standard errors for the bias adjusted estimators. A large-sample small-measurement-error asymptotic distribution theory such as that employed by Wolter and Fuller (1982) and Stefanski and Carroll (1985) suggests that the estimation of standard errors is relatively insensitive to the measurement error and that the usual formulae for asymptotic variances can be employed. However, experience indicates that this is not always acceptable. As all of the bias-adjusted estimators are modified M-estimators or what might be called pseudo M-estimators, a routine though tedious application of the delta method yields standard large-sample (but not necessarily small-measurement-error) estimators of standard errors; see, for example, Stefanski (1985). This approach should be sufficient in most cases since adjustment for bias is not recommended in small samples.

APPENDIX

Let U, Z be independent random vectors such that $U \sim N(\mu_U, \Omega_U)$ and $Z \sim N(0, \Omega_Z)$ and let $X = U + Z$. Then $X|U \sim N(U, \Omega_Z)$. Assume that Ω_U is positive definite, that Y and X are conditionally independent given U with $E(Y|U) = \exp(\alpha_o + U^T\beta_o)$ and let α and β be defined by the equations

$$0 = E\{(Y - e^{\alpha+X^T\beta})e^{\lambda\alpha+\lambda X^T\beta}\begin{bmatrix}1\\X\end{bmatrix}\} \qquad (A.1)$$

for some fixed constant λ. Using the facts that for $W \sim N(\mu, \Omega)$,

$$E(e^{W^T\theta}) = e^{\mu^T\theta + \theta^T\Omega\theta/2}$$

and

$$E(We^{W^T\theta}) = (\mu + \Omega\theta)e^{\mu^T\theta + \theta^T\Omega\theta/2} \quad ,$$

(A.1) can be shown to be equivalent to

$$\begin{Bmatrix} 1 \\ \mu_U + \Omega_U\beta_o + \lambda(\Omega_U + \Omega_Z)\beta \end{Bmatrix} \exp\left[\alpha_o + \mu_U^T\beta_o + \beta_o^T\Omega_U\beta_o/2 + \lambda\beta_o^T\Omega_U\beta\right]$$

$$= \begin{Bmatrix} 1 \\ \mu_U + (\lambda+1)(\Omega_U + \Omega_Z)\beta \end{Bmatrix} \exp\left[\alpha + \mu_U^T\beta + (2\lambda+1)\beta^T(\Omega_U + \Omega_Z)\beta/2\right] . \qquad (A.2)$$

The exponents in (A.2) must be equal which in turn implies that

$$\mu_U + \Omega_U\beta_o + \lambda(\Omega_U + \Omega_Z)\beta = \mu_U + (\lambda+1)(\Omega_U + \Omega_Z)\beta \ ,$$

from which it follows that $\beta = (\Omega_U + \Omega_Z)^{-1}\Omega_U\beta_o$, and

$$\alpha = \alpha_o + \mu_U^T(\Omega_U + \Omega_Z)^{-1}\Omega_Z\beta_o + \beta_o^T\Omega_U(\Omega_U + \Omega_Z)^{-1}\Omega_Z\beta_o/2 \quad .$$

<u>BIBLIOGRAPHY</u>

Amemiya, Y., (1982). Estimators for the errors-in-variables model. Unpublished Ph.D. Thesis, Iowa State University, Ames.

Armstrong, B., (1985). Measurement error in the generalized linear model. *Comm. Statist. B* **14**, 529-44.

Carroll, R. J., Spiegelman, C. H., Lan, K. K., Bailey, K. T., & Abbott, R. D., (1984). On errors-in-variables for binary models. *Biometrika* **71**, 19-25.

Fuller, W. A., (1980). Properties of some estimators for the errors-in-variables model. *Ann. Statist.* **8**, 407-22.

Gauss Version 1-49B (1986). Aptech Systems, Inc., Kent, WA.

Mood, A. M., Graybill, F. A. & Boes, D. C., (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, New York.

Stefanski, L. A., (1985). The effects of measurement error on parameter estimation. *Biometrika* **72**, 583-92.

Stefanski, L. A. & Carroll, R. J., (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13**, 1335-51.

Stefanski, L. A. & Carroll, R. J., (1987). Conditional scores and optimal scores for generated linear measurement-error models. *Biometrika* **74**, 703-16.

Whittemore, A. S. & Keller, J. B., (1987). Approximations for errors-in-variables regression. Stanford University Technical Report.

Wolter, J. M. & Fuller, W. A., (1982). Estimation of nonlinear errors-in-variables models. *Ann. Statist.* **10**, 539-48.

TABLE I

Performance Measures for Estimators of β, Exponential Regression

		Bias	Standard Deviation	Mean Squared Error	$P(\cdot,\hat{\beta})$
n=50	$\hat{\beta}$	-0.09770	0.14483	0.03052	0.000
	$\hat{\beta}_{C,A}$	0.01026	0.18795	0.03543	0.600
	$\hat{\beta}_{C,1}$ (0)	0.02306	0.19718	0.03941	0.600
	$\hat{\beta}_{C,1}$ (2)	0.01666	0.19397	0.03790	0.604
	$\hat{\beta}_{C,1}$ (6)	0.00438	0.18798	0.03536	0.620
	$\hat{\beta}_{C,2}$	-0.01912	0.16928	0.02902	0.642
	$\tilde{\beta}$	-0.00703	0.18107	0.03284	0.630
	$\tilde{\beta}_1$	0.01499	0.19197	0.03708	0.604
n=100	$\hat{\beta}$	-0.10491	0.09578	0.02018	0.000
	$\hat{\beta}_{C,A}$	-0.00220	0.12236	0.01498	0.684
	$\hat{\beta}_{C,1}$ (0)	0.00431	0.12765	0.01631	0.664
	$\hat{\beta}_{C,1}$ (2)	0.00149	0.12669	0.01605	0.674
	$\hat{\beta}_{C,1}$ (6)	-0.00404	0.12483	0.01560	0.698
	$\hat{\beta}_{C,2}$	-0.02687	0.11250	0.01338	0.746
	$\tilde{\beta}$	-0.02039	0.11851	0.01446	0.730
	$\tilde{\beta}_1$	-0.00153	0.12496	0.01562	0.686
n=200	$\hat{\beta}$	-0.09400	0.06771	0.01342	0.000
	$\hat{\beta}_{C,A}$	-0.01392	0.08655	0.00757	0.712
	$\hat{\beta}_{C,1}$ (0)	0.01214	0.08888	0.00805	0.704
	$\hat{\beta}_{C,1}$ (2)	0.01080	0.08858	0.00796	0.706
	$\hat{\beta}_{C,1}$ (6)	0.00813	0.08798	0.00781	0.720
	$\hat{\beta}_{C,2}$	-0.01436	0.07986	0.00658	0.774
	$\tilde{\beta}$	-0.01030	0.08339	0.00706	0.762
	$\tilde{\beta}_1$	0.00730	0.08748	0.00771	0.720

Model: $(Y|U)$, exponential, mean $\exp(\alpha+\beta U)$, $\alpha=0$, $\beta=1/2$; U, $N(0,1)$; measurement error, $N(0,1/4)$.

TABLE II

Performance Measures for Estimators of β, Poisson Regression

		Bias	Standard Deviation	Mean Squared Error	$P(\cdot,\hat{\beta})$
n=50	$\hat{\beta}$	-0.10104	0.13135	0.02746	0.000
	$\hat{\beta}_{C,A}$	0.00644	0.17162	0.02950	0.634
	$\hat{\beta}_{C,1}$ (0)	0.01843	0.18613	0.03498	0.620
	$\hat{\beta}_{C,1}$ (2)	0.01211	0.18267	0.03352	0.628
	$\hat{\beta}_{C,1}$ (6)	-0.00001	0.17624	0.03106	0.650
	$\hat{\beta}_{C,2}$	-0.01704	0.16035	0.02600	0.664
	$\tilde{\beta}$	-0.01119	0.16772	0.02826	0.664
	$\tilde{\beta}_1$	0.01058	0.17980	0.03244	0.630
n=100	$\hat{\beta}$	-0.10315	0.09051	0.01883	0.000
	$\hat{\beta}_{C,A}$	0.00001	0.11519	0.01327	0.692
	$\hat{\beta}_{C,1}$ (0)	0.00550	0.12071	0.01460	0.690
	$\hat{\beta}_{C,1}$ (2)	0.00271	0.11981	0.01436	0.698
	$\hat{\beta}_{C,1}$ (6)	-0.00277	0.11807	0.01395	0.700
	$\hat{\beta}_{C,2}$	-0.02142	0.10796	0.01211	0.738
	$\tilde{\beta}$	-0.01875	0.11212	0.01292	0.730
	$\tilde{\beta}_1$	-0.00012	0.11816	0.01398	0.700
n=200	$\hat{\beta}$	-0.10024	0.06161	0.01384	0.000
	$\hat{\beta}_{C,A}$	0.00128	0.07839	0.00615	0.766
	$\hat{\beta}_{C,1}$ (0)	0.00419	0.08096	0.00657	0.748
	$\hat{\beta}_{C,1}$ (2)	0.00287	0.08068	0.00652	0.754
	$\hat{\beta}_{C,1}$ (6)	0.00024	0.08012	0.00642	0.768
	$\hat{\beta}_{C,2}$	-0.02059	0.07305	0.00576	0.828
	$\tilde{\beta}$	-0.01786	0.07576	0.00606	0.816
	$\tilde{\beta}_1$	-0.00055	0.07958	0.00633	0.770

Model: $(Y|U)$, Poisson, mean $\exp(\alpha+\beta U)$, $\alpha=0$, $\beta=1/2$; U, $N(0,1)$; measurement error, $N(0,1/4)$.

TABLE III

Performance Measures for Estimators of β, Normal Regression

		Bias	Standard Deviation	Mean Squared Error	$P(\cdot,\hat{\beta})$
$\tau = .5$	$\hat{\beta}$	-0.03095	0.10633	0.01227	0.000
	$\hat{\beta}_{C,A}$	0.00131	0.11562	0.01337	0.572
	$\hat{\beta}_{C,1}$ (0)	0.01784	0.14008	0.01994	0.550
	$\hat{\beta}_{C,1}$ (2)	0.01339	0.13558	0.01856	0.562
	$\hat{\beta}_{C,1}$ (6)	0.00484	0.12812	0.01644	0.578
	$\hat{\beta}_{C,2}$	-0.00374	0.11370	0.01294	0.608
	$\tilde{\beta}$	0.01187	0.12881	0.01673	0.558
	$\tilde{\beta}_1$	0.01676	0.13540	0.01861	0.552
$\tau = .25$	$\hat{\beta}$	-0.03122	0.06142	0.00475	0.000
	$\hat{\beta}_{C,A}$	0.00093	0.06750	0.00456	0.594
	$\hat{\beta}_{C,1}$ (0)	0.01466	0.07797	0.00630	0.564
	$\hat{\beta}_{C,1}$ (2)	0.01059	0.07613	0.00591	0.570
	$\hat{\beta}_{C,1}$ (6)	0.00268	0.07278	0.00530	0.590
	$\hat{\beta}_{C,2}$	-0.01798	0.06116	0.00406	0.612
	$\tilde{\beta}$	0.00987	0.07438	0.00563	0.572
	$\tilde{\beta}_1$	0.01407	0.07721	0.00616	0.564
$\tau = .125$	$\hat{\beta}$	-0.03105	0.04436	0.00293	0.000
	$\hat{\beta}_{C,A}$	0.00108	0.04923	0.00242	0.614
	$\hat{\beta}_{C,1}$ (0)	0.01429	0.05756	0.00352	0.564
	$\hat{\beta}_{C,1}$ (2)	0.01028	0.05605	0.00325	0.574
	$\hat{\beta}_{C,1}$ (6)	0.00248	0.05331	0.00285	0.602
	$\hat{\beta}_{C,2}$	-0.04974	0.04227	0.00426	0.382
	$\tilde{\beta}$	0.00968	0.05476	0.00309	0.576
	$\tilde{\beta}_1$	0.01375	0.05705	0.00344	0.564

Model: $(Y|U)$, normal, mean $\exp(\alpha+\beta U)$, variance τ^2, $\alpha=0$, $\beta=1/2$; U, N(0,1); measurement error, N(0,1/16); sample size, n=25.