UNBIASED ESTIMATION OF A NONLINEAR FUNCTION OF A NORMAL MEAN WITH APPLICATION TO MEASUREMENT-ERROR 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; unbiased estimation.

ABSTRACT

Let W be a normal random variable with mean μ and known variance σ^2. Conditions on the function $f(\cdot)$ are given under which there exists an unbiased estimator, $\tilde{f}(W)$, of $f(\mu)$ for all real μ. In particular it is shown that $f(\cdot)$ must be an entire function over the complex plane. Infinite series solutions for $\tilde{f}(\cdot)$ are obtained which are shown to be valid under growth conditions of the derivatives, $f^{(k)}(\cdot)$, of $f(\cdot)$. Approximate solutions are given for the cases in which no exact solution exists. The theory is applied to nonlinear measurement-error models as a means of finding unbiased score functions when measurement error is normally distributed. Relative efficiencies comparing the proposed method to the use of conditional scores (Stefanski and Carroll, 1987) are given for the Poisson regression model with canonical link.

1. INTRODUCTION

1.1 The Estimation Problem

Let W be a normal random variable with mean μ and known variance σ^2. Given a known function $f(\cdot)$, conditions on $f(\cdot)$ are found under which there exists a function $\tilde{f}(\cdot)$ with the property

* Research supported by the National Science Foundation.

that

$$E\{\bar{f}(W)\} = f(\mu) \quad . \tag{1.1}$$

It follows from the results in Section 2 that when $\bar{f}(\cdot)$ exists, $\bar{f}(W)$ is the almost-sure unique unbiased estimator of $f(\mu)$. Note that the maximum likelihood estimator of $f(\mu)$, $f(W)$, is generally biased.

When $f(\mu)$ is a polynomial in μ, a solution to the problem is discussed in Stulajter (1978). Neyman & Scott (1960) present a solultion to (1.1) when $f(\mu) = E\{t(\mu+\tau Z)\}$ where $Z \sim N(0,1)$, $\tau > 0$ is a known constant and $t(\cdot)$ is a known entire function. Other papers relevant to this problem include Parzen (1967) and Ghosh & Singh (1966).

1.2 Application to Nonlinear Measurement-Error Models

The estimation problem outlined above was motivated by the problem of finding unbiased score functions for estimation in nonlinear measurement-error models. Recent work by Stefanski (1985), Stefanski & Carroll (1986), Wolter & Fuller (1982), Carroll et al (1984), Armstrong (1985), Schafer (1987), Whittemore & Keller (1988), and Y. Amemiya in his Iowa State University Ph.D. Thesis indicates how difficult it is to obtain consistent estimators of unknown parameters in nonlinear measurement-error models, even when measurement error is assumed to be normally distributed with known variance. The theory of Section 2 provides a partial solution to this dilemma.

In this paper the problem of measurement error is studied in the context of the general assumptions employed by Stefanski (1985).

Suppose that an unknown parameter Θ_0 is estimated by an M-estimator Θ^* satisfying

$$\sum_{i=1}^{n} \psi(V_i,\Theta^*) = 0. \tag{1.2}$$

The data $V_1,\ldots,V_n$ are assumed to be independent random vectors such that

$$E\{\psi(V_i,\theta_o)\} = 0, \quad (i=1,\ldots,n).$$

Both θ_o and $\psi(\cdot,\cdot)$ are p-dimensional. Generally θ^* is root-n consistent and asymptotically normal (Huber, 1967) with variance $A^{-1}B(A^{-1})^T$,

$$\text{where} \quad A = \lim_n n^{-1} \sum_{i=1}^{n} E\left\{\frac{\partial\psi(V_i,\theta)}{\partial\theta} \,|\theta=\theta_o\right\} , \qquad (1.3)$$

$$B = \lim_n n^{-1} \sum_{i=1}^{n} E\{\psi(V_i,\theta_o)\psi^T(V_i,\theta_o)\} .$$

Let $V_i = (Y_i,U_i)$ and suppose that one gets to observe only $V_i^* = (Y_i,X_i)$ where

$$X_i = U_i+\sigma Z_i, \quad E(Z_i|Y_i,U_i) = 0 , \quad E(Z_iZ_i^T|Y_i,U_i) = I. \qquad (1.4)$$

This paper focuses on the case that σ is known and U_i, X_i and Z_i are all scalars.

Assume for the moment that there exists some function $\bar{\psi}(\cdot,\cdot,\cdot)$ with the property that

$$E\{\bar{\psi}(Y,X,\theta_o)|Y,U\} = \psi(Y,U,\theta_o) \qquad (1.5)$$

for all (Y,U). Recall that in (1.5), $X = U + \sigma Z$ and thus the conditional expectation in (1.5) is only with respect to the measurement-error distribution. It follows from (1.5) that $\bar{\psi}(Y,X,\theta_o)$ is unbiased, i.e.,

$$E\{\bar{\psi}(Y,X,\theta_o)\} = 0,$$

and hence it is to be expected that the M-estimator $\bar{\theta}$, defined as the solution to

$$\sum_{i=1}^{n} \bar{\Psi}(Y_i, X_i, \bar{\theta}) = 0 \tag{1.6}$$

would be consistent for θ_0 and asymptotically normal (Huber, 1967). The problem of course, lies in finding $\bar{\Psi}(\cdot,\cdot,\cdot)$ should it exist. When measurement error is normally distributed with known variance σ^2, this is simply the problem of finding an unbiased estimator of a nonlinear function of a normal mean. The normal variables are $X_i \sim N(U_i, \sigma^2)$ and $\bar{\Psi}(\cdot,\cdot,\cdot)$ must satisfy (1.6) which has the same form as (1.1).

2. UNBIASED ESTIMATION OF $f(\mu)$

2.1 Existence of $\bar{f}(\cdot)$

Assume that (1.1) holds for some function $\bar{f}(\cdot)$, then

$$\frac{1}{\sigma(2\pi)^{1/2}} \int \exp\left\{-\frac{(w-\mu)^2}{2\sigma^2}\right\} \bar{f}(w)dw = f(\mu)$$

for all real μ and hence for all real t

$$\int e^{tw} g(w)dw = e^{t^2\sigma^2/2} f(\sigma^2 t)$$

where

$$g(w) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-w^2/2\sigma^2} \bar{f}(w) .$$

It follows that $e^{t^2\sigma^2/2} f(\sigma^2 t)$ is the bilateral Laplace transform of $g(\cdot)$ and hence $\bar{f}(\cdot)$ is uniquely determined almost surely.

Further appeals to Laplace/Fourier transform theory indicate that if (1.1) holds, then:

(i) $f(i\sigma^2 z)$ is an entire function of the complex variable z; (2.1)

(ii) $\lim_{|t|\to\infty} e^{-t^2\sigma^2/2} f(i\sigma^2 t) = 0$; (2.2)

(iii) if $|e^{-t^2\sigma^2/2} f(i\sigma^2 t)|$ is integrable over $(-\infty,\infty)$ then $\bar{f}$ is continuous, and real valued, $e^{-w^2/2\sigma^2}\bar{f}(w)$ is bounded and

$$\bar{f}(w) = \int_{\infty}^{\infty} \frac{1}{\sigma}\,\phi\left(\frac{t-iw}{\sigma}\right) f(-it)\,dt \tag{2.3}$$

where ϕ is the standard normal density.

In some cases equation (2.3) allows one to find closed form solutions for $\bar{f}(\cdot)$. For example, suppose it is desired to estimate

$$f(\mu) = e^{a\mu^2}$$

for some choice of a. Using (2.2) it follows that an unbiased estimator exists only if $a > -(2\sigma^2)^{-1}$. With this assumption, note that

$$|e^{-t^2\sigma^2/2} e^{a(i\sigma^2 t)^2}| = e^{-\sigma^2 t^2 (a\sigma^2 + 1/2)}$$

is integrable and thus by (2.3) one finds for $\bar{f}(\cdot)$,

$$\bar{f}(w) = \int_{-\infty}^{\infty} \frac{1}{\sigma}\,\phi\left(\frac{t-iw}{\sigma}\right) e^{a(-it)^2}\,dt.$$

This is evaluated by completing the square, and has the value

$$\bar{f}(w) = (1+2a\sigma^2)^{-1/2}\, e^{aw^2/(1+2a\sigma^2)} . \tag{2.4}$$

Thus for $\bar{f}(\cdot)$ as in (2.4), $\bar{f}(W)$ is unbiased for $e^{a\mu^2}$.

Formula (2.3) has an interesting heuristic development. Write $W = \mu + \sigma Z_1$, where Z_1 is a standard normal random variate and let Z_2 be another independent standard normal variable. Define the complex variate Z as

$$Z = \sigma Z_1 + i\sigma Z_2 ,$$

where $i = (-1)^{1/2}$. It is easily seen that all the moments of Z exist and are zero, i.e.,

$$E(Z^k) = 0, \quad (k = 1,2,\ldots) . \tag{2.5}$$

Since $f(\cdot)$ must be an entire function the following series representation is valid,

$$f(\mu + \sigma Z_1 + i\sigma Z_2) = \sum_{k=0}^{\infty} \frac{f^{(k)}(\mu)}{k!} \sigma^k (Z_1 + iZ_2)^k . \tag{2.6}$$

Taking expectations with respect to Z_1 and Z_2 and, if summation and expectation can be interchanged in (2.6), then using (2.5) one finds

$$E\{f(\mu + \sigma Z_1 + i\sigma Z_2)\} = \sum_{k=0}^{\infty} \frac{f^{(k)}(\mu)}{k!} \sigma^k E\{(Z_1 + i\sigma Z_2)^k\} = f(\mu).$$

That is, again assuming the repeated integration makes sense,

$$\begin{aligned} f(\mu) &= E\{f(W + i\sigma Z_2)\} \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\frac{w-\mu}{\sigma}\right) \int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\frac{t-iw}{\sigma}\right) f(-it)dtdw \\ &= E\{\bar{f}(W)\} \end{aligned}$$

where $\bar{f}(\cdot)$ is defined in (2.3). This argument is made rigorous by the following theorem.

THEOREM 1. *If, for all* μ,

$$\sum_{k=0}^{\infty} \frac{|f^{(k)}(\mu)|(2\sigma^2)^{k/2}\Gamma(k/2 + 1)}{k!} < \infty$$

then $|f(W + i\sigma Z_2)|$ *has finite expectation with*

$$E\{f(W + i\sigma Z_2)\} = f(\mu)$$

and (2.3) *holds.*

PROOF. The theorem follows from (2.6), the inequality

$$E(|Z_1 + iZ_2|^k) \le 2^{k/2}\,\Gamma(k/2 + 1)$$

and Fubini's theorem. ////

2.2 A Series Solution for $\bar{f}(\cdot)$

Let $H_k(w)$ be the k^{th} Hermite polynomial (Cramer, 1957, p. 133)

$$H_0(w) = 1,$$
$$H_1(w) = w,$$
$$H_k(w) = wH_{k-1}(w) - (k-1)H_{k-2}(w),$$

and define

$$P_k(w) = \sigma^k H_k(w/\sigma) \ .$$

The following result appears in Stulajter (1978) and it is stated here without proof for reference.

LEMMA 1. *If W is normally distributed with mean μ and variance σ^2 then*

$$E\{P_k(W)\} = \mu^k \ .$$

It follows from the lemma that $\bar{f}(\cdot) = P_k(\cdot)$ when $f(\mu) = \mu^k$.

Since $f(\cdot)$ must be an entire function it has the representation

$$f(\mu) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}\,\mu^k \ .$$

Thus, assuming all operations are justifiable, the function

$$\bar{f}(w) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}\,P_k(w) \qquad (2.7)$$

evaluated at W has expectation

$$E\{\bar{f}(W)\} = E\{\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} P_k(W)\} \quad (2.8)$$

$$= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} E\{P_k(W)\}$$

$$= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \mu^k$$

$$= f(\mu) \ .$$

To make this argument rigorous it is necessary to know when the series in (2.7) is absolutely convergent and when expectation can be carried out termwise in (2.8). The next theorem addresses these issues.

THEOREM 2. *If*

$$\sum_{k=0}^{\infty} \frac{|f^{(k)}(0)|\sigma^k(k!)^{1/2}}{k!} < \infty \quad (2.9)$$

Then the series defining $\bar{f}(\cdot)$ *in (2.7) is absolutely convergent,* $|\bar{f}(W)|$ *has finite expectation, and*

$$E\{\bar{f}(W)\} = f(\mu) \ .$$

PROOF. Using the inequality $H_k(x) < C(k!)^{1/2}e^{x^2/4}$ (Cramer, 1925) where C does not depend on x or k it follows that

$$\sum_{k=0}^{\infty} \frac{|f^{(k)}(0)|}{k!} |P_k(w)| < Ce^{w^2/4\sigma^2} \sum_{k=0}^{\infty} \frac{|f^{(k)}(0)|}{k!} (k!)^{1/2}\sigma^k$$

which is finite by assumption, thus proving the first assertion. The remaining two assertions follow from the fact that $E(e^{W^2/4\sigma^2})$ is finite and Fubini's Theorem. ////

Any function with $|f^{(k)}(0)|$ bounded for all k satisfies (2.9). e.g., e^{μ}, $\sin(\mu)$ and $\cos(\mu)$. For a less transparent application of Theorem 2 consider the function

$$f(\mu) = \frac{\sin^2(\mu/\lambda)}{c\lambda(\mu/\lambda)^2} = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}\mu^k$$

where λ is a fixed positive constant; c is chosen (independently of λ) so that $\int f(t)dt = 1$; and

$$f^{(k)}(0) = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \sum_{i=0}^{r} \frac{(-1)^r (2r)!(1/\lambda)^{2r+1}}{(2i+1)!(2r-2i+1)!} & k = 2r,\ r = 0,1,\ldots\ . \end{cases}$$

Condition (2.9) is satisfied and thus

$$\bar{f}(W) = \sum_{r=0}^{\infty} f^{(2r)}(0)\frac{P_{2r}(W)}{(2r)!}$$

is an unbiased estimator of $f(\mu)$.

This example arises in the context of estimating a density function when the data are measured with $N(0,\sigma^2)$ errors (Stefanski & Carroll, 1987; Carroll & Hall, 1988; Liu & Taylor, 1987). Suppose that U_i i=1,... are iid with unknown density g and let $X_i = U_i + \varepsilon_i$, i=1,..., where the ε_i's are iid $N(0,\sigma^2)$ and are independent of the U_i's. Conditioned on U_i, $X_i - x$ is distributed $N(U_i - x, \sigma^2)$, and thus

$$E\{(1/n)\sum_{i=1}^{n} \bar{f}(X_i)|U_i,\ldots,U_n\}$$
$$= (1/n)\sum_{i=1}^{n} f(U_i - x)\ . \qquad (2.10)$$

The right-hand side of (2.10) is an ordinary kernel density estimator of g and thus $(1/n)\ \Sigma\bar{f}(X_i)$ is a natural method-of-moments estimator of g.

2.3 Some Approximate Solutions

Seldom will the expression in (2.3) be obtainable in closed form and thus (2.7) is useful in that it suggests a method of computing the estimator $\bar{f}(W)$. It is still less than fully satisfactory since in practice only a finite number of terms can be employed and the derivatives $f^{(k)}(0)$ may not be manageable for large values of k. Thus the relevant question is not whether (2.7) is well defined but whether or not the sum of the first few terms in the series provides a reasonable approximation to $\bar{f}(\cdot)$; or more importantly whether or not this finite sum has expectation approximating $f(\mu)$ in some sense. In the analysis of nonlinear measurement-error models useful insights have been obtained by studying estimators when measurement error is small, i.e. when σ tends to zero, see Wolter & Fuller (1982), Stefanski (1985), and Stefanski & Carroll (1986). This same approach, i.e. small-σ asymptotics, is adopted here to study the relative merits of some approximate solutions to (1.1).

Starting from (2.7) a natural candidate for an approximation is

$$\bar{f}_n(w) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} \sigma^k H_k(\frac{w}{\sigma}) \quad . \qquad (2.10)$$

Since, as σ tends to zero

$$|\bar{f}(w) - \bar{f}_n(w)| = |\sum_{k=n+1}^{\infty} \frac{f^{(k)}(0)}{k!} \sigma^k H_k(\frac{w}{\sigma})| = 0(\sigma) \quad ,$$

it is also true that, under sufficient regularity conditions,

$$|E\{\bar{f}(W)\} - E\{\bar{f}_n(W)\}| = 0(\sigma) \quad .$$

This is not very encouraging since it means that taking additional terms in $\bar{f}_{(n)}(\cdot)$ does not improve the approximations as a function of σ.

Consider now the following rearrangement of the series in (2.7). Writing $H_k(w) = \sum_{j=0}^{k} a_j^k w^j$, (2.7) becomes

$$\bar{f}(w) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)\sigma^k}{k!} H_k\left(\frac{w}{\sigma}\right)$$

$$= \sum_{k=0}^{\infty} \frac{f^{(k)}(0)\sigma^k}{k!} \sum_{j=0}^{\infty} a_j^k \sigma^{-j} w^j .$$

Assuming that the order of summation can be interchanged

$$\bar{f}(w) = \sum_{j=0}^{\infty} \left\{ \sum_{k=j}^{\infty} \frac{f^{(k)}(0) a_j^k \sigma^{k-j}}{k!} \right\} w^j$$

$$= \sum_{j=0}^{\infty} \left\{ \sum_{k=0}^{\infty} \frac{f^{(k+j)}(0) a_j^{k+j} \sigma^k}{(k+j)!} \right\} w^j$$

Now using the relationships,

(i) $a_j^{k+j} = \binom{k+j}{j} a_o^k$,

(ii) $a_j^k = 0$ if $k-j$ is odd,

(iii) $a_o^{2k} = (-1)^k \frac{(2k-1)!}{2^{k-1}(k-1)!}$;

one finds

$$\bar{f}(w) = \sum_{j=0}^{\infty} \left\{ \sum_{k=0}^{\infty} \frac{f^{(2k+j)}(0)(-1)^k \sigma^{2k}}{2^k j! k!} \right\} w^j$$

$$= \sum_{k=0}^{\infty} \frac{(-\sigma^2)^k}{2^k k!} \left\{ \sum_{j=0}^{\infty} \frac{f^{(2k+j)}(0)}{j!} w^j \right\} . \quad (2.11)$$

But the expression in brackets in (2.11) is just $f^{(2k)}(w)$ thus

$$\bar{f}(w) = \sum_{k=0}^{\infty} \frac{(-\sigma^2)^k}{2^k k!} f^{(2k)}(w) .$$

Define

$$\tilde{f}_n(w) = \sum_{k=0}^{n} \frac{(-\sigma^2)^k}{2^k k!} f^{(2k)}(w) . \qquad (2.12)$$

and note that

$$|\bar{f}(w)-\tilde{f}_n(w)| = | \sum_{k=n+1}^{\infty} \frac{(-\sigma^2)^k}{2^k k!} f^{(2k)}(w)| = 0(\sigma^{2n+2}) .$$

Thus when σ is small $\tilde{f}_n(w)$ provides a better approximation to $\bar{f}(w)$ than does $\bar{f}_n(w)$. The same should be true of their respective expected values. A few examples are given to illustrate this point.

EXAMPLE 1. Let $f(\mu) = e^{\mu}$; using the moment generating function of $W \sim N(\mu,\sigma^2)$ one finds that $\bar{f}(W) = e^{-\sigma^2/2+W}$ is an unbiased estimator of $f(\mu)$. In this case

$$\bar{f}_n(W) = \sum_{k=0}^{n} \frac{\sigma^k}{k!} H_k(\frac{W}{\sigma})$$

and

$$E\{\bar{f}_n(W)\} = \sum_{k=0}^{n} \frac{\mu^k}{k!}$$

which has an absolute relative error which is independent of σ and small only when μ is small.

On the other hand

$$\tilde{f}_n(W) = \sum_{k=0}^{n} \frac{(-\sigma^2)^k}{2^k k!} e^W$$

and

$$E\{\tilde{f}_n(W)\} = e^{\sigma^2/2}\left\{\sum_{k=0}^{n} \frac{(-\sigma^2)^k}{2^k k!}\right\} e^{\mu}$$

which has an absolute relative error which is uniformly small in μ when σ is small.

EXAMPLE 2. Let $f(\mu) = \sin(\mu)$; from the characteristics function of W it follows that $\bar{f}(W) = e^{\sigma^2/2} \sin(W)$ has expectation $\sin(\mu)$. In this case, setting n = 1 in (2.10) and (2.12) one finds

$$\bar{f}_1(W) = W, \quad \tilde{f}_1(W) = (1 + \sigma^2/2)\sin(W)$$

and

$$E\{\bar{f}_1(W)\} = \mu, \quad E\{\tilde{f}_1(W)\} = (1 + \sigma^2/2)e^{\sigma^2/2}\sin(\mu).$$

Again the expected value of $\bar{f}_1(W)$ approximates $\sin(\mu)$ only when μ is near zero; however, $E\{\tilde{f}_1(W)\}$ is a good approximation to $\sin(\mu)$ for all μ when σ is small.

EXAMPLE 3. In Section 2 it was shown that there exist no unbiased estimators of $f(\mu) = e^{-\mu^2/2\sigma^2}$. In this case the series defined by (2.7) and (2.11) are not absolutely convergent. However, it still makes sense to ask whether $\bar{f}_n(W)$ or $\tilde{f}_n(W)$ have expectations approximating $e^{-\mu^2/2\sigma^2}$ for finite n. When n=1 one finds

$$\bar{f}_1(W) = 1, \quad \tilde{f}_1(W) = e^{-W^2/2\sigma^2}\left(\frac{3}{2} - \frac{W^2}{2\sigma^2}\right)$$

and

$$E\{\bar{f}_1(W)\} = 1, \; E\{\tilde{f}_1(W)\} = 2^{-1/2}e^{-\mu^2/4\sigma^2}\left(\frac{5}{4} - \frac{\mu^2}{8\sigma^2}\right).$$

Obviously $E\{\bar{f}_1(W)\}$ approximates $e^{-\mu^2/2\sigma^2}$ only if μ/σ is near zero. However, $E\{\tilde{f}_1(W)\}$ provides a reasonable approximation to $e^{-\mu^2/2\sigma^2}$ for all μ even though no unbiased estimators exist in this case.

3. APPLICATION TO MEASUREMENT-ERROR MODELS

3.1 Unbiased Score Functions and Objective Functions

As described in the introduction the application of the previous section's theory to construct unbiased score functions is, under appropriate conditions, straightforward. When the non-errors-in-variables score is the gradient of an objective function, e.g., a likelihood function, it may also be possible to find an unbiased objective function. With the notation of Section 1.2 suppose that

$$\Psi(y,x,\Theta) = (\partial/\partial\Theta)\rho(y,x,\Theta)$$

where the estimator $\tilde{\Theta}$ defined in (1.2) is the maximizer of

$$\sum_{i=1}^{n} \rho(Y_i,X_i,\Theta) \ .$$

Suppose now that there exists a function $\bar{\rho}(\cdot,\cdot,\cdot)$ with the property that

$$E\{\bar{\rho}(Y_i,X_i,\Theta)|Y_i,U_i\} = \rho(Y_i,U_i,\Theta) \ . \qquad (3.1)$$

The estimator $\bar{\Theta}$ can be obtained in some cases by maximizing

$$\sum_{i=1}^{n} \bar{\rho}(Y_i,X_i,\Theta) \ . \qquad (3.2)$$

Note that if differentiation and integration can be interchanged in (3.1) then $(\partial/\partial\Theta)\bar{\rho}(\cdot,\cdot,\Theta) = \bar{\Psi}(\cdot,\cdot,\Theta)$.

A number of examples are now given illustrating this approach. As stated in the introduction it is assumed that the normal measurement error variance, σ^2, is known.

EXAMPLE 1. Consider a polynomial regression model in which

$$E(Y|U) = \sum_{j=0}^{k} \beta_{jo}U^j$$

where $\Theta_o = (\beta_{0o},\ldots,\beta_{ko})$ are to be estimated. Let

$$\rho(Y,U,\Theta) = (Y - \sum_{j=0}^{k} \beta_j U^j)^2$$

and

$$\Psi(Y,U,\Theta) = (\partial/\partial\Theta)\rho(Y,U,\Theta) \ .$$

Since

$$\rho(Y,U,\Theta) = Y^2 + \sum_{j=0}^{k} \beta_j^2 U^{2j} + 2 \sum_{j>\ell}\sum \beta_j\beta_\ell U^{j+\ell} - 2Y \sum_{j=0}^{k} \beta_j U^j \ .$$

Lemma 1 indicates that

$$\bar{\rho}(Y,X,\Theta) = Y^2 + \sum_{j=0}^{k} \beta_j^2 P_{2j}(X) + 2 \sum_{j>\ell}\sum \beta_j\beta_\ell P_{j+\ell}(X) - 2Y \sum_{j=0}^{k} \beta_j P_j(X)$$

has expectation

$$E\{\bar{\rho}(Y,X,\Theta)\,|\,Y,U\} = \rho(Y,U,\Theta)$$

and this suggests that $\bar{\Theta}$ be defined as the minimizer of

$$\sum_{i=1}^{n} \bar{\rho}(Y_i,X_i,\Theta) \qquad (3.3)$$

or as the solution to

$$\sum_{i=1}^{n} \bar{\Psi}(Y_i,X_i,\Theta) = 0 \qquad (3.4)$$

where $\bar{\Psi}(Y_i,X_i,\Theta) = (\partial/\partial\Theta)\, \bar{\rho}(Y_i,X_i,\Theta)$.

When $k = 1$, (3.3) attains its minimum value for finite β_0, β_1 only if $\Sigma(X_i-\bar{X})^2 > n\sigma^2$. However, (3.4) admits the unique solution $\hat{\beta}_1 = S_{YX}/(S_{XX}-n\sigma^2)$, $\hat{\beta}_0 = \bar{Y}-\hat{\beta}_1\bar{X}$ which is recognized as the usual method-of-moments correction-for-attenuation estimator. The behaviour of $\bar{\Psi}$ and $\bar{\rho}$ in the linear model is typical in nonlinear models as well. Frequently $\bar{\rho}$ will not have a sensible maximum but $\bar{\Psi}$ will admit a consistent estimator as in the following example.

EXAMPLE 2. Suppose that given U, Y is exponentially distributed with mean $\exp(-\alpha_0 - \beta_0 U)$, i.e., the density of Y|U is

$$f_{Y|U}(y|u) = \exp\{\alpha_0+\beta_0 U - y\exp(\alpha_0+\beta_0 U)\} \ , \ y \geq 0 \ .$$

For maximum likelihood estimation

$$\rho(Y,U,\Theta) = \alpha+\beta U - Ye^{(\alpha+\beta U)}$$

and

$$\psi(Y,U,\Theta) = \{1 - Ye^{(\alpha+\beta U)}\}\binom{1}{U}$$

where $\Theta = (\alpha,\beta)$. Using (2.3) one finds

$$\bar{\rho}(Y,X,\Theta) = \alpha + \beta X - Ye^{(\alpha+\beta X - \frac{\beta^2\sigma^2}{2})}$$

and

$$\bar{\psi}(Y,X,\Theta) = \{1 - Y\exp(\alpha+\beta X - \frac{\beta^2\sigma^2}{2})\}\binom{1}{X} + \begin{bmatrix} 0 \\ Y\beta\sigma^2\exp(\alpha+\beta X - \frac{\beta^2\sigma^2}{2}) \end{bmatrix}$$

as unbiased estimators of $\rho(Y,U,\Theta)$ and $\psi(Y,U,\Theta)$ respectively. It is thus tempting to define $\bar{\Theta}$ as the maximizer of

$$\sum_{i=1}^{n} \bar{\rho}(Y_i,X_i,\Theta) \ . \qquad (3.5)$$

However, because the factor $\exp(-\beta^2\sigma^2/2)$ appears in $\bar{\rho}(Y,X,\Theta)$, (3.5) is unbounded as β approaches positive (negative) infinity if $\Sigma X_i > 0$ ($\Sigma X_i < 0$). There will, however, always be a consistent sequence of solutions to the score equations

$$\sum_{i=1}^{n} \bar{\psi}(Y_i,X_i,\Theta) = 0 \ . \qquad (3.6)$$

The problems of an ill-defined maximum to (3.2) or multiple solutions to (1.6) as encountered in the previous example are not uncommon. Under sufficient regularity conditions unbiasedness of

$\tilde{\Psi}$ implies the existence of a consistent sequence of solutions to (1.6). My experience suggests that for moderate to large sample sizes, a Newton-Raphson iteration of (1.6) starting from the so-called naive estimator (the solution to $\Sigma\psi(Y_i, X_i, \Theta) = 0$) converges to the desired solution nearly all the time. For a modest increase in computational time the root-finding algorithm discussed in Stefanski (1988) can be employed to obtain consistent solutions to (1.6).

Finding $\tilde{\Psi}$ for the exponential model was trivial since ψ depends on U only through U, $e^{\beta U}$ and $Ue^{\beta U}$. Thus determination of $\tilde{\Psi}$ involved only simple manipulations of the normal moment generating function. Similarly $\tilde{\Psi}$ is easily determined for any generalized linear model whose likelihood score depends on U only through exponentials, powers and their cross products. This fact is exploited in Section 5 in connection with the Poisson regression model.

Next an example is given in which $\tilde{\Psi}(\cdot,\cdot,\cdot)$ does not exist.

EXAMPLE 3. Measurement error in logistic regression has been studied by Carroll et. al. (1984) and Stefanski & Carroll (1985). Consider a simple logistic regression model in which

$$pr(Y=1|U) = F(\alpha+\beta U),$$

where

$$F(t) = \frac{1}{1+e^{-t}} .$$

The likelihood score function is

$$\psi(Y,U,\Theta) = \{Y - F(\alpha+\beta U)\}\binom{1}{U}$$

and in order for $\bar{\Psi}(\cdot,\cdot,\cdot)$ to exist it is necessary to estimate $F(\alpha+\beta U)$ and $UF(\alpha+\beta U)$ unbiasedly. In Section 2 it was noted that in order for a function of a normal mean to be estimated unbiasedly it is necessary for that function to be an entire function of the complex variable z. Since

$$F(z) = \frac{1}{1+e^{-z}}$$

has singularities at multiplies of $i\pi$ there is no unbiased estimator of $F(\alpha+\beta U)$ for all (α,β), i.e., $\bar{\Psi}(\cdot,\cdot,\cdot)$ does not exist for all (α,β) in this case. However, in the next section an approximate unbiased score is found for this model.

3.2 Approximate Unbiased Score Functions

In the event that $\bar{\Psi}(\cdot,\cdot,\cdot)$ cannot be found in closed form or does not exist, e.g., as in the case of logistic regression, it is useful to consider approximate unbiased scores using the approximation developed in Section 2.3. Given $\Psi(\cdot,\cdot,\cdot)$ let

$$\Psi_{22}(Y,x,\Theta) = \frac{\partial^2}{\partial x^2}\Psi(Y,x,\Theta)$$

and define

$$\tilde{\Psi}(Y,U,\Theta) = \Psi(Y,U,\Theta) - \frac{\sigma^2}{2}\Psi_{22}(Y,U,\Theta) \ .$$

The approximate score $\tilde{\Psi}(\cdot,\cdot,\cdot)$ is obtained by taking the first two terms in the series representation for $\bar{\Psi}(\cdot,\cdot,\cdot)$ analogous to (2.12). Now define $\tilde{\Theta}$ as the estimator satisfying

$$\sum_{i=1}^{n} \tilde{\Psi}(Y_i,X_i,\tilde{\Theta}) = 0 \ , \qquad (3.7)$$

realizing that care must be taken in defining $\tilde{\Theta}$ if (3.7) has more than one root.

It is now shown that the estimator $\tilde{\Theta}$ converges, under regularity conditions, to a quantity $\Theta_u(\sigma)$ where

$$\Theta_u(\sigma) = \Theta_o + o(\sigma^2) \ , \qquad (3.8)$$

i.e., the estimator $\tilde{\Theta}$ enjoys the same desirable small-measurement-error properties as the bias-corrected estimator in Stefanski (1985).

The quantity $\Theta_u(\sigma)$ satisfies (Huber, 1967)

$$\lim_n \frac{1}{n} \sum_{i=1}^{n} E[\tilde{\Psi}\{Y_i, X_i, \Theta_u(\sigma)\}] = 0, \qquad (3.9)$$

cf. equation (1.5). For notational convenience (3.9) will be written as

$$E[\tilde{\Psi}\{Y, U+\sigma Z, \Theta_u(\sigma)\}] = 0 \ . \qquad (3.10)$$

Following the approach used by Stefanski (1985), $\Theta_u(\sigma)$ is expanded in a Maclaurin series; note that $\Theta_u(0) = \Theta_o$.

Differentiation of (3.10) with respect to σ shows that

$$0 = E[\Psi_2 Z + \Psi_3 \Theta_u^{(1)}(\sigma) - \sigma\Psi_{22} - \frac{\sigma^2}{2} \{\Psi_{22} Z + \Psi_{223}\Theta_u(\sigma)\}]. \qquad (3.11)$$

In (3.11) Ψ_2, Ψ_3, Ψ_{22}, etc. denote partial derivatives of $\Psi(Y,U,\Theta)$ and are all evaluated at $(Y, U+\sigma Z, \Theta_u(\sigma))$. Evaluating (3.11) at $\sigma = 0$ and using (1.3) and (1.4) implies that $\Theta_u^{(1)}(0) = 0$. A second differentiation of (3.11) with respect to σ results in

$$0 = E[\Psi_{22} Z^2 + 2\Psi_{23} Z \Theta_u^{(1)}(\sigma) + \Psi_{33}\{\Theta_u^{(1)}(\sigma)\}^2 +$$

$$\Psi_3 \Theta_u^{(2)}(\sigma) - \Psi_{22} - 2\sigma\{\Psi_{22} Z + \Psi_{223}\Theta_u^{(1)}(\sigma)\}$$

$$- \frac{\sigma^2}{2} (\frac{d}{d\sigma}) \{\Psi_{222} Z + \Psi_{223}\Theta_u^{(1)}(\sigma)\}] \ . \qquad (3.12)$$

Evaluating (3.12) at $\sigma = 0$ and using (1.3) and (1.4) implies that $\Theta_u^{(2)}(0) = 0$. Thus (3.8) holds as claimed.

Note that in this argument it was not necessary to have normally distributed errors since only the first two moments of the error distribution come into play. Thus while $\tilde{\Psi}$ is only approximately unbiased in the sense of (3.8) it enjoys this property irrespective of the error law, provided of course that the first two moments are finite. Thus unless an assumption of normal measurement error is known to be valid, $\tilde{\Psi}$ may be preferable to $\bar{\Psi}$ even when the latter is obtainable in closed form.

The approximately unbiased score, $\tilde{\Psi}$, for logistic regression is now derived.

EXAMPLE 3 (continued). Since

$$\Psi(Y,U,\Theta) = \{Y - F(\alpha+\beta U)\}\binom{1}{U} ,$$

$$\Psi_{22}(Y,U,\Theta) = -\beta^2 F^{(2)}(\alpha+\beta U)\binom{1}{U} - 2\beta F^{(1)}(\alpha+\beta U)\binom{0}{1}$$

and

$$\tilde{\Psi}(Y,X,\Theta) = \{Y - F(\alpha+\beta X) + \frac{\beta^2\sigma^2}{2} F^{(2)}(\alpha+\beta X)\}\binom{1}{X}$$

$$+ \sigma^2\beta\, F^{(1)}(\alpha+\beta X)\binom{0}{1} .$$

4.0 ASYMPTOTIC DISTRIBUTION THEORY

Suppose that $\bar{\Psi}$ exists and that $\bar{\Theta}$ is a consistent solution to (1.6). Since $\bar{\Theta}$ is an M-estimator, $n^{1/2}(\bar{\Theta}-\Theta_0)$ is, under regularity conditions, asymptotically normal with mean zero and covariance matrix $\bar{A}^{-1}\bar{B}(\bar{A}^{-1})^T$ where $\bar{A}$ and $\bar{B}$ are consistently estimated by

$$\hat{A} = n^{-1} \sum_{i=1}^{n} \left\{ \frac{\partial\bar{\Psi}(Y_i,X_i,\Theta)}{\partial\Theta} \Big|_{\Theta=\bar{\Theta}} \right\}, \tag{4.1}$$

$$\hat{B} = n^{-1} \sum_{i=1}^{n} \bar{\Psi}(Y_i,X_i,\bar{\Theta})\bar{\Psi}^T(Y_i,X_i,\bar{\Theta}) \tag{4.2}$$

respectively.

If $\tilde{\psi}$ is employed, then the corresponding estimator $\tilde{\Theta}$ converges to $\Theta_u(\sigma)$ (see Section 3.2) and $n^{1/2}\{\tilde{\Theta}-\Theta_u(\sigma)\}$ will have a limiting normal distribution with mean zero and covariance matrix $\tilde{A}^{-1}\tilde{B}(\tilde{A}^{-1})^T$ where $\tilde{A}$ and $\tilde{B}$ are estimated as in (4.1) and (4.2) with $\tilde{\psi}$ and $\tilde{\Theta}$ replacing $\bar{\Psi}$ and $\bar{\Theta}$ respectively.

5.0 EFFICIENCY OF $\bar{\Psi}$

In this section we consider the application of the theory in Section 2 to the semiparametric models studied in Stefanski & Carroll (1987). It will be assumed that $\bar{\Psi}$ exists, although it is evident that this need not be the case for all generalized linear measurement error models. A question of some interest is the asymptotic relative efficiency of $\bar{\Psi}$ to the conditional scores obtained by Stefanski & Carroll (1987). It follows from Lemma 3.1 and Theorem 3.1 of that paper that $E(\bar{\Psi}|\Delta) = 0$ where $\Delta = X+Y\sigma^2\beta$, and that $V_{\bar{\Psi}} > V_{\psi_c}$ where V_{ψ} denotes the asymptotic covariance matrix associated with ψ and ψ_c is the 'optimal' conditional score (denoted by ψ^* in Stefanski & Carroll (1987)). Thus whenever both $\bar{\Psi}$ and ψ_c exist, ψ_c is to be preferred based on asymptotic efficiency. However, for certain models ψ_c is not available in closed form whereas $\bar{\Psi}$ is. Thus a case can be made for employing $\bar{\Psi}$ based on computational considerations. A factor involved in this decision is the efficiency loss incurred by using

$\bar{\Psi}$. In the remainder of this section the relative magnitude of $V_{\bar{\Psi}}$ to V_{Ψ_c} is investigated at a particular structural generalized linear model.

We now compare the efficiency of $\bar{\Psi}$ to the conditional scores in the framework of a structural Poisson measurement error model.

Let

$$pr(Y=k|U) = e^{-\lambda}\lambda^{k}/k! \quad k=0,1,\ldots \tag{5.1}$$

where

$$\lambda = \exp(\alpha+\beta U) \ ,$$

and

$$U \sim N(\mu_U,\sigma_U^2) \ . \tag{5.2}$$

For this model

$$\bar{\Psi}(Y,X,\alpha,\beta) = \begin{bmatrix} Y - \exp(\alpha+\beta X - \beta^2\sigma^2/2) \\ YX - (X-\sigma^2\beta)\exp(\alpha+\beta X-\beta^2\sigma^2/2) \end{bmatrix}$$

and the corresponding asymptotic covariance matrix,

$$V_{\bar{\Psi}} = E\{\partial\bar{\Psi}/\partial(\alpha,\beta)\}^{-1}E(\bar{\Psi}\,\bar{\Psi}^T)\ E\{\partial\bar{\Psi}/\partial(\alpha,\beta)\}^{-T}$$

can be computed analytically under (5.2). The conditional scores considered have the form

$$\Psi_t(Y,X,\alpha,\beta) = \{Y-E(Y|\Delta)\}\{\begin{matrix}1\\ t(\Delta)\end{matrix}\} \tag{5.3}$$

where

$$\Delta = X + Y\sigma^2\beta$$

and $t(\cdot)$ is a function not depending on the data. When $t(\Delta) = E(U|\Delta)$ the corresponding score, designated ψ_C, is efficient in the structural mixture model setting discussed in Stefanski & Carroll (1987). The corresponding asymptotic covariance matrix $V_{\psi_C} = E(\psi_C \psi_C^T)$ is a lower bound on the asymptotic covariance matrices of root-n consistent estimators of $(\alpha,\beta)^T$ for this model. Since $E(U|\Delta)$ depends on the unknown distribution of U so too does ψ_C. Although it is possible to construct estimators achieving this lower bound, doing so involves preliminary estimation of $E(U|\Delta = \cdot)$ and is computationally complicated; see for example Bickel & Ritov (1987). For this reason the computationally simpler score, ψ_{CL}, defined by taking $t(\Delta) = \Delta$ in (4.3) is also studied. The corresponding asymptotic covariance matrix is denoted $V_{\psi_{CL}}$.

Neither ψ_C nor ψ_{CL} have closed form expressions and thus both V_{ψ_C} and $V_{\psi_{CL}}$ must be calculated numerically.

It follows from Theorem 3.1 in Stefanski & Carroll (1987) that $V_{\bar{\Psi}} > V_{\psi_C}$ and $V_{\psi_{CL}} > V_{\psi_C}$ in the sense of positive definiteness.

Table I gives values of $100V_{\psi}(2,2)/V_{\bar{\Psi}}(2,2)$ and $100V_{\psi_C}(2,2)/V_{\psi_{CL}}(2,2)$ (where $M(ij)$ denotes the i,jth entry of M) for model (5.1) with $\alpha = 0$, $\beta = 1/2$, $\sigma^2 = 1/8, 1/4, 1/2, 1$, $\mu_U = 0$ and $\sigma_U^2 = 1$. The fact that all entries in the table are less than 100 follows from the optimality of ψ_C. The loss of efficiency in ψ_{CL} is hardly significant although the same is not true of $\bar{\Psi}$ for values of $\sigma^2 > 1/4$. The loss of efficiency for both scores increases with increasing σ^2 although the effect is substantial only for $\bar{\Psi}$.

TABLE 1

Relative Efficiencies of $V_{\Psi_{cL}}$ and $V_{\bar{\Psi}}$ to V_{Ψ_c} for model (5.1) with $\alpha = 0$, $\beta = 1/2$, σ^2 as indicated, $\mu_U = 0$ and $\sigma_U^2 = 1$.

	Measurement Error Variance (σ^2)			
Score	1/8	1/4	1/2	1
Ψ_{cL}	99.98	99.95	99.90	99.84
$\bar{\Psi}$	96.06	92.75	86.30	71.30

For the Poisson regression model Ψ_{cL} is more efficient than $\bar{\Psi}$. This is likely to be the case for any generalized linear model for which both Ψ_{cL} and $\bar{\Psi}$ exist. However, the conditional scores exist only for generalized linear models in canonical form and thus $\bar{\Psi}$ or its approximate version, $\tilde{\Psi}$, is more generally applicable.

6.0 SUMMARY

Interest in nonlinear measurement-error models is evident from the numerous recent published articles in this area. The present paper differs from preceding papers by employing a method-of-(nonlinear) moments estimation procedure. This technque yields consistent estimators for a large number of models when measurement errors are normally distributed. The class of models covered by this theory can be expanded by use of the approximately unbiased scores corresponding to (2.12). Since the technique preserves the conditional role of the true covariates, it is appropriate for both functional and structural models. The method is particularly suited to generalized linear models whose likelihood scores depend on the covariates only through exponentials and powers.

REFERENCES

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

Bickel, P. J. & Ritov, Y. (1987). Efficient estimation in the errors-in-variables model. *Ann. Statist.* **15**, 513-40.

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

Carroll, R. J. & Hall, P. J. (1988). Optimal rates of convergence for deconvolving a density. *JASA*, to appear.

Cramèr, H. (1925). On some classes of series used in mathematical statistics. *Sixth Scandinavian Congr. of Math.*, p. 399, Kobenhaven.

Cramèr, H. (1957). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

Ghosh, J. K. and Singh, R. (1966). Unbiased estimation of location and scale parameters. *Ann. Math. Statist.* **37**, 1671-75.

Huber, P. J. (1967). The behavior of maximum likelihood estimators under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, Ed. L. M. LeCam & J. Neyman, 221-33, University of California Press.

Liu, M. C. & Taylor, R. L. (1987). A consistent nonparametric density estimator for the deconvolution problem. University of Georgia, Department of Statistics Technical Report, STA 73.

Neyman, J. and Scott, E. L. (1960). Correction for bias induced by a transformation of variables. *Ann. Math. Statist.* **31**, 643-55.

Parzen, E. (1967). *Time Series Analysis Papers*. San Francisco: Holden-Day.

Schafer, D. W. (1987). Covariate measurement error in generalized linear models. *Biometrika* **74**, 385-91.

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). Deconvoluting kernel density estimators. Institute of Statistics Mimeo Series No. 1909, North Carolina State University, Raleigh, NC.

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

Stefanski, L. A. (1988). Measurement errors in generalized linear model explanatory variables. Institute of Statistics Mimeo Series No. 1926, North Carolina State University, Raleigh, NC.

Stulajter, F. (1978). Nonlinear estimators of polynomials in mean values of a Guassian stochastic process. *Kybemetrika* **14**, 206-20.

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

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