# EFFICIENT ESTIMATION IN A CLASS OF GENERALIZED LINEAR MEASUREMENT-ERROR MODELS

by

Leonard A. Stefanski[1]
Department of Statistics
North Carolina State University
Raleigh, North Carolina 27695-8203

[1] Research supported by National Science Foundation Grant No. DMS-8013681.

AMS 1980 *subject classification*. Primary 62F12; secondary 62J02, 62B99.

*Key words and phrases*. Efficient estimation, generalized linear models, measurement-error models, structural models.

SUMMARY

Efficient estimation of the regression parameters in a structural generalized linear measurement-error model is studied. Using the efficient scores derived in Stefanski and Carroll (1987), the existence of asymptotically efficient estimators is established by employing a one-step construction similar to that used by Bickel (1982). The construction assumes the availability of an $n^{\frac{1}{2}}$-consistent estimator which is argued to exist quite generally for the models under investigation.

## 1. INTRODUCTION.

For a generalized linear model in canonical form, the conditional density of the response variable Y given the p×1-dimensional covariate U = u can be written as

$$f_{Y|U}(y|u;\theta,\phi) = f(y;\ \alpha+\beta^T u,\phi), \tag{1.1}$$

where

$$f(y;\eta,\phi) = \exp\left\{\frac{y\eta - b(\eta)}{a(\phi)} + c(y,\phi)\right\} . \tag{1.2}$$

In (1.1) and (1.2), $\theta^T = (\alpha,\beta^T)$, $a(\cdot)$, $b(\cdot)$ and $c(\cdot,\cdot)$ are known functions and $f(\cdot;\ \eta,\phi)$ is a density with respect to a σ-finite measure $m(\cdot)$. The mean and variance of (1.2) are $b'(\eta)$ and $a(\phi)b''(\eta)$ respectively. We assume $b''(\eta) > 0$ and thus $b'(\eta)$ is strictly increasing. For the regression model (1.1), the conditional mean and variance of $Y|U = u$ are $b'(\alpha + \beta^T u)$ and $a(\phi)b''(\alpha + \beta^T u)$.

There is a substantial literature on the theory and practice of fitting generalized linear models given independent observations $(Y_1,U_1),\cdots,(Y_n,U_n)$. Much of this is summarized in McCullugh & Nelder (1983). In this paper we present some theory concerning estimation of the regression parameter, θ, when the covariate U is measured with error. In particular we assume that one observes an independent measurement X, of U which is normally distributed given U, i.e.,

$$f_{X|U}(x|u;\bar{\Omega}) = \frac{(2\pi)^{-p/2}}{|\bar{\Omega}|^{\frac{1}{2}}} \exp\left\{- \tfrac{1}{2}(x-u)^T\bar{\Omega}^{-1}(x-u)\right\} . \tag{1.3}$$

There is a vast literature on this problem in the special case that (1.2) is a normal density. This dates back to Adcock (1878), has been reviewed by Moran (1971) and Anderson (1976) and is the topic of a recent book by Fuller (1987). Current interest in nonlinear measurement error models is evident from the work of Prentice (1982), Wolter & Fuller (1982a, 1982b), Amemiya (1982), Clark (1982), Carroll *et al.* (1984), Stefanski (1985), Armstrong (1985) and Stefanski and Carroll (1985, 1987). The latter six papers deal entirely or in part with generalized linear models.

We limit our investigation to the so-called structural version of the measurement-error model in which the unobserved covariable U has an unknown distribution $G(\cdot)$. This means that the joint density of an observed pair (Y,X) has the form

$$f_{Y,X}(y,x;\theta,\phi,\bar{\Omega},G) = \int f_{Y|U}(y|u;\theta,\phi)f_{X|U}(x|u;\bar{\Omega})G(du) \qquad (1.4)$$

The mixture density $f_{Y,X}$ depends on $\theta,\phi$ and $\bar{\Omega}$ as well as on the unknown mixing distribution, $G(\cdot)$. Our interest is in efficient estimation of $\theta$ in the presence of the nuisance function $G(\cdot)$. We will work with a subclass of models of the form (1.2) which is delineated in Section 2. In this context efficiency is defined in the sense of Stein (1956), Pfanzagl (1983), Begun *et al.* (1983) and Lindsay (1983, 1985).

The efficient score function and corresponding information bound for model (1.4) is derived in Stefanski & Carroll (1987) and these results are summarized in Section 2.3. In Section 3 the existence of efficient estimators is established using a construction similar to that employed by Bickel (1982). (See also Lindsay (1983, 1985).) Technical details appear in Section 4.

## 2. THE MODEL, IDENTIFIABILITY AND EFFICIENCY.

### 2.1 *Model assumptions.*

We now impose the following restrictions on the basic model introduced in Section 1.

(M1) The conditional distribution of $Y|U = u$ is concentrated on $[0,\infty]$, i.e., the response variable Y is nonnegative.

(M2) $\phi$ is known.

(M3) $\bar{\Omega}$ is known.

Of the more popular generalized linear models (McCullugh & Nelder, 1983) (M1) excludes only the normal model, for which Bickel & Ritov (1986) have already constructed efficient estimators under various identifiability assumptions. For Poisson regression and the important case of logistic regression, $a(\phi) = 1$ and thus (M2) imposes no restrictions on these models. However, for the gamma and inverse gaussian models (M2) imposes restrictions on the usual parameterizations. Assumption (M3) assures identifiability of the mixing distribution $G(\cdot)$. It is not an uncommon assumption in theoretical studies of nonlinear/nonnormal measurement-error models (Wolter & Fuller, 1982a & b; Stefanski, 1985; Stefanski & Carroll, 1985) and is a reasonable working assumption in many practical applications (Fuller, 1987).

Of these three assumptions only (M1) plays a crucial role in our one-step procedure and this via the Laplace transform of (1.2) (see Section 3). We believe that with conditions (M2) & (M3) replaced by any assumptions (regarding $\phi$, $\bar{\Omega}$ and/or replicate measurements) which guarantee identifiability of both the parametric and nonparametric components of (1.4), efficient one-step estimators can be constructed using the basic approach in Section 3, although we have not worked out the details.

Under (M2) and (M3) we assume without loss of further generality that $a(\phi) = 1$ and $\bar{\Omega}$ is the p×p identity matrix. To emphasize that θ and G are the only unknown components of $f_{Y,X}$ we write $f_{Y,X}(y,x;\theta,G)$ for the left hand side of (1.4). Also in (1.2) we shorten $c(y,\phi)$ to $c(y)$. The joint density $f_{Y,X}$ is determined by $\theta \in \Theta$, which we assume is an open subset of $\mathbb{R}^{p+1}$, and $G \in \mathcal{G}$, a set of distributions on $\mathbb{R}^p$. We assume that $\mathcal{G}$ contains only distribution which are absolutely continuous with respect to Lebesque measure on $\mathbb{R}^p$, a natural assumption in view of the normal error structure. For many models Θ will be all of $\mathbb{R}^{p+1}$ and there need be no additional restrictions on $\mathcal{G}$. However, for some generalized linear models there are natural constraints on the quantity $\eta=\alpha+\beta^T u$, usually of the form $\eta<0$, and for such models Θ and $\mathcal{G}$ must be restricted accordingly.

In what follows we write $F_{\theta,G}$ for the joint distribution determined by $f_{Y,X}(\cdot,\cdot;\theta,G)$, and $E_{\theta,G}$, $P_{\theta,G}$ and $L_{\theta,G}$ for expectation, probability and law when $(\theta,G)$ holds. Also we take $L(y,x,\theta,G) = \log f_{Y,X}(y,x;\theta,G)$ and $\ell(y,x,\theta,G) = (\partial/\partial\theta)L(y,x,\theta,G)$.

## 2.2 *Identifiability.*

Under (M3) the marginal distribution of X is the convolution of a known normal distribution with $G(\cdot)$ and thus $G(\cdot)$ is identified. The joint characteristic function of (Y,X) under $(\theta,G)$, $C(s,t;\theta,G) = E_{\theta,G}\{\exp(isY+it^TX)\}$, can be written as

$$C(s,t;\theta,G) = \int \exp\left\{-\frac{tt^T}{2} + itu + b(\alpha+\beta^Tu+is) - b(\alpha+\beta^Tu)\right\}G(du).$$

Equating $C(s,t;\theta_1,G)$ and $C(s,t,\theta_2,G)$ one finds that for all s and almost all u in the support of $G(\cdot)$,

$$b(\alpha_1+\beta_1^T u+is) - b(\alpha_1+\beta_1^T u) = b(\alpha_2+\beta_2^T u+is) - b(\alpha_2+\beta_2^T u).$$

Taking derivatives with respect to s and evaluating at s = 0 results in the equality

$$b'(a_1+\beta_1^T u) = b'(\alpha_2+\beta_2^T u),$$

which since $b'(\cdot)$ is strictly increasing implies

$$\alpha_1+\beta_1^T u = \alpha_2+\beta_2^T u. \tag{2.1}$$

And this must hold for almost all u in the support of $G(\cdot)$. Since $G(\cdot)$ is absolutely continuous, the closure of its support contains an open ball in $\mathbb{R}^p$ and by continuity (2.1) must hold in this open ball. This implies $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, thus the model is identified.

### 2.3 *Information bounds and efficient scores.*

This section reviews the results in Stefanski & Carroll (1987). Note that $f_{Y,X}(y,x;\theta,G)$ can be written as

$$f_{Y,X}(y,x;\theta,G) = \int\cdots\int q(\delta,\theta,u)r(y,x,\theta)G(du) \tag{2.2}$$

where

$$q(\delta,\theta,u) = \exp\left\{u^T\delta - \frac{u^Tu + 2b(\alpha+\beta^Tu)}{2}\right\}; \tag{2.3}$$

$$r(y,x,\theta) = \exp\left\{\frac{2\alpha y - x^Tx}{2} + C(y) - \frac{p}{2}\log 2\pi\right\}; \tag{2.4}$$

$$\delta = \delta(y,x,\theta) = x+y\beta . \tag{2.5}$$

Writing

$$\Delta = \Delta(Y,X,\theta) = X+Y\beta \tag{2.6}$$

it follows from Stefanski & Carroll (1987) that the efficient score for $\theta$ can be written as

$$\ell_c(y,x,\theta,G) = \ell(y,x,\theta,G) - E_{\theta,G}\left\{\ell(Y,X,\theta,G)\,|\,\Delta = \delta\right\}_{\delta=x+y\beta}$$

$$= \{y - E(Y_1|\Delta=\delta)\}\begin{Bmatrix} 1 \\ R(\delta,\theta,G)\end{Bmatrix}\Bigg|_{\delta=x+y\beta}, \tag{2.7}$$

where $R(\delta,\theta,G) = E(U|\Delta=\delta)$.

For short we let

$$\ell_{c,i}(\theta) = \ell_c(Y_i,X_i,\theta,G) \ . \tag{2.8}$$

In (2.7)

$$R(\delta,\theta,G) = \dot{w}(\delta)/w(\delta) \tag{2.9}$$

where

$$w(\delta) = \int\ldots\int q(\delta,\theta,u)G(du) \tag{2.10}$$

and

$$\dot{w}(\delta) = (\partial/\partial\delta)w(\delta) = \int\ldots\int u\ q(\delta,\theta,u)G(du). \tag{2.11}$$

Note that $\ell_c(y,x,\theta,G)$ depends on G only through $R(\delta,\theta,G)$ and to indicate this we sometimes write $\ell_c(y,x,\theta,R)$ in place of $\ell_c(y,x,\theta,G)$, i.e., both $\ell_c(y,x,\theta,G)$ and $\ell_c(y,x,\theta,R)$ are used to represent the right hand side of (2.7).

Finally, the lower bound for the variance of the asymptotic distribution of regular consistent estimators of $\theta$ is given by $I_c^{-1}(\theta,G)$ where

$$I_c(\theta,G) = E_{\theta,G}\{\ell_c\ell_c^T\}$$

$$= E_{\theta,G}\left[\mathrm{var}_{\theta,G}\left\{Y_1|\Delta_1(\theta)\right\}\bar{R}\bar{R}^T(\Delta_1,\theta,G)\right] \tag{2.12}$$

and $\bar{R}^T = (1,R^T)$.

$I_c(\theta,G)$ represents an information upper bound for the model (1.4) provided the family of distributions G is complete in the sense of assumption (C) in Stefanski & Carroll (1987).

## 2.4 *Existence of $n^{1/2}$-consistent estimators.*

The efficient estimator presented in Section 3 assumes the availability of a discretized $n^{1/2}$-consistent estimator of $\Theta$ which we now argue exists quite generally. Suppose in (2.7) $R(\delta,\Theta,G)$ is replaced by $\delta$. Denote the resulting estimating equation by $\Psi_{CL}(y,x,\Theta)$. Then $\Psi_{CL}(y,x,\Theta) = \{y-E(Y|\Delta=\delta)\}(1,\delta^T)^T$ where $\delta$ is evaluated at $x+y\beta$. Since $E_{\Theta,G}\{\Psi_{CL}(Y,X,\Theta)\} = 0$ for all $(\Theta,G)$, the system of equations $\Sigma\Psi_{CL}(Y_i,X_i,\Theta) = 0$ possesses a $n^{1/2}$-consistent sequence of solutions under minimal regularity conditions (Huber, 1967). Denote this estimator by $\hat{\Theta}_{CL}$. (There are even conditions under which $\hat{\Theta}_{CL}$ is optimal, namely when $R(\delta,\Theta,G)$ is a linear function of $\delta$; see Stefanski & Carroll (1987) for examples.) Now let $R_n^{p+1} = \left\{n^{-1/2}(i_1,\ldots,i_{p+1})^T: i_1,\ldots,i_{p+1}\right.$ are arbitrary integers$\}$ and define $\bar{\Theta}$ = the point in $R_n^{p+1}$ closest to $\hat{\Theta}_{CL}$. Then $\bar{\Theta}$ is the type of preliminary estimator required in Section 3. The construction of $\bar{\Theta}$ is identical to that employed by Bickel (1982) for a similar purpose.

## 3. AN EFFICIENT ESTIMATOR

### 3.1 *The basic approach.*

Although a number of details differ, our basic approach to constructing efficient estimators is similar to Bickel's (1982). Define

$$S_n(\theta) = n^{-1}\sum_{i=1}^{n} \ell_{c,i}(\theta),$$

and starting with the discretized $n^{\frac{1}{2}}$-consistent estimator $\bar{\theta}$, form the one-step "estimator"

$$\hat{\theta}_T = \bar{\theta} + I_c^{-1}(\bar{\theta},G)S_n(\bar{\theta}) \ . \tag{3.1}$$

Under (R1)-(R5) below it is shown (Theorem 3.1) that

$$L_{\theta_o,G}\{n^{\frac{1}{2}}(\hat{\theta}_T-\theta_o)\} \longrightarrow N\{0,I_c^{-1}(\theta_o,G)\},$$

i.e., $\hat{\theta}_T$ is asymptotically efficient. But, $\hat{\theta}_T$ is not a proper estimator as it depends on G which is unknown. We construct estimators $\hat{S}_n(\cdot)$ and $\hat{I}_n(\cdot)$ of $S_n(\cdot)$ and $I_c(\cdot,G)$ respectively and show (Theorem 3.2) that under mild assumptions the one-step estimator

$$\hat{\theta} = \bar{\theta} + \hat{I}_n^{-1}(\bar{\theta})\hat{S}_n(\bar{\theta})$$

is $n^{\frac{1}{2}}$-equivalent to $\hat{\theta}_T$, i.e. $n^{\frac{1}{2}}(\hat{\theta}_T - \hat{\theta}) = o_{P_{\theta_o,G}}(1)$ and thus $\hat{\theta}$ is efficient as well.

### 3.2. *The one-step construction.*

For convenience we assume a sample $(Y_i, X_i)$ $(i=1,\ldots,n)$ of even size, say $n = 2k$. Let $\mathcal{F}_{n,1} = \sigma\{(Y_1,X_1),\ldots,(Y_k,X_k)\}$ and $\mathcal{F}_{n,2} = \sigma\{(Y_{k+1},X_{k+1}),\ldots,(Y_n,X_n)\}$. For sequences $(t_n)$, $(\varepsilon_n)$ and $(\tau_n)$ of positive constants to be specified later define the following functions; here $|\cdot|$ denotes the Euclidean norm.

$$Q_{n,i}(\delta) = \left(\frac{1+\varepsilon_n}{\varepsilon_n}\right)^{p/2} \exp\left\{\frac{|\delta|^2}{2} - b(-t_n) - t_n Y_i - \frac{|X_i - \delta|^2}{2\varepsilon_n}\right\}; \quad (3.2)$$

$$\dot{Q}_{n,i}(\delta) = (\partial/\partial\delta) Q_{n,i}(\delta); \quad (3.3)$$

$$q_n(\delta,\theta,u) = \exp\left\{\frac{|\delta|^2}{2} + b(\alpha+\beta^T u - t_n) - b(-t_n) - b(\alpha+\beta^T u) - \frac{|u-\delta|^2}{2(1+\varepsilon_n)}\right\}; \quad (3.4)$$

$$W_{n,j}(\delta) = k^{-1} \sum_{i=(j-1)k+1}^{jk} Q_{n,i}(\delta), \quad j=1,2; \quad (3.5)$$

$$\dot{W}_{n,j}(\delta) = (\partial/\partial\delta) W_{n,j}(\delta), \quad j=1,2; \quad (3.6)$$

and

$$R_{n,j}(\delta) = \frac{(1+\varepsilon_n)\,\dot{W}_{n,j}(\delta)}{W_{n,j}(\delta) + \tau_n D_\lambda(\delta)}, \quad j=1,2, \quad (3.7)$$

with

$$D_\lambda(\delta) = \exp\left\{\frac{(1+\lambda)|\delta|^2}{2}\right\}. \quad (3.7a)$$

The quantity $\tau_n D_\lambda(\delta)$ in (3.7) insures that $|R_{n,j}(\delta)|^2$ has finite expectation. Our proof of Theorem 3.2 requires that $\lambda$ be greater than the $\lambda^*$ guaranteed by assumption (T) below. We assume this from now on. Note that by construction $R_{n,2}(\cdot)$ is independent of $\mathcal{F}_{n,1}$ and $R_{n,1}(\cdot)$ is independent of $\mathcal{F}_{n,2}$.

Next define

$$\hat{\ell}_{c,i}(\theta) = \begin{cases} \ell_c[Y_i, X_i, \theta, R_{n,2}\{\Delta_i(\theta)\}], & 1 \le i \le k \\ \ell_c[Y_i, X_i, \theta, R_{n,1}\{\Delta_i(\theta)\}], & k < i \le n \end{cases} \tag{3.8}$$

As estimators of $S_n(\theta)$ and $I_c(\theta,G)$ we propose

$$\hat{S}_n(\theta) = n^{-1} \sum_{i=1}^{n} \hat{\ell}_{c,i}(\theta), \quad \hat{I}_n(\theta) = n^{-1} \sum_{i=1}^{n} \hat{\ell}_{c,i}(\theta) \hat{\ell}^T_{c,i}(\theta) \tag{3.9}$$

resulting in the one-step estimator

$$\hat{\theta} = \bar{\theta} + \hat{I}_n^{-1}(\bar{\theta})\, \hat{S}_n(\bar{\theta}).$$

The construction is motivated by the facts that (Proposition 4.4)

$$E_{\theta,G}\{Q_{n,i}(\delta) | U_{i\cdot} = u\} = q_n(\delta,\theta,u)$$

and

$$E_{\theta,G}\{Q_{n,i}(\delta)\} = \int \cdots \int q_n(\delta,\theta,u) g(u) du.$$

Now under the conditions of Theorem 3.2, $q_n(\delta,\theta,u) \to q(\delta,\theta,u)$ and $E\{Q_{n,i}(\delta)\} \to w(\delta)$ given in (2.10) as $t_n \to \infty$, $\varepsilon_n \to 0$. Thus for $\tau_n \to 0$, $R_{n,j}(\delta) \to R(\delta,\theta)$ given in (2.9) and hence $\hat{\ell}_{c,i}(\theta)$ is approximating $\ell_{c,i}(\theta)$ which means that $\hat{S}_n(\theta)$ and $\hat{I}_n(\theta)$ are approximating $S_n(\theta)$ and $I_c(\theta,G)$ respectively. Roughly speaking, the proof that the construction works entails showing that for suitable rates of convergence of the sequences $t_n \to \infty$, $\varepsilon_n \to 0$ and $\tau_n \to 0$, $\hat{\ell}_{c,i}(\theta)$ converges to $\ell_{c,i}(\theta)$ in quadratic mean under $P_{\theta,G}$.

The most interesting feature of the construction is its simplicity. Splitting the sample into two equal halves is a modification of an idea due to Bickel (1982) and has the advantage of simplifying some of the proofs. The factor in the denominator of $R_{n,j}(\delta)$ involving $\tau_n$ is a convenient device insuring finite second moments of $|\hat{\ell}_i(\theta)|$.

To prove efficiency of our one-step estimator some regularity conditions on the model are necessary. For each fixed $G \in \mathcal{G}$ we assume the parametric submodel $f_{Y,X}(y,x;\theta,G)$ satisfies, for all $\theta \in \Theta$:

(R1) $L(Y,X,\theta,G)$ is differentiable in $\theta$ a.e. $P_{\theta,G}$ and $\ell(Y,X,\theta,G) = (\partial/\partial\theta)L(Y,X,\theta,G)$;

(R2) The Fisher information matrix $I(\theta,G) = E_{\theta,G}\{\ell\ell^T(Y,X,\theta,G)\}$ is finite;

(R3) The square root likelihood is differentiable in quadratic mean, i.e., as $\rho \to 0$,

$$E_{\theta,G}\left[\left\{\frac{f_{Y,X}(Y,X;\theta+\rho,G)}{f_{Y,X}(Y,X;\theta,G)}\right\}^{\frac{1}{2}} - 1 - \frac{\rho^T}{2}\ell(Y,X,\theta,G)\right]^2 = o(|\rho^2|),$$

and

$$P_{\theta+\rho,G}\left\{f_{Y,X}(Y,X;\theta,G) = 0\right\} = o(|\rho|^2) .$$

Conditions (R1) - (R3) are identical to Bickel's (1982) R(i) - R(iii). We use them as a means for insuring contiguity of the product measures of $(Y_i,X_i),\cdots,(Y_n,X_n)$ under $P_{\theta,G}$ and $P_{\theta_n,G}$ when $\theta_n = \theta_o + \rho_n n^{-\frac{1}{2}}$ for convergent deterministic sequences $(\rho_n)$.

In addition to (R1) - (R3) we also assume the following concerning the conditional score (2.7).

(R4) For all $\theta_o,\theta \in \Theta$ and $G \in \mathcal{G}$, the vector $H_G(\theta,\theta_o)$ defined as $H_G(\theta,\theta_o) = E_{\theta_o,G}\{\ell_c(Y_1,X_1,\theta,G)\}$ is differentiable in $\theta$ with $\dot{H}_G(\theta,\theta_o) = (\partial/\partial\theta)H_G(\theta,\theta_o)$, $\dot{H}_G(\theta,\theta_o)$ is continuous in $\theta$ and $\theta_o$ and $\dot{H}_G(\theta_o,\theta_o) = - I_c(\theta_o,G)$ which is nonsingular.

(R5) $\ell_c(Y,X,\theta,G)$ is continuous in quadratic mean, i.e., as $\rho \to 0$

$$E_{\theta,G}\{|\ell_c(Y_1,X_1,\theta+\rho,G) - \ell_c(Y_1,X_1,\theta,G)|^2\} = o(1).$$

These last two conditions enable us to show (Theorem 3.1) that the one-step "estimator" $\hat{\theta}_T$ in (3.1) based on the true conditional score is asymptotically efficient. Thus (R4) and (R5) play a role similar to Bickel's (1982) UR(iii).

Finally we make a technical assumption which for many special cases of interest is easily verified or is implied by moment conditions on G. First note that when $Y \geq 0$ a.s. it follows that $b'(-t) \geq 0$ for all t and $b'(-t)$ is decreasing in $t \geq 0$. Therefore $b'(-t)$ has a limit as $t \longrightarrow \infty$ which we now assume to be zero. This will be the case whenever Y is not bounded away from zero with probability one. Next define

$$A_\eta(t) = \frac{\exp\{b(\eta-t) - b(-t)\} - 1}{b'(-t)}, \tag{3.10}$$

$$A(\eta) = \sup_{t>0} |A_\eta(t)| \tag{3.11}$$

and

$$B_{\theta,G}(\delta,\varepsilon) = \int\cdots\int \{1+A(\alpha+\beta^T u\}\exp\left\{- b(\alpha+\beta^T u) - \frac{|u-\delta|^2}{2(1+\varepsilon)}\right\}g(u)du. \tag{3.11a}$$

We assume

(T) There exist $\varepsilon^* > 0$ and $\lambda^* < \infty$ such that for all $\theta \in \Theta$ and $G \in \mathcal{G}$

$$B_G(\theta) = 1 + \sup_\delta \left\{\exp\left(- \frac{\lambda^*|\delta|^2}{2}\right) B_{\theta,G}(\delta,\varepsilon^*)\right\} \tag{3.12}$$

is bounded in a neighborhood of $\theta$.

Note that since $B_{\theta,G}(\delta,\varepsilon)$ is increasing in $\varepsilon$ we can take $\varepsilon^* < 1$ without loss of generality. The addition of the constant one to the right hand side of (3.12) is for convenience only and serves to simplify some of the inequalities derived in Section 4. In Section 4 we make frequent use of the obvious inequality

$$B_{\theta,G}(\delta,\varepsilon^*) \le B_G(\theta)\exp\left(-\frac{\lambda^*|\delta|^2}{2}\right). \tag{3.12a}$$

### 3.3 *Asymptotic efficiency of* $\hat{\theta}$.

In this section we state and prove the major results concerning the efficiency of our one-step procedure. We start with Theorem 3.1.

Theorem 3.1. Under conditions (R1)-(R5)

$$L_{\theta_o,G}\left\{n^{\frac{1}{2}}(\hat{\theta}_T - \theta_o)\right\} \to N\left\{0,\ I_c^{-1}(\theta_o,G)\right\}.$$

**Proof.** Write

$$\hat{\theta}_T - \theta_o = I_c^{-1}(\theta_o,G)S_n(\theta_o) + \left\{I_c^{-1}(\bar{\theta},G) - I_c^{-1}(\theta_o,G)\right\}S_n(\bar{\theta})$$
$$+ [\bar{\theta} - \theta_o + I_c^{-1}(\theta_o,G)\{S_n(\bar{\theta}) - S_n(\theta_o)\}]. \tag{3.13}$$

By the Central Limit Theorem $L_{\theta_o,G}\{n^{\frac{1}{2}}I_c^{-1}(\theta_o,G)S_n(\theta_o)\} \to N\{0,I_c^{-1}(\theta_o,G)\}$ and thus to prove the theorem it suffices to show that the remaining two terms in (3.13) converge in $P_{\theta_o,G}$-probability to zero. Since $\bar{\theta}_n$ is discretized and $n^{\frac{1}{2}}$-consistent we need only establish this with $\bar{\theta}_n$ replaced by $\theta_n = \theta_o + \rho_n n^{-\frac{1}{2}}$, where $(\rho_n)$ is an arbitrary convergent deterministic sequence. Having made the substitution, (R4) and contiguity are used to show $\{I_c^{-1}(\theta_n,G) - I_c^{-1}(\theta_o,G)\}n^{\frac{1}{2}}S_n(\theta_n) = o_{P_{\theta_o,G}}(1)$. As for the third term in (3.13) note that it is enough to show $W_n = o_{P_{\theta_o,G}}(1)$, where

$$W_n = I_c(\theta_o,G)\rho_n + n^{\frac{1}{2}}\{S_n(\theta_o+\rho_n n^{-\frac{1}{2}}) - S_n(\theta_o)\}.$$

Using (R4)

$$E_{\theta_o,G}(W_n) = I_c(\theta_o,G)\rho_n + n^{-\frac{1}{2}}\{H_G(\theta_o+\rho_n n^{-\frac{1}{2}},\theta_o)\},$$

and for arbitrary but fixed $\gamma \in \mathbb{R}^p$ a Taylor series expansion shows that

$$\gamma^T E_{\theta_o,G}(W_n) = \gamma^T I_c(\theta_o,G)\rho_n + n^{-\frac{1}{2}}\{\gamma^T H_c(\theta_o+\rho_n n^{-\frac{1}{2}},\theta_o)\}$$

$$= \gamma^T I_c(\theta_o,G)\rho_n + \gamma^T \dot{H}_G(\tilde{\theta}_n,\theta_o)\rho_n$$

where $\tilde{\theta}_n$ is on the line segment joining $\theta_n$ and $\theta_o$. Since $(\rho_n)$ converges, (R4) implies $\gamma^T E_{\theta_o,G}(W_n) \to 0$ which, since $\gamma$ is arbitrary, implies $E(W_n) \to 0$. Finally we have

$$\text{var}_{\theta_o,G}(W_n) = \text{var}_{\theta_o,G}\{\ell_c(Y_1,X_1,\theta_n,G)-\ell_c(Y_1,X_1,\theta_o,G)\}$$

$$\leq E_{\theta_o,G}\{\ell_c(Y_1,X_1,\theta_n,G) -\ell_c(Y_1,X_1,\theta_o,G)\}^2$$

which converges to zero in light of (R5). Thus $W_n = o_{P_{\theta_o,G}}(1)$ concluding the proof. ////

The proof of our main result depends on the following lemma whose proof uses results from a number of propositions proved in Section 4. First we need some additional notation. We let

$$\Lambda_n = \exp\{b(-2t_n) - 2b(-t_n)\}; \qquad \Pi_n=2\{b'(-t_n) + \varepsilon_n\}.$$

<u>Lemma 3.1</u>. If $\lambda>\lambda^*$ and $\varepsilon_n,t_n$ and $\tau_n$ are chosen so that

$$\varepsilon_n \to 0,\ t_n \to \infty,\ \tau_n \to 0 \tag{3.14}$$

and

$$\max\left(\frac{\Pi_n}{\tau_n^2}, \frac{\Lambda_n}{n\varepsilon_n^{p/2}\tau_n^2}\right) \to 0 \tag{3.15}$$

then, as $n \to \infty$,

$$E_{\theta_n,G}\left(\mathrm{var}_{\theta_n,G}\{Y_1|\Delta_1(\theta_n)\}E_{\theta_n,G}\left[|R_{n,2}\{\Delta_1(\theta_n)\}-R\{\Delta_1(\theta_n),\theta_n,G\}|^2\Big|\Delta_1(\theta_n)\right]\right)\to 0. \tag{3.16}$$

<u>Proof</u>. After some rearrangement of terms we have

$$R_{n,2}(\delta) - R(\delta,\theta_n,G) = \frac{(1+\varepsilon_n)\dot{W}_{n,2}(\delta) - \dot{w}(\delta,\theta_n,G)}{W_{n,2}(\delta) + \tau_n D_\lambda(\delta)}$$

$$- R(\delta,\theta_n,G)\left\{\frac{W_{n,2}(\delta)-w(\delta,\theta_n,G)+\tau_n D_\lambda(\delta)}{W_{n,2}(\delta)+\tau_n D_\lambda(\delta)}\right\}.$$

Repeated use of the inequality $|a+b|^2 \le 2(|a|^2+|b|^2)$ is used to show

$$|R_{n,2}(\delta) - R(\delta,\theta_n,G)|^2 \le 2\frac{|(1+\varepsilon_n)\dot{W}_{n,2}(\delta) - \dot{w}(\delta,\theta_n,G)|^2}{\{W_{n,2}(\delta) + \tau_n D_\lambda(\delta)\}^2}$$

$$+ 4|R(\delta,\theta_n,G)|^2\left[\left\{1 - \frac{W_{n,2}(\delta)}{W_{n,2}(\delta)+\tau_n D_\lambda(\delta)}\right\}^2 + \left\{\frac{W_{n,2}(\delta)-w(\delta,\theta_n,G)}{W_{n,2}+\tau_n D_\lambda(\delta)}\right\}^2\right]$$

$$\le 2\,\frac{|(1+\varepsilon_n)\dot{W}_{n,2}(\delta)-\dot{w}(\delta,\theta_n,G)|^2}{\tau_n^2 D_\lambda^2(\delta)}$$

$$+ 4|R(\delta,\theta_n,G)|^2\left[\left\{1- \frac{W_{n,2}(\delta)}{W_{n,2}(\delta)+\tau_n D_\lambda(\delta)}\right\}^2 + \left\{\frac{W_{n,2}(\delta)-w(\delta,\theta_n,G)}{\tau_n D_\lambda(\delta)}\right\}^2\right]$$

Now using (4.11) and (4.16) we are able to conclude that

$$E_{\theta_n,G}|R_{n,2}(\delta) -R(\delta,\theta_n,G)|^2 \le$$

$$\frac{2\kappa_6\left(\dfrac{\Lambda_n}{n\varepsilon_n^{p/2+1}} + \Pi_n^2\right)B_G^2(\theta_n)(1+|\delta|)^2\exp\left\{\dfrac{(2+2\lambda^*)|\delta|^2}{2}\right\}}{\tau_n^2 D_\lambda^2(\delta)}$$

$$+ \quad 4|R(\delta,\theta_n,G)|^2\,\frac{\kappa_3\left(\Pi_n^2+ \dfrac{\Lambda_n}{n\varepsilon_n^{p/2}}\right)B_G^2(\theta_n)\exp\left\{\dfrac{(2+2\lambda^*)|\delta|^2}{2}\right\}}{\tau_n^2 D_\lambda^2(\delta)}$$

$$+ 4|R(\delta,\theta_n,G)|^2 C_n(\delta) \tag{3.17}$$

where

$$C_n(\delta) = E_{\theta_n,G}\left\{1 - \frac{W_{n,2}(\delta)}{W_{n,2}(\delta) + \tau_n D_\lambda(\delta)}\right\}^2 . \tag{3.18}$$

Now for $\lambda > \lambda^*$, $(1+|\delta|)^2\exp\{(1+\lambda^*)|\delta|^2\}$ is bounded by a constant multiple of $D_\lambda^2(\delta)$ and thus there exists a finite constant, $\kappa_7$, such that

$$E_{\theta_n,G}\left|R_{n,2}(\delta) - R(\delta,\theta_n,G)\right|^2 \le \frac{\kappa_7\left(\frac{\Lambda_n}{n\varepsilon_n^{p/2+1}} + \Pi_n^2\right)B_G^2(\theta_n)}{\tau_n^2}$$

$$+ 4\left|R(\delta,\theta_n,G)\right|^2\left\{\frac{\kappa_7\left(\frac{\Lambda_n}{n\varepsilon_n^{p/2}} + \Pi_n^2\right)B_G^2(\theta_n)}{\tau_n^2} + C_n(\delta)\right\} . \tag{3.19}$$

It follows from Proposition 4.6 that $W_{n,2}(\delta) - w(\delta,\theta,G) = o_{P_{\theta_n,G}}(1)$ and thus the bracketed quantity in (3.18) converges to zero in $P_{\theta_n,G}$-probability. As it is bounded, it also converges to zero in mean square, i.e., $C_n(\delta)$ converges to zero for each fixed $\delta \in \mathbf{R}$. Using the inequality in (3.19), noting the definition of $I_c(\theta,G)$ in (2.12) and recalling the independence of $R_{n,2}$ and $\Delta_1(\theta_n)$ we can now bound the quantity in (3.16) by

$$\left\{\frac{\kappa_7\left(\Pi_n^2 + \frac{\Lambda_n}{n\varepsilon_n^{p/2+1}}\right)B_G^2(\theta_n) + 4\kappa_7\left(\Pi_n^2 + \frac{\Lambda_n}{n\varepsilon_n^{p/2}}\right)B_G^2(\theta_n)}{\tau_n^2}\right\}\mathrm{tr}\left\{I_c(\theta_n,G)\right\} + 4T_n, \tag{3.20}$$

where

$$T_n = E_{\theta_n,G}\Big[\mathrm{var}_{\theta_n,G}\{Y_1|\Delta_1(\theta_n)\}|R\{\Delta_1(\theta_n),\theta_n,G\}|^2\ C_n\{\Delta_1(\theta_n)\}\Big]. \qquad (3.21)$$

Assumptions (R2), (R4) and (T) show that the first term in (3.20) converges to zero when (3.15) holds. Note that $T_n \geq 0$ and furthermore

$$tr^{(1)}\Big\{I_c(\theta_n,G)\Big\} - T_n = E_{\theta_n,G}\Big(\mathrm{var}_{\theta_n,G}\{Y_1|\Delta_1(\theta_n)\}|R\{\Delta_1(\theta_n),\theta_n,G\}|^2 \times[1 - C_n\{\Delta_1(\theta_n)\}]\Big), \qquad (3.22)$$

where $tr^{(1)}(M)$, for a square matrix $M = (m_{ij})$, is defined as $tr^{(1)}(M) = tr(M)-m_{11}$. If we let $f(\delta;\theta_n)$ denote the density of $\Delta_1(\theta_n)$ under $P_{\theta_n,G}$, then (3.22) can be written as

$$tr^{(1)}\Big\{I_c(\theta_n,G)\Big\}-T_n=\int\cdots\int \mathrm{var}_{\theta_n,G}\{Y_1|\Delta_1(\theta_n)=\delta\}|R\{\delta,\theta_n,G\}|^2\{1-C_n(\delta)\}f(\delta;\theta_n)d\delta.$$

Now as $n \to \infty$, $I_c(\theta_n,G) \to I_c(\theta_o,G)$, $\mathrm{var}_{\theta_n,G}\{Y_1|\Delta_1(\theta_n) = \delta\} \to \mathrm{var}_{\theta_o,G}\{Y_1|\Delta_1(\theta_o) = \delta\}$, $R(\delta,\theta_n,G) \to R(\delta,\theta_o,G)$, $f(\delta;\theta_n) \to f(\delta;\theta_o)$ and $C_n(\delta) \to 0$.

Thus, by an application of Fatou's Lemma we find

$$I_c(\theta_o,G) - \limsup T_n \geq \int\cdots\int \mathrm{var}_{\theta_o,G}\{Y_1|\Delta_1(\theta_o)=\delta\}|R(\delta,\theta_o G)|^2 f(\delta;\theta_o)d\delta = I_c(\theta_o,G),$$

which since $T_n \geq 0$ implies $T_n \to 0$. This completes the proof. ////

<u>Theorem 3.2</u>. Under (M1)-(M3), (R1)-(R5), (T) and (3.14) and (3.15),

$$n^{\frac{1}{2}}\{\hat{S}_n(\bar{\theta}) - S_n(\bar{\theta})\} = o_{P_{\theta_o,G}}(1); \qquad (3.23)$$

$$\hat{I}_n(\bar{\theta}) - I_c(\bar{\theta},G) = o_{P_{\theta_o,G}}(1); \qquad (3.24)$$

$$L_{\theta_o,G}\{n^{\frac{1}{2}}(\hat{\theta} - \theta_o)\} \rightarrow N\{0, I_c^{-1}(\theta_o, G)\}. \tag{3.25}$$

<u>Proof</u>. By the discretization of $\bar{\theta}$ it suffices to prove (3.23) and (3.24) with $\bar{\theta}$ replaced by $\theta_n = \theta_o + \rho_n n^{-\frac{1}{2}}$ where $\rho_n$ is an arbitrary deterministic convergent sequence. Also by appealing to contiguity it is enough to establish (3.23) and (3.24) in $P_{\theta_n,G}$ - probability.

Note that by the manner in which $\hat{S}_n(\cdot)$ is constructed, to prove (3.23) it need only be shown that

$$n^{-\frac{1}{2}} \sum_{i=1}^{k} T_{n,i} = o_{P_{\theta_n,G}}(1), \tag{3.26}$$

where

$$T_{n,i} = \{Y_i - M_i(\theta_n)\}[R_{n,2}\{\Delta_i(\theta_n)\} - R\{\Delta_i(\theta_n), \theta_n, G\}]$$

and

$$M_i(\theta_n) = E_{\theta_n,G}\{Y_i | \Delta_i(\theta_n)\} \quad .$$

Now it follows from Lemma 3.1 that $|T_{n,i}|^2$ has finite expectation under $P_{\theta_n,G}$. Thus, using the fact that $\mathscr{F}_{n,2}$ and $\{Y_i, \Delta_i(\theta_n)\}$ are independent for $1 \le i \le k$, we have

$$E_{\theta_n,G}(T_{n,i}) = E_{\theta_n,G}\left[E_{\theta_n,G}\{T_{n,i} | \Delta_i(\theta_n), \mathscr{F}_{n,2}\}\right]$$

$$= E_{\theta_n,G}\left(\left[R_{n,2}\{\Delta_i(\theta_n)\} - R\{\Delta_i(\theta_n), \theta_n, G\}\right] E_{\theta_n,G}\{Y_i - M_i(\theta_n) \big| \Delta_i(\theta_n)\}\right)$$

$$= 0.$$

Also for $i \neq j$, $1 \le i,j \le k$ a similar conditioning argument shows that $E_{\theta_n,G}(T_{n,i}T_{n,j}^T) =$ 0 and thus the normed sum in (3.26) has a zero mean and a covariance matrix bounded in norm by

$$E_{\theta_n,G}\Big[n^{-1}\sum_{i=1}^{k}\{Y_i-M_i(\theta_n)\}^2\,|R_{n,2}\{\Delta_i(\theta_n)\} - R\{\Delta_i(\theta_n),\theta_n,G\}|^2\Big]. \quad (3.27)$$

Showing (3.27) converges to zero will complete the proof of (3.23). Since the summands in (3.27) are identically distributed it is enough to show

$$E_{\theta_n,G}\Big[\{Y_1-M_1(\theta_n)\}^2\,|R_{n,2}\{\Delta_1(\theta_n)\}-R\{\Delta_1(\theta_n),\theta_n,G\}|^2\Big] \quad (3.28)$$

converges to zero. After conditioning on $\Delta_1(\theta_n)$ and using the independence of $\{Y_1,\Delta_1(\theta_n)\}$ and $\mathcal{F}_{n,2}$, (3.28) can be written as

$$E_{\theta_n,G}\Big(\text{var}_{\theta_n,G}\{Y_1|\Delta_1(\theta_n)\}E_{\theta_n,G}[\,|R_{n,2}\{\Delta_1(\theta_n)\}-R\{\Delta_1(\theta_n),\theta_n,G\}|^2\,\big|\Delta_1(\theta_n)]\Big). \quad (3.29)$$

which converges to zero in light of Lemma 3.1.

Now since

$$I_c(\theta_n,G) = E_{\theta_n,G}\{\ell_{c,1}(\theta_n)\ell_{c,1}^T(\theta_n)\} \quad (3.30)$$

and

$$n^{-1}\sum_{i=1}^{n}\ell_{c,i}(\theta_n)\ell_{c,i}^T(\theta_n) - E_{\theta_n,G}\Big\{\ell_{c,1}(\theta_n)\ell_{c,1}^T(\theta_n)\Big\} = o_{P_{\theta_n,G}}(1), \quad (3.31)$$

to show (3.24) converges to zero it need only be demonstrated that

$$n^{-1}\sum_{i=1}^{k}\Big\{\ell_{c,i}(\theta_n)\ell_{c,i}^T(\theta_n) - \hat{\ell}_{c,i}(\theta_n)\hat{\ell}_{c,i}^T(\theta_n)\Big\} = o_{P_{\theta_n,G}}(1). \quad (3.32)$$

Now in light of (3.30), (3.31), and (R4), (3.32) is implied by the fact that (3.28) converges to zero, which has already been established.

Finally for (3.25), note that (3.23 ) and (3.24) imply $n^{\frac{1}{2}}(\hat{\theta}_T - \hat{\theta}) = o_{P_{\theta_o,G}}(1)$ and thus $\hat{\theta}$ and $\hat{\theta}_T$ have the same limiting distribution which is given by Theorem 3.1. ////

### 3.4 *Illustration*

Consider the logistic regression model in which $P(Y=1|U=u) = F(\alpha+\beta^T u)$ where $F(t) = 1/(1+e^{-t})$. Assumptions M1 and M2 both hold and M3 is a reasonable working assumption in the absence of replicated measurements. For this model $b(\eta) = \log(1+e^{\eta})$, $A_{\eta}(t) = e^{\eta} - 1$, $A(\eta) = |e^{\eta}-1|$ and $e^{-b(\eta)}A(\eta) = |e^{\eta}-1|/(1+e^{\eta}) \leq 1$, which implies that assumption (T) is satisfied with $\lambda^* = 0$ and for all $\varepsilon^* > 0$ (the technical assumption (T) imposes no additional constraints). (R1)-(R5) hold under natural and reasonable integrability conditions on $\mathcal{b}$.

To see that (3.15) holds note first that

$$\Lambda_n = \{1 + \exp(-2t_n)\}/\{1 + \exp(-t_n)\}^2 \to 1$$

and $\Pi_n = 2\{F(-t_n)+ \varepsilon_n\} \to 0$ as $t_n \to \infty$ and $\varepsilon_n \to 0$. For (3.15) to hold it is required that

$$\text{(a)} \quad \{F(-t_n) + \varepsilon_n\}/\tau_n^2 \to 0$$

and

$$\text{(b)} \quad 1/(n\, \tau_n^2\, \varepsilon_n^{p/2}) \to 0$$

simultaneously as $t_n \to \infty$, $\varepsilon_n \to 0$ and $\tau_n \to 0$. There are many choices of $(\varepsilon_n)$, $(t_n)$ and $(\tau_n)$ which accomplish this. For example, take $\varepsilon_n = n^{-1/p}$ and $t_n = n$ then (a) and (b) hold for the choice $\tau_n = \max[\{F(-n) + n^{-1/p}\}^{1/4}, n^{-1/8}]$. Finally in the construction of $R_{n,j}(\delta)$ in (3.7) we can pick any $\lambda > 0$ in $D_{\lambda}(\delta)$ (since (T) holds with $\lambda^* = 0$), e.g., $D_{\lambda}(\delta) = \exp(|\delta|^2)$ will do.

## 4. THEORETICAL DETAILS.

In this section we derive the results used to establish Lemma 3.1. For notational convenience we write E for $E_{\theta,G}$, drop the subscripts on $t_n$ and $\varepsilon_n$ and let $\eta = \alpha+\beta^T u$ and $d = |u-\delta|/2^{\frac{1}{2}}$. Also we take (Y,X,U) to be random variables having joint density (1.4) under $P_{\theta,G}$. Thus for instance, we can write $E(Y|U=u) = b'(\eta)$. We also assume t is sufficiently large so that $b'(-t) < 1$. We make frequent reference to the quantities $B_{\theta,G}$ and $B_G$ defined in (3.11a) and (3.12) and to the inequality in (3.12a.) Whenever it is convenient and causes no confusion we write $q_n$ and q for $q_n(\delta,\theta,u)$ and $q(\delta,\theta,u)$ respectively. Finally $\kappa_1,\cdots,\kappa_6$ denote finite positive constants which will serve as bounds for certain quantities we encounter. We start by obtaining some conditional expectations.

PROPOSITION 4.1. If $t>0$ and $\varepsilon>0$ then,

$$E\left(e^{-tY}|U=u\right) = \exp\{b(\eta-t)-b(\eta)\}; \tag{4.1}$$

$$E\left\{\left(\frac{1+\varepsilon}{\varepsilon}\right)^{p/2} \exp\left(-\frac{|X-\delta|^2}{2\varepsilon}\right)\middle|U=u\right\} = \exp\left\{\frac{-d^2}{(1+\varepsilon)}\right\} \quad ; \tag{4.2}$$

$$E\left\{\frac{|X-\delta|^2}{2} \exp\left(-\frac{|X-\delta|^2}{2\varepsilon}\right)\middle|U=u\right\} =$$

$$\left(\frac{\varepsilon}{1+\varepsilon}\right)^{p/2+1}\left(\frac{p}{2}+\frac{\varepsilon d^2}{1+\varepsilon}\right)\exp\left\{\frac{-d^2}{(1+\varepsilon)}\right\} \quad . \tag{4.3}$$

PROOF. The first claim follows from the fact that the density of $Y|U=u$ is of exponential-family type. Derivation of the latter two identities is a routine exercise in manipulating normal integrals. ////

For future reference we state two easy inequalities. Note that (4.4) implies that $\exp\{b(\eta-t)-b(-t)\} \le A(\eta)+1$ provided $b'(-t) < 1$.

PROPOSITION 4.2. For $t>0$ and $\varepsilon>0$,

$$\left|\exp\{b(\eta-t)-b(-t)\}-1\right| \le A(\eta)b'(-t) \tag{4.4}$$

$$\left|\exp\left(\frac{\varepsilon d^2}{1+\varepsilon}\right) - 1\right| \le \varepsilon d^2 \exp\left(\frac{\varepsilon d^2}{1+\varepsilon}\right) \tag{4.5}$$

PROOF. The first follows from the definition of $A(\eta)$ while the second is a variation of a familiar inequality. ////

Propositions 4.3-4.5 result in a bound (Proposition 4.6) on the mean squared error of $W_{n,2}$ as an estimator of $w(\delta)$. Recall that $\varepsilon^* < 1$.

PROPOSITION 4.3. If $\varepsilon<\frac{1}{2}\varepsilon^*$ then

$$\max(1,d)\left|q_n - q\right| \le \kappa_1 \Pi_n \{1+A(\eta)\}\exp\left\{\frac{|\delta|^2}{2} - b(\eta) - \frac{d^2}{1+\varepsilon^*}\right\} . \tag{4.6}$$

PROOF. For fixed $\delta,\eta$ and $d$, let the function $f(\cdot)$ be defined as

$$f(\varepsilon) = \exp\left\{\frac{|\delta|^2}{2} - b(\eta) - \frac{d^2}{1+\varepsilon}\right\} .$$

Note that $f(\cdot)$ increases over $(0,\infty)$ and furthermore that for $\varepsilon < \frac{1}{2}\varepsilon^* < 1$,

$$f(\varepsilon) \le f(\tfrac{1}{2}\varepsilon^*) = f(\varepsilon^*)\exp\left\{\frac{-\varepsilon^* d^2}{(1+\varepsilon^*)(2+\varepsilon^*)}\right\}$$

$$\le f(\varepsilon^*)\exp\left(-\frac{\varepsilon^* d^2}{6}\right) .$$

Using this fact, the definitions of $q_n$ and $q$, an obvious inequality, (4.4), (4.5) and the definition of $\Pi_n$ we derive the following inequality.

$$|q_n - q| = f(\varepsilon)\left[\exp\{b(\eta-t)-b(-t)\}-\exp\left(\frac{-\varepsilon d^2}{1+\varepsilon}\right)\right]$$

$$\le f(\varepsilon)\left[\,|\exp\{b(\eta-t)-b(-t)\}-1|+\left|1-\exp\left(\frac{-\varepsilon d^2}{1+\varepsilon}\right)\right|\right]$$

$$\le f(\varepsilon^*)\exp\left(\frac{-\varepsilon^* d^2}{6}\right)\{A(\eta)b'(-t) + \varepsilon d^2\}$$

$$\le \tfrac{1}{2}\Pi_n f(\varepsilon^*)\exp\left(\frac{-\varepsilon^* d^2}{6}\right)\{A(\eta)+d^2\} \ .$$

The existance of a constant $\kappa_1$ such that

$$\max(1,d)\exp\left(\frac{-\varepsilon^* d^2}{6}\right)\{A(\eta)+d^2\} \le 2\kappa_1\{1+A(\eta)\}$$

is easily established and this completes the proof of the proposition.////

<u>PROPOSITION 4.4.</u> If $\varepsilon<\varepsilon^*$ and (T) holds for some $\lambda^* < \infty$, then

$$E\left\{Q_{i,n}(\delta)\,\middle|\,U_i=u\right\} = q_n(\delta,\theta,u); \tag{4.7}$$

$$E\left\{Q_{i,n}(\delta)\right\} = \int\cdots\int q_n(\delta,\theta,u)g(u)du$$

$$\le B_G(\theta)\ \exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\}; \tag{4.8}$$

$$E\left\{Q_{i,n}^2(\delta)\right\} \le \kappa_2\ \frac{A_n}{\varepsilon^{p/2}}\ B_G(\theta)\exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\} \ . \tag{4.9}$$

<u>PROOF</u>. The identity in (4.7) follows from (4.1), (4.2) and the conditional independence of Y and X given U.

The equality in (4.8) is obtained by integrating the conditional expectation in (4.7). As for the inequality in (4.8) note that from (4.4) it follows that $\exp\{b(\eta-t)-b(-t)\} \le A(\eta) + 1$, if $b'(-t) \le 1$. Thus we have

$$q_n(\delta,\theta,u) = \exp\left\{\frac{|\delta|^2}{2} + b(\eta-t)-b(-t)-b(\eta)- \frac{d^2}{1+\varepsilon}\right\}$$

$$\leq \{A(\eta)+1\}\exp\left\{\frac{|\delta|^2}{2} - b(\eta) - \frac{d^2}{1+\varepsilon}\right\} .$$

It follows that for $\varepsilon<\varepsilon^*$

$$E\{Q_{i,n}(\delta)\} \leq \exp\left(\frac{|\delta|^2}{2}\right) \int \cdots \int \{A(\eta)+1\}\exp\left\{-b(\eta)- \frac{d^2}{1+\varepsilon}\right\}g(u)du$$

$$= \exp\left(\frac{|\delta|^2}{2}\right)B_{\theta,G}(\delta,\varepsilon)$$

$$\leq \exp\left(\frac{|\delta|^2}{2}\right)B_{\theta,G}(\delta,\varepsilon^*)$$

$$\leq \exp\left\{\frac{(\lambda^*+1)|\delta|^2}{2}\right\}B_G(\theta),$$

where we have used the properties of the quantities $B_{\theta,G}$ and $B_G$ defined in (3.11a) and (3.12) and the inequality in (3.12a).

To prove (4.9) note that

$$E\left\{Q_{i,n}^2(\delta)\right\} = \int \cdots \int E\left[\left(\frac{1+\varepsilon}{\varepsilon}\right)^p \exp\left\{|\delta|^2-2b(-t)-2tY_i - \frac{|X_i-\delta|^2}{\varepsilon}\right\}\middle| U_i = u\right]g(u)du.$$

The conditional expectation in the integral is evaluated using (4.1) and (4.2) (with t replaced by 2t and $\varepsilon$ replaced by $\frac{1}{2}\varepsilon$), and the conditional independence of Y and X given U resulting in the identity,

$$E\left\{Q_{i,n}^2(\delta)\right\} = \left(\frac{1+\varepsilon}{\varepsilon}\right)^p\left(\frac{\varepsilon}{2+\varepsilon}\right)^{p/2}\exp(|\delta|^2)\Lambda_n\int \cdots \int \exp\left\{b(\eta-2t)-b(-2t)-b(\eta)- \frac{d^2}{1+\frac{1}{2}\varepsilon}\right\}g(u)du.$$

Again we use (4.4) to claim $\exp\{b(\eta-2t)-b(-2t)\} \leq A(\eta)+1$ and arrive at

$$E\left\{Q_{i,n}^2(\eta)\right\} \leq \kappa_2 \frac{\Lambda_n}{\varepsilon^{\frac{1}{2}p}} \exp(|\delta|^2)\ B_{\theta,G}(\delta,\tfrac{1}{2}\varepsilon)$$

$$\leq \kappa_2 \frac{\Lambda_n}{\varepsilon^{\frac{1}{2}p}} \exp(|\delta|^2)\ B_{\theta,G}(\delta,\varepsilon^*)$$

$$\leq \kappa_2 \frac{\Lambda_n}{\varepsilon^{\frac{1}{2}p}} \exp\left\{\frac{(\lambda^*+2)|\delta|^2}{2}\right\} B_G(\theta)\ ,$$

where $\kappa_2$ is a bound on $(1+\varepsilon)^p(2+\varepsilon)^{-\frac{1}{2}p}$ and $B_{\theta,G}$ and $B_G$ are as defined in (3.11a) and (3.12). ////

PROPOSITION 4.5. If $\varepsilon<\frac{1}{2}\varepsilon^*$ and (T) holds for some $\lambda^*<\infty$, then

$$\left|E\left\{W_{n,2}(\delta)-w(\delta,\theta,G)\right\}\right| \leq \kappa_1 \Pi_n B_G(\theta)\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\}. \qquad (4.10)$$

PROOF. Using 4.6, (4.8) and some obvious inequalities we find that

$$\left|E\left\{W_{n,2}(\delta)-w(\delta,\theta,G)\right\}\right| \leq \int\cdots\int \left|q_n - q\ \right| g(u)du$$

$$\leq \kappa_1 \Pi_n \exp\left(\frac{|\delta|^2}{2}\right) B_{\theta,G}(\delta,\varepsilon^*)$$

$$\leq \kappa_1 \Pi_n B_G(\theta)\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\} \qquad ////$$

Finally we have

PROPOSITION 4.6. If $\varepsilon<\frac{1}{2}\varepsilon^*$ and (T) holds for some $\lambda^*<\infty$, then

$$E\{W_{n,2}(\delta)-w(\delta,\theta,G)\}^2 \leq$$

$$\kappa_3\left(\Pi_n^2 + \frac{\Lambda_n}{n\varepsilon^{p/2}}\right)\left\{B_G(\theta)\right\}^2 \exp\left\{\frac{(2+2\lambda^*)|\delta|^2}{2}\right\}\ . \qquad (4.11)$$

PROOF. Since $W_{n,2}(\delta) = k^{-1}\sum Q_{n,i}(\delta)$ and the $Q_{n,i}$ are independent and identically distributed, (4.9) provides a bound on $k\mathrm{var}\{W_{n,2}(\delta)\}$. Inequality (4.10) bounds the bias of $W_{n,2}(\delta)$ as an estimator of $w(\delta,\theta,G)$. Combining these facts we get, with $k = \frac{1}{2}n$,

$$E\{W_{n,2}(\delta) - w(\delta,\theta,G)\}^2 \le \kappa_2\left(\frac{2\Lambda_n}{n\varepsilon^{p/2}}\right)B_G(\theta)\exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\}$$

$$+ \kappa_1^2\Pi_n^2B_G^2(\theta)\exp\left\{\frac{(2+2\lambda^*)|\delta|^2}{2}\right\}$$

$$\le \kappa_3\left(\frac{\Lambda_n}{n\varepsilon^{p/2}} + \Pi_n^2\right)B_G^2(\theta)\exp\left\{\frac{(2+2\lambda^*)|\delta|^2}{2}\right\},$$

where in the last inequality we have used the fact that $B_G(\theta) \le B_G^2(\theta)$.

////

Next Propositions 4.7 and 4.8 result in a bound (Proposition 4.9) on the mean squared error of $(1+\varepsilon)\dot{W}_{n,2}(\delta)$ as an estimator of $\dot{w}(\delta)$.

PROPOSITION 4.7 If $\varepsilon<\varepsilon^*/2$ and (T) holds for some $\lambda^*<\infty$, then

$$E\left\{\dot{Q}_{n,i}(\delta)\middle|U_i=u\right\} = \frac{u+\varepsilon\delta}{1+\varepsilon}\, q_n(\delta,\theta,u); \tag{4.12}$$

$$E\left\{\dot{Q}_{n,i}(\delta)\right\} = \int\cdots\int\frac{u+\varepsilon\delta}{1+\varepsilon}\, q_n(\delta,\theta,u)g(u)du; \tag{4.13}$$

$$E\left\{|\dot{Q}_{n,i}(\delta)|^2\right\} \le \kappa_4\left(\frac{\Lambda_n}{\varepsilon^{p/2+1}}\right) B_G(\theta)\left(1+|\delta|^2\right)\exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\}. \tag{4.14}$$

PROOF. Equation (4.12) can be verified by differentiating both sides of (4.7) with respect to $\delta$, and (4.13) follows directly from (4.12).

As for (4.14) note that

$$\dot{Q}_{i,n}(\delta) = \left(\delta + \frac{X_i-\delta}{\varepsilon}\right)Q_{i,n}(\delta)$$

and thus

$$E\left\{\left|\dot{Q}_{i,n}(\delta)\right|^2\right\} \le 2|\delta|^2\, E\left\{Q^2_{i,n}(\delta)\right\} + \frac{4}{\varepsilon^2}\, E\left\{\frac{|X_i - \delta|^2}{2} Q^2_{i,n}(\delta)\right\}. \qquad (4.14a)$$

We first find bounds for each of the quantities on the right hand side of (4.14a) and then combine these to arrive at (4.14).

Using (4.9) we get immediately that

$$2|\delta|^2 E\left\{Q^2_{i,n}(\delta)\right\} \le \frac{2\kappa_2\Lambda_n}{\varepsilon^{p/2}} B_G(\theta)|\delta|^2 \exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\}. \qquad (4.14b)$$

Next note that

$$\frac{4}{\varepsilon^2} E\left\{\frac{|X_i - \delta|^2}{2} Q^2_{i,n}(\delta)\right\} = \frac{4}{\varepsilon^2}\left(\frac{1+\varepsilon}{\varepsilon}\right)^p \exp\left\{|\delta|^2 - 2b(-t)\right\} E\left[\frac{|X_i-\delta|^2}{2} \exp\left\{-\frac{|X_i-\delta|^2}{\varepsilon} - 2tY_i\right\}\right].$$

Conditioning on $U_i$, appealing to (4.1) and (4.3) (with t replaced by 2t and $\varepsilon$ by $\frac{1}{2}\varepsilon$) and using the definition of $\Lambda_n$ we obtain the following expression for the right hand side above:

$$\frac{4\Lambda_n}{\varepsilon^2}\left(\frac{1+\varepsilon}{\varepsilon}\right)^p \left(\frac{\varepsilon}{2+\varepsilon}\right)^{\frac{1}{2}p+1} \exp(|\delta|^2)$$

$$\times \int\cdots\int \left(\frac{p}{2} + \frac{\varepsilon d^2}{2+\varepsilon}\right) \exp\left\{-\left(\frac{2d^2}{2+\varepsilon}\right) + b(\eta-2t) - b(-2t) - b(\eta)\right\} g(u)du.$$

Note that when $\varepsilon < \frac{1}{2}\varepsilon^*$ there exists a constant $\kappa$, depending on p and $\varepsilon^*$, such that for all d,

$$\left(\frac{p}{2} + \frac{\varepsilon d^2}{2+\varepsilon}\right) \exp\left(-\frac{2d^2}{2+\varepsilon}\right) \le \kappa \exp\left(\frac{-d^2}{1+\varepsilon^*}\right).$$

Using this inequality and the fact that by (4.4), $\exp\{b(\eta-2t) - b(-2t)\} \le A(\eta)+1$ we find that

$$\frac{4}{\varepsilon^2}E\left\{\frac{|X_i-\delta|^2}{2}\; Q^2_{i,n}(\delta)\right\} \le \kappa^*\left(\frac{\Lambda_n}{\varepsilon^{\frac{1}{2}p+1}}\right)\exp(|\delta|^2)B_{\theta,G}(\delta,\varepsilon^*)$$

$$\le \kappa^* \; \frac{\Lambda_n}{\varepsilon^{\frac{1}{2}p+1}}B_G(\theta)\exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\}, \qquad (4.14c)$$

where the last inequality follows from (3.12a) and $\kappa^*$ bounds $4\kappa(2+\varepsilon)^p(2+\varepsilon)^{-\frac{1}{2}p-1}$.

Finally combining (4.14a), (4.14b) and (4.14c) we find that for a suitable constant, $\kappa_4$,

$$E\left\{|\dot{Q}_{i,n}(\delta)|^2\right\} \le \frac{\kappa_4\Lambda_n}{\varepsilon^{\frac{1}{2}p+1}}B_G(\theta)(1+\varepsilon|\delta|^2)\exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\},$$

from which (4.14) follows since $\varepsilon<\varepsilon^*<1$.

__PROPOSITION 4.8__. If $\varepsilon<\frac{1}{2}\varepsilon^*$ and (T) holds for some $\lambda^*<\infty$, then

$$\left|E\left\{(1+\varepsilon)\dot{W}_{n,2}(\delta) - \dot{w}(\delta,\theta,G)\right\}\right| \le \kappa_5\Pi_n B_G(\theta)(1+|\delta|)\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\}. \qquad (4.15)$$

__PROOF__. Usng the definitions of $\dot{W}_{n,2}(\delta)$ and $\dot{w}(\delta,\theta,G)$, equation (4.13) and some standard inequalities, we find that

$$\left|E\left\{(1+\varepsilon)\dot{W}_{n,2}(\delta) - \dot{w}(\delta,\theta,G)\right\}\right| \le \int\cdots\int\left|(u+\varepsilon\delta)q_n - uq\right|g(u)du$$

$$\le \int\cdots\int\Big(|u|\,|q_n-q|+\varepsilon|\delta|q_n\Big)g(u)du$$

$$\le \int\cdots\int\Big(|u-\delta|\,|q_n-q|+|\delta|\,|q_n-q|+\varepsilon|\delta|q_n\Big)g(u)du$$

$$= \int\cdots\int\Big(2^{\frac{1}{2}}d|q_n-q|+|\delta|\,|q_n-q|+\varepsilon|\delta|q_n\Big)g(u)du.$$

From (4.6), (3.11a) and (3.12a) we derive the inequality

$$\int\cdots\int 2^{\frac{1}{2}}d|q_n-q|g(u)du \le 2^{\frac{1}{2}}\kappa_1\Pi_n\exp\left(\frac{|\delta|^2}{2}\right)B_{\theta,G}(\delta,\varepsilon^*)$$

$$\le 2^{\frac{1}{2}}\kappa_1\Pi_n B_G(\theta)\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\}. \qquad (4.15a)$$

Work done in the proof of Proposition 4.5 shows that

$$\int\cdots\int |\delta||q_n-q|g(u)du \le \kappa_1\Pi_n B_G(\theta)|\delta|\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\} \qquad (4.15b)$$

and from (4.8) we get that

$$\int\cdots\int \varepsilon|\delta|q_n g(u)du \le \varepsilon B_G(\theta)|\delta|\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\}. \qquad (4.15c)$$

Combining (4.15a), (4.15b) and (4.15c) we find that the normed bias of $(1+\varepsilon)\dot{W}_{n,2}(\delta)$ as an estimator of $\dot{w}(\delta,\theta,G)$ is bounded by

$$B_G(\theta)\exp\left\{\frac{(1+\lambda^*)|\delta|^2}{2}\right\}(2^{\frac{1}{2}}\kappa_1\Pi_n + \kappa_1\Pi_n|\delta| + \varepsilon|\delta|).$$

Using the fact that $\varepsilon < \Pi_n$ we obtain (4.15) by choosing $\kappa_5$ so that $2^{\frac{1}{2}}\kappa_1+(1+\kappa_1)|\delta| \le \kappa_5(1+|\delta|)$; this holds whenever $\kappa_5 \ge \max(2^{\frac{1}{2}}\kappa_1, 1+\kappa_1)$.////

We now give a bound on the mean squared error of $(1+\varepsilon)\dot{W}_{n,2}(\delta)$ as an estimator of $\dot{w}(\delta)$.

<u>PROPOSITION 4.9</u>. If $\varepsilon<\frac{1}{2}\varepsilon^*$ and (T) holds for some $\lambda^*<\infty$, then

$$E\left\{\left|(1+\varepsilon)\dot{W}_{n,2}(\delta)-\dot{w}(\delta,\theta,G)\right|^2\right\} \le \kappa_6\left(\frac{\Lambda_n}{n\varepsilon^{p/2+1}}+\Pi_n^2\right)\left\{B_G(\theta)\right\}^2$$

$$\times\left(1+|\delta|\right)^2\exp\left\{\frac{(2+2\lambda^*)|\delta|^2}{2}\right\}. \qquad (4.16)$$

PROOF. The left hand side of (4.16) can be written as

$$\mathrm{tr}\left[\mathrm{var}\{(1+\varepsilon)\dot{W}_{n,2}(\delta)\}\right] + \left|E\left\{(1+\varepsilon)\dot{W}_{n,2}(\delta)-\dot{w}(\delta,\theta,G)\right\}\right|^2 .$$

Since $\dot{W}_{n,2}(\delta) = k^{-1}\sum\dot{Q}_{i,n}(\delta)$ where the $\dot{Q}_{i,n}(\delta)$ are independent and identically distributed we have, with $k = \frac{1}{2}n$,

$$\mathrm{tr}\left[\mathrm{var}\{(1+\varepsilon)\dot{W}_{n,2}(\delta)\}\right] = \frac{2(1+\varepsilon)^2}{n}\,\mathrm{tr}\left[\mathrm{var}\{\dot{Q}_{i,n}(\delta)\}\right]$$

$$\le \frac{2(1+\varepsilon)^2}{n}\,E\left\{|\dot{Q}_{i,n}(\delta)|^2\right\}.$$

Using (4.14) and (4.15) we can now bound the left hand side of (4.16) by

$$\frac{2(1+\varepsilon)^2}{n}\left(\frac{\kappa_4\Lambda_n}{\varepsilon^{\frac{1}{2}p+1}}\right)B_G(\theta)(1+|\delta|^2)\exp\left\{\frac{(2+\lambda^*)|\delta|^2}{2}\right\}$$

$$+\ \kappa_5^2\Pi_n^2B_G^2(\theta)(1+|\delta|)^2\exp\left\{\frac{(2+2\lambda^*)|\delta|^2}{2}\right\} . \qquad (4.16a)$$

Using the facts that $B_G(\theta) \le B_G^2(\theta)$ and $(1+|\delta|^2) \le (1+|\delta|)^2$, it can be shown that (4.16a) is bounded by the right hand side of (4.16) for any $\kappa_6 \ge \max\{\kappa_5^2,\ 2\kappa_4(1+\varepsilon^*)^2\}$. ////

## 5. CONCLUDING REMARKS

The existence proof in Section 3 is constructive and can be used as a recipe for calculating efficient estimators. It has the undesirable properties of relying on: (1) a discretized initial estimator and (2) an arbitrary splitting of the sample into two equal halves. The effect of discretizing is likely to be minor. Splitting the data may prove to be more crucial although this can be checked by considering a finite number of partitions and computing the one-step estimator for each.

References

Adcock, R. J. (1878). A problem in least squares. *The Analyst* 5, 53-4.

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

Anderson, T. W. (1976). Estimation of lienar functional relationships with discussion). *J. Roy. Statist. Soc. Ser. B* 38, 1-36.

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

Begun, J. M., Hall, W. J., Huang, W. M. & Wellner, J. A> (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* 11, 432-52.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* 10, 647-71.

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

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.

Clark, R. R. (1982). The errors-in-variables problem in the logistic regression model. Unpublished Ph.D. thesis, University of North Carolina, Chapel Hill.

Fuller, W. A. (1987). *Measurement Error Models.* To be published by Wiley.

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.

Lindsay, B. G. (1983). Efficiency of the conditional score in a mixture setting. *Ann. Statist.* 11, 486-97.

Lindsay, B. G. (1985). Using empirical partially Bayes inference for increased efficiency. *Ann. Statist.* 13, 914-32.

McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.

Moran, P. (1971). Estimating structural and functional relationships. *J. Multi. Anal.* 1, 232-55.

Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory.* New York: Springer-Verlag.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331-42.

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* (to appear).

Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, 187-96, University of California Press.

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

Wolter, J. M. & Fuller, W. A. (1982b). Estimation of the quadratic errors-in-variables model. *Biometrika* **69**, 174-82.