Orthogonal Locally Ancillary Estimating Functions for Matched-Pair Studies and Errors-in-Covariates

Molin Wang

*Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, USA*

and John J. Hanfelt

*Rollins School of Public Health, Emory University, Atlanta, USA*

**Summary.** We propose an estimating function method for two related applications, matched-pair studies and studies with errors-in-covariates under a functional model, where a mismeasured unknown scalar covariate is treated as a fixed nuisance parameter. Our method addresses the severe inferential problem posed by an abundance of nuisance parameters in these two applications. We propose orthogonal locally ancillary estimating functions for the above two applications that depend on merely the mean model and partial modeling of the variances of the observations (and observed mismeasured covariate, if applicable), and achieve first-order bias correction of inferences under a “small dispersion and large sample size” asymptotic. Simulation results confirm that the proposed estimator is largely improved over that using a regular profile estimating function. We apply the proposed approach to a coronary heart disease study with a mismeasured blood pressure covariate.

**keywords:** Ancillarity; Estimating function; Matched pair study; Measurement error; Nuisance parameter; Orthogonality.

**Address for correspondence:** Molin Wang, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA.

E-mail: mwang@jimmy.harvard.edu
1. Introduction

This paper focuses on two applications that are closely related mathematically, matched-pair studies and studies with errors-in-covariates, and that pose a severe inferential challenge, as a result of an abundance of nuisance parameters and lack of full specification of the likelihood.

Example 1. Consider the matched-pairs example of Rathouz and Liang (1999), which involves regression with a reciprocal link function. The outcomes of the \(i\)th pair are \(Y_{i1}\) and \(Y_{i2}\), \(i = 1, \ldots, n\). Assume that the mean of \(Y_{ij}\) is given by \(\mu_{ij} = (\lambda_i + \theta_1 z_{ij})^{-1}\), with variance \(\phi \mu_{ij}^2\), where \(\phi\) is an unknown dispersion parameter common to all observations, the pair-specific intercepts \(\lambda_i\)s are nuisance parameters, and \(z_{ij}\) is a scalar covariate. Of interest is the regression coefficient \(\theta_1\).

We now generalize the problem in Example 1 by the following application.

Application 1. Matched-pairs study. Let \(\{Y_{i1}, Y_{i2}, i = 1, \ldots, n\}\) be independent matched pairs such that \(E(Y_{ij}) = \mu_{ij} = \mu_j(\theta, \lambda_i; z_{ij})\), where \(z_{ij}\) is a vector of known covariates and \(\mu_j(\cdot)\) is a known, real-valued function. Of scientific interest is \(\theta\), a parameter vector of fixed length \(p\) that relates \(z_{ij}\) to the expected response. The stratum-specific intercepts, \(\lambda_i, i = 1, \ldots, n\), are scalar nuisance parameters. Assume that \(\text{var}(Y_{ij}) = \phi_j v_j(\theta, \lambda_i; z_{ij}) < \infty\), where unless otherwise noted the dispersion parameters, \(\phi_1\) and \(\phi_2\), and the positive-valued variance functions, \(v_1(\cdot)\) and \(v_2(\cdot)\), are unknown. If the dispersion parameters \(\phi_j\) and variance functions \(v_j(\cdot)\) were known, then the \(\theta\)- and \(\lambda\)-quasi-scores would be available and given by, respectively,

\[
s(\theta; \lambda_1, \ldots, \lambda_n) = \sum_{i=1}^n s_i(\theta, \lambda_i) = \sum_{i=1}^n \sum_{j=1}^2 \left( \frac{\partial \mu_{ij}}{\partial \theta} \right)^T \frac{Y_{ij} - \mu_{ij}}{\phi_j v_j(\theta, \lambda_i; z_{ij})},
\]

(1)
\[
h_i(\lambda_i; \theta) = \sum_{j=1}^{2} \frac{\partial \mu_{ij}}{\partial \lambda_i} \frac{Y_{ij} - \mu_{ij}}{\phi_j v_j(\theta, \lambda_i; z_{ij})}, \quad i = 1, \ldots, n.
\]

Here the number of nuisance parameters increases at the same rate as the number of observations, as in the well-known Neyman and Scott (1948) problem, which hampers inference on \( \theta \). An analytical tool to investigate the effects of fitting nuisance parameters in this application, relative to the dispersion of the data, is provided by the examination of the limiting behavior of a profile estimating function under a “small dispersion and large sample size” asymptotic. That is, we assume that \( n \to \infty, \phi_1 \to 0, \) and \( \phi_2 = O(\phi_1) \). This asymptotic might be relevant with matched pairs of observations with comparable dispersion as one often expects the estimator performs well at least in the situation where the dispersion parameters are small and the sample size is large. Under this asymptotic scheme, the appropriate standardization is \( n^{-1/2} \phi_1^{1/2} s = O_p(1) \). Although \( s \) is an unbiased estimating function, in the sense that \( E\{s(\theta, \lambda_1, \ldots, \lambda_n); \theta, \lambda_1, \ldots, \lambda_n\} = 0 \), the act of forming a profile quasi-score \( \hat{s} \), by substituting \( h_i \)-based estimators \( (\hat{\lambda}_1, \ldots, \hat{\lambda}_n) \) for unknown \( (\lambda_1, \ldots, \lambda_n) \) in \( s \), will introduce plug-in bias, that is \( E(\hat{s}) \neq 0 \). Specifically, \( n^{-1/2} \phi_1^{1/2}(\hat{s} - s) = O_p(1) + O_p(n^{-1/2} \phi_1^{1/2}) \), where the \( O_p(1) \) term has mean zero, and the plug-in bias is \( E(n^{-1/2} \phi_1^{1/2} \hat{s}) = O(n^{1/2} \phi_1^{1/2}) \). This is of concern because bias of an estimating function usually translates into bias of the resulting estimator of \( \theta \) (e.g., Yanagimoto and Yamamoto, 1991; Liang and Zeger, 1995). Let \( \hat{\theta}_s \) be the root of \( \hat{s} \). Under our asymptotic scheme, typically, \( \hat{\theta}_s = \theta + O_p(n^{-1/2} \phi_1^{1/2}) + O_p(\phi_1) \), and the bias of \( \hat{\theta}_s \) is of an order \( n^{-1/2} \phi_1^{1/2} \) times the bias of standardized profile score \( n^{-1/2} \phi_1^{1/2} \hat{s} \).

Next we consider another example that is related mathematically to the first
one.

**Example 2.** Economics and public health studies often include covariates measured with error. Let the dependent variable $Y_i$, $i = 1, \ldots, n$, be distributed as Bernoulli with mean given by $\mu_{iy} = 1/{[1 + \exp(-\theta_0 - \theta_1 \lambda_i - \theta_2 z_{i2})]}$, where $z_i = (1, z_{i2})$ is a vector including an intercept term and a correctly measured covariate, $\lambda_i$ is an unknown covariate, the observed mismeasured version of this covariate is $X_i$ with mean $\mu_{ix} = E(X_i | \lambda_i, z_i) = \lambda_i$ and var$(X_i | \lambda_i, z_i) = \phi_x v_x(\lambda_i, z_i)$, and $\phi_x$ and $v_x(\cdot)$ are unknown. In this example, $\lambda'_i$s are nuisance parameters. Of interest are the regression coefficients $\theta_0$, $\theta_1$ and $\theta_2$.

We can generalize Example 2 by the following application.

**Application 2. Study with errors-in-covariates.** Let $Y_i$ be a response with mean and variance given by

$$E(Y_i | \lambda_i, z_i) = \mu_{iy}(\theta, \lambda_i; z_i), \quad \text{var}(Y_i | \lambda_i, z_i) = \phi_y v_y(\theta, \lambda_i; z_i),$$

where $z_i$ is a $(p - 1)$-dimensional vector of correctly measured covariates, $\lambda_i$ is a scalar unknown covariate, and $\theta$ is a $p$-dimensional vector of regression parameters relating covariate vector $(\lambda_i, z_i)$ to the expected response. Let $X_i$ be an observed mismeasured version of unknown covariate, $\lambda_i$, with a mean and variance

$$E(X_i | \lambda_i, z_i) = \mu_{ix}(\lambda_i; z_i), \quad \text{var}(X_i | \lambda_i, z_i) = \phi_x v_x(\lambda_i; z_i)$$

that typically do not depend on $\theta$. We assume the $n$ copies of pairs $(Y_i, X_i)$, $i = 1, \ldots, n$, are independently distributed, and the pair $(X_i, Y_i)$ are conditionally independent given $(\lambda_i, z_i)$. Unless otherwise stated, the dispersion parameters, $\phi_y$ and $\phi_x$, and variance functions, $v_y(\cdot)$ and $v_x(\cdot)$, are unknown. Under the so-called functional model, we regard $\lambda_i$, $i = 1, \ldots, n$, as stratum-specific nuisance
parameters (e.g., Stefanski and Carroll, 1987; Rathouz and Liang, 2001). Assuming
that surrogate covariate $X_i$ is a scalar, the functional model can be put into the
framework of a matched-pair study by adopting the notation $(X_i, Y_i) = (Y_{i1}, Y_{i2})$,
$(\mu_{i\epsilon}, \mu_{iy}) = (\mu_{i1}, \mu_{i2})$, $(\phi_x, \phi_y) = (\phi_1, \phi_2)$, and $(v_x, v_y) = (v_1, v_2)$, but here $\mu_{i1} =
\mu_{i\epsilon}(\lambda_i; z_i)$ typically does not depend on $\theta$. We consider an asymptotic setting sim-
ilar to the one in Application 1, with $n \to \infty$ and $\phi_x \to 0$, but here we assume
that $\phi_y = O(1)$, as might be relevant if the focus is on studying the effect of small
measurement error, keeping the variance of response $Y_i$ bounded. It follows that
$n^{-1/2}s = O_p(1)$, $n^{-1/2}(\hat{s} - s) = O_p(\phi_x^{1/2}) + O_p(n^{1/2}\phi_x)$, where the $O_p(\phi_x^{1/2})$ term has
mean zero, and the plug-in bias is $E(n^{-1/2}\hat{s}) = O(n^{1/2}\phi_x)$. It is straightforward to
obtain that $\hat{\theta}_x = \theta + O_p(\phi_x) + O_p(n^{-1/2})$.

In matched-pair studies, the classical method to reduce the effects of fitting
stratum-specific intercepts has been conditional likelihood (e.g., Kalbfleisch and
Sprott, 1970). This approach, unfortunately, is limited almost exclusively to applica-
tions of the exponential family with canonical link.

Several methods have been developed for the errors-in-covariates problem under
a functional model; outstanding issues remain, however, when we lack replicated
mismeasured covariates and lack knowledge of the full probability mechanism that
generated the data. The conditional score method of Stefanski and Carroll (1987)
and the semiparametric locally efficient method of Tsiatis and Ma (2004) both re-
quire complete specification of the probability model for the observations $Y$ and
surrogate covariates $X$; these assumptions can be relaxed somewhat if replicated
data are available. The corrected score method of Nakamura (1990) requires full
specification of the probability model for $Y$ but not necessarily for $X$; it exists
for only certain types of regression problems (Stefanski, 1989). The simulation-extrapolation approach (SIMEX) of Stefanski and Cook (1995) provides a nonparametric adjustment for measurement error, but does not yield a consistent estimator of regression parameters \( \theta \). The Monte Carlo corrected score method of Novick and Stefanski (2002), which is closely related to the SIMEX, is a promising method, but the theory is lacking when the errors are not normally distributed. The corrected estimating function approach of Huang and Wang (2000, 2001) is attractive because it accommodates both nonparametric adjustment of measurement error and consistent estimation of \( \theta \), under a large sample size asymptotic; unfortunately, this approach currently is limited to applications of logistic regression or proportional-hazards regression, and can require replicated mismeasured covariates. The instrumental variable method for the errors-in-covariates problem (Carroll et al., 1995, P107-121) applies even if knowledge of the measurement error variance is not available, but this method is not available if the study fails to include an instrumental variable.

A useful alternative to the above methods is an approach that seeks to reduce the sensitivity of an estimating function to nuisance parameters by considerations of local ancillarity. The idea is to adjust an original estimating function so that the resulting estimating function, if evaluated at an arbitrary perturbation of \( \lambda \), has a bias that is negligible, relative to the magnitude of the perturbation of \( \lambda \). Let \( Y = (y_1, \ldots, y_n) \) be a collection of independent random vectors whose distribution depends upon parameters of interest \( \theta \) and nuisance parameters \( \lambda \). A \textit{second-order locally ancillary estimating function} (SOLAEF), \( g \), is defined as satisfying (Small
and McLeish, 1989; 1994, P81-84)

\[ E\{g(\theta, \lambda; Y); \theta, \lambda^*\} = o\{||\lambda^* - \lambda||^2\}, \quad \text{as} \quad \lambda^* \to \lambda, \]

for all \( \theta \) and \( \lambda \). See Section 2.1 for a review.

In fully parametric problems, Small and McLeish (1989) and Waterman and Lindsay (1996a, b) present a SOLAEF that possesses certain optimality properties. In semiparametric problems where the probability model for observations is not completely available, if the scientific interest is on parameters related to the expected response, then Rathouz and Liang (1999, 2001) propose a different SOLAEF, known as the second-order locally ancillary quasi-score (SOLAQS), that requires specification of only the first two moments of the data; see Section 2.2 for a brief review.

The SOLAQS approach has been applied to matched pair studies (Rathouz and Liang 1999) and studies with errors-in-covariates (Rathouz and Liang 1999, 2001). In the latter type of study, if all components of the variance models, namely, \( \phi_x, \phi_y, v_x(\cdot) \) and \( v_y(\cdot) \) are known, and the asymptotic scenario presented in Application 2 is considered, then the profile SOLAQS, \( \hat{s}_2 \), say, for \( \theta \) achieves first-order bias correction of the profile quasi-score (Rathouz and Liang, 2001). That is, \( n^{-1/2}(\hat{s}_2 - s_2) = O_p(\phi_x^{1/2}) + O_p(n^{1/2}\phi_x^{3/2}) \), and its bias is only \( O(n^{1/2}\phi_x^{3/2}) \). Unfortunately, in the realistic situation where one or more of the variance components \( \phi_x, \phi_y, v_x(\cdot) \) and \( v_y(\cdot) \) is unknown, such as might occur if internal replicates or validation data either do not exist or are very limited, then the order of the bias correction of \( \hat{s}_2 \) is not guaranteed.

To overcome this limitation of the SOLAQS approach, we consider a class of second-order locally ancillary estimating functions for Applications 1 and 2 that do
not depend on full specification of the variances of the observations, and achieve first-order reduction of plug-in bias, which is the same order of bias reduction provided by the SOLAQS method, under a “small dispersion and large sample size” asymptotic.

To accomplish this goal, it is necessary to highlight an issue that heretofore has not received much attention: local ancillarity by itself is insufficient to ensure that a general estimation function has reduced order of plug-in bias. A key finding of this paper is that SOLAEFs $g$ for $\theta$ that satisfy the following orthogonality condition

$$E \left( h \frac{\partial g}{\partial \lambda} \right) = 0,$$

where $h$ is an estimating function for $\lambda$, typically achieve first-order bias reduction of profile estimating function $\hat{g}$. In fact, it can easily be seen that the above orthogonality condition is satisfied if $g$ is a SOLAEF and $h$ is the likelihood score for $\lambda$. The optimal SOLAEF of Small and McLeish (1989) satisfies the above orthogonality condition. By contrast, Rathouz and Liang’s SOLAQS approach satisfies the desired orthogonality condition only approximately. We present a class of SOLAEFs for Applications 1 and 2 that achieve exactly the above orthogonality condition and require remarkably few modeling assumptions. Specifically, in Application 1 we do not require that any of the variance components, $\phi_1$, $\phi_2$, $v_1(\cdot)$ and $v_2(\cdot)$, are known. We will see later that in Application 2 we often need to specify the variance function $v_2(\cdot)$ for the response variable, but we do not require that any of the remaining variance components $\phi_1$, $\phi_2$, and $v_1(\cdot)$ are known.

This article is organized as follows. In Section 2, we give background material on local ancillarity and SOLAQS. In Section 3, we present in detail our approach for the two applications of interest. We summarize asymptotic results under a “small
dispersion and large sample size” asymptotic setting, and present a brief simulation study and a data analysis example in Section 4. We conclude with a brief discussion in Section 5.

2. Background

2.1. Local ancillarity

Let a vector of observations $Y$ be a realization from a joint density function $f(y; \theta, \lambda)$, where $(\theta, \lambda) \in \Theta \times \Lambda \subset \mathbb{R}^p \times \mathbb{R}$. In this section we will omit using subscript $i$ to denote stratum. A general estimating function $g(\theta, \lambda; Y)$ is defined as $r$th-order locally ancillary (Small and McLeish, 1989) for nuisance parameter $\lambda$ if

$$E\{g(\theta, \lambda; Y); \theta, \lambda^*\} = o(|\lambda^* - \lambda|^r), \quad \text{as} \quad \lambda^* \to \lambda,$$

for all $\theta$ and $\lambda$. The order of the local ancillarity of $g$ provides a measure of the degree of robustness of $g$ to $\lambda$. The second order provides much of the insensitivity to nuisance parameters that is achievable through second- and higher-order local ancillarity, and as a result is considered the most important in practice (Rathouz and Liang, 1999, 2001; Waterman and Lindsay, 1996a, b; Small and McLeish, 1989).

Rathouz and Liang (1999) have shown that, under regularity conditions, for an unbiased estimating function $g$, $E(g') = E(g'') = 0$ implies second-order local ancillarity. Throughout this paper, we use a prime to denote the first derivative with respect to the nuisance parameter, two primes to denote the second derivative, and so on.

2.2. Second-order locally ancillary quasi-score

Using the notation in Application 1, Rathouz and Liang (1999 and 2001)’s SOLAQS
for the $i$th matched pair $Y_i = (Y_{i1}, Y_{i2})$ is given by

$$s_2(\theta; \lambda_i, Y_i) = s_i(\theta; \lambda_i, Y_i) - d_{l1}(\theta, \lambda_i)\hat{h}_i(\lambda_i; \theta, Y_i) - d_{2l}(\theta, \lambda_i)\tilde{h}_i(\lambda_i; \theta, Y_i), \quad i = 1, \ldots, n, \quad (3)$$

where $s_i$ and $h_i$ are defined in (1) and (2), respectively, $\tilde{h}_i = h_i^2 + h_i'$, and $d_{l1}$ and $d_{2l}$ are unique nonrandom vectors chosen to satisfy the second-order local ancillarity property $E(s_{2l}^t) = E(s_{2l}^\prime) = 0$. The SOLAQs estimator for $\theta$ is obtained by solving the equation $\hat{s}_2 = \sum_i s_{2i}|_{\lambda_i = \lambda_0} = 0$, where $\lambda_0$ is the root of $h_i$ for fixed $\theta$.

SOLAQS achieves first-order reduction of plug-in bias in the situation where, using the notation of Section 1, variance functions $v_1(\cdot)$ and $v_2(\cdot)$ and dispersion parameters $\phi_1$ and $\phi_2$ are known. Next we propose a class of SOLAEFs that provides the same order of plug-in bias reduction as SOLAQS and requires only partial modeling of the variances of the data.

3. Main results

The SOLAQS of Rathouz and Liang (1999 and 2001), as given by (3), consists of a linear combination of marginal terms, $Y_{ij} - \mu_{ij}$, a cross-product term, $(Y_{i1} - \mu_{i1})(Y_{i2} - \mu_{i2})$, and quadratic terms $(Y_{ij} - \mu_{ij})^2$. The coefficients corresponding to these terms are uniquely determined and require the correct specification of the variances of the observations. To address this concern, our proposal is to omit at least one of the quadratic terms from $s_{2i}$ and consider a class of SOLAEFs for Applications 1 and 2 that do not require complete specification of the variances of the observations. Below we present two types of SOLAEF, one of which omits all quadratic terms $(Y_{ij} - \mu_{ij})^2$, and the other of which omits only $(Y_{i1} - \mu_{i1})^2$, as might be appropriate in a study with errors-in-covariates if the response $Y_i \equiv Y_{i2}$ has known variance function $v_2(\cdot)$ and little is known about the variance of the observed
mismeasured covariate $X_i \equiv Y_{i1}$.

3.1. Nonquadratic orthogonal locally ancillary estimating function

We begin with a class of unbiased estimating functions $g_{2i}$ for the $i$th stratum taking the form

$$g_{2i} = w_i(\theta, \lambda_i)\{a_i(\theta, \lambda_i)(Y_{i1} - \mu_{i1}) + b_i(\theta, \lambda_i)(Y_{i2} - \mu_{i2}) + (Y_{i1} - \mu_{i1})(Y_{i2} - \mu_{i2})\}, \quad (4)$$

where scalars $a_i$ and $b_i$ are arbitrary nonrandom weights, $w_i$ is an arbitrary nonrandom $p$-dimensional vector, and $a_i$, $b_i$ and $w_i$ may depend on $\theta$ and $\lambda_i$. Let an unbiased estimating function $h_i$ for unknown $\lambda_i$ take the form

$$h_i = t_{1i}(\theta, \lambda_i)(Y_{i1} - \mu_{i1}) + t_{2i}(\theta, \lambda_i)(Y_{i2} - \mu_{i2}), \quad i = 1, \ldots, n, \quad (5)$$

where scalars $t_{1i}$ and $t_{2i}$ are arbitrary nonrandom weights, which may depend on $\theta$ and $\lambda_i$. To motivate our approach, we consider the plug-in bias of $\hat{g}_{2i} = g_{2i}(\theta, \hat{\lambda}_{i\theta})$, where $\hat{\lambda}_{i\theta}$ is the root of $h_i$ in (5) for a fixed value of $\theta$, under a small dispersion asymptotic. We briefly assume that the above choices of $g_{2i}$ and $h_i$ are standardized to be of order $O_p(1)$, and $a_i/b_i = O(1)$, as $\phi_1 \to 0$, $\phi_2/\phi_1 = O(1)$. Also assume that the orders of $a_i'$, $a_i''$ and $a_i'''$ ($b_i'$, $b_i''$ and $b_i'''$) are the same as or lower than that of $a_i$ ($b_i$). One can show by a Taylor expansion argument that we typically have, assuming the existence of the observed and expected derivatives,

$$\hat{g}_{2i} - g_{2i} = h_i g_{2i}' O(\phi_1^{1/2}) + E(g_{2i}') O_p(\phi_1) + E(g_{2i}'') O_p(\phi_1) + O_p(\phi_1), \quad (6)$$

where $E(g_{2i}') = O(\phi_1^{-1/2})$, $E(g_{2i}'') = O(\phi_1^{-1/2})$, $h_i g_{2i}' = O_p(\phi_1^{-1/2})$, and $E(h_i g_{2i}''') = O(1)$. A sketch of the proof of (6) can be found in the Appendix. If $E(g_{2i}') = 0$ then $h_i g_{2i}' = O_p(1)$, but the bias of this term generally remains of order $O(1)$. Hence,
we typically have $E(\hat{g}_{2i} - g_{2i}) = O(\phi_1^{1/2})$, unless both the following orthogonality condition,

$$E(h_i g''_{2i}) = 0, \quad (7)$$

and the second-order locally ancillary condition, $E(g'_{2i}) = E(g''_{2i}) = 0$, are met, in which case we achieve first-order bias reduction, $E(\hat{g}_{2i} - g_{2i}) = O(\phi_1)$. Note that an orthogonality condition studied by Godambe (1991), namely $E(h_i g_{2i}) = 0$, is not directly relevant to the goal of achieving this first-order bias reduction. We call an SOLAEF satisfying orthogonality condition (7) an orthogonal second-order locally ancillary estimating function (OSOLAEF), and call the class of OSOLAEFs taking form (4), which are free of quadratic terms $(Y_{ij} - \mu_{ij})^2$, the class of nonquadratic OSOLAEFs.

Note that if orthogonality criterion (7) were relaxed to $E(h_i g'_{2i}) \to 0$ as $\phi_1 \to 0$, as is the case with the SOLAQS approach (3) considered by Rathouz and Liang (1999, 2001), then first-order bias reduction of $\hat{g}_{2i}$ would still hold. Intuitively, however, one would expect that when conducting inference in realistic situations, there would be an advantage to meeting the orthogonality condition exactly.

### 3.2. Construction and properties of nonquadratic OSOLAEF

We can choose nonrandom weights $(w_i, a_i, b_i, t_{1i}, t_{2i})$ in (4) and (5) such that $g_{2i}$ is an OSOLAEF. Assuming that $\mu'_{ij} \neq 0$ and $\mu''_{i1}/\mu'_{i1} - \mu''_{i2}/\mu'_{i2} \neq 0$, it follows that $a_i$ and $b_i$ are uniquely determined in (4), and depend on only the mean model for the data

$$a_i = -2\mu'_{i2}/\left(\frac{\mu''_{i1}}{\mu'_{i1}} - \frac{\mu''_{i2}}{\mu'_{i2}}\right), \quad b_i = 2\mu'_{i1}/\left(\frac{\mu''_{i1}}{\mu'_{i1}} - \frac{\mu''_{i2}}{\mu'_{i2}}\right), \quad (8)$$
whereas the solution to the remaining weights \((w_i, t_{i1}, t_{i2})\) is non-unique and need only satisfy

\[
t_{2i} \{(b_i w_i) - w_i \mu_{i1}'\} + t_{1i} \{(a_i w_i) - w_i \mu_{i2}'\} r_{vi} = 0,
\]

where \(r_{vi} = \text{var}(Y_{i1})/\text{var}(Y_{i2}) = \phi_1 v_1(\theta, \lambda_i)/\{\phi_2 v_2(\theta, \lambda_i)\}\). Note that, by (8), we have \(\mu_{i2}' \{(b_i w_i) - w_i \mu_{i1}'\} + \mu_{i1}' \{(a_i w_i) - w_i \mu_{i2}'\} = 0\). We make two recommendations for \((w_i, t_{i1}, t_{i2})\), depending on whether or not the ratio of variances of \(Y_{i1}\) and \(Y_{i2}\) can be specified correctly.

**Efficient choice:** If the variance functions \(v_1(\cdot)\) and \(v_2(\cdot)\) are known, as is the ratio \(\phi_1/\phi_2\) of dispersion parameters, then choosing \(h_i\) in (5) as proportional to the \(\lambda_i\)-quasi-score

\[
h_i \propto \mu_{i1}'(Y_{i1} - \mu_{i1}) + \mu_{i2}'(Y_{i2} - \mu_{i2}) r_{vi},
\]

for arbitrary \(w_i(\theta, \lambda_i)\) in (4), achieves two desirable orthogonality conditions, \(E(h_i g'_{2i}) = 0\) and \(E(h_i g_{2i}) = 0\), just as would the optimal SOLAEF by Small and Mdeish (1989) if the full probability model for the observations were available. An efficient choice of weight \(w_i\) in (4) is given by

\[
w_i^{\text{eff}} = \phi_2 \frac{\phi \partial g_{2i}/\partial \theta}{\text{var}(g_{2i})} = \frac{a_i(\partial \mu_{i1}/\partial \theta) + b_i(\partial \mu_{i2}/\partial \theta)}{\phi_1/\phi_2 a_i^2 v_1(\theta, \lambda_i) + b_i^2 v_2(\theta, \lambda_i) + \phi_1 v_1(\theta, \lambda_i)v_2(\theta, \lambda_i)},
\]

where \(g_{2i} = a_i(Y_{i1} - \mu_{i1}) + b_i(Y_{i2} - \mu_{i2}) + (Y_{i1} - \mu_{i1})(Y_{i2} - \mu_{i2})\). Under this choice of \(w_i\), the scaled estimating function \(\hat{\phi}_2^{-1} g_{2i}\) is information unbiased in the sense that (Lindsay, 1982) \(E\{(\hat{\phi}_2^{-1} g_{2i})^2 + \hat{\phi}_2^{-1} g'_{2i}\} = 0\), a property shared by the fully parametric optimal SOLAEF. In the special case where there is a common scale parameter, \(\hat{\phi}_1 = \hat{\phi}_2 = \phi\), and \(\phi\) is unknown, note that the efficient choice of \(h_i\), as given by (10), is free of \(\phi\) and, if \(\phi\) is small, the third term in the denominator of
(11) is smaller than the other two terms and can safely be ignored, in which case a nearly efficient weight that does not depend on \( \phi \) is

\[
w_i^* = \frac{a_i(\partial \mu_{i1}/\partial \theta) + b_i(\partial \mu_{i2}/\partial \theta)}{a_i^2v_1(\theta, \lambda_i) + b_i^2v_2(\theta, \lambda_i)}.
\]  

(12)

**Robust Choice:** Alternatively, if one desires robustness of inferences to second moment assumptions, then inspection of (9) reveals that it suffices to choose \( w_i \) such that \( (a_iw_i)' - w_i\mu_{i2}' = 0 \), and consequently, \( (b_iw_i)' - w_i\mu_{i1}' = 0 \), which leads to

\[
g_{2i} = c_i(\theta)\frac{(\mu_{i2}'/\mu_{i1})^{1/2}}{a_i}\left\{a_i(Y_{i1} - \mu_{i1}) + b_i(Y_{i2} - \mu_{i2}) + (Y_{i2} - \mu_{i2})(Y_{i1} - \mu_{i1})\right\},
\]  

(13)

where \( c_i(\theta) \) is an arbitrary nonrandom p-dimensional vector that may depend on \( \theta \), and \( a_i \) and \( b_i \) are defined as in (8). It follows that \( g_{2i} \) in (13) is a nonquadratic OSOLAEF for arbitrary choices of weights \( t_{i1} \) and \( t_{2i} \) in estimating function \( h_i \) for \( \lambda_i \). A natural choice of \( h_i \) is one that mimics the efficient choice (10)

\[
\tilde{h}_i \propto \mu_{i1}'(Y_{i1} - \mu_{i1}) + \mu_{i2}'(Y_{i2} - \mu_{i2})\tilde{r}_{wi},
\]

(14)

where \( \tilde{r}_{wi} = \tilde{r}_{wi}(\theta, \lambda_i) \), which is allowed to depend on \( \theta \) and \( \lambda_i \), is an arbitrary “working variance ratio” of \( Y_{i1} \) and \( Y_{i2} \). Here, we do not require that \( \tilde{r}_{wi} = r_{wi} \), although one would expect that the better \( \tilde{r}_{wi} \) approximates \( r_{wi} \), the greater the efficiency of inferences.

Combining all strata, the nonquadratic OSOLAEF for \( \theta \) is \( g_2 = \sum_i g_{2i} \) and the profile estimating function is \( \hat{g}_2 = \sum_i g_{2i} \big|_{\lambda_i = \hat{\lambda}_i} \), where \( \hat{\lambda}_i \) is the solution to \( h_i = 0 \). We evaluate the plug-in bias of \( g_2 \) under an asymptotic setting where \( \phi_1 \to 0 \) and \( n \to \infty \). We first consider the situation where observations have comparable dispersion, that is \( \phi_2/\phi_1 = O(1) \), as might be relevant in matched-pair studies. Note that the order of \( g_2 \) will depend to some extent on the order of the
weights $w_i$, which can be rather arbitrary. In general, if weight $w_i = O(\phi_i^\beta)$, say, for some constant $\beta$, then $g_{2i}$ is of order $O_p(\phi_i^{\beta + 1/2})$. To standardize the statement of the following theorem, we simply assume that there exists a constant $\alpha$ such that $\phi_i^{-\alpha} g_{2i} = O_p(1), i = 1, \ldots, n$.

**Conclusion 1.** Let $g_2 = \sum_i g_{2i}$ and $\hat{g}_2 = \sum_i g_{2i} |_{\lambda_i = \lambda_0}$, where $g_{2i}$ is given by (4) with $a_i$ and $b_i$ defined by (8), and $\lambda_0$ is the root of $h_i$ in (5) with $t_{1i}$, $t_{2i}$ and $w_i$ satisfying (9). Let $\hat{\theta}$ be the root of $\hat{g}_2$. Let $\phi_2/\phi_1 = O(1)$ as $\phi_1 \to 0$. Under this asymptotic scheme, assume that each weight $w_i$ in (4) is of the same order, $O(\phi_1^{\alpha - 1/2})$, say, for arbitrary constant $\alpha$, so that $\phi_i^{-\alpha} g_{2i} = O_p(1), i = 1, \ldots, n$. Assume $E(\partial g_{2i} / \partial \theta) \neq 0$.

Under appropriate regularity conditions, as $\phi_1 \to 0$ and $n \to \infty$, $n^{-1/2} \phi_i^{-\alpha} (\hat{g}_2 - g_2) = Z + O_p(n^{1/2} \phi_1), \text{ where } Z = O_p(\phi_1^{1/2}) \text{ and } E(Z) = 0$, and $E(n^{-1/2} \phi_i^{-\alpha} \hat{g}_2) = O(n^{1/2} \phi_1)$. Furthermore $\hat{\theta} = \theta + O_p(\phi_1^{3/2}) + O_p(n^{-1/2} \phi_1^{1/2})$.

The proof is in the Appendix. Conclusion 1 implies that $\hat{\theta}$ converges to $\theta$ at a faster rate than $\hat{\theta}_*$, the estimator of the usual profile quasi-score approach, provided that $n^{-1} = o(\phi_1)$. Note that, if either the orthogonality condition or the second-order local ancillarity condition fails, then the bias of $n^{-1/2} \phi_i^{-\alpha} \hat{g}_2$ is $O(n^{1/2} \phi_1^{1/2})$.

Next we consider a different asymptotic situation where $\phi_2 = O(1)$ as $\phi_1 \to 0$, as might be relevant in an errors-in-covariates problem if one desires to study the effect of small measurement error, keeping the variance of the response bounded. Here, some care is needed in the choice of weights $t_{1i}$ and $t_{2i}$ in (5), since the terms $Y_{1i} - \mu_{1i}$ and $Y_{2i} - \mu_{2i}$ are now of discrepant orders, $O_p(\phi_1^{1/2})$ and $O_p(1)$, respectively, and the second term by itself would lead to poor inferences. If $t_{2i} \neq 0$, then it is imperative to choose weights $t_{1i}$ and $t_{2i}$ such that $t_{2i}/t_{1i} = O(\phi_1^{1+\beta})$, say, for some arbitrary nonnegative constant $\beta \geq 0$, so that the first term in (5) dominates the
latter term as φ₁ → 0, φ₂ = O(1). Note that the efficient choice (10) meets this criterion, while for the robust choice (14), with unknown φ₁, we would resort to omitting the second term from hᵢ, that is, setting t₂i = 0.

**Conclusion 2.** Define g₂, ˆφ₂, ˆθ, and hᵢ, i = 1,...,n as in Conclusion 1. Let φ₂ = O(1) as φ₁ → 0. Under this asymptotic scheme, assume that each weight wᵢ in (4) is of the same order, O(φᵢ⁻¹), say, for arbitrary constant α, so that φ⁻α₁g₂i = O_p(1), i = 1,...,n. If t₂i ≠ 0, further assume that t₂i/t₁i in (5) is of order O(φ₁^(1+β)) as φ₁ → 0, for some arbitrary nonnegative constant β ≥ 0. Assume E(∂g₂i/∂θ) ≠ 0.

Under appropriate regularity conditions, as φ₁ → 0 and n → ∞, n⁻¹/₂φ⁻¹₁(ˆg₂ - g₂) = Z + O_p(n⁻¹/₂φ₁⁻³/₂), where Z = O_p(φ₁¹/₂) and E(Z) = 0, and E(n⁻¹/₂φ⁻α₁g₂) = O(n⁻¹/₂φ⁻³/₂). Furthermore, ˆθ = θ + O_p(φ₁³/₂) + O_p(n⁻¹/2).

The proof is in the Appendix. From Conclusion 2, it follows that, similar to the SOLAQS approach, ˆθ converges to θ at a faster rate than the naive θ-estimator from the usual profile quasi-score approach, under the asymptotic scheme considered, provided that n⁻¹ = O(φᵢ²). Note that, if either the orthogonality condition or the second-order local ancillarity condition fails, then we typically have n⁻¹/₂φ⁻³₁(ˆg₂ - g₂) = O_p(φ₁¹/₂) + O_p(n⁻¹/2φ₁) with bias O(n⁻¹/2φ₁).

In Application 1, a natural choice of cᵢ(θ) in the robust OSOLAEF (13) might be based on the p-dimensional vector of regression covariates zᵢj. For example, one can set cᵢ(θ) as either zᵢ1, zᵢ2, or their average (z₁i + zᵢ2)/2. These natural choices of weight are not available in Application 2, however, since in the errors-in-covariates problem the p-dimensional vector of covariates is given by (λᵢ, zᵢ), and it would be unacceptable to set cᵢ(θ) = (λᵢ, zᵢ) as we require that cᵢ(θ) be free of λᵢ in (13) and in Conclusions 1 and 2. Moreover, the naive strategy of substituting surrogate
covariate $X_i$ for $\lambda_i$ in the weight, that is, setting $c_i(\theta) = (X_i, z_i)$, is also unacceptable, since $X_i$ is a random variable and the resulting estimating function (13) generally would no longer be unbiased. It follows that, as a practical matter, one would resort to the simple specification $c_i(\theta) = (1, z_i)$ for the robust OSOLAEF in Application 2. But if $z_i$ includes an intercept term then, with this simple choice of weight $c_i(\theta)$, the rank of estimating function (13) reduces from $p$ to $p - 1$. Hence, in the absence of replicates of $X_i$, we face a potential non-identifiability problem for $\theta$ under the robust nonquadratic OSOLAEF approach for Application 2.

Note that if independent replicates $\{X_{i1}, X_{i2}\}$ of $X_i$ were available, then this identifiability problem could be resolved, since one could use a robust, unbiased, rank-$p$ OSOLAEF of the form $g_{2i} = (X_{i1}, z_i)g_{3i}(Y_i, X_{i2}; \theta, \lambda_i) + (X_{i2}, z_i)g_{3i}(Y_i, X_{i1}; \theta, \lambda_i)$, where $g_{3i}$ is the scalar-valued kernel of estimating function (13), that is,

$$
g_{3i}(Y_i, X_i; \theta, \lambda_i) = \left(\frac{\mu'_{iy}/\mu'_{ix}}{a_i}\right)^{1/2}\left\{a_i(Y_i - \mu_{iy}) + b_i(X_i - \mu_{ix}) + (X_i - \mu_{ix})(Y_i - \mu_{iy})\right\}.
$$

When replicates of surrogate covariate $X_i$ are not available and one faces the above identifiability problem for $\theta$ in Application 2, we propose the following alternative class of OSOLAEFs.

### 3.3. Quadratic OSOLAEF

Assume in Application 2 that the variance function $v_y(\cdot)$ for the response is known, but the remaining variance components in the model, namely $\phi_y$, $\phi_x$ and $v_x(\cdot)$, possibly are unknown. Consider a type of OSOLAEF that includes the quadratic term $(Y_i - \mu_{iy})^2$, given by

$$
g_{2i}^\dagger = e_{1i}(\theta, \lambda_i)(X_i - \mu_{ix}) + e_{2i}(\theta, \lambda_i)(Y_i - \mu_{iy}) + e_{3i}(\theta, \lambda_i)(X_i - \mu_{ix})(Y_i - \mu_{iy}) + e_{4i}(\theta, \lambda_i)((Y_i - \mu_{iy})^2 - \phi_y v_y(\theta, \lambda_i; z_i)),
$$

(15)
where $e_{1i}$, $e_{2i}$, $e_{3i}$ and $e_{4i}$ are arbitrary nonrandom weights that may depend on $\theta$ and $\lambda_i$. Let $h_i = X_i - \mu_{ix}$. We will consider two cases depending upon whether or not dispersion parameter $\phi_y$ for the response is known. If $\phi_y$ is known then, in the above quadratic OSOLAEF, one can define the weights, $e_{1i}$, $e_{2i}$, $e_{3i}$ and $e_{4i}$ as scalars, and estimate $\theta$ by solving the rank $p$ mixed system of equations

$$\left( \sum_i g_{2i}^\perp(\theta; \hat{\lambda}_i \theta), \sum_i g_{2i}(\theta; \hat{\lambda}_i \theta) \right) = 0,$$

(16)

where $g_{2i}(\theta; \hat{\lambda}_i \theta)$ is the rank $(p-1)$ robust, nonquadratic OSOLAEF (13) with the simple choice of weight $c_i(\theta) = (1, z_i)$, where $z_i$ is assumed to include an intercept term. In the case where $\phi_y$ is unknown, one can define $e_{1i}$, $e_{2i}$, $e_{3i}$ and $e_{4i}$ as two-dimensional vectors, and estimate $\theta$ and $\phi_y$ by solving the rank $(p+1)$ mixed system of equations (16).

One can choose $e_{1i}$, $e_{2i}$, $e_{3i}$ and $e_{4i}$ such that the system of equations $E(g_{2i}^\perp) = 0$, $E(g_{2i}^\perp h_i) = 0$ and $E(g_{2i}^\perp h_i^2) = 0$, which do not require knowledge of the second moment of $X_i$, hold. After plugging in the expression (15) of $g_{2i}^\perp$, this system of equations can be written as

$$I(v_y^\prime \neq 0) \left\{ 2e_{1i} \mu_{ix}^\prime + e_{1i} \mu_{ix}^\prime + e_{2i} \mu_{iy}^\prime - (e_{1i} \mu_{ix}^\prime + e_{2i} \mu_{iy}^\prime) \frac{2\mu_{iy}^2}{\phi_y v_y^2} \right\} + I(v_y^\prime = 0) \{ e_{1i} + \mu_{iy}^\prime / \mu_{ix}^\prime e_{2i} \} = 0,$$

$$e_{3i} = e_{1i} / \mu_{iy}^\prime, \quad e_{4i} = I(v_y^\prime \neq 0) \left\{ -e_{1i} \mu_{ix}^\prime - e_{2i} \mu_{iy}^\prime \right\} + I(v_y^\prime = 0) \left\{ -e_{1i} \mu_{ix}^\prime + e_{2i} \mu_{iy}^\prime - 2e_{1i} \mu_{ix}^\prime \right\},$$

where $I(v_y^\prime = 0)$ is 1 if $v_y^\prime = 0$ and is 0 if $v_y^\prime \neq 0$, and $I(v_y^\prime \neq 0) = 1 - I(v_y^\prime = 0)$. Since there are three equations and four unknown weights, one set of solutions is obtained by fixing $e_{2i}$ and determining the other three weights by solving the above equations. If $\phi_y$ is known, we recommend setting scalar $e_{2i} = (\partial \mu_{iy} / \partial \theta_1) \text{var}^{-1}(Y_i)$, where $\theta_1$ denotes the regression coefficient of the mismeasured covariate $\lambda_i$. This choice $e_{2i}$ is an optimal weight (Godambe, 1976) for an estimating function taking
the form $\sum_i e_{2i}(Y_i - \mu_i)$. If $\phi_y$ is unknown, one may use the two-dimensional choice $e_{2i} = \{\partial\mu_i/\partial(\theta_0, \theta_1)\} \text{var}^{-1}(Y_i)$, where $\theta_0$ denotes the intercept. It is straightforward to see that the asymptotic results in Conclusion 2 hold for rank $p$ or rank $(p + 1)$ OSOLAEFs of mixed form (16).

For the OSOLAEFs presented in Sections 3.2 and 3.3, the variance of the resulting solution $\hat{\theta}$ can be estimated by the bootstrap method; for results on an analytical variance estimator, see the appendix.

4. Examples

Example 1 (Continued.) To implement an OSOLAEF approach for this matched-pairs example, note that $g_{2i}$ as given by (4), with $a_i$ and $b_i$ determined by (8), takes the form

$$g_{2i} = w_i(\theta, \lambda_i) \left\{ \frac{-\mu_{i2}^2}{\mu_{i1} - \mu_{i2}}(y_{i1} - \mu_{i1}) + \frac{\mu_{i1}^2}{\mu_{i1} - \mu_{i2}}(y_{i2} - \mu_{i2}) + (y_{i1} - \mu_{i1})(y_{i2} - \mu_{i2}) \right\}$$

$$= w_i(\theta, \lambda_i) \left\{ \frac{y_{i1} - y_{i2}}{\theta_1(z_{i1} - z_{i2})} + y_{i1}y_{i2} \right\}.$$  \hspace{1cm} (17)

It follows that $g_{2i}$ is free of nuisance parameters, with the possible exception of the presence of $\lambda_i$ in the choice of weight $w_i(\theta_1, \lambda_i)$, and the result in (17) depends upon only the model for the expected responses. Here we consider the simple choice $w_i = 1$, and we denote the resulting estimating function $g_2 = \sum_{i=1}^n g_{2i}$ as OEF-1.

We also consider a potentially more efficient choice of weight as given by (12), which depends on $\lambda_i$ and the variance functions $v_1(\cdot)$ and $v_2(\cdot)$, but does not depend on unknown common dispersion parameter $\phi$. We obtain

$$g_{2i}^* = (z_{i1} - z_{i2}) \frac{\mu_{i1} - \mu_{i2}}{\mu_{i1}^2 + \mu_{i2}^2} \left\{ \frac{y_{i1} - y_{i2}}{\theta_1(z_{i1} - z_{i2})} + y_{i1}y_{i2} \right\}.$$  \hspace{1cm} (18)

We denote as OEF-2 the profile estimating function $\hat{g}_{2i}^* = \sum_{i=1}^n g_{2i}^*|_{\lambda_i = \hat{\lambda}_{\theta_1}}$, where $\hat{\lambda}_{\theta_1}$ is the root of the scaled $\lambda_i$-quasi-score $h_i = -\phi\sum_{j=1}^\mu \mu_{ij}^2 \text{var}^{-1}(y_{ij})(y_{ij} - \mu_{ij})$.
\[ \sum_{j=1}^{2} (y_{ij} - \mu_{ij}). \] By contrast, the SOLAQS approach for this application requires knowledge of dispersion parameter \( \phi \) in addition to the variance functions \( v_1(\cdot) \) and \( v_2(\cdot) \).

We conducted a brief simulation study to compare the performance of OEF-1 and OEF-2 with other approaches. As in Rathouz and Liang (1999), we set \( z_{i1} = 1 - z_{i2} = 0 \). The model has the canonical link and the mean-variance relationship of a gamma regression model, but to show that our method is robust to higher moment assumptions, we generated data from log-normal distributions with means and variances as stated above. Shown in Figure 1 are the comparisons of point estimates of \( \theta_1 \) by OEF-1 and OEF-2, with those obtained by the profile quasi-score (PQS), SOLAQS substituting a second-order locally ancillary estimator for unknown \( \phi \) (AQS-2) and SOLAQS assuming \( \phi \) known (AQS-1). The coverage rates of estimated 95% confidence interval were also obtained, but, due to a space limitation, are not displayed here. As expected, SOLAQS with estimated \( \phi \) yielded an estimator of \( \theta_1 \) with larger bias and poorer 95% confidence interval coverage rate than that obtained by AQS-1. As seen in Figure 1, OEF-1 yielded an approximately unbiased estimator with large variance. By contrast, OEF-2 provided a \( \theta_1 \) estimator that was satisfactory in terms of both precision and small bias, achieving a variance as small as that from AQS-1. In terms of coverage rate of the 95% confidence interval for \( \theta_1 \), OEF-1 and OEF-2 were uniformly better than AQS-1 and AQS-2. We conjecture that this result is at least partially explained by the absence of nuisance parameters in the kernel of estimating function (17) and the benefit derived from meeting the orthogonality criterion, \( E(h_1 g'_{21}) = 0 \), exactly.

Example 2 (Continued.) A robust, nonquadratic OSOLAEF for regression pa-
rameters $\theta = (\theta_0, \theta_1, \theta_2)$ obtained according to (13) is
\[
g_{2i} = \frac{z_i(Y_i - \mu_i)}{\{\mu_i(1 - \mu_i)\}^{1/2}} - z_i \theta_1 \{\mu_i(1 - \mu_i)\}^{1/2} (X_i - \lambda_i) - \frac{z_i \theta_1(1 - 2\mu_i)(Y_i - \mu_i)(X_i - \lambda_i)}{2 \{\mu_i(1 - \mu_i)\}^{1/2}}, \tag{18}
\]
which has a rank of only 2, since here $z_i = (1, z_{i2})$. To avoid issues of nonidentifiability of $\theta$ in (18), we consider augmenting (18) using a scalar-valued quadratic OSOLAEF taking the form
\[
g_{2i}^\dagger = e_{i2}(Y_i - \mu_i) - d_i \theta_1 (X_i - \lambda_i) - \frac{d_i (1 - 2\mu_i) \theta_1}{2 \{\mu_i(1 - \mu_i)\}^{1/2}} (Y_i - \mu_i)(X_i - \lambda_i)
+ \frac{d_i - e_{i2} \{\mu_i(1 - \mu_i)\}^{1/2}}{\{\mu_i(1 - \mu_i)\}^{1/2}(1 - 2\mu_i)} \{(Y_i - \mu_i)^2 - \mu_i(1 - \mu_i)\},
\]
where $d_i$ and $e_{i2}$ are nonrandom terms satisfying $d_i^2 = 0$ and $d_i \neq 0$. The last inequality is due to the regularity condition $E(\partial g_{2i}^\dagger / \partial \theta) \neq 0$ in Conclusion 2. As recommended in Section 3.3, let $e_{i2} = \partial \mu_i / \partial \theta_1 \text{var}^{-1}(Y_i) = \lambda_i$. Also set $d_i = 1$. We have
\[
g_{2i}^\dagger = \lambda_i(Y_i - \mu_i) - \{\mu_i(1 - \mu_i)\}^{1/2} \theta_1 (X_i - \lambda_i) - \frac{(1 - 2\mu_i) \theta_1}{2 \{\mu_i(1 - \mu_i)\}^{1/2}} (Y_i - \mu_i)(X_i - \lambda_i)
+ \left[ \frac{1}{\{\mu_i(1 - \mu_i)\}^{1/2}(1 - 2\mu_i)} - \frac{\lambda_i}{(1 - 2\mu_i)} \right] \{(Y_i - \mu_i)^2 - \mu_i(1 - \mu_i)\}. \tag{19}
\]
Recall from the discussion in Section 3.2 that, if the variance of $X_i$ is unknown in the context of an errors-in-covariates problem, then ordinarily it is preferred to set $t_{2i} = 0$ in equation (5); accordingly, here we set $h_i = X_i - \lambda_i$. Note that after plugging $\lambda_i = X_i$ into the rank 3 system of OSOLAEFs $(\sum g_{2i}(\theta; \lambda_i), \sum g_{2i}^\dagger(\theta; \lambda_i))$, the terms containing $X_i - \lambda_i$ will vanish.

In a brief simulation study, we generated $\lambda_i$ from $N(0,1)$, and $z_i$ from $\text{Unif}(0.1,2.1)$, $i = 1, \ldots, 200$. We generated $Y_i$, $i = 1, \ldots, 200$, from Bernoulli distributions with the correctly specified means and generated $X_i$ from $N(\lambda_i, \phi_x)$ with $\phi_x = 0.1$. We
compared estimation of $\theta$ under the OSOLAEF ($\sum_i g_{2i}, \sum_i g^1_{2i}$) (OEF-1), which does not require knowledge of the variance model components $\phi_x$ and $v_x(\cdot)$ of the surrogate covariate, and the naive $\theta$-quasi-score (PQS) $g = \sum (\partial \hat{\mu}_i / \partial \theta ) \hat{\vartheta}^{-1}(\hat{\mu}_i)(Y_i - \hat{\mu}_i)$ with $\hat{\mu}_i = 1 / \{1 + \exp(-\theta_0 - \theta_1 X_i - \theta_2 z_i)\}$ and $\hat{\vartheta}(\hat{\mu}_i) = \hat{\mu}_i(1 - \hat{\mu}_i)$, where we ignored the measurement error. We also obtained $\theta$-estimates from the nonquadratic OSOLAEF with efficient choice of weights (OEF-2), the SOLAQS method (AQS), and the SIMEX method (Carroll et al., 1995, Chapter 4), assuming in the implementation of these methods that $\phi_x$ is specified either correctly as $\phi_x = 0.1$ or misspecified as 0.13. Shown in Table 1 are the averages of the estimates of $\theta$ over 1000 simulated replicates, along with their empirical standard errors. The estimator using OEF-1, which does not require specification of the second moment of $X_i$, is satisfactory in term of bias, and OEF-2, AQS and SIMEX were all sensitive to misspecification of the variance of the measurement error.

Example 3. Framingham Heart Study. We used the proposed method to reanalyze data on $n = 1615$ men from the Framingham Heart study (e.g., Carroll et al., 1995, P4-5), which had a binary response, denoted as $Y_i$, representing the occurrence of coronary heart disease. A main predictor was long-term systolic blood pressure (LSBP). Since blood pressure has daily, as well as seasonal, variation, it is difficult to measure long-term systolic blood pressure (Carroll et al., 1995, P5). Instead, the blood pressure measurement taken during a clinic visit was used as a surrogate. For the $i$th subject, denote the unknown long-term systolic blood pressure as $\lambda_i$ and the mismeasured version of the unknown true long-term systolic blood as $X^{(1)}_i$. There was also an extra measurement of blood pressure, $X^{(2)}_i$, taken in a clinic visit 4 years before $X^{(1)}_i$ was observed. The measurement error variance, $\phi_x$, can be estimated...
by \( \hat{\phi}_x = \sum_i \sum_k (X^{(k)}_i - X_i)^2 / n \), where \( X_i \) is the average of \( X^{(1)}_i \) and \( X^{(2)}_i \). The other important predictor was smoking status \( (z^{(2)}_i) \), coded as 1 for smokers and 0 for nonsmokers, and considered to be measured correctly.

As in Carroll et al. (1995, P88), the values of \( X^{(k)}_i \) were transformed to \( \log(X^{(k)}_i - 50) \) for \( k = 1, 2, i = 1, \ldots, n \). For convenience, we still use the notation \( X^{(k)}_i \) to denote the transformed values of blood pressure.

We assumed a logistic regression model for the binary response \( Y_i \); that is, \( \mu_{iy} = \{1 + \exp(-\theta_0 - \theta_1 \lambda_i - \theta_2 z^{(2)}_i)\}^{-1} \), which includes an intercept term, so the vector of correctly measured predictors can be denoted \( z_i = (1, z^{(2)}_i) \), in keeping with notation used in Application 2. We also assumed \( E(X^{(1)}_i) = \lambda_i \) and \( \text{var}(X^{(1)}_i) = \phi_x \). To show that our approach works for a study where we lack knowledge of the variance of the measurement error, we fitted a model (OEF-1) not using knowledge of the measurement error variance and ignoring \( X^{(2)}_i \). This was accomplished using a mixed equations OSOLAEF \((\sum g_{2i}; \sum g_{2i}^\dagger)\), where \( g_{2i} \) is the rank 2, robust, nonquadratic OSOLAEF (18), scalar \( g_{2i}^\dagger \) is the quadratic OSOLAEF (19), and \( h_i = X^{(1)}_i - \lambda_i \).

We also applied a potentially more efficient choice of OSOLAEF for \( \theta \) (OEF-2). Specifically, we estimated \( \lambda_i \) using the efficient choice (10)

\[
h_i = (X^{(1)}_i - \lambda_i) + \theta \phi_x (Y_i - \mu_{iy}),
\]

and we estimated \( \theta \) using the nonquadratic OSOLAEF (4)

\[
g_2 = \sum_{i=1}^{1615} w_i(\theta, \lambda) \{a_i(X^{(1)}_i - \lambda_i) + b_i(Y_i - \mu_{iy}) + (Y_i - \mu_{iy})(X^{(1)}_i - \lambda_i)\},
\]

where \( a_i \) and \( b_i \) are defined in (8), and \( w_i \) takes the efficient form (11). In addition, we applied the SOLAQS method (AQS) and SIMEX. Note that OEF-2, AQS and
SIMEX require estimation of $\phi_{\alpha}$ using the replicates $X_{i}^{(2)}$. The estimated coefficients with bootstrap standard errors, along with the naive profile quasi-score estimate (PQS) that used $X_{i}^{(1)}$ as long-term blood pressure and ignored measurement error, are presented in Table 2.

In the original data set, there was another measurement, $X_{i}^{(3)}$, of blood pressure in the same visit as that when $X_{i}^{(1)}$ was taken. The instrumental variable method (Carroll et al., 1995, P107-121) can be applied as $X_{i}^{(3)}$ can be regarded as an instrumental variable. This method, however, requires regressing $X_{i}^{(3)}$ on $X_{i}^{(1)}$, and additional assumptions.

5. Discussion

A key advantage of the proposed approach for addressing nuisance parameter problems in matched pair studies and studies with errors-in-covariates is that it requires minimal modeling assumptions. Specifically, for studies with errors-in-covariates, the approach does not require modeling the second moment of the surrogate. This is useful in studies where knowledge about the variance of the surrogate covariate is incomplete or absent, which frequently occurs in practice, when internal replicates or validation data either do not exist or are very limited. Currently the approach is limited to applications with a scalar mismeasured covariate, however.

Rathouz and Liang (2001) found that SOLAQS is “reasonably efficient”. Our simulation study of matched pair data suggests that the proposed OSOLAEF method with the efficient choice of weights, which accommodates an unknown dispersion parameter, is at least as efficient as the SOLAQS method. For studies with errors-in-covariates where the measurement error variance can be specified or well estimated, the OSOLAEF method with the efficient choice of weights has similar efficiency to
the SOLAQS method. However, when information about the measurement error variance is not available, SOLAQS method is not applicable and one would resort to the robust OSOLAEF method presented in this paper. A more in-depth study on the relative efficiency of OSOLAEF estimators is a subject of future research.

The nonquadratic type OSOLAEF presented in Section 3.2 requires the following condition: $\mu_{i1}''/\mu_{i1}' - \mu_{i2}''/\mu_{i2}' \neq 0$. Hence, the proposed method does not apply in the situation where, for example, the mean models for both $Y_{i1}$ and $Y_{i2}$ are linear in $\lambda_i$. For instance, in a linear errors-in-covariates problem where $E(Y_i) = \theta_0 + \lambda_i \theta_1 + z_{i2} \theta_2$, $E(X_i) = \lambda_i$, the proposed nonquadratic OSOLAEF method is not available since $\mu_{i1}'' = \mu_{i2}'' = 0$. In this special case, note that a two-dimensional unbiased estimating function that is free of nuisance parameter $\lambda_i$ is given by $g(\theta) = \sum g_i(\theta) = \sum_i (1, z_{i2})(Y_i - \theta_0 - \theta_1 X_i - \theta_2 z_{i2})$, which presumably can be augmented with a scalar, quadratic OSOLAEF of type (15) to form a rank 3 system of estimating functions for interest parameters $\theta = (\theta_0, \theta_1, \theta_2)$, as in (16). In addition, an important regularity condition for the asymptotic results in Conclusions 1 and 2 is that $E(\partial g_{2i}/\partial \theta) \neq 0$, which ensures $g_{2i}$ is sensitive to $\theta$ in some sense. One needs to make sure that this condition is satisfied when choosing the weights in $g_{2i}$.

Simulation studies indicate that the choice of weights in $g_{2i}$ and $g_{2i}^\dagger$ may affect the precision and, to a lesser extent, the bias, of the proposed estimator. We recommend the efficient choice of weights if available. If the efficient choice is not available in Application 2 and one faces an identifiability problem, then we recommend using the mixed system of equations with the weights presented in Section 3.3 and illustrated in Examples 2 and 3.

Godambe (1991) defined a certain type of orthogonality property, $E(gh) = 0$,
when $g$ and $h$ are what he terms ‘extended’ quasi-score functions, and further noted that $g(\theta, \hat{\lambda}_\theta)$, where $\hat{\lambda}_\theta$ is the root of $h$ given $\theta$, achieves large-sample approximate optimality (Godambe, 1991). However, Godambe (1991)’s approach is limited to extended quasi-score functions, requires correct specification of $\text{var}(g)$, $\text{var}(h)$ and $\text{cov}(g, h)$, and his asymptotic result applies only in the case where there are a bounded number of nuisance parameters as $n \to \infty$, in contrast to the infinite number of nuisance parameters’ problem posed in our Applications 1 and 2.

**Acknowledgements**

The authors would like to thank Dr. Raymond J. Carroll for kindly providing the coronary heart disease data. We are also grateful to the editor, an associate editor and three referees for their helpful comments that improved the paper. This research was partially supported by a grant from the U.S. National Institutes of Health.

**Appendix**

*An analytical estimator of the variance of $\hat{\theta}$ obtained from the OSOLAEFs presented in Sections 3.2 and 3.3.* For notational simplicity, we assume $\theta$ is a scalar. It is straightforward to extend the results to the cases where $\theta$ is a vector. Under either the asymptotic scenario defined in Conclusion 1 or 2, applying Taylor expansion, we have $-\dot{\gamma}_2 = \left( \frac{d\dot{\gamma}_2}{d\theta} \right) (\hat{\theta} - \theta) + \text{Rem}$, where we assume the orders of the second and the subsequent derivatives of $\hat{\gamma}_2$ are the same as or lower than that of $d\dot{\gamma}_2/d\theta$, and we use $\text{Rem}$ to represent the remaining lower order terms. Since $d\dot{\gamma}_2/d\theta = \text{E}(\partial \gamma_2/\partial \theta) + \text{Rem}$,

$$\hat{\theta} - \theta = - \left( \frac{d\dot{\gamma}_2}{d\theta} \right)^{-1} \dot{\gamma}_2 + \text{Rem} = -E^{-1} \left\{ \frac{\partial \gamma_2}{\partial \theta} \right\} g_2 + \text{Rem}. \quad (20)$$
The details are in a technical report available through the first author. It follows that \( \hat{\theta} \) has asymptotic variance \( \text{var}(g_2)/E^2(\partial g_2/\partial \theta) \), where, unfortunately, \( \text{var}(g_2) \) depends on the variance models of both \( Y_{i1} \) and \( Y_{i2} \). This asymptotic variance can be estimated at \( \theta = \hat{\theta} \) and \( \lambda_i = \hat{\lambda}_{i\theta} \) in practice.

**Proof of Conclusion 2.** We assume \( t_{2i} \neq 0 \) and \( \beta = 0 \); that is, \( t_{2i}/t_{1i} = O(\phi_1) \). The proofs when \( t_{2i} = 0 \) or \( \beta \neq 0 \) are similar. Note that \( h_i \) is proportional to \( h_i^\dagger = h_{i1} + h_{i2} \), where \( h_{i1} = t_{3i}(Y_{i1} - \mu_{i1})/\text{var}(Y_{i1}) \) and \( h_{i2} = t_{4i}(Y_{i2} - \mu_{i2}) \), with \( t_{3i} \) and \( t_{4i} \) of order \( O(1) \). The arguments of Lemmas B.1. and B.2. in Rathouz and Liang (2001) are still valid when replacing \( T_{1y} \) and \( T_{1x} \) by \( h_{i2} \) and \( h_{i1} \), respectively, and \( \phi_x \) by \( \phi_1 \). Hence,

\[
\hat{\lambda}_{i\theta} - \lambda_i = \frac{h_i^\dagger}{-E(h_i)^{\prime}} + Z_1 + Z_2 = O_p(\phi_1^{1/2})
\]

where \( Z_1 = Z_1(Y_{i1}) = O_p(\phi_1) \) and \( Z_2 = Z_2(Y_{i1}, Y_{i2}) = O_p(\phi_1^{3/2}) \).

By Taylor expansion, there exists \( \lambda_i^* \) such that \( |\lambda_i^* - \lambda_i| \leq |\hat{\lambda}_{i\theta} - \lambda_i| \), and

\[
\hat{g}_{2i} - g_{2i} = (\hat{\lambda}_{i\theta} - \lambda_i)g_2' + \frac{1}{2}(\hat{\lambda}_{i\theta} - \lambda_i)^2g_2'' + \frac{1}{6}(\hat{\lambda}_{i\theta} - \lambda_i)^3g_2'''(\lambda_i^*).
\]

Denote the three terms on the right-hand side as \( A_i \), \( B_i \) and \( C_i \), respectively.

For \( g_{2i} \) in (4), we have \( g_2' = q_0 + q_1(Y_{i1} - \mu_{i1}) + q_2(Y_{i2} - \mu_{i2}) + q_3(Y_{i1} - \mu_{i1})(Y_{i2} - \mu_{i2}) \), where \( q_j, j = 0, \ldots, 3 \), are non-random. Since \( E(g_2') = 0, q_0 = 0 \). It follows that

\[
g_2' = q_2(Y_{i2} - \mu_{i2}) + O_p(\phi_1^{1/2+\alpha}) = O_p(\phi_1^0).
\]

Therefore,

\[
A_i = \frac{h_i^\dagger}{-E(h_i)^{\prime}} + Z_1(Y_{i1})q_2(Y_{i2} - \mu_{i2}) + O_p(\phi_1^{3/2+\alpha}),
\]

where the first term is of \( O_p(\phi_1^{1/2+\alpha}) \) and mean-zero, the second term is of \( O_p(\phi_1^{1+\alpha}) \) and mean-zero due to independence of \( Y_{i1} \) and \( Y_{i2} \).
Since $E(g''_{2i}) = 0$, we have $g''_{2i} = q_i(Y_{i2} - \mu_i) + O_p\{\phi_1^{(1/2+\alpha)}\} = O_p(\phi_1^\alpha)$, where $q_i$ is non-random and of $O(\phi_1^\alpha)$. It follows that $B_i = q_i(Y_{i2} - \mu_i)h_i^2/E^2(h_i^*)/2 + O_p\{\phi_1^{(3/2+\alpha)}\}$, where the first term is mean-zero and of $O_p\{\phi_1^{(1+\alpha)}\}$.

We have $g'''_{2i}(\lambda_i) = O_p(\phi_1^\alpha)$. Under smoothness of $g'''_{2i}$ in $\lambda_i$, $g'''_{2i}(\lambda^*) = O_p(\phi_1^\alpha)$. It follows from $(\hat{\lambda}_i - \lambda_i)^3 = O_p(\phi_1^{3/2})$ that $C_i = O_p\{\phi_1^{(3/2+\alpha)}\}$.

It follows that $\phi_1^{-\alpha}(\hat{g}_2 - g_2) = f_{i1} + f_{i2}$, where $f_{i1} = O_p(\phi_1^{1/2})$, $f_{i2} = O_p(\phi_1^{3/2})$ and $E(f_{i1}) = 0$, as $\phi_1 \to 0$. Under uniform integrability, $E(f_{i2}) = O(\phi_1^{3/2})$. Using similar arguments to those in the Appendix of Sartori (2003), under mild regularity conditions, $\sum_{i=1}^n f_{ij} = O\{\sum_{i=1}^n E(f_{ij})\} + O_p\{\sum_{i=1}^n \text{var}(f_{ij})\}^{1/2},$ for $j = 1, 2$. It follows that $\sum_{i=1}^n f_{i1} = O_p(n^{1/2}\phi_1^{1/2})$ and $\sum_{i=1}^n f_{i2} = O_p(n\phi_1^{3/2})$. The result for $\hat{\theta} - \theta$ follows from (20).

Proof of Conclusion 1. We still have $\hat{\lambda}_i - \lambda_i = O_p(\phi_1^{1/2})$. But now all of $g_2$, $g'_{2i}$ and $g''_{2i$ are at the order of $O_p(\phi_1^\alpha)$. Plugging in the order of each term in formula (21), we have $\phi_1^{-\alpha}(\hat{g}_2 - g_2) = O_p(\phi_1^{1/2})$. Under uniform integrability, since $E(g'_2\hat{h}_i) = 0$, $E\{\phi_1^{-\alpha}(\hat{g}_2 - g_2)\} = O(\phi_1)$. The result in Conclusion 1 follows from similar arguments to those at the end of the proof of Conclusion 2.

Proof of (6). Plugging $\hat{\lambda}_i - \lambda_i = -h_i/E(h_i') + O_p(\phi_1) = O_p(\phi_1^{1/2})$ into (21), we have $\hat{g}_2 - g_2 = -h_i g'_2/E(h_i') + g'_2 O_p(\phi_1) + g''_{2i} O_p(\phi_1) + O_p(\phi_1)$. Formula (6) follows from $E(h_i') = O_p(\phi_1^{-1/2})$ and $g''_{2i} - E(g''_{2i}) = O_p(1)$.

References


Fig.1. Simulation study of estimators of $\theta_1$ for matched pair model based on 1000 replicates. $\lambda_i$ was generated from $\text{unif}(0.1+\theta_0, 1+\theta_0)$. Number of strata $n=400$. The long horizontal lines are at the true values of $\theta_1$. The short horizontal bars are at the averages of the $\theta_1$-estimates over all simulated replicates, obtained using the five methods PQS, AQS-1, AQS-2, OEF-1 and OEF-2, respectively. The average values of the estimates are also written beside the vertical bars. Lengths of the solid vertical lines represent values of empirical standard errors, and the dotted vertical lines are for the averages of bootstrap standard errors over all simulated replicates.
Table 1. Simulation Study for Errors-in-Covariates Logistic Regression Model, 1000 Replicates.

<table>
<thead>
<tr>
<th></th>
<th>PQS</th>
<th>OEF-1</th>
<th>OEF-2</th>
<th>AQS</th>
<th>SIMEX</th>
<th>OEF-2</th>
<th>AQS</th>
<th>SIMEX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(φ₀ unknown)</td>
<td>(φ₀ = 0.1)</td>
<td>(φ₀ = 0.1)</td>
<td>(φ₀ = 0.13)</td>
<td>(φ₀ = 0.13)</td>
<td>(φ₀ = 0.13)</td>
<td>(φ₀ = 0.13)</td>
</tr>
<tr>
<td>θ₀</td>
<td>1.0</td>
<td>1.01(0.43)</td>
<td>1.01(0.44)</td>
<td>1.02(0.43)</td>
<td>1.02(0.43)</td>
<td>1.02(0.43)</td>
<td>1.03(0.43)</td>
<td>1.03(0.43)</td>
</tr>
<tr>
<td>θ₁</td>
<td>-0.5</td>
<td>-0.46(0.20)</td>
<td>-0.48(0.32)</td>
<td>-0.51(0.22)</td>
<td>-0.51(0.22)</td>
<td>-0.53(0.23)</td>
<td>-0.53(0.23)</td>
<td>-0.52(0.22)</td>
</tr>
<tr>
<td>θ₂</td>
<td>0.5</td>
<td>0.51(0.37)</td>
<td>0.52(0.37)</td>
<td>0.51(0.37)</td>
<td>0.51(0.37)</td>
<td>0.53(0.37)</td>
<td>0.53(0.38)</td>
<td>0.53(0.37)</td>
</tr>
<tr>
<td>θ₀</td>
<td>1.0</td>
<td>1.02(0.40)</td>
<td>1.02(0.42)</td>
<td>1.02(0.40)</td>
<td>1.02(0.40)</td>
<td>1.02(0.40)</td>
<td>1.02(0.40)</td>
<td>1.02(0.40)</td>
</tr>
<tr>
<td>θ₁</td>
<td>0.5</td>
<td>0.47(0.19)</td>
<td>0.49(0.31)</td>
<td>0.52(0.22)</td>
<td>0.52(0.22)</td>
<td>0.54(0.22)</td>
<td>0.54(0.22)</td>
<td>0.53(0.22)</td>
</tr>
<tr>
<td>θ₂</td>
<td>0.5</td>
<td>0.51(0.34)</td>
<td>0.51(0.36)</td>
<td>0.51(0.34)</td>
<td>0.51(0.34)</td>
<td>0.52(0.34)</td>
<td>0.52(0.36)</td>
<td>0.51(0.34)</td>
</tr>
<tr>
<td>θ₀</td>
<td>1.0</td>
<td>0.97(0.39)</td>
<td>0.99(0.44)</td>
<td>1.01(0.40)</td>
<td>1.01(0.40)</td>
<td>1.00(0.40)</td>
<td>1.02(0.40)</td>
<td>1.02(0.40)</td>
</tr>
<tr>
<td>θ₁</td>
<td>1.0</td>
<td>0.93(0.21)</td>
<td>0.98(0.47)</td>
<td>1.05(0.25)</td>
<td>1.05(0.22)</td>
<td>1.03(0.25)</td>
<td>1.09(0.27)</td>
<td>1.09(0.27)</td>
</tr>
<tr>
<td>θ₂</td>
<td>0.5</td>
<td>0.54(0.33)</td>
<td>0.54(0.37)</td>
<td>0.53(0.34)</td>
<td>0.53(0.34)</td>
<td>0.53(0.34)</td>
<td>0.53(0.34)</td>
<td>0.53(0.36)</td>
</tr>
</tbody>
</table>

NOTE: Number of strata n = 200. The second column presents the parameter values used to generate data. Displayed in columns 3-10 are the average values of the estimates over 1000 replicates, along with the empirical standard errors in parentheses.
### Table 2. Results for Framingham data

<table>
<thead>
<tr>
<th></th>
<th>PQS</th>
<th>OEF-1</th>
<th>OEF-2</th>
<th>AQS</th>
<th>SIMEX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>s(\hat{e})</td>
<td>Coeff</td>
<td>s(\hat{e})</td>
<td>Coeff</td>
</tr>
<tr>
<td>LSBP</td>
<td>2.055</td>
<td>0.247</td>
<td>3.314</td>
<td>0.616</td>
<td>2.837</td>
</tr>
<tr>
<td>SMOKE</td>
<td>0.471</td>
<td>0.197</td>
<td>0.543</td>
<td>0.205</td>
<td>0.523</td>
</tr>
</tbody>
</table>

Note: s\(\hat{e}\) is bootstrap standard error.