Biometrika (2004), **91**, 4, *pp.* 835–848

Printed in Great Britain

Locally efficient semiparametric estimators for functional measurement error models

By ANASTASIOS A. TSIATIS

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

tsiatis@stat.ncsu.edu

AND YANYUAN MA

Centre for Research in Scientific Computation, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

yma@unity.ncsu.edu

Summary

A class of semiparametric estimators are proposed in the general setting of functional measurement error models. The estimators follow from estimating equations that are based on the semiparametric efficient score derived under a possibly incorrect distributional assumption for the unobserved 'measured with error' covariates. It is shown that such estimators are consistent and asymptotically normal even with misspecification and are efficient if computed under the truth. The methods are demonstrated with a simulation study of a quadratic logistic regression model with measurement error.

Some key words: Efficient score; Functional measurement error; Semiparametric estimator.

1. Introduction

A common problem is that of making inference about the relationship of a response variable Y to predictor variables, some of which may be measured with error. Using the notation of Carroll et al. (1995), we denote by Z the predictor variables that are measured precisely and by X those that are not. Instead, variables W, which are related to X, are observed.

To be specific, we consider parametric models with conditional densities $p(y|x, z; \beta)$, where the conditional distribution of Y given X and Z is described through a finite-dimensional parameter β. For example, with a binary response Y, a popular model is the logistic regression model that assumes $\text{logit}\{\text{pr}(Y=1|X, Z)\} = \beta_0 + \beta_1^{\mathrm{T}} X + \beta_2^{\mathrm{T}} Z$, where $\text{logit}(p) = \log\{p/(1-p)\}$. If the variables (Y, X, Z) were available for a sample of observations, then an estimator for β can easily be derived using, say, maximum likelihood. In our problem, the variables X are not measured directly, but rather we measure W, a surrogate for X; thus the observed data are (Y, W, Z). The objective of this paper is to derive estimators for β with a sample of data (Y_i, W_i, Z_i) for $i = 1, \ldots, n$.

We denote the conditional density of W given X and Z by $p(w|x, z)$. For example, a popular model is additive normally-distributed measurement error, where $W = X + \varepsilon$ and ε, the measurement error, is normally distributed with mean zero and variance σ_ε^2

independent of (Y, X, Z). For this model, $p(w|x, z) = (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} \exp\{-(w-x)^2/2\sigma_\varepsilon^2\}$. It will be assumed throughout that $p(w|x, z)$ is known; for example, this would correspond to the variance σ_ε^2 of the additive normally distributed measurement error being known. However, this assumption can be weakened to allow for unknown parameters. For example, with replicate measurements on W, the variance σ_ε^2 can be treated as an unknown parameter and estimated along with β. We also make the usual assumptions of surrogacy; that is, Y and W are conditionally independent given X and Z, so that $p(y|w, x, z) = p(y|x, z)$.

Traditionally, the literature makes the distinction between classical functional models, where the unobserved true X's are regarded as a sequence of unknown fixed constants, and classical structural models, in which the X's are regarded as random variables following some parametric model. In this paper, we consider measurement error models where X is considered a random variable but no restriction is made on the conditional distribution of X given Z. Using the terminology of Carroll et al. (1995, § 7.2), we refer to such models as 'functional' measurement error models. As such, these functional measurement error models are semiparametric models with a finite-dimensional parameter β of interest and an infinite-dimensional nuisance parameter corresponding to the non-parametric conditional distribution of X given Z. Such functional measurement error models are also examples of mixture models as summarised by Lindsay & Lesperance (1995) and van der Vaart (1996).

For some specific situations, semiparametric estimators have been proposed for the parameter β in functional measurement error models. These generally include models that admit a statistic, $\psi_\beta(Y, W, Z)$, which is complete and sufficient for the nuisance parameter corresponding to the nonparametric conditional density $p(x|z)$ given a fixed value of β. As described by Carroll et al. (1995), complete sufficient statistics occur naturally for models where the distribution of Y given X and Z is a canonical generalised linear model and there is additive normally distributed measurement error. A simple example is the linear logistic regression model with a single covariate X given by $\text{logit}\{\text{pr}(Y=1|X)\} = \beta_0 + \beta_1 X$ and additive normal measurement error. Estimators for such models were studied in detail by Stefanski & Carroll (1985, 1987).

For models where such a complete sufficient statistic exists, semiparametric estimators have been derived and the efficient estimator has been characterised. For a survey of these methods see Carroll et al. (1995, § 6). In more general problems where no such complete sufficient statistic exists, to the best of our knowledge, little is known about how to derive feasible semiparametric estimators for β. For example, the existing theory could not be used to derive an estimator for the parameters in a logistic regression model with a single covariate that includes a quadratic term, i.e.

$$\text{logit}\{\text{pr}(Y=1|X)\} = \beta_0 + \beta_1 X + \beta_2 X^2, \tag{1}$$

and normal additive measurement error. This model will be used later to illustrate the proposed methods.

In this paper, we construct estimators for the parameter β in functional measurement error models by defining estimating equations using the efficient score derived as the residual after projecting the score vector with respect to the parameter β on to the nuisance tangent space for the nonparametric conditional distribution of X given Z. However, the projection on to the nuisance tangent space involves the unknown distribution of X given Z. Although the distribution of X given Z can be estimated nonparametrically using deconvolution, the rate of convergence may be slow. The innovation of this paper is to show that the residual, derived under an incorrectly specified model for the distribution of X given Z, has mean zero. This allows us to compute the residuals from a sample of data using a simpler, but possibly incorrect, model for the distribution of X given Z to form estimating equations whose solution will still yield consistent semiparametric estimators of β. Moreover, if the correct model is used to derive the projection to the nuisance tangent space, then the estimator will be semiparametric efficient. Such an estimator is referred to as locally efficient.

2. Notation and preliminaries

We consider regular asymptotically linear estimators for the q-dimensional parameter vector β, derived from a sample of data, represented by the independent and identically distributed random vectors O_i $(i = 1, \ldots, n)$, where $O_i = (Y_i, W_i, Z_i)$; that is, the estimator minus the estimand can be approximated asymptotically by a sum of independent and identically distributed zero-mean random vectors. To be specific, an estimator $\hat{\beta}_n$ for the parameter β is asymptotically linear if

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

where $\phi(O_i)$ $(i = 1, \ldots, n)$ are independent and identically distributed zero-mean q-dimensional random vectors and $o_p(1)$ denotes a term that converges in probability to zero. The random vector $\phi(O_i)$ is referred to as the ith influence function of the estimator $\hat{\beta}_n$. The restriction to regular estimators is a technical condition imposed to exclude estimators that have undesirable local properties; for details, see Newey (1990). It is clear from the representation above that the asymptotic variance of a regular asymptotically linear estimator is equal to the variance of its influence function. Consequently, the optimal estimator among a class of regular asymptotically linear estimators is the one whose influence function has the smallest variance.

A geometric point of view is taken in which influence functions of a single observation for regular asymptotically linear estimators of β lie in the Hilbert space $\mathscr{H}$ of all L_2 q-dimensional zero-mean measurable functions, $h(O)$, of the observed data, with finite variance, equipped with the inner product $\langle h_1, h_2 \rangle = E\{h_1^{\mathrm{T}}(O)h_2(O)\}$, where $h_1, h_2 \in \mathscr{H}$ and norm $\|h\| = \{E(h^{\mathrm{T}}h)\}^{\frac{1}{2}}$. Note that when we refer to functions of a single observation we often suppress the subscript i. This geometric representation is useful because it will enable us to identify influence functions of regular asymptotically linear estimators, which, in turn, will motivate estimating equations that will yield semiparametric estimators for β. According to the theory of semiparametrics, see for example Bickel et al. (1993), influence functions for regular asymptotically linear estimators belong to the linear space orthogonal to the nuisance tangent space, as we now describe.

Consider a parametric model where the density of the data is given by $p(o; \beta, \eta)$, where β is the q-dimensional parameter of interest, and η is an r-dimensional nuisance parameter. Denote by $S_\beta(O)$ and $S_\eta(O)$ the score vectors with respect to β and η respectively, where the score vector is defined as the partial derivative of the loglikelihood with respect to the indicated parameters. We denote by β_0 and η_0 the true values of the parameters, and, unless otherwise stated, the score vectors are evaluated at the true values. For a parametric model, the nuisance tangent space, denoted by Λ, is the linear space in $\mathscr{H}$ that is spanned by the nuisance score vector $S_\eta(O)$; that is, $\Lambda = \{BS_\eta(O)$ for all $q \times r$ matrices $B\}$. For semiparametric models, in which the nuisance parameter is infinite-dimensional, the nuisance tangent space Λ is defined as the mean squared closure of all parametric submodel nuisance tangent spaces. Roughly speaking, a parametric submodel is a parametric model contained in the semiparametric model that contains the truth, i.e. contains (β_0, η_0). By the projection theorem for Hilbert spaces (Luenberger, 1969, p. 51), the projection of $h \in \mathscr{H}$ on to a closed linear subspace Λ of $\mathscr{H}$ is the unique element in Λ, denoted by $\Pi(h|\Lambda)$, such that $\|h - \Pi(h|\Lambda)\|$ is minimised and the residual $h - \Pi(h|\Lambda)$ is orthogonal to all $\lambda \in \Lambda$; that is, $E[\{h - \Pi(h|\Lambda)\}^{\mathrm{T}}\lambda] = 0$ for all $\lambda \in \Lambda$. The efficient score $S_{\mathrm{eff}}(O)$ for the semiparametric model is the residual of the score vector S_β after projecting it on to the nuisance tangent space Λ, denoted by $S_{\mathrm{eff}} = S_\beta - \Pi(S_\beta|\Lambda)$, and the efficient influence function is $\phi_{\mathrm{eff}}(O) = (E[\{S_{\mathrm{eff}}(O)S_{\mathrm{eff}}^{\mathrm{T}}(O)\}])^{-1}S_{\mathrm{eff}}(O)$. In general, the variance of ϕ_{eff}, given by $(E[\{S_{\mathrm{eff}}(O)S_{\mathrm{eff}}^{\mathrm{T}}(O)\}])^{-1}$, achieves the so-called semiparametric efficiency bound, i.e. the supremum over all parametric submodels of the Cramér–Rao lower bounds for β.

3. The geometry of functional measurement error models

3·1. *Under the truth*

Data from measurement error models can be viewed as coarsened data; that is, if we denote the full or complete, but possibly unobserved, data for the ith observation by D_i then the coarsened observed data O_i are defined by a many-to-one function of D_i. In our case the full data are $D_i = (Y_i, X_i, W_i, Z_i)$ and the observed data are $O_i = (Y_i, W_i, Z_i)$. The probability model for the observed data is induced from the underlying probability model for the full data. Under the assumptions previously stated, the probability density of the full data is given by

$$p(d; \beta, \eta_1, \eta_2) = p(y|x, z; \beta)p(w|x, z)\eta_1(x|z)\eta_2(z), \tag{2}$$

where β, the parameter of interest, is q-dimensional, whereas the nuisance parameters $\eta = (\eta_1, \eta_2)$ are infinite-dimensional, reflecting the fact that no restriction is put on the conditional distribution of X given Z or the marginal distribution of Z. Influence functions of regular asymptotically linear estimators for β, based on the full data D, lie in the Hilbert space $\mathscr{H}^F$ corresponding to all q-dimensional zero-mean functions of the full data with finite variance and inner product $\langle h_1, h_2 \rangle = E\{h_1^{\mathrm{T}}(D)h_2(D)\}$ for $h_1, h_2 \in \mathscr{H}^F$. As a result of the factorisation given in (2), the nuisance tangent space, Λ^F, is the mean squared closure of the tangent spaces of parametric submodels $\eta_1(x|z; \xi_1)$ and $\eta_2(z; \xi_1)$ given by the sum of two subspaces; that is $\Lambda^F = \Lambda_1^F \oplus \Lambda_2^F$, where (Newey, 1990)

$$\Lambda_1^F = [h_1(X, Z) \in \mathscr{H}^F \text{ such that } E\{h_1(X, Z)|Z\} = 0],$$
$$\Lambda_2^F = [h_2(Z) \in \mathscr{H}^F \text{ such that } E\{h_2(Z)\} = 0].$$

The nuisance tangent space can also be represented as

$$\Lambda^F = [h(X, Z) \in \mathscr{H}^F \text{ such that } E\{h(X, Z)\} = 0].$$

The full-data score vector $S_\beta^F(Y, X, Z)$ is the q-dimensional vector $\partial/\partial\beta\{\log p(Y|X, Z; \beta)\}$. Again, because of the factorisation in (2), it is easy to show that Λ_1^F is orthogonal to Λ_2^F and that $S_\beta^F(Y, X, Z)$ is orthogonal to $\Lambda_1^F \oplus \Lambda_2^F$.

Since the score vector $S_\eta(O)$ for coarsened data O is equal to $E\{S_\eta(D)|O\}$, see for example Rao (1973, p. 330), the observed (coarsened) data nuisance tangent space is given by $\Lambda = \Lambda_1 \oplus \Lambda_2$, where

$$\begin{aligned}\Lambda &= E(\Lambda^F | Y, W, Z) \\ &= [E\{h(X, Z)|Y, W, Z\}, \text{ where } h(X, Z) \in \Lambda^F; \text{ that is } E\{h(X, Z)\} = 0], \end{aligned} \tag{3}$$

$\Lambda_1 = E(\Lambda_1^F | Y, W, Z)$ and $\Lambda_2 = \Lambda_2^F$. Also with coarsened data, the score vector with respect to β is given by

$$S_\beta(Y, W, Z) = E\{S_\beta^F(Y, X, Z)|Y, W, Z\}. \tag{4}$$

It is important to keep in mind the fact that a Hilbert space depends on the probability distribution that generates the data, both in terms of identifying the elements of the space of zero-mean random vectors and of defining the inner product. Consequently, different probability distributions define different Hilbert spaces. In working with these Hilbert space representations, it is generally assumed that the Hilbert space is defined with respect to the true distribution that generated the data, which we denote by $p(o; \beta_0, \eta_{10}, \eta_{20})$.

3·2. *Under misspecification*

As we will see shortly, it is useful to explore the consequences of having assumed parts of the model incorrectly. To be specific, we allow the conditional density $\eta_1(x|z)$ to be incorrectly specified and we denote the incorrect conditional density by $\eta_1^*(x|z)$. We will continue to assume that the conditional distribution of Y given X and Z is generated at the truth with conditional density $p(y|x, z; \beta_0)$. Consequently, we will consider the Hilbert space $\mathscr{H}^*$ with respect to the partially misspecified probability distribution for (Y, W, Z) that is induced by the distribution of the full data given by the density $p(y, w, x, z) = p(y|x, z; \beta_0)p(w|x, z)\eta_1^*(x|z)\eta_2(z)$. We also use the convention that expectations or conditional expectations computed under the incorrectly specified distribution are denoted by $E_*(.)$, and expectations computed under the correctly specified distribution will be denoted by $E_0(.)$ or just $E(.)$.

Similarly, the incorrectly specified nuisance tangent space will be denoted by $\Lambda^* = \Lambda_1^* \oplus \Lambda_2^*$, where Λ^*, Λ_1^* and Λ_2^* are defined as before but replacing the expectations $E(.)$ by $E_*(.)$. Note that Λ_1^* is orthogonal to Λ_2^* in $\mathscr{H}^*$. We now demonstrate that, although the nuisance tangent space $\Lambda^* \subset \mathscr{H}^*$ depends on the possibly incorrectly specified conditional density $\eta_1^*(x|z)$, the space orthogonal to the nuisance tangent space is invariant, almost surely, to misspecification.

THEOREM 1. *For the Hilbert space $\mathscr{H}^*$ that is induced by a possibly incorrect probability distribution of the observed data with conditional density $\eta_1^*(x|z)$, the space orthogonal to the nuisance tangent space $\Lambda^* \subset \mathscr{H}^*$ is given by*

$$[h(Y, W, Z) \text{ such that } E\{h(Y, W, Z)|X, Z\} = 0 \text{ almost everywhere}],$$

as long as the possibly incorrect conditional density $\eta_1^*(x|z)$ has the same support as the true conditional density $\eta_{10}(x|z)$.

Proof. By the definition of $\Lambda^* \subset \mathscr{H}^*$ given by (3), $h(Y, W, Z)$ is orthogonal to Λ^* if

$$E_*[h^{\mathrm{T}}(Y, W, Z)E_*\{g(X, Z)|Y, W, Z\}] = 0 \tag{5}$$

for all $g(X, Z)$ such that $E_*\{g(X, Z)\} = 0$. Using a series of iterated conditional expectations, we note that

$$\begin{aligned} E_*[h^{\mathrm{T}}(Y, W, Z)E_*\{g(X, Z)|Y, W, Z\}] &= E_*[E_*\{h^{\mathrm{T}}(Y, W, Z)g(X, Z)|\,Y, W, Z\}] \\ &= E_*\{h^{\mathrm{T}}(Y, W, Z)g(X, Z)\} \\ &= E_*[E_*\{h^{\mathrm{T}}(Y, W, Z)g(X, Z)|X, Z\}] \\ &= E_*[E_*\{h^{\mathrm{T}}(Y, W, Z)|X, Z\}g(X, Z)]. \end{aligned}$$

Consequently, (5) is equivalent to

$$E_*[E_*\{h^{\mathrm{T}}(Y, W, Z)|X, Z\}g(X, Z)] = 0 \tag{6}$$

for all $g(X, Z)$ such that $E_*\{g(X, Z)\} = 0$, which implies that

$$E_*\{h(Y, W, Z)|X, Z\} = 0, \tag{7}$$

almost everywhere ($*$) where almost everywhere ($*$) denotes almost everywhere with respect to the distribution of the observed data induced by the possibly incorrect model $p(o; \beta_0, \eta_1^*, \eta_2)$. Since the conditional distribution of (Y, W, Z) given (X, Z) depends on the conditional densities $p(y|x, z; \beta_0)$ and $p(w|x, z)$, both of which are assumed correctly specified, then $E_*\{h(Y, W, Z)|X, Z\} = 0$ almost everywhere ($*$) is equivalent to $E_0\{h(Y, W, Z)|X, Z\} = 0$ almost everywhere ($*$). If, in addition, the possibly incorrect conditional density $\eta_1^*(x|z)$ has the same support as the true conditional density $\eta_{10}(x|z)$, then (7) is equivalent to $E_0\{h(Y, W, Z)|X, Z\} = 0$ almost everywhere which completes the proof. □

This result is important because we may start with an incorrectly specified conditional distribution of X given Z, and the corresponding incorrectly specified Hilbert space $\mathscr{H}^*$, but, as long as we can define a statistic $h(Y, W, Z)$ of the observed data which is orthogonal to the nuisance tangent space Λ^*, then by Theorem 1 this statistic will have conditional expectation, with respect to the truth, equal to zero almost surely; hence, it will have unconditional expectation equal to zero, with respect to the truth. We will exploit this to derive unbiased estimating equations for the parameter β. To be specific, the strategy we propose is to derive the efficient score, for the possibly incorrectly specified model, by finding the residual after projecting the observed data score vector $S_\beta^*(Y, W, Z) \in \mathscr{H}^*$ on to the nuisance tangent space $\Lambda^* \subset \mathscr{H}^*$ using the possibly incorrect conditional density $\eta_1^*(x|z)$. Note that the observed data score vector given by equation (4) may also be misspecified; therefore, we denote the possibly misspecified score vector by $S_\beta^*(Y, W, Z)$. Even though we used the incorrect model, the efficient score derived above will be orthogonal to $\Lambda^* \subset \mathscr{H}^*$ and hence will have mean zero, under the truth. The efficient score will serve as the basis for constructing unbiased estimating equations whose solution will be a semiparametric estimator for β. Moreover, if the underlying model is correctly specified, then the estimator will be semiparametric efficient.

3·3. *The efficient score under misspecification*

The observed-data score vector $S^*_\beta(Y, W, Z)$, defined by (4), is given by

$$\frac{\int S^F_\beta(Y, x, Z)p(Y|x, Z; \beta_0)p(W|x, Z)\eta^*_1(x|Z)d\mu(x)}{\int p(Y|x, Z; \beta_0)p(W|x, Z)\eta^*_1(x|Z)d\mu(x)}, \tag{8}$$

where, throughout, we use $d\mu(.)$ to denote the dominating measure, which in our examples is either Lebesgue or counting measure for continuous and discrete random variables, respectively. We showed earlier that the nuisance tangent space can also be written as $\Lambda^*_1 \oplus \Lambda^*_2$, where Λ^*_1 is orthogonal to Λ^*_2. Since $S^*_\beta(Y, W, Z)$ is a score vector, this implies that $E_*\{S^*_\beta(Y, W, Z)|Z\} = 0$; hence, $S^*_\beta(Y, W, Z)$ is orthogonal to Λ^*_2. Consequently, to find the projection of $S^*_\beta(Y, W, Z)$ on to Λ^*, it suffices to project on to Λ^*_1. The projection of $S^*_\beta(Y, W, Z)$ on to Λ^*_1 is the unique element in Λ^*_1, that is $E_*\{a(X, Z)|Y, W, Z\}$, such that $S^*_\beta(Y, W, Z) - E_*\{a(X, Z)|Y, W, Z\}$ is orthogonal to Λ^*_1 or, equivalently, orthogonal to Λ^*. By Theorem 1, the function $a(X, Z)$ must therefore satisfy the integral equation

$$E[S^*_\beta(Y, W, Z) - E_*\{a(X, Z)|Y, W, Z\}|X, Z] = 0,$$

which implies that

$$E\{S^*_\beta(Y, W, Z)|X, Z\} = E[E_*\{a(X, Z)|Y, W, Z\}|X, Z]. \tag{9}$$

Remark 1. The outer conditional expectations on both sides of equation (9) are functions of (Y, W, Z) given (X, Z), are hence derived under the truth and are thus denoted by $E(.)$.

Remark 2. In order that $E_*\{a(X, Z)|Y, W, Z\}$ be the projection on to Λ^*_1, the function $a(X, Z)$ must satisfy $E_*\{a(X, Z)|Z\} = 0$. However, if the function $a(X, Z)$ satisfies equation (9) then it must also satisfy $E_*\{a(X, Z)|Z\} = E_*\{S^*_\beta(Y, W, Z)|Z\} = 0$.

Remark 3. Although the projection $E_*\{a(X, Z)|Y, W, Z\}$ is unique, the function $a(X, Z)$ may not be. However, because of the projection theorem, there must exist at least one solution to the integral equation (9), and any solution will lead to the same projection.

After we have found a solution $a(X, Z)$ to the integral equation, the efficient score is given as

$$S^*_{\text{eff}}(Y, W, Z) = S^*_\beta(Y, W, Z) - E_*\{a(X, Z)|Y, W, Z\}. \tag{10}$$

The efficient score, although computed under the incorrect model $\eta^*_1(x|z)$, has conditional expectation $E\{S^*_{\text{eff}}(Y, W, Z)|X, Z\} = 0$.

Remark 4. In order to ensure that the efficient score is not identically equal to zero, at the least, we need the space orthogonal to the nuisance tangent space to contain nontrivial elements, i.e. functions other than zero. In the Appendix, we show this to be the case for the quadratic logistic regression models (1) with normal and exponential measurement error which are used for illustration in § 5. Depending on the model, it may or may not be difficult to prove the existence of nontrivial elements. However, in § 4·2 we describe methods for computing the projection orthogonal to the nuisance tangent space numerically. Consequently, if the projection, computed numerically, were close to zero, then this would be an indication that the problem does not have a good semiparametric estimator.

Remark 5. To take advantage of these results, we need methods for solving, or at least for finding an approximate numerical solution to, the integral equation (9). A simple approximation by discretising X is discussed later.

4. The locally efficient semiparametric estimator

4·1. *Asymptotic properties*

The observed-data score vector with respect to β, the projection and the efficient score vector, derived in § 3·3, were computed assuming the correct density for $p(y|x, z; \beta_0)$ and the possibly incorrect density $\eta_1^*(x|z)$. Consequently, we suppressed the relationship of these quantities to the unknown parameters in the model. In this section, we need to make such relationships explicit in order to derive the appropriate estimating equations. We also consider a parametric model for the possibly misspecified conditional density of X given Z, given by $\eta_1^*(x|z; \xi)$, where ξ denotes an unknown r-dimensional parameter. The true conditional density $\eta_{10}(x|z)$ may or may not belong to this parametric model, but we assume that the support is the same. Using the observed data O_i $(i = 1, \ldots, n)$, we denote the observed data score vector (8) by $S_\beta^*(O_i, \beta, \xi)$, the solution to the integral equation (9) by $a(X_i, Z_i, \beta, \xi)$, and the efficient score (10) by $S_{\mathrm{eff}}^*(O_i, \beta, \xi)$, where these quantities are computed using the conditional densities $p(y|x, z; \beta)$ and $\eta_1^*(x|z; \xi)$.

Based on the theory developed in the previous section, we propose to estimate β by solving the estimating equation

$$\sum_{i=1}^{n} S_{\mathrm{eff}}^*(O_i, \beta, \hat{\xi}_n) = 0, \tag{11}$$

where $\hat{\xi}_n$ is an estimator for ξ that is root-n consistent; that is, there exists a constant ξ^* such that $n^{\frac{1}{2}}(\hat{\xi}_n - \xi^*)$ is bounded in probability. We denote the resulting estimator by $\hat{\beta}_n$. Under suitable regularity conditions, we now argue heuristically that the estimator $\hat{\beta}_n$ will be consistent and asymptotically normal.

One way to show consistency is by the use of the inverse function theorem (Foutz, 1977). In this regard, we need to show that $E\{S_{\mathrm{eff}}^*(O, \beta_0, \xi^*)\} = 0$. We also need the estimating function $n^{-1}\sum_{i=1}^{n} S_{\mathrm{eff}}^*(O_i, \beta, \xi)$ and its expectation $E\{S_{\mathrm{eff}}^*(O, \beta, \xi)\}$ to be sufficiently smooth as functions of β and ξ, in a neighbourhood $\mathcal{N}$ of (β_0, ξ^*), so that the $q \times q$ matrix of partial derivatives,

$$J_n(\beta, \xi) = n^{-1} \sum_{i=1}^{n} \frac{\partial S_{\mathrm{eff}}^*(O_i, \beta, \xi)}{\partial \beta},$$

converges uniformly, in probability, to the matrix $B(\beta, \xi) = E[\partial/\partial\beta\{S_{\mathrm{eff}}^*(O, \beta, \xi)\}]$ for (β, ξ) in $\mathcal{N}$. The matrix $B(\beta, \xi)$ must also be smooth in β and ξ in $\mathcal{N}$ and $B(\beta_0, \xi^*)$ nonsingular.

That $E\{S_{\mathrm{eff}}^*(O, \beta_0, \xi^*)\} = 0$ follows from equation (9), which is a direct consequence of Theorem 1. In the Appendix we show that the expected values of the matrix of partial derivatives is

$$B(\beta_0, \xi^*) = -E\{S_{\mathrm{eff}}^*(O, \beta_0, \xi^*) S_{\mathrm{eff}}^{\mathrm{T}}(O, \beta_0, \xi^*)\}. \tag{12}$$

The other technical smoothness conditions, necessary to verify Foutz's theorem, have to be considered separately for each model and may be technically challenging. From here on, we assume that these conditions hold.

To prove asymptotic normality, we first expand the estimating function, given in (11), as a function of β, about β_0 and keeping $\hat{\xi}_n$ fixed, to obtain

$$n^{\frac{1}{2}}(\hat{\beta}_n - \beta_0) = \{-J_n(\tilde{\beta}_n, \hat{\xi}_n)\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} S_{\mathrm{eff}}^*(O_i, \beta_0, \hat{\xi}_n), \tag{13}$$

where $\tilde{\beta}_n$ is a value between β_0 and $\hat{\beta}_n$. Another expansion of $\hat{\xi}_n$ about ξ^*, in the summand on the right-hand side of equation (13), yields that $n^{\frac{1}{2}}(\hat{\beta}_n - \beta_0)$ equals

$$\{-J_n(\tilde{\beta}_n, \hat{\xi}_n)\}^{-1}\left\{n^{-\frac{1}{2}} \sum_{i=1}^{n} S^*_{\text{eff}}(O_i, \beta_0, \xi^*) + G_n(\beta_0, \tilde{\xi}_n) n^{\frac{1}{2}}(\hat{\xi}_n - \xi^*)\right\}, \tag{14}$$

where $G_n(\beta_0, \xi)$ is the $q \times r$ matrix of partial derivatives

$$G_n(\beta_0, \xi) = n^{-1} \sum_{i=1}^{n} \frac{\partial S^*_{\text{eff}}(O_i, \beta_0, \xi)}{\partial \xi},$$

and $\tilde{\xi}_n$ is a value between ξ^* and $\hat{\xi}_n$. We now make the additional technical assumption that the efficient score, $S^*_{\text{eff}}(O_i, \beta_0, \xi)$, is sufficiently smooth in ξ so that $G_n(\beta_0, \xi)$ converges to

$$E\left\{\frac{\partial S^*_{\text{eff}}(O_i, \beta_0, \xi)}{\partial \xi}\right\} \tag{15}$$

uniformly in ξ in a neighbourhood of ξ^*. In the Appendix we prove that (15) is equal to zero, which, together with (14), the technical assumptions and the assumption that $\hat{\xi}_n$ is root-n consistent, yields

$$n^{\frac{1}{2}}(\hat{\beta}_n - \beta_0) = \{-B(\beta_0, \xi^*)\}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} S^*_{\text{eff}}(O_i, \beta_0, \xi^*) + o_p(1). \tag{16}$$

Since $n^{-\frac{1}{2}} \sum_{i=1}^{n} S^*_{\text{eff}}(O_i, \beta_0, \xi^*)$ is a normalised sum of zero-mean random vectors, this will converge in distribution to a multivariate normal with mean zero and covariance matrix

$$V(\beta_0, \xi^*) = E\{S^*_{\text{eff}}(O_i, \beta_0, \xi^*) S^{*\text{T}}_{\text{eff}}(O_i, \beta_0, \xi^*)\}.$$

It then follows immediately from (16) that $n^{\frac{1}{2}}(\hat{\beta}_n - \beta_0)$ is asymptotically normal with mean zero and covariance matrix

$$\{-B(\beta_0, \xi^*)\}^{-1} V(\beta_0, \xi^*) \{-B^{\text{T}}(\beta_0, \xi^*)\}^{-1}. \tag{17}$$

We suggest estimating the asymptotic covariance matrix (17) using the sandwich estimator, namely,

$$\{-J_n(\hat{\beta}_n, \hat{\xi}_n)\}^{-1} \hat{V}_n(\hat{\beta}_n, \hat{\xi}_n) \{-J_n^{\text{T}}(\hat{\beta}_n, \hat{\xi}_n)\}^{-1}, \tag{18}$$

where $J_n(\hat{\beta}_n, \hat{\xi}_n)$ is computed using numerical derivatives and

$$\hat{V}_n(\hat{\beta}_n, \hat{\xi}_n) = n^{-1} \sum_{i=1}^{n} S^*_{\text{eff}}(O_i, \hat{\beta}_n, \hat{\xi}_n) S^{*\text{T}}_{\text{eff}}(O_i, \hat{\beta}_n, \hat{\xi}_n).$$

Finally, if the true conditional density for X given Z is contained in the model $\eta_1^*(x|z; \xi)$, $\eta_{10}(x|z) = \eta_1^*(x|z; \xi_0)$ say, and the estimator $\hat{\xi}_n$ converges in probability to ξ_0, then $S^*_{\text{eff}}(O_i, \beta_0, \xi^*)\}$ is the efficient score and $\hat{\beta}_n$ is the semiparametric efficient estimator for β.

4·2. *Solving the integral equation for discrete X*

In order to obtain estimators using (11), we need to compute $S^*_{\text{eff}}(O_i, \beta, \xi)$. This entails solving the integral equation (9). We now give a simple approximation to this solution by taking X to be discrete with mass at m points, $x_1, \ldots, x_m$, spread across the support of X. This approximation is easily implemented and works well in the examples described later. We denote the discrete conditional probability density of X given Z by $\eta_1^*(x|z)=\sum_{j=1}^m c_j(z)I(x=x_j)$, with $\sum_{j=1}^m c_j(z)=1$ for all z in the support of Z. Then

$$E[E_*\{a(X,Z)|Y,W,Z\}|X=x_i,Z] \\ = \int \left\{\frac{\sum_{j=1}^m a_j(Z)q_j(y,w,Z)c_j(Z)}{\sum_{j=1}^m q_j(y,w,Z)c_j(Z)}\right\} q_i(y,w,Z)d\mu(y)d\mu(w),$$

where $q_j(y,w,Z)=p(y|x_j,Z)p(w|x_j,Z)$ and $a_j(Z)=a(x_j,Z)$. Also, by the definition of the observed score given by (8), we have

$$E_*\{S_\beta^*(Y,W,Z)|X=x_i,Z\} \\ = \int \left\{\frac{\sum_{j=1}^m S_\beta^F(y,x_j,Z)q_j(y,w,Z)c_j(Z)}{\sum_{j=1}^m q_j(y,w,Z)c_j(Z)}\right\} c_i(y,w,Z)d\mu(y)d\mu(w). \tag{19}$$

Consequently, the solution to the integral equation reduces to the linear equations,

$$A(Z)a^{\mathrm{T}}(Z)=b^{\mathrm{T}}(Z), \tag{20}$$

where $a(Z)$ is the $q\times m$ matrix $\{a_1(Z),\ldots,a_m(Z)\}$, corresponding to the solution of the integral equation, $A(Z)$ is an $m\times m$ matrix whose (i,j)th element is given by

$$A_{ij}(Z)=\int \left\{\frac{q_j(y,w,Z)c_j(Z)}{\sum_{j=1}^m q_j(y,w,Z)c_j(Z)}\right\} q_i(y,w,Z)d\mu(y)d\mu(w), \tag{21}$$

and $b(Z)$ is an $q\times m$ matrix whose ith column is $E_*\{S_\beta^*(Y,W,Z)|X=x_i,Z\}$ defined in (19).

The efficient score is given by

$$S_{\mathrm{eff}}^*(Y,W,Z)=\frac{\sum_{j=1}^m \{S_\beta^F(Y,x_j,Z)-a_j(Z)\}q_j(Y,W,Z)c_j(Z)}{\sum_{j=1}^m q_j(Y,W,Z)c_j(Z)}. \tag{22}$$

Since we put mass only at points $x_1\ldots,x_m$, the theory only guarantees that

$$E\{S_{\mathrm{eff}}^*(Y,W,Z)|X,Z\}=0$$

when $X=x_j$ $(j=1,\ldots,m)$. However, if $E\{S_{\mathrm{eff}}^*(Y,W,Z)|X=x,Z\}$ is a smooth function of x and if the grid points are sufficiently dense along the support of X, then $|E\{S_{\mathrm{eff}}^*(Y,W,Z)|X=x,Z\}|$ can be made arbitrarily small for all x in the support of X and the unconditional mean $E\{S_{\mathrm{eff}}^*(Y,W,Z)\}\simeq 0$.

5. EXAMPLE USING A QUADRATIC LOGISTIC REGRESSION MODEL

We illustrate the proposed methods by considering a specific example where a binary response Y is related to a single covariate X, measured with error, through a quadratic logistic regression model. In particular, we consider the model $\mathrm{logit}\{\mathrm{pr}(Y=1|X)\}=\beta_0+\beta_1X+\beta_2X^2$, where we only observe $W=X+\varepsilon$ with ε independent of Y and X following a known distribution. We conducted several simulation experiments to evaluate the properties of the locally efficient estimator for $\beta=(\beta_0,\beta_1,\beta_2)$.

We report here on simulations where we took $X\sim N(-1,1)$. We considered two different measurement error models, one with $\varepsilon\sim N(0,\sigma_\varepsilon^2)$ and the other in which ε was taken to have an exponential distribution with mean μ_ε. The latter allows investigation of the consequences of an asymmetric measurement error model. In both scenarios, there is no complete sufficient statistic for the nonparametric nuisance parameter $\eta_1(x)$; consequently, we know of no published method for obtaining semiparametric estimators for β.

We considered the estimator given as the solution to equation (11) using two models for the possibly incorrect density of X, $\eta_1^*(x; \xi)$, both of which depend on a parameter ξ, which we took to be the first two moments of X. To be specific, letting μ_X and σ_X^2 denote the mean and variance of X, we took $\eta_1^*(x; \mu_X, \sigma_X^2)$ to be the normal density with mean μ_X and variance σ_X^2; in this case, this coincides with the true distribution that was used to generate X. We also considered $\eta_1^*(x; \mu_X, \sigma_X^2)$ to be the uniform density on $\mu_X \pm 3\sigma_X$, to consider the effect of misspecification. Simple moment estimators were used to estimate μ_X and σ_X^2. For example, with normal measurement error, the estimator for μ_X was $\hat{\mu}_X = \bar{W} = \sum_{i=1}^n W_i/n$, and the estimator for σ_X^2 was $\hat{\sigma}_X^2 = s_W^2 - \sigma_\varepsilon^2$, where s_W^2 is the sample variance for W. Similar estimators were obtained for the exponential error model.

In constructing the estimator for β, we used the discrete approximation given in (22), based on a grid of 15 points equally spaced between $\hat{\mu}_X \pm 3\hat{\sigma}_X$ and with probabilities proportional to the density $\eta_1^*(x; \hat{\mu}_X, \hat{\sigma}_X^2)$. We found the results to be insensitive to the number of such grid points as long as that number exceeds 7 for the normal measurement error model and 15 for the exponential measurement error model. In order to compute the values a_j in (22), which are the solution to the linear system of equations in (20), we computed the integrals in (19) and (21) using Hermite quadrature for normal measurement error and Laguerre quadrature for exponential measurement error.

For each scenario, we conducted 1000 simulations, each with sample size $n = 500$. The true values for $(\beta_0, \beta_1, \beta_2)$ were taken to be $(-1, 0{\cdot}7, 0{\cdot}7)$. A substantial amount of measurement error was considered; for the normal measurement error, $\sigma_\varepsilon = 0{\cdot}40$, and for the exponential measurement error, $\mu_\varepsilon = 0{\cdot}40$. The covariance matrix of the estimator was estimated using the sandwich variance (18), and 95% confidence intervals were constructed using the estimate $\pm 1{\cdot}96$ estimated standard error. For comparison, we also considered naive and regression calibration estimators. For the naive estimator, the parameters were estimated using standard logistic regression maximum likelihood where W was used instead of X in the quadratic logistic regression model with normal measurement error and $W - \mu_\varepsilon$ was substituted for X with exponential measurement error. For the regression calibration estimators, we replaced X and X^2 by $E(X|W)$ and $E(X^2|W)$ when fitting the quadratic logistic regression model. To compute the conditional expectations for the regression calibration estimators, one needs the marginal distribution of X as well as the conditional distribution of W given X. We considered two extremes, one in which the distribution of X was correctly specified, $N(-1, 1)$, and the other where X was taken to be $\mathrm{Un}[-4, 2]$. We wanted to consider the consequences of such misspecification for the distribution of X even though, in practice, the misspecification is unlikely to be as extreme as this with proper data analysis.

The results of the simulations are summarised in Table 1. As expected, the naive estimators for β are severely biased whereas the locally efficient estimators all give good results. These estimators exhibit little bias, the average of the estimated variances closely approximates the Monte Carlo variance, and the proportion of times that the estimated 95% confidence interval covers the true value is close to the nominal level. Of particular interest is that the results are insensitive to misspecification of $\eta_1^*(x)$. The efficiency when using the misspecified uniform distribution is virtually identical to that when using the correct normal distribution for the normal measurement error example, and there is only a slight loss of efficiency for the exponential error example. We found this to be the case

Table 1: *Simulation study. Bias, variance and coverage probabilities of the naive, regression calibration and locally efficient semiparametric estimators for the quadratic logistic regression model with normal and exponential measurement error*

Estimator		Normal errors $\beta_0(-1)$	Normal errors $\beta_1(0{\cdot}7)$	Normal errors $\beta_2(0{\cdot}7)$	Exponential errors $\beta_0(-1)$	Exponential errors $\beta_1(0{\cdot}7)$	Exponential errors $\beta_2(0{\cdot}7)$
naive	mean	−0·97	0·40	0·48	−0·99	0·37	0·48
	emp var	0·021	0·028	0·007	0·022	0·033	0·008
	est var	0·020	0·027	0·007	0·020	0·022	0·006
	emp cov	0·94	0·53	0·24	0·93	0·40	0·25
reg cal (true X)	mean	−0·97	0·63	0·64	−0·99	0·64	0·66
	emp var	0·023	0·046	0·012	0·022	0·048	0·013
	est var	0·022	0·046	0·012	0·022	0·045	0·012
	emp cov	0·94	0·94	0·89	0·95	0·93	0·90
reg cal (mis X)	mean	−1·03	0·44	0·49	−1·07	0·37	0·48
	emp var	0·021	0·027	0·007	0·022	0·031	0·008
	est var	0·020	0·029	0·007	0·021	0·022	0·006
	emp cov	0·94	0·64	0·31	0·92	0·40	0·24
semipar (true X)	mean	−1·00	0·72	0·72	−1·01	0·69	0·70
	emp var	0·026	0·068	0·022	0·024	0·064	0·017
	est var	0·026	0·070	0·022	0·026	0·070	0·018
	emp cov	0·95	0·96	0·95	0·95	0·94	0·94
semipar (mis X)	mean	−1·00	0·72	0·72	−1·00	0·68	0·69
	emp var	0·027	0·068	0·022	0·023	0·073	0·020
	est var	0·026	0·070	0·022	0·024	0·077	0·021
	emp cov	0·94	0·96	0·95	0·95	0·94	0·93

reg cal (true X) and reg cal (mis X) denote the regression calibration estimators derived under the true distribution and misspecified distribution of X, respectively; semipar (true X) and semipar (mis X) denote the locally efficient semiparametric estimators derived under the true distribution and misspecified distribution of X, respectively; emp var is the empirical Monte Carlo variance of the estimators; est var is the average of the estimated variances; emp cov is the proportion of the simulations whose estimated 95% confidence intervals cover the true value of the parameters.

in all of the many simulations we conducted. When we used the correct distribution for X, the regression calibration estimators showed a slight bias. However, use of the same misspecification for the distribution of X as that for the locally efficient estimators resulted in regression calibration estimators that were severely biased.

6. Concluding remarks

Throughout, we have assumed that the distribution of the measurement error was known. However, this can be relaxed to allow the measurement error model to be specified up to a few parameters. For example, one may assume that the distribution of the measurement error is normal with mean zero and variance σ_ε^2, left unspecified. In this case, we would need replicate measurements of W to obtain a reasonable estimate of σ_ε^2. With such replicates, one can then easily modify the score equations allowing the parameter σ_ε^2 to be part of β and then proceed to derive locally efficient estimators for all the parameters.

In the reported simulation experiments, we did not include additional variables Z in our model that are not measured with error. This simplified the computations in that, for each dataset, only one system of linear equations (20) needs to be solved. Solution of the estimating equations was very fast with 1000 simulations programmed in Fortran taking about four minutes on a PC with a 2 GHz processor. If we include additional covariates Z in the model, then we must solve the linear system of equations for each data point. We conducted several simulations, not shown here, using additional variables Z taken as linear predictors in the logistic regression model. The results were all similar to those presented in § 4, but each simulation took about 50 seconds for each dataset with sample size 500. Finally, the methods outlined above were used for measurement error models, but we believe that they are more broadly applicable to general mixture models.

Acknowledgement

This research was supported by grants from the National Institute of Allergy and Infectious Disease and the National Cancer Institute.

Appendix

Technical details

Nontrivial elements orthogonal to the nuisance tangent space for the quadratic logistic regression model. By Theorem 1, if we can find a nonzero function $h(Y, W)$ such that $E\{h(Y, W)|X\} = 0$, then the orthogonal complement of the nuisance tangent space is nontrivial. We now consider the examples of quadratic logistic regression models with normal and exponential additive measurement error that were studied in § 5.

When the conditional distribution of W given $X = x$ is $N(x, \sigma^2)$, then standard calculations for normal densities yield that, for any β_1 and for $\beta_2 \geqslant -\sigma^2/2$,

$$E[\exp\{\gamma_1(\beta_1, \beta_2)W + \gamma_2(\beta_2)W^2\}|X] = c(\beta_1, \beta_2)\exp(\beta_1 X + \beta_2 X^2), \tag{A1}$$

where $\gamma_2(\beta_2) = \beta_2/(1 + 2\beta_2\sigma^2)$, $\gamma_1(\beta_1, \beta_2) = \beta_1/(1 + 2\beta_2\sigma^2)$ and $c(\beta_1, \beta_2)$ is a positive constant that depends on β_1 and β_2.

Since the response variable Y is binary, any function of Y and W can be written as $h(Y, W) = Yh_1(W) - h_2(W)$. As a result of the surrogacy assumption, i.e. that Y and W are conditionally independent given X, we have $E\{h(Y, W)|X\} = E(Y|X)E\{h_1(W)|X\} - E\{h_2(W)|X\}$. Consequently, when the distribution of Y given X follows the quadratic logistic regression model (1), the conditional expectation,

$$E\{h(Y, W)|X\} = \{1 + \exp - (\beta_0 + \beta_1 X + \beta_2 X^2)\}^{-1}E\{h_1(W)|X\} - E\{h_2(W)|X\},$$

is equal to zero if $E\{h_1(W)|X\} = E\{h_2(W)|X\}\{1 + \exp - (\beta_0 + \beta_1 X + \beta_2 X^2)\}$. If we use (A1) and if $\beta_2 \leqslant 0$, then a nontrivial solution exists by choosing $h_2(W) = 1$ and

$$h_1(W) = 1 + \exp(-\beta_0)\{c(-\beta_1, -\beta_2)\}^{-1}\exp\{\gamma_1(-\beta_1, -\beta_2)W + \gamma_2(-\beta_2)W^2\}.$$

On the other hand, if $\beta_2 \geqslant 0$, then a nontrivial solution exists by choosing

$$h_2(W) = \exp(\beta_0)\{c(\beta_1, \beta_2)\}^{-1}\exp\{\gamma_1(\beta_1, \beta_2)W + \gamma_2(\beta_2)W^2\}, \quad h_1(W) = h_2(W) + 1.$$

When the conditional distribution of W given $X = x$ follows an exponential distribution with density $\lambda\exp - \{\lambda(w - x)\}$, for $w > x$, then $E\{h(W)|X = x\} = \lambda\int_x^\infty h(w)\exp - \{\lambda(w - x)\}dw$. Consequently, to find the solution to the equation $E\{h(W)|X\} = q(X)$, or

$$\lambda\int_x^\infty h(w)\exp - \{\lambda(w - x)\}dw = q(x), \tag{A2}$$

where $q(x)$ is differentiable in x with derivative $q'(x)$, we differentiate both sides of (A2), and, after solving, we find that $h(w) = q(w) - q'(w)/\lambda$. If we let $q(x) = \{1 + \exp - (\beta_0 + \beta_1 x + \beta_2 x^2)\}^{-1}$, then $E\{Y - h(W)|X\} = 0$ when $h(W) = q(W) - q'(W)/\lambda$.

Proof of (12). As a result of Theorem 1, the conditional expectation of the efficient score $S^*_{\text{eff}}(O, \beta, \xi^*)$, given X and Z, is equal to zero when computed with respect to $p(y|x, z; \beta)$ and $p(w|x, z)$; that is,

$$\int S^*_{\text{eff}}(O, \beta, \xi^*) p(y|X, Z; \beta) p(w|X, Z) d\mu(y) d\mu(w) = 0 \tag{A3}$$

for all β. Taking the derivative of (A3) with respect to β, interchanging integration and differentiation and evaluating at β_0, we obtain

$$E\left\{\left.\frac{\partial S^*_{\text{eff}}(O, \beta_0, \xi^*)}{\partial \beta}\right| X, Z\right\} + E\{S^*_{\text{eff}}(O, \beta_0, \xi^*) S^{F\text{T}}_{\beta}(Y, X, Z, \beta_0, \xi^*)|X, Z\} = 0. \tag{A4}$$

Computing the unconditional expectation of (A4), under the true distribution, we obtain

$$E\left\{\frac{\partial S^*_{\text{eff}}(O, \beta_0, \xi^*)}{\partial \beta}\right\} + E\{S^*_{\text{eff}}(O, \beta_0, \xi^*) S^{\text{T}}_{\beta}(Y, X, Z, \beta_0, \xi^*)\} = 0. \tag{A5}$$

Evaluating the second term in (A5) using iterated conditional expectations, first conditioning on O, we obtain

$$E\left\{\frac{\partial S^*_{\text{eff}}(O, \beta_0, \xi^*)}{\partial \beta}\right\} + E\{S^*_{\text{eff}}(O, \beta_0, \xi^*) S^{\text{T}}_{\beta}(O, \beta_0, \xi^*)\} = 0.$$

Since $S^*_{\text{eff}}(O, \beta_0, \xi^*)$ is orthogonal to the nuisance tangent space Λ in $\mathscr{H}$, it follows that

$$E\left\{\frac{\partial S^*_{\text{eff}}(O, \beta_0, \xi^*)}{\partial \beta}\right\} + E\{S^*_{\text{eff}}(O, \beta_0, \xi^*) S^{\text{T}}_{\text{eff}}(O, \beta_0, \xi^*)\} = 0,$$

leading to (12).

Proof that (15) *equals zero*. Since the efficient score $S^*_{\text{eff}}(O, \beta_0, \xi)$, computed under the correct distributions for $p(y|x, z; \beta_0)$ and $p(w|x, z)$ but for the possibly incorrect distribution for $\eta^*_1(x|z, \xi)$, has mean zero, as long as the support for $\eta^*_1(x|z, \xi)$ is the same as that for the truth $\eta_{10}(x|z)$, this implies that

$$E\{S^*_{\text{eff}}(O, \beta_0, \xi)\} = 0 \tag{A6}$$

for all ξ. Taking the derivative of (A6) with respect to ξ and interchanging expectation and differentiation gives us the desired result.

References

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

CARROLL, R. J., RUPPERT, D. & STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. New York: Chapman and Hall.

FOUTZ, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Am. Statist. Assoc.* **72**, 147–8.

LINDSAY, B. G. & LESPERANCE, M. L. (1995). A review of semiparametric mixture models. *J. Statist. Plan. Infer.* **47**, 29–39.

LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods*. New York: Wiley.

NEWEY, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Economet.* **5**, 99–135.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: Wiley.

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 generalized linear measurement-error models. *Biometrika* **74**, 703–16.

VAN DER VAART, A. W. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24**, 862–78.

[*Received October* 2003. *Revised February* 2004]