Statistica Sinica **16**(2006), 183-193

ON CLOSED FORM SEMIPARAMETRIC ESTIMATORS FOR MEASUREMENT ERROR MODELS

Yanyuan Ma and Anastasios A. Tsiatis

Texas A&M University and North Carolina State

Abstract: We examine the locally efficient semiparametric estimator proposed by Tsiatis and Ma (2004) in the situation when a sufficient and complete statistic exists. We derive a closed form solution and show that when implemented in generalized linear models with normal measurement error, this estimator is equivalent to the efficient score estimator in Stefanski and Carroll (1987). We also demonstrate how other consistent semiparametric estimators naturally emerge. The method is used in an extension of the usual generalized linear models.

Key words and phrases: Measurement error models, semiparametric estimator.

1. Introduction

Measurement error models commonly arise in making inference on the relationship of a response variable Y and predictor variables when some of the latter may be measured with error. Following the notation of Carroll, Ruppert and Stefanski (1995), we denote by Z the predictor variables that are measured precisely and X those which cannot be measured precisely. Instead of X, the variables W, which are related to X, are observed. Hence, a typical measurement error model problem is to estimate and make inference on an unknown parameter θ involved in a parametric model $p(y|x, z; \theta)$, based on the observed data set $\{O_i = (Y_i, W_i, Z_i), i = 1, \ldots, n\}$. The relation between W and X, Z is modeled by $p(w|x, z)$ and is often assumed to be known completely or up to a parametric form. The most typical assumption for measurement error is normal error: $W|(X, Z) \sim N(X, \bar{\Omega})$, where $\bar{\Omega}$ is a known or unknown constant covariance matrix. In this paper, we regard X as a random variable whose conditional distribution on Z, denoted by $\eta(x|z)$, is completely unrestricted. Hence, we work in a non-classical functional measurement error model setting established in Carroll et al. (1995, Chap.7.2).

Tsiatis and Ma (2004) proposed a class of semiparametric estimators for θ in such functional measurement error model setting. They view the problem as a semiparametric model of the form $p(y, w|z; \theta, \eta) = \int p(y|x, z; \theta)p(w|x, z)\eta(x|z)\, d\mu(x)$, where the finite dimensional parameter θ is the parameter of interest, the infinite dimensional parameter $\eta(x|z)$ is a nuisance parameter. Specifically, they gave a locally efficient estimator through solving a Fredholm integral equation of the first type. In general, solving such equations numerically is an ill-posed problem and hence requires regularization schemes (see Kress (1999)).

For measurement error models where the measurement W is normally distributed with its center at the unobservable explanatory variable X, its variance a constant, and where the response variable Y relates to X through a generalized linear model, Stefanski and Carroll (1985, 1987) proposed a number of estimators which are referred to as the sufficient score, conditional score and efficient score estimators. A crucial condition which enables the derivation of such estimators is the existence of a complete sufficient statistic for X.

We examine the relation between the two proposals in the presence of a complete sufficient statistic and conclude that they agree. This does not come as a surprise since both programs yield consistent semiparametric estimators that are optimal in terms of their estimation variance. By ignoring some terms in the resulting estimator from the proposal by Tsiatis and Ma (2004), while preserving consistency, the remaining estimators of the proposal by Stefanski and Carroll (1987) emerge naturally.

Studying the nuisance tangent space and its orthogonal complement directly in the framework established in Bickel, Klaassen, Ritov and Wellner (1993) can produce these estimators as well.

In Section 2, we state the problem and summarize the known estimators and some relevant results. We derive the closed form of these estimators when a complete and sufficient statistic exists, in Section 3. In Section 4, we implement the closed form estimator in generalized linear models with normal additive measurement error with constant variance, and verify that they result in the estimators given by Stefanski and Carroll (1985, 1987). We extend the generalized linear model to include covariates measured without error and an error structure depending on these covariates, in Section 5. The computation details are given in the Appendix.

2. Problem Statement and Summary of Relevant Results

We summarize the problem and relevant results in this section (for details, see Tsiatis and Ma (2004)). Consider the measurement error model where $p(y|x,z;\theta)$ is known up to the parameter θ. The measurement distribution $p(w|x,z)$ is known and the distribution of X conditional on Z, $\eta(x|z)$, is unspecified. Assume W is a surrogate of X (see Carrol, Ruppert and Stefanski (1995)), i.e., $p(y|w,x,z) = p(y|x,z)$. We consider semiparametric estimators of θ derived from independent and identically distributed random vectors $O_i, i = 1,\ldots,n$, where, for the i-th individual, $O_i = (Y_i, W_i, Z_i)$. We are interested in regular asymptotically linear (RAL) estimators $\hat{\theta}_n$ of θ (Newey (1990)). That is, the estimator minus the estimand can be approximated asymptotically by a sum of identically and independently distributed (i.i.d.) mean zero random vectors. Specifically, an estimator $\hat{\theta}_n$ of the parameter θ is asymptotically linear if

$$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi(O_i, \theta) + o_p(1),$$

where $\phi(O_i, \theta), i = 1, \ldots, n$ are i.i.d. mean zero q-dimensional random vectors and $o_p(1)$ denotes a term that converges in probability to zero. The random vector $\phi(O_i, \theta)$ is referred to as the i-th influence function of the estimator $\hat{\theta}_n$, which satisfies $\mathrm{E}(\phi) = 0$ and $\mathrm{E}(\phi\phi^T)$ finite and nonsingular (See Bickel et al. (1993)). It can be verified that $\hat{\theta}_n$ can be obtained through solving $\sum_{i=1}^{n} \phi(O_i, \hat{\theta}) = 0$, hence one way to construct RAL estimators is through finding the influence functions. The restriction to regular estimators is a technical condition imposed to exclude estimators that have undesirable local properties. For details regarding the regularity conditions, the readers are referred to Newey (1990). It is clear from the representation above that the asymptotic variance of an RAL estimator is equal to the variance of its influence function $\phi(O, \theta)$. Consequently, the optimal estimator among a class of RAL estimators is the one whose influence function has the smallest variance. This is referred to as an efficient semiparametric estimator. If the estimator depends on $\eta(x|z)$ it is termed as locally efficient, otherwise it is globally efficient.

For a complete study of influence functions and their associated geometry, readers are referred to Bickel et al. (1993). Influence functions lie in a space $\Lambda^{\perp}$ given by $[h(Y, W, Z) : E\{h(Y, W, Z)|X, Z\} = 0$ a.e.], the orthogonal complement to the nuisance tangent space $\Lambda = [E\{h(X, Z)|Y, W, Z\} : E\{h(X, Z)|Z\} = 0]$. Every function in $\Lambda^{\perp}$ can be properly normalized to yield an influence function. Thus, given an estimator $\sum_{i=1}^{n} h(Y_i, W_i, Z_i) = 0$, we only need to verify that $E\{h(Y, W, Z)|X, Z\} = 0$ for consistency.

A locally efficient semiparametric estimator S_{eff} is constructed in two steps:

- Propose a distribution for η, say we adopt $\eta(x|z) = p(x|z)$.
- Solve the estimating equation

$$\sum_{i=1}^{n} S_{eff}(Y_i, W_i, Z_i) = \sum_{i=1}^{n} S_{\theta}(Y_i, W_i, Z_i) - E\{a(X, Z)|Y_i, W_i, Z_i\} = 0, \quad (1)$$

where $S_{\theta}(Y, W, Z) = E\{S_{\theta}^F(Y, X, Z)|Y, W, Z\}$ with $S_{\theta}^F(Y, X, Z) = [\partial \log\{p(Y|X, Z; \theta)\}]/(\partial\theta)$, and $a(X, Z)$ satisfies

$$E[S_{\theta}(Y, W, Z) - E\{a(X, Z)|Y, W, Z\}|X, Z] = 0. \quad (2)$$

Here, conditional expectations are taken either with respect to the truth (for $p(w,y|x,z)$) or to the proposed distribution (for $p(x|z)$).

The estimator is locally efficient in the sense that if the posited model $p(X|Z)$ is correct, the estimator is efficient; however, even if the posited model $p(X|Z)$ is incorrect, the estimator is still consistent.

3. The Locally Efficient Estimator in the Presence of a Complete Sufficient Statistic

To find $a(X,Z)$ in (2) involves solving an ill-posed type I integral equation. This is often difficult and presents numerical challenges. However, when a sufficient and complete statistic for X exists, a closed form solution for $E\{a(X,Z)|Y_i, W_i, Z_i\}$ can be derived without solving for $a(X,Z)$ explicitly in (2). Note that it is $E\{a(X,Z)|Y_i,W_i,Z_i\}$ that appears in (1). We denote the complete sufficient statistic for X by $\Delta(w,y,z;\theta)$. Because Δ is a function of W,Y,Z, we have $S_\theta(Y,\Delta,Z) = E\{S_\theta^F(Y,X,Z)|Y,\Delta,Z\} = E\{S_\theta^F(Y,X,Z)|Y,W,Z\} = S_\theta(Y,W,Z)$. Substituting the random variable W with Δ in (2), we obtain

$$\begin{aligned}&E\{S_\theta(Y,\Delta,Z)|X,Z\} - E[E\{a(X,Z)|Y,\Delta,Z\}|X,Z]\\ &= E[E\{S_\theta(Y,\Delta,Z)|\Delta,X,Z\}|X,Z] - E[E\{a(X,Z)|Y,\Delta,Z\}|X,Z] = 0.\end{aligned}$$

Since Δ is sufficient, conditional on Δ and Z, X and Y are independent. Hence the inner expectation can be simplified to yield

$$E[E\{S_\theta(Y,\Delta,Z)|\Delta,Z\} - E\{a(X,Z)|\Delta,Z\}|X,Z] = 0.$$

Because Δ is also complete, the inner expectation, as a function of Δ and Z, is identically zero. Hence we obtain $E\{a(X,Z)|\Delta,Z\} = E\{S_\theta(Y,\Delta,Z)|\Delta,Z\}$. This provides a closed form expression for S_{eff}:

$$S_{eff}(Y,\Delta,Z) = S_\theta(Y,\Delta,Z) - E\{S_\theta(Y,\Delta,Z)|\Delta,Z\}. \quad (3)$$

In fact, in the presence of a complete and sufficient statistic, the closed form estimator in (3) can be derived directly by characterizing the orthogonal complement of the nuisance tangent space $\Lambda^\perp$ without forming the integral equation. Since the nuisance tangent space Λ consists of elements of the form $E\{h(X,Z)|Y,W,Z\}$ and Δ is sufficient, where $h(X,Z)$ satisfies $E\{h(X,Z)|Z\} = 0$, we can write

$$E\{h(X,Z)|Y,W,Z\} = E\{h(X,Z)|Y,\Delta,Z\} = E\{h(X,Z)|\Delta,Z\}$$

which is a function of Δ and Z only. Denoting $E\{h(X,Z)|Y,W,Z\}$ by $f(\Delta,Z)$, and noticing that $E\{h(X,Z)|Z\} = E[E\{h(X,Z)|Y,W,Z\}|Z] = E\{f(\Delta,Z)|Z\}$, we obtain $\Lambda \subset [f(\Delta, Z) : E\{f(\Delta, Z)|Z\} = 0]$. On the other hand, assume $g(\Delta, Z)$ is the projection of $f(\Delta, Z)$ to Λ. Then $E[\{f(\Delta, Z) - g(\Delta, Z)\}E\{h(X, Z|\Delta, Z)\}|Z] = 0$ for any $h(X, Z)$ such that $E\{h(X, Z)|Z\} = 0$. Because $E[\{f(\Delta, Z) -g(\Delta, Z)\}E\{h(X, Z|\Delta, Z)\}|Z] = E(E[\{f(\Delta, Z) - g(\Delta, Z)\}h(X, Z)|\Delta, Z]|Z) = E[\{f(\Delta, Z) - g(\Delta, Z)\}h(X, Z)|Z] = E(E[\{f(\Delta, Z) - g(\Delta, Z)\}|X, Z]h(X, Z)|Z)$, we obtain $E[\{f(\Delta, Z) - g(\Delta, Z)\}|X, Z] = 0$ a.e.. Because Δ is complete, this means that $f(\Delta, Z) - g(\Delta, Z) = 0$ a.e.. Thus $f(\Delta, Z)$ itself is in Λ, so $\Lambda = [f(\Delta, Z) : E\{f(\Delta, Z)\} = 0]$.

Consequently, the projection of an arbitrary function $g(y, w, z)$ to Λ is $E\{g(Y, W, Z)|\Delta, Z\}$, and its projection to $\Lambda^{\perp}$,

$$S_{[g]}(Y, W, Z) = g(Y, W, Z) - E\{g(Y, W, Z)|\Delta, Z\}, \tag{4}$$

suggests a consistent RAL estimator. Note that (4) gives a general form of an arbitrary RAL estimator, i.e., for every mean zero function $g(Y, W, Z)$, $\sum_{i=1}^{n} S_{[g]}(Y_i, W_i, Z_i) = 0$ is an estimating equation that yields a consistent estimator for θ. Specifically, if we take $g(Y, W, Z) = S_\theta(Y, \Delta, Z)$ in (4), we obtain a locally efficient estimator which is the same as in (3).

The locally efficient estimator in (3) can be calculated explicitly. A change of variable yields

$$p(y, w|x, z; \theta) = p(y, \delta|x, z; \theta)/J(\delta, y, z; \theta) = p(\delta|x, z; \theta)\{p(y|\delta, z; \theta)/J(\delta, y, z; \theta)\},$$

where the last equality is due to the sufficiency of Δ. Here $J(\delta, y, z; \theta)$ stands for the Jacobian of the transformation $\partial(w, y)/\partial(\delta, y)$. Using this relation, we obtain

$$\begin{aligned} S_{eff}(Y, W, Z) = &\frac{\partial}{\partial\theta} \log p(Y|\Delta, Z; \theta) \\ &+ \left[\frac{d\Delta}{d\theta}\frac{\partial}{\partial\Delta} \log p(Y|\Delta, Z; \theta) - E\left\{\frac{d\Delta}{d\theta}\frac{\partial}{\partial\Delta} \log p(Y|\Delta, Z; \theta)|\Delta, Z\right\}\right] \\ &+ \left\{\frac{d\Delta}{d\theta} - E\left(\frac{d\Delta}{d\theta}|\Delta, Z\right)\right\}\frac{\partial}{\partial\Delta} \log p(\Delta|Z; \theta) \\ &- \left\{\frac{d}{d\theta} \log J - E\left(\frac{d}{d\theta} \log J|\Delta, Z; \theta\right)\right\}. \end{aligned} \tag{5}$$

Details of the calculation are presented in the Appendix.

Implementing the estimator $\sum_{i=1}^{n} S_{eff}(Y_i, \Delta_i, Z_i) = 0$, where $S_{eff}(Y, W, Z)$ is given in (3) or (5), does not require solving any integral equation, hence there is no need to take into account the ill-posedness associated with using the more general form of the estimator in (1). The only technical challenge lies in finding a sufficient and complete statistic, but this is not an easy task in general. In the measurement error model setting, it requires finding a function of W, Z and Y which can depend on θ, say $\Delta(W, Y, Z; \theta)$, such that the following two conditions are satisfied.

1. $p(W, Y|X, Z)$ can be decomposed into a product of a function of Y, Δ, Z and a function of X, Δ, Z: $p(W, Y|X, Z) = h(Y, \Delta, Z) f(\Delta, X, Z)$.
2. $\int g(\Delta, Z) f(\Delta, X, Z) d\Delta = 0 \Longrightarrow g(\Delta, Z) = 0$.

The second condition is often difficult to verify. One general class where such complete sufficient statistic exists is the generalized linear model with a generalized linear error model for $p(W|X, Z)$ as well. A special case where the measurement error is additive normal has been studied, and a series of estimators have been proposed by Stefanski and Carroll (1985, 1987). In the next section, we proceed from our general result and show that these estimators can be obtained as special cases of (3) (or (5)) and (4).

4. Generalized Linear Model with Normal Measurement Error

In this section Z does not appear and we consider a generalized linear model of the form

$$p(y|x; \theta) = \exp\left\{\frac{y(\alpha + \beta^T x) - b(\alpha + \beta^T x)}{a(\gamma)} + c(y, \gamma)\right\},$$

where the parameter of interest is $\theta = (\alpha, \beta^T, \gamma)^T$. We also assume the measurement error w has

$$p(w|x) = \frac{(2\pi)^{-\frac{r}{2}}}{|\bar{\Omega}|^{\frac{1}{2}}} \exp\{-(w - x)^T \bar{\Omega}^{-1} \frac{w - x}{2}\},$$

where r is the dimension of β. It is known that $\delta(w, y; \theta) = w + y\Omega\beta$ is a complete sufficient statistic, where $\Omega = \bar{\Omega}/a(\gamma)$ is assumed known. Then $J(\delta, y; \theta) = 1$, $d\delta/d\alpha = d\delta/d\gamma = 0$, $d\delta/d\beta = y\Omega$, and

$$p(y|\delta; \theta) = \exp[\xi y - \frac{1}{2} y^2 \beta^T \Omega \beta / a(\gamma) + c(y, \gamma) - \log\{s(\xi, \beta, \gamma)\}], \tag{6}$$

where $\xi = (\alpha + \delta^T \beta)/a(\gamma)$ and $s(\xi, \beta, \gamma) = \int \exp\{\xi y - (1/2) y^2 \beta^T \Omega \beta / a(\gamma) + c(y, \gamma)\} d\mu(y)$. Since $d\delta/d\alpha = 0$, (5) implies

$$\begin{aligned} S_{eff}(Y, W; \theta)_\alpha = \frac{\partial}{\partial \alpha} \log p(Y|\Delta; \theta) &= \frac{Y}{a(\gamma)} - \frac{1}{a(\gamma)} \frac{d}{d\xi} \log\{s(\xi, \beta, \gamma)\} \\ &= \{Y - E(Y|\Delta)\}/a(\gamma), \end{aligned} \tag{7}$$

where we use $S_{eff}(Y, W; \theta)_\alpha$ to denote the component corresponding to α in $S_{eff}(Y, W; \theta)$. The relation $(d/d\xi) \log\{s(\xi, \beta, \gamma)\} = E(Y|\Delta)$ used in (7) can be verified through straightforward algebra. Similarly,

$$\begin{aligned} S_{eff}(Y,W;\theta)_\gamma &= \frac{\partial}{\partial\gamma}\log p(Y|\Delta;\theta) \\ &= -\frac{Y(\alpha+\Delta^T\beta)a'(\gamma)}{a^2(\gamma)} + \frac{Y^2\beta^T\Omega\beta a'(\gamma)}{2a^2(\gamma)} + \frac{dc(Y,\gamma)}{d\gamma} \\ &\quad -E\left\{-\frac{Y(\alpha+\Delta^T\beta)a'(\gamma)}{a^2(\gamma)} + \frac{Y^2\beta^T\Omega\beta a'(\gamma)}{2a^2(\gamma)} + \frac{dc(Y,\gamma)}{d\gamma}|\Delta;\theta\right\}. \quad (8) \end{aligned}$$

Because $d\delta/d\beta \neq 0$, the component $S_{eff}(Y,W;\theta)_\beta$ is more complex. Specifically,

$$\begin{aligned} S_{eff}(Y,W;\theta)_\beta &= \frac{\partial}{\partial\beta}\log p(Y|\Delta;\theta) + \{Y - E(Y|\Delta)\}\,\Omega\frac{\partial}{\partial\Delta}\log p(\Delta;\theta) \\ &\quad +\Omega\left[Y\frac{\partial}{\partial\Delta}\log p(Y|\Delta;\theta) - E\left\{Y\frac{\partial}{\partial\Delta}\log p(Y|\Delta;\theta)|\Delta\right\}\right]. \quad (9) \end{aligned}$$

Because of the consistency of the estimator, $E\{S_{eff}(Y,W;\theta)_\beta\} = E\{S_{eff}(Y,W;\theta)_\beta|X\} = 0$. However, the second and third terms in (9) both lie in $\Lambda^\perp$. This means that the first term $(\partial/\partial\beta)\log p(Y|\Delta;\theta)$ itself is in $\Lambda^\perp$, hence suggests a consistent estimator. Thus a possible $S(Y,W;\theta)_\beta$ is

$$\begin{aligned} &\frac{\partial}{\partial\beta}\log p(Y|\Delta;\theta) \\ &= \frac{\Delta}{a(\gamma)}Y - Y^2\frac{\Omega\beta}{a(\gamma)} - \frac{\Delta}{a(\gamma)}\frac{\partial}{\partial\xi}\log\{s(\xi,\beta,\gamma)\} - E\{Y^2\frac{\Omega\beta}{a(\gamma)}|\Delta;\theta\} \\ &= \frac{\Delta}{a(\gamma)}\{Y - E(Y|\Delta,\theta)\} - \frac{\Omega\beta}{a(\gamma)}\{Y^2 - E(Y^2|\Delta,\theta)\}, \quad (10) \end{aligned}$$

which, combined with (7) and (8), gives the sufficient score estimator of Stefanski and Carroll (1987). In (10), we used the relation $(\partial/\partial\beta)\log\{s(\xi,\beta,\gamma)\} = E(Y^2[\Omega\beta/a(\gamma)]|\Delta)$, which again follows from straightforward algebra.

As pointed out at (4), any function of the form $f(Y,\Delta) - E\{f(Y,\Delta)|\Delta;\theta\}$ will yield a candidate for $S(Y,W;\theta)_\beta$. In particular, $f(Y,\Delta) = Yt(\Delta)$ is a natural choice, where $t(\Delta)$ is an arbitrary function of Δ. This choice yields

$$S(Y,W;\theta)_\beta = \{Y - E(Y|\Delta)\}t(\Delta), \quad (11)$$

which, combined with (7) and (8), gives the conditional score estimator of Stefanski and Carroll (1987) ($S(Y,W;\theta)_\beta = \{Y - E(Y|\Delta)\}t(\Delta)$ and $S(Y,W;\theta)_\beta = \{Y - E(Y|\Delta)\}t(\Delta)/a(\gamma)$ lead to the same estimator, the latter one is the precise form used in Stefanski and Carroll (1987)).

Finally, the locally efficient estimator term $S_{eff}(Y,W;\theta)_\beta$ itself can be calculated explicitly and results in the expression

$$S_{eff}(Y,W;\theta)_\beta = \{Y - E(Y|\Delta)\}E(X|\Delta)/a(\gamma), \quad (12)$$

which, combined with (7) and (8), gives the efficient score estimator of Stefanski and Carroll (1987). Detail of the computation is given in the Appendix. Note that in order to implement this estimator, we need to propose a distribution model for X, say $p(X)$, for computing $E(X|\Delta)$. If the true distribution of X, say $\eta(X)$, is the same as $p(X)$, then the resulting estimator is efficient. Otherwise, the resulting estimator is no longer efficient, although it is still consistent. Hence "efficient score estimator" is indeed a locally efficient estimator.

Although regularity conditions ensure that consistent sequence of estimators exist as solutions of estimating equations, a multiple roots issue inevitably arises, hence an inconsistent sequence of estimates can occur. Since this is an issue for general M-estimators, we do not comment further on the problem but refer the readers to Stefanski and Carroll (1987) for practical techniques to get around it.

The completeness of the statistic Δ enables one to derive $E\{a(X;\theta)|Y,W\}$ directly, and permits a direct characterization of the nuisance tangent space Λ. The sufficiency of Δ provides the framework to separate the conditional distribution $p(Y|\Delta)$ from X. These are crucial conditions and we believe a closed form solution does not exist under any other conditions.

5. An Extension

Consider the generalized linear model above when the covariates Z, measured without error, are also present, and Y has a generalized linear form with respect to X and Z. One way to deal with such situation is to concatenate X and Z and treat it as X^N, concatenate W and Z and treat it as W^N. Then treat the model of Y, W^N and X^N as a special case of the model in Section 4 with a degenerate covariance matrix $\bar{\Omega}^N$, where $\bar{\Omega}^N$ is a block diagonal matrix, with its $(1,1)$ block being $\bar{\Omega}$, the covariance matrix of $W|(X,Z)$, and all other blocks being zero. Here the superscript N is used to stand for the new model/variables. Similar results as in Section 4 can be derived despite the singularity of $\bar{\Omega}^N$. However, the method does not survive if the measurement error variance $\bar{\Omega}$ depends on the error-free measurement Z, for example.

Still, using the estimating equations proposed in (3) and (4), this can be handled easily. In fact, a much broader class of problems can be solved. Suppose the model dependence of Y on X has a generalized linear model form which depends on the observed covariate Z,

$$p(y|x,z;\theta) = \exp\left\{\frac{y(\alpha+\beta^T x) - b(z)(\alpha+\beta^T x)}{a(z,\gamma)} + c(y,z,\gamma)\right\},$$

and the measurement error model $p(w|x,z)$ is normal with mean x and covariance matrix $\bar{\Omega}(z)$. Combining X and Z and treating the combination X^N as the predictive variables will yield a model $p(y|x,z)$ that is very complex and does not have a complete sufficient statistic. But, comparing this model with the one in Section 4, it is tempting to treat z simply as a known constant so as to apply the result there. In fact, this works. The estimators given in (3) (or (5)) and (4) precisely serve the purpose of justifying such approach. In constructing the estimators, all the expectations are calculated conditional on Z, thus treating Z as a random variable or as a constant is equivalent. Setting $\Delta(W, Y, Z; \theta) = W + Y\Omega(Z)\beta$, where $\Omega(Z) = \bar{\Omega}(Z)/a(Z, \gamma)$, replacing $a(\gamma)$, $c(Y, \gamma)$, Ω and $t(\Delta)$ in (7), (8), (10), (11) and (12) with $a(Z, \gamma)$, $c(Y, Z, \gamma)$, $\Omega(Z)$ and $t(\Delta, Z)$ respectively, and replacing the conditional expectations $E(\cdot|\Delta)$ in these equations with the corresponding $E(\cdot|\Delta, Z)$, we would obtain the sufficient score estimator, conditional score estimator, and efficient score estimator, respectively, in this more general setting.

Acknowledgements

The authors thank Ray Carroll for suggesting the extension of the generalized linear model and an anonymous referee for helpful comments.

Appendix

a. To get (5), note that from

$$
\begin{aligned}
&S_\theta^F(Y, x, Z) = \frac{d \log p(Y|x, Z; \theta)}{d\theta} = \frac{d \log p(W, Y|x, Z; \theta)}{d\theta} \\
&= \frac{d}{d\theta} \log p(\Delta|x, Z; \theta) + \frac{d}{d\theta} \log p(Y|\Delta, Z; \theta) - \frac{d}{d\theta} \log J(\Delta, Y, Z; \theta) \\
&= \frac{d\Delta}{d\theta} \frac{\partial}{\partial \Delta} \log p(\Delta|x, Z; \theta) + \frac{\partial}{\partial\theta} \log p(\Delta|x, Z; \theta) + \frac{d\Delta}{d\theta} \frac{\partial}{\partial \Delta} \log p(Y|\Delta, Z; \theta) \\
&\quad + \frac{\partial}{\partial\theta} \log p(Y|\Delta, Z; \theta) - \frac{d}{d\theta} \log J(\Delta, Y, Z; \theta),
\end{aligned}
$$

$$
\begin{aligned}
&S_\theta(Y, \Delta, Z) = E\{S_\theta^F(Y, X, Z)|Y, \Delta, Z\} \\
&= E\left\{\frac{d\Delta}{d\theta} \frac{\partial}{\partial \Delta} \log p(\Delta|X, Z; \theta) + \frac{\partial}{\partial\theta} \log p(\Delta|X, Z; \theta)|Y, \Delta, Z\right\} \\
&\quad + \frac{d\Delta}{d\theta} \frac{\partial}{\partial \Delta} \log p(Y|\Delta, Z; \theta) + \frac{\partial}{\partial\theta} \log p(Y|\Delta, Z; \theta) - \frac{d}{d\theta} \log J(\Delta, Y, Z; \theta).
\end{aligned}
$$

Because X and Y are independent conditional on Δ and Z, we get

$$
\begin{aligned}
&E\left\{\frac{d\Delta}{d\theta} \frac{\partial}{\partial \Delta} \log p(\Delta|X, Z; \theta) + \frac{\partial}{\partial\theta} \log p(\Delta|X, Z; \theta)|Y, \Delta, Z\right\} \\
&= \frac{d\Delta}{d\theta} E\left\{\frac{\partial}{\partial \Delta} \log p(\Delta|X, Z; \theta)|\Delta, Z\right\} + E\left\{\frac{\partial}{\partial\theta} \log p(\Delta|X, Z; \theta)|\Delta, Z\right\} \\
&= \frac{d\Delta}{d\theta} \frac{\partial}{\partial \Delta} \log p(\Delta|Z; \theta) + \frac{\partial}{\partial\theta} \log p(\Delta|Z; \theta).
\end{aligned}
$$

Then we find

$$S_\theta(Y,\Delta,Z) = \frac{\partial}{\partial\theta}\log p(Y|\Delta,Z;\theta) + \frac{d\Delta}{d\theta}\frac{\partial}{\partial\Delta}\log p(Y|\Delta,Z;\theta) + \frac{d\Delta}{d\theta}\frac{\partial}{\partial\Delta}\log p(\Delta|Z;\theta) + \frac{\partial}{\partial\theta}\log p(\Delta|Z;\theta) - \frac{d}{d\theta}\log J(\Delta,Y,Z;\theta), \quad (13)$$

$$\begin{aligned} &E\{S_\theta(Y,\Delta,Z)|\Delta,Z\} \\ &= 0 + E\left\{\frac{d\Delta}{d\theta}\frac{\partial}{\partial\Delta}\log p(Y|\Delta,Z;\theta)|\Delta,Z\right\} + E\left(\frac{d\Delta}{d\theta}|\Delta,Z\right)\frac{\partial}{\partial\Delta}\log p(\Delta|Z;\theta) \\ &\quad + \frac{\partial}{\partial\theta}\log p(\Delta|Z;\theta) - E\left\{\frac{d}{d\theta}\log J(\Delta,Y,Z;\theta)|\Delta,Z\right\}. \end{aligned} \quad (14)$$

Apply (13) and (14) to the estimator given in (3) to get (5).

b. For the calculation of $S_{eff}(Y,W;\theta)_\beta$, note that the last two terms of (9) are

$$\begin{aligned} \Omega Y\frac{\partial}{\partial\Delta}\log p(Y|\Delta;\theta) &= \Omega Y\left\{Y\frac{\beta}{a(\gamma)} - \frac{\partial}{\partial\xi}\log s(\xi,\beta,\gamma)\frac{\beta}{a(\gamma)}\right\} \\ &= \frac{\Omega\beta}{a(\gamma)}\{Y^2 - YE(Y|\Delta)\}, \end{aligned}$$

$$\begin{aligned} \Omega E\left\{Y\frac{\partial}{\partial\Delta}\log p(Y|\Delta;\theta)|\Delta\right\} &= \Omega E\left[Y\left\{Y\frac{\beta}{a(\gamma)} - E(Y|\delta)\frac{\beta}{a(\gamma)}\right\}|\Delta\right] \\ &= \frac{\Omega\beta}{a(\gamma)}[E(Y^2|\Delta) - \{E(Y|\Delta)\}^2]. \end{aligned}$$

Also note that

$$\begin{aligned} &\log p(\Delta|x;\theta) = \log p(W|x) + \log(Y|x;\theta) - \log p(Y|\Delta;\theta) \\ &= \log\{(2\pi)^{-p/2}/|\bar\Omega|^{1/2}\} - \frac{1}{2}(W-x)^T\bar\Omega^{-1}(W-x) + \frac{Y(\alpha+\beta^Tx) - b(\alpha+\beta^Tx)}{a(\gamma)} \\ &\quad + c(Y,\gamma) - Y\xi + \frac{Y^2\beta^T\Omega\beta}{2a(\gamma)} - c(Y,\gamma) + \log\{s(\xi,\beta,\gamma)\} \\ &= \log\{(2\pi)^{-p/2}/|\bar\Omega|^{1/2}\} - \frac{1}{2}(\Delta - Y\Omega\beta - x)^T\bar\Omega^{-1}(\Delta - Y\Omega\beta - x) \\ &\quad + \frac{Y(\alpha+\beta^Tx) - b(\alpha+\beta^Tx)}{a(\gamma)} - Y\xi + \frac{Y^2\beta^T\Omega\beta}{2a(\gamma)} + \log\{s(\xi,\beta,\gamma)\}. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial\log p(\Delta|x;\theta)}{\partial\Delta} &= -\bar\Omega^{-1}(\Delta - Y\Omega\beta - x) - \frac{Y\beta}{a(\gamma)} + \frac{\partial\log\{s(\xi,\beta,\gamma)\}}{\partial\xi}\frac{\beta}{a(\gamma)} \\ &= -\bar\Omega^{-1}\Delta + \bar\Omega^{-1}x + E(Y|\Delta)\frac{\beta}{a(\gamma)}, \end{aligned}$$

$$\Omega\frac{\partial \log p(\Delta;\theta)}{\partial\Delta} = \Omega E\left\{\frac{\partial \log p(\Delta|X;\theta)}{\partial\Delta}|\Delta\right\} = \frac{-\Delta + E(X|\Delta) + E(Y|\Delta)\Omega\beta}{a(\gamma)}.$$

Here we used the fact that $\bar{\Omega} = a(\gamma)\Omega$ and

$$\begin{aligned}\frac{\partial \log p(\Delta)}{\partial\Delta} &= \frac{\frac{\partial \int p(\Delta|x)p(x)d\mu(x)}{\partial\Delta}}{p(\Delta)} = \int \frac{\partial p(\Delta|x)}{\partial\Delta\, p(\Delta|x)}\frac{p(\Delta|x)p(x)}{p(\Delta)}d\mu(x)\\ &= E\left\{\frac{\partial \log p(\Delta|X)}{\partial\Delta}|\Delta\right\}.\end{aligned}$$

Combining these relations, we obtain

$$\begin{aligned}S_{eff}(Y,W;\theta)_\beta &= \frac{\Delta}{a(\gamma)}\{Y - E(Y|\Delta)\} - \frac{\Omega\beta}{a(\gamma)}\{Y^2 - E(Y^2|\Delta)\}\\ &\quad+\{Y - E(Y|\Delta)\}\frac{-\Delta + E(X|\Delta) + E(Y|\Delta)\Omega\beta}{a(\gamma)}\\ &\quad+\frac{\Omega\beta}{a(\gamma)}\{Y^2 - YE(Y|\Delta)\} - \frac{\Omega\beta}{a(\gamma)}[E(Y^2|\Delta) - \{E(Y|\Delta)\}^2]\\ &= \{Y - E(Y|\Delta)\}\frac{E(X|\Delta)}{a(\gamma)}.\end{aligned}$$

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* The Johns Hopkins University Press, Baltimore.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman & Hall, New York.

Kress, R. (1999). *Linear Integral Equations.* Springer-Verlag, New York.

Newey, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Econom.* **5**, 99-135.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13**, 1335-1351.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703-716.

Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika.* **91**, 835-848.

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, U.S.A.

E-mail: ma@stat.tamu.edu

Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, U.S.A.

E-mail: tsiatis@stat.ncsu.edu

(Received April 2004; accepted September 2004)