Score tests in a generalized linear model with surrogate covariates

J.H. Sepanski

Department of Statistics and Management Science, School of Business Administration, University of Michigan, Ann Arbor, MI, USA

Abstract: We consider generalized linear models where a predictor is measured with error. The efficient score test for the effect of that predictor depends on the regression of the true predictor on its observed surrogate. Using validation data, we estimate the regression by nonparametric techniques. The resulting semiparametric score test is shown to be nearly asymptotically efficient.

Keywords: Generalized linear model; nonparametric regression; score test.

1. Background

Let $X = (R, Z^T)^T$ be the $(q + 1) \times 1$ vector of true covariate where R is a scalar and let $\beta = (\beta_1, \beta_2^T)^T$ where β_1 and β_2^T are the regression parameters associated with R and Z respectively. Assume that given X, the response follows a generalized linear model taking the form

$$\exp[\{y\eta - b(\eta)\}/\gamma + c(y, \gamma)], \tag{1.1}$$

for some function $b(\cdot)$ and $c(\cdot)$. The natural parameter η satisfies that

$$\frac{\partial}{\partial \eta} b(\eta) = u(\beta_0 + x^T\beta) \quad \text{and} \quad \frac{\partial^2}{\partial^2 \eta} b(\eta) = g(\beta_0 + x^T\beta),$$

where $E(Y \mid X = x) = u(\beta_0 + x^T\beta)$ and $\text{Var}(Y \mid X = x) = \gamma g(\beta_0 + x^T\beta)$. Frequently some of the components of the true predictor X can not be observed exactly. In particular, we consider the case when one of the predictors, say R, is measured with error and a d-variate surrogate W of R is observed. By a surrogate we mean a variable W such that given X, Y is independent of W. We wish to test the hypothesis H_0: $\beta_1 = 0$ based on a sample from (Y, W, Z) instead of (Y, X).

When there is only one predictor, i.e., $X = R$, the score test statistic based on a sample of size n_p is defined by

$$T_1 = \frac{1}{n_p^{1/2}} \sum_{j=1}^{n_p} m(W_j)(Y_j - \bar{Y})/\{S_m S_Y\}, \tag{1.2}$$

where $m(w) = E(R \mid W = w)$ and S_m and S_Y are the usual sample variances of the $m(W)$'s and Y's

Correspondence to: J.H. Sepanski, Department of Statistics and Management Science, University of Michigan, Ann Arbor, MI 48109-1234, USA.

respectively. Under the null hypothesis the statistic T_1 has an asymptotic standard normal distribution. In model (1.1), for local alternatives $\beta_1 = \Delta/n_p^{1/2}$, Tosteson and Tsiatis (1988) show that T_1 is asymptotically normally distributed with variance 1 and mean

$$\Lambda(\Delta) = \Delta\dot{u}(\beta_0)\left[\mathrm{Var}\{m(W)\}/\{\gamma g(\beta_0)\}\right]^{1/2}, \tag{1.3}$$

where $\dot{u}(v) = \partial u(v)/\partial v$, $\dot{u}(\beta_0) = \dot{u}(\beta_0 + x^{\mathrm{T}}\beta)|_{\beta=0}$ and $g(\beta_0) = g(\beta_0 + x^{\mathrm{T}}\beta)|_{\beta=0}$.

Suppose that W is scalar. The 'naive' score test is obtained by ignoring measurement error, replacing R by W, and computing (1.2). The preceding discussion shows that the naive test equals the score test whenever the regression of R on W is known to be linear.

In cases that $m(w)$ is neither linear nor known, Caroll and Stefanski (1990) investigated two test statistics appropriate when $W = R + \varepsilon$, where ε is the measurement error. The first one is a Wald test based on estimators corrected for measurement error. It is shown that this test has the same local power as the naive test which substitutes W for $m(W)$. The second one employs an estimate $\hat{q}(w)$ of $q(w) = w + \sigma^2 f'_W(w)/f_W(w)$ as an approximate estimate of $m(w)$, where σ^2 is the variance of ε, f_W is the density function of W, and $f'(w) = \partial f(w)/\partial w$. The test is then based on (1.2) with $m(w)$ replaced by $\hat{q}(w)$. They show that this test is approximately efficient when the measurement error is small.

When there is an auxiliary covariate Z besides R, let $m(w, z) = E(R \mid W = w, Z = z)$ and $\hat{\beta}_0$ and $\hat{\beta}_2$ be the constrained maximum likelihood estimates satisfying

$$\sum_{j=1}^{n_p} \left(1, Z_j^{\mathrm{T}}\right)^{\mathrm{T}} \dot{u}\left(\hat{\beta}_0 + Z_j^{\mathrm{T}}\hat{\beta}_2\right)\left\{Y_j - u\left(\hat{\beta}_0 + Z_j^{\mathrm{T}}\hat{\beta}_2\right)\right\}/g\left(\hat{\beta}_0 + Z_j^{\mathrm{T}}\hat{\beta}_2\right) = 0,$$

where $\dot{u}(\beta_0 + z^{\mathrm{T}}\beta_2) = \dot{u}(\beta_0 + z^{\mathrm{T}}\beta)|_{\beta_1=0}$ and similarly for u and g. Tosteson and Tsiatis (1988) show that under the null hypothesis $\beta_1 = 0$, the statistic

$$L_2 = \frac{1}{n_p^{1/2}} \sum_{j=1}^{n_p} m(W_j, Z_j)\dot{u}\left(\hat{\beta}_0 + Z_j^{\mathrm{T}}\hat{\beta}_2\right)\left\{Y_j - u\left(\hat{\beta}_0 + Z_j^{\mathrm{T}}\hat{\beta}_2\right)\right\}/g\left(\hat{\beta}_0 + Z_j^{\mathrm{T}}\hat{\beta}_2\right) \tag{1.4}$$

is asymptotically normal with mean 0 and covariance $D = \gamma\Sigma$, where

$$\Sigma = E\left\{\frac{\dot{u}^2(\beta_0 + Z^{\mathrm{T}}\beta_2)}{g(\beta_0 + Z^{\mathrm{T}}\beta_2)}m^2(W, Z)\right\} - A^{\mathrm{T}}\left[E\left\{\frac{\dot{u}^2(\beta_0 + Z^{\mathrm{T}}\beta_2)}{g(\beta_0 + Z^{\mathrm{T}}\beta_2)}(1, Z^{\mathrm{T}})^{\mathrm{T}}(1, Z^{\mathrm{T}})\right\}\right]^{-1}A, \tag{1.5}$$

$$A = E\left\{\frac{\dot{u}^2(\beta_0 + Z^{\mathrm{T}}\beta_2)}{g(\beta_0 + Z^{\mathrm{T}}\beta_2)}(1, Z^{\mathrm{T}})^{\mathrm{T}}m(W, Z)\right\}.$$

The efficient score test statistic is defined by $T_2 = L_2/\hat{D}^{1/2}$, where $\hat{D}$ is a consistent estimator of D, see (2.3). Under the alternative $\beta_1 = \Delta/n_p^{1/2}$, the statistic T_2 is normally distributed with variance 1 and mean $\Delta\Sigma/D^{1/2}$. They show that the score test statistics T_1 and T_2 yield tests which are asymptotically equivalent to the efficient score test. When the conditional mean m of R given (W, Z) is a linear function of W, the naive test statistics will be asymptotically locally optimal. However, the more usual case is that the regression of X on (W, Z) depends on Z also, in which case the naive test will not be locally optimal.

The main purpose of this paper is to use validation data to construct locally optimal tests for the general function $m(w, z)$.

We will assume that in addition to the primary data set of size n_p which contains observations $\{(Y_j, W_j)_{j=1}^{n_p}\}$ or $\{(Y_j, W_j, Z_j)_{j=1}^{n_p}\}$, another validation data set of size n_v which contains $\{(R_i, W_i)_{i=n_p+1}^{n_p+n_v}\}$ or

$\{(R_i, W_i, Z_i)_{i=n_p+1}^{n_p+n_v}\}$ is available. Examples can be found in Rosner et al. (1989 and 1990), Pepe and Flemming (1991) and Carroll and Wand (1991), etc. Since $m(w)$ or $m(w, z)$ is usually unknown, the estimation of $m(w)$ or $m(w, z)$ is then important in constructing an efficient score test.

Based on the validation data, we can use a parametric regression or use a nonparametric regression to estimate $m(w)$ or $m(w, z)$. We will discuss the score tests based on two estimates of m which will be described in Section 2. In Section 2.1, we will consider the parametric problem, where the regression function is linear, and we will construct efficient tests. In Section 2.2, we consider general regression functions estimated by nonparametric regression. Here we will construct a version of the score test which is almost fully locally efficient, the loss of efficiency being due to edge effects. In Section 3, we consider some simple extensions.

2. Score tests

To implement the following tests, we do need a validation data set which is not required by the tests of Carroll and Stefanski. We assume that W is a d-variate surrogate for the true scalar predictor R. It is also assumed that the conditional mean of R given W or (W, Z) is the same for the validation and primary study populations. In this section, we state the asymptotic results under the alternatives. The proofs of Theorems 1 and 3 are given in Section 4. Theorems 2 and 4 can be obtained by similar arguments.

2.1. Based on linear regression estimate

Suppose that the regression of R on W is linear, so that $R = \alpha_0 + W^T \alpha_1 + \varepsilon$, where W and ε are independent and ε has mean 0 and constant variance σ^2. With n_v independent observations on (R, W) in the validation data, the model becomes

$$\mathscr{R}_{v,n_v \times 1} = \mathscr{W}_{v,n_v \times (d+1)} \alpha_{(d+1) \times 1} + \varepsilon_{v,n_v \times 1},$$

where $\alpha = (\alpha_0, \alpha_1^T)^T$. Then an estimator $\hat{m}_{1l}(w)$ of $m(w)$ is given by $(1, w)\hat{\alpha}$, where $\hat{\alpha} = (\mathscr{W}_v^T \mathscr{W}_v)^{-1} \mathscr{W}_v^T \mathscr{R}_v$. Define

$$\hat{T}_{1l} = \frac{1}{n_p^{1/2}} \sum_{j=1}^{n_p} \hat{m}_{1l}(W_j)(Y_j - \bar{Y}) / \{\hat{S}_{1l} S_Y\}, \qquad (2.1)$$

where $\hat{S}_{1l}$ is the sample variance of $\hat{m}_{1l}(w)$.

Theorem 1. *Under the local alternatives* H_a: $\beta_1 = \Delta/n_p^{1/2}$, *the score test based on* $\hat{T}_{1l}$ *is asymptotically normally distributed with variance* 1 *and mean*

$$\Lambda_1(\Delta) = \Delta \dot{u}(\beta_0) E[\alpha_1^T \text{Var}(W) \alpha_1 / \{\gamma g(\beta_0)\}]^{1/2}.$$

Next, we consider in addition to R which is observed with error, that we have an auxiliary q-variate covariate Z which is observed without error. We have a validation data set containing (R, W, Z) and a primary data set containing (Y, W, Z). Again, it is assumed that there is a linear relationship between R

and (W, Z). An efficient score test statistic is defined by $\hat{T}_{2l} = \hat{L}_{2l}/\hat{D}^{1/2}$, where

$$\hat{L}_{2l} = \frac{1}{n_p^{1/2}} \sum_{j=1}^{n_p} \hat{m}_{2l}(W_j, Z_j)\dot{u}\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)\left\{Y_j - u\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)\right\}/g\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right), \tag{2.2}$$

$$\hat{D} = \hat{\gamma}\hat{\Sigma}, \tag{2.3}$$

$$\hat{\gamma} = \frac{1}{n_p} \sum_{j=1}^{n_p} \frac{\left\{Y_j - u\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)\right\}^2}{g\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)},$$

$$\hat{\Sigma} = \frac{1}{n_p} \sum_{j=1}^{n_p} \frac{\dot{u}^2\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)}{g\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)}\hat{m}_{2l}^2(W_j, Z_j) - \hat{A}^T\left[\frac{1}{n_p} \sum_{j=1}^{n_p} \frac{\dot{u}^2\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)}{g\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)}\left(1, Z_j^T\right)^T\left(1, Z_j^T\right)\right]^{-1}\hat{A},$$

$$\hat{A} = \frac{1}{n_p} \sum_{j=1}^{n_p} \frac{\left\{Y_j - u\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)\right\}^2}{g\left(\hat{\beta}_0 + Z_j^T\hat{\beta}_2\right)}\left(1, Z_j^T\right)^T\hat{m}_{2l}(W_j, Z_j),$$

and $\hat{m}_{2l}(w, z)$ is obtained by linear regression of R on (W, Z) using the validation data. Note that $m(W, Z) = \alpha_0 + W^T\alpha_1 + Z^T\alpha_2$ in this case.

Theorem 2. *Under local alternatives* H_a: $\beta_1 = \Delta/n_p^{1/2}$, *the score test based on* $\hat{T}_{2l}$ *is asymptotically normally distributed with variance 1 and mean* $\Delta\Sigma/D^{1/2}$. $\square$

2.2. Based on nonparametric regression estimate

We first consider the case of a single covariate in the model, and then proceed to the general case.

First suppose that there is only a single covariate in the model, so that $X = R$. We are going to use nonparametric (kernel) regression technique to estimate $m(w)$ using the validation data, and then plug in the estimated regression function into (1.2). In order to avoid problems with edge effects, we will truncate the data to a fixed compact set Ω. The proposed score test statistic defined in (2.4) below is shown to be nearly optimal whether $m(w)$ is a linear function of w or not and whether the measurement error is small or not.

Let Ω be a set and $I(w \in \Omega)$ be an indicator function. Let K be a symmetric second order density function and h be a window or bandwidth. Define a nonparametric regression estimate $\hat{m}_{1n}(w)$ of $m(w)$ applicable on the set Ω:

$$\hat{k}(w) = \frac{1}{n_v h^d} \sum_{i=1}^{n_v} R_i K\left(\frac{W_i - w}{h}\right), \qquad \hat{f}_W(w) = \frac{1}{n_v h^d} \sum_{i=1}^{n_v} K\left(\frac{W_i - w}{h}\right),$$

$$\hat{m}_{1n}(w) = \hat{k}(w)/\hat{f}_W(w).$$

Based on the estimate, define a test statistic

$$\hat{T}_{1n} = \frac{1}{n_p^{1/2}} \sum_{j-n_v+1}^{n} \hat{m}_{1n}(W_j)\left(Y_j - \bar{Y}\right)I(W_j \in \Omega)/\left\{\hat{S}_{1n}S_Y\right\}, \tag{2.4}$$

where $\hat{S}_{1n}$ is the sample variance of $\hat{m}_{1n}(W)I(W \in \Omega)$. Note that in (2.4), we have restricted W to the set Ω, mainly for technical reasons. However, in practice, the validation study is smaller than the primary study. If we tried to compute $\hat{m}$ on the range of the validation study, we would likely be extrapolating the

kernel estimate outside the (usually) smaller range of the validation study. Thus, restriction to the set Ω has practical importance as well.

In the results, we consider two cases. In the first, we use the compactness of Ω explicitly. In the second, we do not necessarily require restriction to a compact set, although we believe that it is desirable in practice to make the truncation.

Let $f_*(w) = (1 - \alpha_w)\hat{f}_W(w) + \alpha_w f_W(w)$, where $\alpha_w \in (0, 1)$. Let $w = (w_1, w_2, \ldots, w_d)^{\mathrm{T}}$ and $z = (z_1, z_2, \ldots, z_d)^{\mathrm{T}}$. Define

$$M_1(w) = \{\partial m(w)/\partial w_i\}_{d \times 1}, \qquad M_2(w) = \{\partial^2 m(w)/\partial w_i\, \partial w_j\}_{d \times d},$$

$$F_1(w) = \{\partial f_W(w)/\partial w_i\}_{d \times 1}, \quad \text{and} \quad F_2(w) = \{\partial^2 f_W(w)/\partial w_i\, \partial w_j\}_{d \times d}.$$

Define

$$g(w, z) = 2z^{\mathrm{T}}M_1(w)F_1^{\mathrm{T}}(w)z + m(w)z^{\mathrm{T}}F_2(w + \zeta zh)z + f_W(w)z^{\mathrm{T}}M_2(w + \xi zh)z$$

$$+ \tfrac{1}{2}h^2 z^{\mathrm{T}}M_2(w + \xi zh)zz^{\mathrm{T}}F_2(w + \zeta zh)z, \tag{2.5}$$

where $\xi, \zeta \in (0, 1)$. Let $H(w, z) = \sup_{|\zeta| < 1} z^{\mathrm{T}}F_2(w + \zeta zh)z$ and let $G(w, z) = \sup_{|\xi| < 1} g(w, z)$.

Theorem 3. *Assume that the density f_W of W and the conditional mean m of R given W have second continuous derivatives and $\inf_{w \in \Omega} f_*(w) > 0$. Then, under local alternatives $\beta_1 = \Delta/n_p^{1/2}$, the test statistic $\hat{T}_{1n}$ is asymptotically normal with variance 1 and mean*

$$\Delta \dot{u}(\beta_0)\left[\mathrm{Var}\{m(W)I(W \in \Omega)\}/\{\gamma g(\beta_0)\}\right]^{1/2}$$

as $n_p, n_v \to \infty$ and $h \to 0$, if one of the following holds:
Case I (*Truncation*). *Let Ω be a compact set such that*
- $\sup_{w \in \Omega} E(X^2 \mid W = w) < \infty$;
- $n_v \to \infty$, $h \to 0$, and $n_v h^d \to \infty$.

Case II. *Note the set Ω need not be compact. Assume that*
- *m has finite second moment;*

- $$E\left[\frac{1}{n_v h^d}\frac{1}{f_W^p(W)}I(W \in \Omega)\int x^p K(z)f(x, W + zh)\,\mathrm{d}x\,\mathrm{d}z\right] \to 0, \quad p = 0, 1, 2,$$

$$E\left[\left\{h^2\int K(z)H(W, z)\,\mathrm{d}z\right\}^2 I(W \in \Omega)\right] \to 0,$$

$$E\left[\left\{\frac{h^2}{f_W(W)}\int K(z)G(W, z)\,\mathrm{d}z\right\}^2 I(W \in \dot{\Omega})\right] \to 0.$$

We next consider the case when there is an auxiliary covariate Z which is observed exactly. If the distribution of R given W is independent of Z, the score test statistic is obtained by replacing $\hat{m}_{2l}$ with $\hat{m}_{1n}$ in (2.2). If the distribution of R given W is not independent of Z, we estimate $m(w, z)$ by nonparametric multiple regression of R on $V = (W, Z)$. Let $v = (w, z)$ and define

$$\hat{m}_{2n}(v) = \sum_{i=1}^{n_v} R_i K\left(\frac{V_i - v}{h}\right)\bigg/ \sum_{i=1}^{n_v} K\left(\frac{V_i - v}{h}\right).$$

Let $\mathscr{C}$ be the set in which $\hat{m}_{2n}$ is well defined. Define

$$\Sigma_{\mathscr{C}} = E\left\{ \frac{\dot{u}^2(\beta_0 + Z^{\mathsf{T}}\beta_2)}{g(\beta_0 + Z^{\mathsf{T}}\beta_2)} m^2(W, Z) I(V \in \mathscr{C}) \right\}$$

$$- A_{\mathscr{C}}^{\mathsf{T}} \left[E\left\{ \frac{\dot{u}^2(\beta_0 + Z^{\mathsf{T}}\beta_2)}{g(\beta_0 + Z^{\mathsf{T}}\beta_2)} (1, Z^{\mathsf{T}})^{\mathsf{T}} (1, Z^{\mathsf{T}}) I(V \in \mathscr{C}) \right\} \right]^{-1} A_{\mathscr{C}}, \qquad (2.6)$$

$$A_{\mathscr{C}} = E\left\{ \frac{\dot{u}^2(\beta_0 + Z^{\mathsf{T}}\beta_2)}{g(\beta_0 + Z^{\mathsf{T}}\beta_2)} (1, Z^{\mathsf{T}})^{\mathsf{T}} m(W, Z) I(V \in \mathscr{C}) \right\}.$$

An efficient score test statistic is defined by $\hat{T}_{2n} = \hat{L}_{2n}/\hat{D}_{\mathscr{C}}^{1/2}$, where $\hat{D}_{\mathscr{C}}$ is defined as $\hat{D}$ in (2.3) with the sums restricted to the set $\mathscr{C}$ and $\hat{m}_{2l}$ replaced by $\hat{m}_{2n}$ and

$$\hat{L}_2 = \frac{1}{n_p^{1/2}} \sum_{j=1}^{n_p} \hat{m}_{2n}(V_j) I(V_j \in \mathscr{C}) \dot{u}(\hat{\beta}_0 + Z_j^{\mathsf{T}}\hat{\beta}_2) \left\{ Y_j - u(\hat{\beta}_0 + Z_j^{\mathsf{T}}\hat{\beta}_2) \right\} / g(\hat{\beta}_0 + Z_j^{\mathsf{T}}\hat{\beta}_2).$$

We here use the definitions in (2.4) with d replaced by $d + q$.

As in Theorem 3, we consider two cases. The first uses the compactness of the set $\mathscr{C}$ explicitly, while the second allows possible relaxation of that condition.

Theorem 4. *Assume that the density f_V of V and the conditional mean m of R given V have second continuous derivatives and $\inf_{v \in \mathscr{C}} f_*(v) > 0$. Then, under local alternatives $\beta_1 = \Delta/n_p^{1/2}$, the test statistic $\hat{T}_{2n}$ is asymptotically normal with variance 1 and mean $\Delta \Sigma_{\mathscr{C}}/D_{\mathscr{C}}^{1/2}$ as $n_p, n_v \to \infty$ and $h \to 0$, if one of the following holds:*

Case I (*Truncation*). *Let $\mathscr{C}$ be a compact set such that*
- $\sup_{v \in \mathscr{C}} E(X^2 \mid V = v) < \infty$;
- $n_v \to \infty$, $h \to 0$, *and* $n_v h^{(d+q)} \to \infty$.

Case II. *The set $\mathscr{C}$ need not be compact in $\mathbb{R}^{(d+q)}$ and assume that*
- *m has finite second moment;*

- $$E\left[\frac{1}{n_v h^{(d+q)}} \frac{1}{f_V^p(V)} I(V \in C) \int x^p K(z) f(x, V + zh) \, \mathrm{d}x \, \mathrm{d}z \right] \to 0, \quad p = 0, 1, 2,$$

$$E\left[\left\{ h^2 \int K(z) H(V, z) \, \mathrm{d}z \right\}^2 I(V \in C) \right] \to 0,$$

$$E\left[\left\{ \frac{h^2}{f_V(V)} \int K(z) G(V, z) \, \mathrm{d}z \right\}^2 I(V \in C) \right] \to 0. \qquad \square$$

3. Discussion

We have shown that validation data enable the construction of efficient score tests for the effect of a scalar covariate measured with error, using both parametric and nonparametric techniques.

We have assumed that validation data are available in which R is measured. In practice, it will more often be the case that instead of observing R, we will observe $R_* = R + \xi$, where ξ is independent of W and Z, see for example Rosner et al. (1990). Our results apply without change to this case by replacing R by R_*. We consider the test of association for the case when a scalar covariate R is measured with error.

When a validation data set is available, we can estimate the mean function m of the true R given (W, Z) based on the validation data. It is shown that the estimated score tests yield asymptotically local optimal tests.

4. Proofs of Theorems 1 and 3

Proof of Theorem 1. From (1.3), the test statistics $T_l = n_p^{-1/2} \sum_{j=n_v+1}^{n} (\alpha_0 + W_j^{\mathrm{T}} \alpha_1)(Y_j - \bar{Y}) / \{S_m S_Y\}$ is asymptotically normally distributed with variance 1 and mean $\Lambda_1(\Delta)$ under the local alternatives $\beta_1 = \Delta / n_p^{1/2}$. To show Theorem 1 it suffices to show that under the null hypothesis $\beta_1 = 0$, $\hat{T}_{1l} - T_l \to^{\mathrm{P}} 0$. We will show that

$$D_n = \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^{n} \left\{ \hat{m}_{1l}(W_j) - \left(\alpha_0 + W_j^{\mathrm{T}} \alpha_1 \right) \right\} \left(Y_j - \bar{Y} \right) \xrightarrow{\mathrm{P}} 0. \tag{4.1}$$

Similarly, we can show that $S_m - \hat{S}_{1l} \to^{\mathrm{P}} 0$.

Let $\lambda = n_p / n_v$ and $\mathscr{W}_p$ be the $n_p \times (d+1)$ matrix containing the observations from W in the primary data and with 1 as the first column. Now

$$\mathrm{Var}\left[\frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^{n} \left\{ \hat{m}_{1l}(W_j) - \left(\alpha_0 + W_j^{\mathrm{T}} \alpha_1 \right) \right\} (Y_j - u_Y) \right]$$

$$= \mathrm{Var}(Y) E\left[\frac{1}{n_p} \sum_{j=n_v+1}^{n} \left\{ \hat{m}_{1l}(W_j) - \left(\alpha_0 + W_j^{\mathrm{T}} \alpha_1 \right) \right\}^2 \right]$$

$$= \frac{\mathrm{Var}(Y)}{n_p} E\left[\left\{ \mathscr{W}_p (\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1} \mathscr{W}_v^{\mathrm{T}} \mathscr{R}_v - \mathscr{W}_p \alpha \right\}^{\mathrm{T}} \left\{ \mathscr{W}_p (\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1} \mathscr{W}_v^{\mathrm{T}} \mathscr{R}_v - \mathscr{W}_p \alpha \right\} \right]. \tag{4.2}$$

By substituting $\mathscr{R}_v$ by $\mathscr{W}_v \alpha + \varepsilon_v$ and $E(\mathscr{W}_p^{\mathrm{T}} \mathscr{W}_p) = n_p E(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v) / n_v$, the expectation in (4.2) is given by

$$E\left[\lambda \, \mathrm{trace}\left\{ \varepsilon_v^{\mathrm{T}} \mathscr{W}_v (\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1} E(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1} \mathscr{W}_v^{\mathrm{T}} \varepsilon_v \right\} \right]$$

$$= E\left[\lambda \sigma^2 \, \mathrm{trace}\left\{ (\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1} E(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v) \right\} \right].$$

Note the diagonal of the matrix $E\{(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1}(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)\} - E\{(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)^{-1}\} E^{\mathrm{T}}\{(\mathscr{W}_v^{\mathrm{T}} \mathscr{W}_v)\} > 0$. Therefore, $(4.2) < \lambda(d+1)\sigma^2(\mathrm{Var}(Y)) / n_p$ which converges to 0 as $n_p \to \infty$. Hence, $(4.2) = o(1)$ as $n_p \to \infty$. Similarly, we can show that

$$n_p^{-1/2} \sum_{j=n_v+1}^{n} \left\{ \hat{m}_{1l}(W_j) - \left(\alpha_0 + W_j^{\mathrm{T}} \alpha_1 \right) \right\} (\bar{Y} - u_Y) \xrightarrow{\mathrm{P}} 0.$$

Thus $D_n \to^{\mathrm{P}} 0$ as $n_p \to \infty$. $\square$

Proof of Theorem 3. Define

$$A_{1n} = n_p^{-1/2} \sum_{j=n_v+1}^{n} \left\{ \hat{m}_{1n}(W_j) - m(W_j) \right\} (Y_j - \mu_Y) I(W_j \in \Omega)$$

and

$$
A_{2n} = n_p^{-1/2} \sum_{j=n_v+1}^{n} \{\hat{m}_{1n}(W_j) - m(W_j)\}(\bar{Y} - \mu_Y) I(W_j \in \Omega).
$$

Then,

$$
n_p^{-1/2} \sum_{j=n_v+1}^{n} \{\hat{m}_{1n}(W_j) - m(W_j)\}(Y_j - \bar{Y}) I(W_j \in \Omega) = A_{1n} - A_{2n}.
$$

To show that $A_{1n} = o_P(1)$ and $A_{2n} = o_P(1)$, note that $E(A_{1n}) = E(A_{2n}) = 0$ and

$$
\text{Var}(A_{1n}) = \text{Var}(Y) E\left[\{\hat{m}_{1n}(W) - m(W)\}^2 I(W \in \Omega)\right],
$$

$$
\text{Var}(A_{2n}) = \frac{1}{n_p} \text{Var}(Y) E\left[\{\hat{m}_{1n}(W) - m(W)\}^2 I(W \in \Omega)\right].
$$

Since $\hat{f}(w) \to^P f(w)$ for all $w \in \Omega$, by a first order Taylor series, we have that

$$
\hat{m}_{1n}(W) = \hat{k}(W)\left[\frac{1}{f_W(W)} - \frac{1}{f_*^2(W)}\{\hat{f}_W(W) - f_W(W)\}\right],
$$

where $f_*(w) = \alpha_w \hat{f}_W(w) + (1 - \alpha_w)f_W(w)$, $\alpha_w \in (0, 1)$. Therefore,

$$
E\left[\{\hat{m}_{1n}(W) - m(W)\}^2 I(W \in \Omega)\right]
$$

$$
= E\left[\left\{\frac{\hat{k}(W)}{f_W(W)} - m(W)\right\}^2 I(W \in \Omega) + \frac{1}{f_*^4(W)}\{\hat{f}_W(W) - f_W(W)\}^2 I(W \in \Omega)\right.
$$

$$
\left. - \frac{2}{f_*^2(W)}\{\hat{f}_W(W) - f_W(W)\}\left\{\frac{\hat{k}(W)}{f_W(W)} - m(W)\right\} I(W \in \Omega)\right]
$$

$$
= B_{1n} + B_{2n} - B_{3n}.
$$

Note that the following expectations are taken over the set Ω. By change of variables and a second order Taylor series, we have $E\{\hat{k}(W)\}$ given by

$$
E\left[\int (\{m(W) + hM_1^T(W)z + \tfrac{1}{2}h^2 z^T M_2(W + \xi zh)z\}K(z)\right.
$$

$$
\left. \cdot \{f_W(W) + hF_1^T(W)z + \tfrac{1}{2}h^2 z^T F_2(W + \zeta zh)z\})\, dz\right],
$$

$$
= E\left\{m(W)f_W(w) + \tfrac{1}{2}h^2 \int g(W, z)K(z)\, dz\right\},
$$

where M_1, M_2, F_1 and F_2 are defined in (2.5). Also not that $E\{\hat{k}^2(W)\}$ is given by

$$
\frac{1}{n_v h^d}\frac{1}{f^2(W)}\int x^2 K(z)f(x, W + zh)\, dx\, dz
$$

$$
- \frac{n_v(n_v - 1)}{n_v}\left\{m(W) + \frac{h^2}{2f_W(W)}\int K(z)g(W, z)\, dz\right\}^2.
$$

Therefore,

$$E(B_{1n}) = E\left[\frac{1}{n_v h^d} \frac{1}{f^2(W)} \int x^2 K(z) f(x, W + zh) \, dx \, dz\right.$$

$$- \frac{1}{n_v} \left\{m(W) + \frac{h^2}{2 f_W(W)} \int K(z) g(W, z) \, dz\right\}^2$$

$$\left. + \left\{\frac{h^2}{2 f_W(W)} \int K(z) g(W, z) \, dz\right\}^2\right]. \tag{4.3}$$

By assumptions, there exists a constant c such that

$$E(B_{2n}) \leqslant cE\left[\left\{\hat{f}_W(W) - f_W(W)\right\}^2 I(W \in \Omega)\right]$$

$$= cE\left[\frac{1}{n_v h^d} \int K^2(z) f_W(w + zh) \, dz + \left\{\tfrac{1}{2} h^2 \int K(z) z^T F_2(W + \xi zh) z \, dz\right\}^2\right.$$

$$\left. - \frac{1}{n_v} \left\{f_W(W) + \tfrac{1}{2} h^2 \int K(z) z^T F_2(W + zh) z \, dz\right\}^2\right], \tag{4.4}$$

where $\xi \in (0, 1)$. Similarly,

$$E(B_{3n}) \leqslant 2cE\left[\frac{1}{n_v h^d} \frac{1}{f_W(W)} \int x K^2(z) f(x, W + zh) \, dx \, dz\right.$$

$$+ \left\{\frac{h^2}{2 f_W(W)} \int K(z) g(W, z) \, dz\right\} \left\{\tfrac{1}{2} h^2 \int K(z) z^T F_2(W + \xi zh) z \, dz\right\}$$

$$- \frac{1}{n_v} \left\{m(W) + \frac{h^2}{2 f_W(W)} \int K(z) g(W, z) \, dz\right\}$$

$$\left. \times \left\{f_W(W) + \tfrac{1}{2} h^2 \int K(z) z^T F_2(W + \xi zh) z \, dz\right\}\right]. \tag{4.5}$$

Case I. Note that by assumption f_W and m have bounded second derivatives on the set Ω. By bounded convergence, we can easily show that (4.3), (4.4) and (4.5) converge to 0 as $n_v \to \infty$, $h \to 0$, and $n_p, n_v h^d \to \infty$.

Case II. By the assumptions, we see that (4.3), (4.4) and (4.5) converge to 0 as $n_p, n_v \to \infty$, $h \to 0$. $\square$

Acknowledgement

References

Armstrong, B. (1985), Measurement error in the generalized linear model, *Comm. Statist.—Simulation Comput.* **14**, 529–544.

Carroll, R.J. and L.A. Stefanski (1990), Score test in generalized linear measurement error models, *J. Roy. Statist. Soc. Ser. B* **52**, 345–459.

Carroll, R.J. and M.P. Wand (1991), Semiparametric estimation in logistic measurement error models, to appear in: *J. Roy. Statist. Soc. Ser. B.*

Clark, R. (1982), Logistic regression with measurement error in predictors, Unpublished Ph.D. dissertation, Department of Biostatistics, University of North Carolina (Chapel Hill, NC).

Eubank, R.L. (1988), *Spline Smoothing and Nonparametric Regression* (Dekker, New York).

Fuller, W.A. (1987), *Measurement Error Models* (Wiley, New York).

Pepe, M.S. and T.R. Fleming (1991), A general nonparametric method for dealing with errors in missing or surrogate covariate data, *J. Amer. Statist. Assoc.* **86**, 108–113.

Rosner, B., D. Spiegelman and W.C. Willett (1990), Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error, *Amer. J. Epidem.* **132**, 734–745.

Rosner, B., W.C. Willett and D. Spiegelman (1989), Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error, *Statist. Medicine* **8**, 1051–1070.

Tosteson, T.D. and A.A. Tsiatis (1988), The asymptotic relative efficiency of the score test in a generalized linear model with surrogate covariates, *Biometrika* **75**, 507–514.