Nonparametric correction to errors in covariates

Yijian Huang and C. Y. Wang

Fred Hutchinson Cancer Research Center

Abstract: With accurate covariates unascertainable, existing approaches to account for errors in covariates generally require distributional assumptions on the errors, if not in addition to those on the unobserved true covariates. In this article, we propose a nonparametric method without assumptions on both of them. Our procedure uses replicated mismeasured covariates and exploits their independence structure to construct corrected estimating functions. The resulting estimator is consistent and asymptotically normal. This methodology is illustrated through commonly used generalized linear models. Numerical results for Poisson regression indicate that the proposed estimator has good efficiency as compared to the corrected-score estimator, when the error assumption required for the latter holds. Further, we show that the proposed methodology can be extended to instrumental variable estimation.

Key words and phrases: Corrected score; Estimating function; Generalized linear model; Instrumental variable; Measurement error; Nonlinear model; Poisson regression; Replication data

1. Introduction

Many medical applications involve covariates that are not accurately ascertainable; common examples include CD4 lymphocyte count and viral load in HIV/AIDS research and systolic blood pressure in cardiovascular disease studies. Typically, such a measurement is subject to both error from the instrument and biological diurnal fluctuation, with the intended covariate being the average of an underlying process during a certain period of time. As is well known, ignoring errors in covariates and naively adopting conventional

inference procedures may result in substantial bias in regression analysis. Statistical issues and recent developments on this topic have been well summarized in two monographs by Fuller (1987) and Carroll, Ruppert and Stefanski (1995); the former focuses on linear models and the latter on nonlinear ones. However, existing approaches to consistent estimation for nonlinear models generally require distributional assumptions on the errors, if not in addition to those on the underlying true covariates. In this article we propose a method to yield consistent estimators without assumptions on both of them, using replication data on the mismeasured covariates.

Several functional methods have been proposed for a few generalized linear models; they do not require distributional assumptions on the underlying true covariates (Carroll et al., 1995). Stefanski and Carroll (1987) suggested the conditional score function on the basis of certain sufficient statistics. However, as pointed out by these authors, the expectation of such an estimating function is not guaranteed to admit a unique zero-crossing and thus the resulting estimator may not converge. Stefanski (1989) and Nakamura (1990) proposed the corrected score function which has the same expectation as the likelihood score function based on the true covariates; see also Buonaccorsi (1996) for related work. They showed that the corrected-score estimator is consistent under certain regularity conditions. Note that both conditional- and corrected-score approaches require distributional assumptions on the measurement error. In the literature, normal distribution is often adopted.

The situation under consideration precludes the availability of a validation subset, which is regarded as the ideal data source for measurement error analysis (Carroll et al., 1995, p. 12). Instead, the commonly attainable are replicated mismeasured covariate data. Often, they are employed to obtain the replicate average, as a less error-contaminated surrogate covariate, and/or the (estimated) error variance, to which methods such as the corrected-score approach are subsequently applied. In this case, fully parametric assumption is required on the error distribution except for finitely many pa-

rameters. As one exception, Buzas (1997) showed that unbiased estimating functions can be constructed with weaker error assumption. Nevertheless, the errors are still required to be zero-unbiased and symmetric in distribution.

In this article, we show that replicated mismeasured covariate data can be more effectively used to construct estimating functions that yield consistent estimators, without distributional assumption on both the true covariates and the measurement errors. We term our proposal the *nonparametric-correction approach*. Similarly to the corrected-score approach, we start with an estimating function based on the underlying true covariates; that is, the *original estimating function* as is called. One difference, however, is that we do not necessarily restrict to the score function. Most importantly, our procedure performs correction without distributional assumption on the errors. To contrast, we shall refer to the corrected-score method (Stefanski, 1989; Nakamura, 1990) as the *parametric-correction approach*. In Section 2, we introduce some basic results and construction techniques on estimating functions, with focus on the original estimating functions. In the presence of covariate measurement error, we propose the nonparametric-correction approach in Section 3. Numerical results for Poisson regression are presented in Section 4 to illustrate our methodology and compare with the parametric-correction approach. Further remarks are made and extensions of the proposal discussed in Section 5. Specifically, its relationship to other existing methods, particularly Buzas (1997), is examined. Also, we show that our methodology can be applied to instrumental variable estimation.

2. Preliminaries on estimating function

We are concerned with estimating functions that are sums of independent and identically distributed components. In this section, we introduce a few basic results, focusing on generalized linear models (in the absence of measurement error). The intention here is not to be comprehensive but rather to set the stage for our later development.

Suppose that $\mathcal{D}$ is a family of random variables, $D_{\boldsymbol{\gamma}_0} \in \mathcal{D}$ is a member indexed by

$\boldsymbol{\gamma}_0 \in \boldsymbol{\Gamma}$, and $\boldsymbol{\Gamma}$ is the parameter space of interest. Given that nuisance parameters might exist, $\boldsymbol{\gamma}_0$ may not fully determine the distribution of $D_{\boldsymbol{\gamma}_0}$. Let data $\{d_i : i = 1, \cdots, n\}$ be n independent realizations of $D_{\boldsymbol{\gamma}_0}$. A typical estimating function for $\boldsymbol{\gamma}_0$ can be written as of the normalized form

$$\widehat{\mathcal{E}}\{\boldsymbol{\rho}(.; D_{\boldsymbol{\gamma}_0})\} \equiv \frac{1}{n} \sum_{i=1}^{n} \{\boldsymbol{\rho}(.; d_i)\} \tag{1}$$

for some *estimating kernel* $\boldsymbol{\rho}(.;.)$, where $\widehat{\mathcal{E}}$ denotes the empirical mean. Apparently, (normalized) score and quasi-score functions are special cases. An estimator for $\boldsymbol{\gamma}_0$ is a zero-crossing of the estimating function, if attainable; in case of multiple zero-crossings, the estimator can be any one of them. We define a *root-consistent estimating function* as such that every zero-crossing is consistent. Denote expectation by $\mathcal{E}$.

Lemma 1. Let $\boldsymbol{\Gamma}$ be compact. For every $\boldsymbol{\gamma}_0 \in \boldsymbol{\Gamma}$, suppose that $\mathcal{E}\{\boldsymbol{\rho}(\boldsymbol{\gamma}; D_{\boldsymbol{\gamma}_0})\}$ exists, as being continuous in $\boldsymbol{\gamma}$, and, in probability,

$$\sup_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} \left| \widehat{\mathcal{E}}\{\boldsymbol{\rho}(\boldsymbol{\gamma}; D_{\boldsymbol{\gamma}_0})\} - \mathcal{E}\{\boldsymbol{\rho}(\boldsymbol{\gamma}; D_{\boldsymbol{\gamma}_0})\} \right| \to \mathbf{0}. \tag{2}$$

Then, $\widehat{\mathcal{E}}\{\boldsymbol{\rho}(.; D_{\boldsymbol{\gamma}_0})\}$ is root-consistent if and only if $\mathcal{E}\{\boldsymbol{\rho}(.; D_{\boldsymbol{\gamma}_0})\}$ admits a unique zero-crossing at $\boldsymbol{\gamma}_0$.

Proof. This assertion follows Theorems 2.1 and 3.1 of Crowder (1986). □

The result is indeed rather self-evident. Note that the estimator does not converge if $\mathcal{E}\{\boldsymbol{\rho}(.; D_{\boldsymbol{\gamma}_0})\}$ admits multiple zero-crossings. In the literature, the notion of *unbiased estimating function* is often encountered. However, an unbiased estimating function is not necessarily root-consistent.

Note that condition (2) is the uniform consistency of $\widehat{\mathcal{E}}\{\boldsymbol{\rho}(.; D_{\boldsymbol{\gamma}_0})\}$. As well known, for any fixed $\boldsymbol{\gamma}$, the law of large numbers asserts the pointwise consistency. With monotone estimating functions, including most score and quasi-score functions, the uniform consistency then follows from convex function theory (cf. Andersen and Gill, 1982, Theorem II.1). However, non-monotone estimating functions will be of concern in this article.

4

In this case, mild regularity conditions are needed and the reader is referred to, for example, Crowder (1986). To focus on main ideas, throughout this article we assume that the uniform consistency holds under sufficient regularity conditions.

Now consider a regression model. Let $\mathcal{Z}$ be a class of scalar random variables indexed by scalar $\theta_0 \in \Theta$ for parameter space Θ. Suppose that Y is a scalar dependent variable and $\mathbf{X}$ is a covariate vector. A generalized linear model is often formulated as such that, conditional on $\mathbf{X}$, Y is a random variable belonging to $\mathcal{Z}$ with index

$$\theta_0(\mathbf{X}) = \alpha_0 + \boldsymbol{\beta}_0^T \mathbf{X}, \tag{3}$$

where α_0 is a scalar intercept and $\boldsymbol{\beta}_0$ is a slope vector. The regression coefficients $(\alpha_0, \boldsymbol{\beta}_0^T)^T$ are the parameters of interest, particularly the slope vector $\boldsymbol{\beta}_0$. To estimate $(\alpha_0, \boldsymbol{\beta}_0^T)^T$, one can take advantage of the structure of the generalized linear model (3) and build upon estimating functions for $\theta_0 \in \Theta$.

Lemma 2. Suppose that scalar function $\phi(.;.)$ is the estimating kernel of a root-consistent estimating function for $\theta_0 \in \Theta$. Further, $\delta(\theta, \theta_0) \equiv \mathcal{E}\{\phi(\theta; Z_{\theta_0})\}$, for $Z_{\theta_0} \in \mathcal{Z}$ with index θ_0, satisfies that $\delta(\theta, \theta_0)(\theta - \theta_0) \neq 0$ is of the same sign for all $\theta, \theta_0 \in \Theta$ such that $\theta \neq \theta_0$. Under the generalized linear model (3) and subject to regularity conditions, estimating function

$$\widetilde{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta}) \equiv \widehat{\mathcal{E}} \left\{ \phi(\alpha + \boldsymbol{\beta}^T \mathbf{X}; Y) \begin{pmatrix} 1 \\ \mathbf{X} \end{pmatrix} \right\} \tag{4}$$

is root-consistent for $(\alpha_0, \boldsymbol{\beta}_0^T)^T$ if the variance matrix of $\mathbf{X}$ is positive definite.

Proof. Note

$$\mathcal{E}\{\widetilde{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})\}^T \begin{pmatrix} \alpha - \alpha_0 \\ \boldsymbol{\beta} - \boldsymbol{\beta}_0 \end{pmatrix} = \mathcal{E}\{\delta(\alpha + \boldsymbol{\beta}^T \mathbf{X}, \alpha_0 + \boldsymbol{\beta}_0^T \mathbf{X})(\alpha + \boldsymbol{\beta}^T \mathbf{X} - \alpha_0 - \boldsymbol{\beta}_0^T \mathbf{X})\}.$$

Under the given conditions, the above expression is 0 if and only if $(\alpha, \boldsymbol{\beta}^T)^T = (\alpha_0, \boldsymbol{\beta}_0^T)^T$. Therefore, $\mathcal{E}\{\widetilde{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})\}$ has a unique zero-crossing at $(\alpha_0, \boldsymbol{\beta}_0^T)^T$. The desired result then follows from Lemma 1. □

Lemma 2 presents an estimating-function construction technique for the generalized linear model (3). Meanwhile, note that an estimating function with form (4), as being rather typical, has a desirable property: A linear reparameterization of $(\alpha_0, \boldsymbol{\beta}_0^T)^T$ in (3) results in the same linear transformation of the corresponding estimators.

In anticipation of the development in next section, we now restrict our attention to two classes of functions for $\phi(.;.)$; namely, $\phi_{\mathrm{I}}(\theta; z) = \sum_{k=0}^{K} \varphi_k(z)\theta^k$ and $\phi_{\mathrm{II}}(\theta; z) = \sum_{k=0}^{K} \varphi_k(z)\exp(ck\theta)$ for some functions $\varphi_k(.)$, $k = 0, \cdots, K$, and constants K and c. Correspondingly,

$$
\widetilde{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta}) =
\begin{cases}
\widehat{\mathcal{E}}\left\{ \sum_{k=0}^{K} \varphi_k(Y)(\alpha + \boldsymbol{\beta}^T\mathbf{X})^k \begin{pmatrix} 1 \\ \mathbf{X} \end{pmatrix} \right\} & \text{Class I} \\
\widehat{\mathcal{E}}\left[\sum_{k=0}^{K} \varphi_k(Y)\exp\{ck(\alpha + \boldsymbol{\beta}^T\mathbf{X})\} \begin{pmatrix} 1 \\ \mathbf{X} \end{pmatrix} \right] & \text{Class II}
\end{cases}
. \tag{5}
$$

These two classes cover commonly used estimating functions for several important generalized linear models, including the score functions of Normal, Poisson, Gamma, and inverse Gaussian models as studied in Nakamura (1990).

Example 1. Normal regression model.

Conditional on $\mathbf{X}$, $Y \sim$ Normal with mean $\theta_0(\mathbf{X}) = \alpha_0 + \boldsymbol{\beta}_0^T\mathbf{X}$. The score function takes $\phi(\theta; z) = z - \theta$, belonging to Class I with $K = 1$.

Example 2. Poisson regression model.

Conditional on $\mathbf{X}$, $Y \sim$ Poisson with mean $\exp\{\theta_0(X)\} = \exp(\alpha_0 + \boldsymbol{\beta}_0^T\mathbf{X})$. The score function corresponds to $\phi(\theta; z) = z - \exp(\theta)$, as a member of Class II with $K = 1$ and $c = 1$.

As seen, root-consistent estimating functions of form (5) are readily available for these generalized linear models. Referred to as the original estimating functions, they will be our building blocks for estimation in the presence of errors in covariates.

3. Nonparametric correction to measurement error

In many practical situations, some of the covariates $\mathbf{X}$ are not accurately measurable and are subject to measurement error. Write $\mathbf{X} = (\mathbf{X}_e^T, \mathbf{X}_a^T)^T$ for error-prone $\mathbf{X}_e$ and accurately measured $\mathbf{X}_a$. We do not observe $\mathbf{X}_e$ directly but through its surrogate $\mathbf{W}$. Under the additive measurement error model,

$$\mathbf{W} = \mathbf{X}_e + \boldsymbol{\varepsilon}, \tag{6}$$

where $\boldsymbol{\varepsilon}$ is independent of $\mathbf{X}$, as well as Y. Note that elements in $\boldsymbol{\varepsilon}$ may be mutually dependent. Group $\mathbf{X}_e$ into, say M, clusters, $\mathbf{X}_e \equiv (\mathbf{X}_1^T, \cdots, \mathbf{X}_M^T)^T$, such that the corresponding measurement errors $\boldsymbol{\varepsilon} \equiv (\boldsymbol{\varepsilon}_1^T, \cdots, \boldsymbol{\varepsilon}_M^T)^T$ are independent across these clusters, but possibly dependent within each one of them. Note that replication data on the surrogates are available at the level of dependence cluster. Corresponding to each $\mathbf{X}_m$, $m = 1, \cdots, M$, finite R_m surrogates $\mathbf{W}_{mj} = \mathbf{X}_m + \boldsymbol{\varepsilon}_{mj}$, $j = 1, \cdots, R_m$, are observed. These errors, $\boldsymbol{\varepsilon}_{mj}$, are independent and identically distributed replicates of $\boldsymbol{\varepsilon}_m$. The number of replications, R_m, $m = 1, \cdots, M$, may be random and can be mutually dependent, as well as dependent on $\mathbf{X}$ and Y. With respect to the original estimating function (5), we shall require $R_m \geq K + 1$. For the Normal and Poisson regression models, $K = 1$ and the number of replicates R_m can be so small as 2. Suppose that the data sample consists of n independent realizations of $\{Y, \mathbf{X}_a, \mathbf{W}_{mj} : j = 1, \cdots, R_m, m = 1, \cdots, M\}$.

To estimate $(\alpha_0, \boldsymbol{\beta}_0^T)^T$, Lemma 1 indicates that a desirable estimating function would have its limit admitting a unique zero-crossing at $(\alpha_0, \boldsymbol{\beta}_0^T)^T$. Intuitively the original estimating function, possessing such a property, may serve as a building block. Stefanski (1989) and Nakamura (1990) suggested the parametric-correction approach along this line. In the following, we relieve the distributional assumption requirement on the measurement error and propose the nonparametric-correction approach.

From the replication data, pick $K + 1$ replicates for each dependence cluster, $m = 1, \cdots, M$, to form $K + 1$ surrogate covariate vectors $\mathbf{W}^{(r)} : r = 1, \cdots, K + 1$. With $\mathbf{X}_e$

7

unobserved, replacing it with its surrogates seems to be a natural approach to modifying the original estimating function (5). However, the challenge is to make a sensible replacement. We proceed under the guidance from expectation of the original estimating function. Note that the particular difficulty is on items in (5) with multiple appearances of $\mathbf{X}_e$; naively replacing them with the same surrogate would render an intractable expectation (conditioning on $\mathbf{X}$ and Y). This problem is due to that expectation is in general not interchangeable with multiplication of random variables. The crucial step of our proposal is to recognize and take advantage of one exception when these random variables are independent. We suggest to use the conditionally independent surrogate replicates (given $\mathbf{X}$ and Y) in place of the multiple appearances of $\mathbf{X}_e$ in a product. The resulting nonparametrically corrected estimating function is given as,

$$
\widehat{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta}) = \begin{cases} \widehat{\mathcal{E}}\left\{ \sum_{k=0}^{K} \varphi_k(Y) \prod_{r=1}^{k} (\alpha + \boldsymbol{\beta}_e^T \mathbf{W}^{(r)} + \boldsymbol{\beta}_a^T \mathbf{X}_a) \begin{pmatrix} 1 \\ \mathbf{W}^{(K+1)} \\ \mathbf{X}_a \end{pmatrix} \right\} & \text{Class I} \\ \widehat{\mathcal{E}}\left[\sum_{k=0}^{K} \varphi_k(Y) \prod_{r=1}^{k} \exp\{c(\alpha + \boldsymbol{\beta}_e^T \mathbf{W}^{(r)} + \boldsymbol{\beta}_a^T \mathbf{X}_a)\} \begin{pmatrix} 1 \\ \mathbf{W}^{(K+1)} \\ \mathbf{X}_a \end{pmatrix} \right] & \text{Class II} \end{cases} ,
$$
(7)

where $\prod_{r=1}^{0} . \equiv 1$. Note that the surrogate vectors $\mathbf{W}^{(r)}$, $r = 1, \cdots, K+1$ are chosen arbitrarily from the replication data and $\prod_{m=1}^{M} R_m!/(R_m - K - 1)!$ different sets can be formed. For definitiveness and efficiency, we have specified that the empirical expectation $\widehat{\mathcal{E}}$ first operates on each individual to average over these different replication sets, and then takes the empirical mean over the individuals in the sample. As one might have noticed, the expectation of the nonparametrically corrected estimating function (7) is, however, not the same as that of the original one (5). Nevertheless, the following result shows that it yields a consistent estimator for $\boldsymbol{\beta}_0$. Let $(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}^T)^T$ be a zero-crossing of $\widehat{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})$.

8

Theorem 1. Denote the parameter spaces of interest by compact sets $\mathcal{A}$ and $\mathcal{B}$, where $(\alpha_0, \boldsymbol{\beta}_0^T)^T \in \mathcal{A} \times \mathcal{B}$. Suppose that the original estimating function $\widetilde{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})$ in (5) is root-consistent for $(\alpha_0, \boldsymbol{\beta}_0^T)^T$. Subject to regularity conditions, nonparametrically corrected estimating function $\widehat{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})$, given in (7), is root-consistent for $(\alpha_*, \boldsymbol{\beta}_0^T)^T$ under the additive measurement error model (6), where

$$\alpha_* \equiv \alpha_0 - \begin{cases} \boldsymbol{\beta}_{e0}^T \mathcal{E}(\boldsymbol{\varepsilon}) & \text{Class I} \\ c^{-1} \log[\mathcal{E}\{\exp(c\boldsymbol{\beta}_{e0}^T \boldsymbol{\varepsilon})\}] & \text{Class II} \end{cases}.$$

Furthermore, $n^{1/2}\{(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}^T)^T - (\alpha_*, \boldsymbol{\beta}_0^T)^T\}$ is asymptotically normal with mean $\mathbf{0}$.

Proof. Write

$$g(\boldsymbol{\beta}_e) \equiv \begin{cases} \boldsymbol{\beta}_e^T \mathcal{E}(\boldsymbol{\varepsilon}) & \text{Class I} \\ c^{-1} \log[\mathcal{E}\{\exp(c\boldsymbol{\beta}_e^T \boldsymbol{\varepsilon})\}] & \text{Class II} \end{cases}.$$

Based on the independence structure of the replicated surrogate covariates, algebraic derivation shows that

$$\mathcal{E}\{\widehat{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})\} = \mathcal{E}\left[\phi\{\alpha + g(\boldsymbol{\beta}_e) + \boldsymbol{\beta}^T \mathbf{X}; Y\} \begin{pmatrix} 1 \\ \mathbf{X}_e + \mathcal{E}(\boldsymbol{\varepsilon}) \\ \mathbf{X}_a \end{pmatrix} \right].$$

The root-consistency for $(\alpha_*, \boldsymbol{\beta}_0^T)^T$ is then established since the above right-hand side has the same zero-crossing with or without $\mathcal{E}(\boldsymbol{\varepsilon})$.

The asymptotic normality of $(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}^T)^T$ follows that of $\widehat{\boldsymbol{\Phi}}(\alpha_*, \boldsymbol{\beta}_0)$, by the central limit theorem, and the asymptotic linearity of $\widehat{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta})$ at $(\alpha_*, \boldsymbol{\beta}_0^T)^T$. The argument is standard and details omitted. □

Remarkably, the slope $\boldsymbol{\beta}_0$ is consistently estimated without any assumptions on the measurement error, even with allowance for unknown drift $\mathcal{E}(\boldsymbol{\varepsilon})$. This, however, is not true for α_0. In fact, α_0 is not identifiable unless $\mathcal{E}(\boldsymbol{\varepsilon})$ is known, which is evident from the linearity of regression model (3) and the additivity of measurement error model (6). This issue may contribute to the fact that α_0 is not so much of interest as $\boldsymbol{\beta}_0$ is in most practical situations.

In case that $\mathcal{E}(\varepsilon)$ is known, a consistent estimator for α_0 can be easily obtained for Class I. However, for Class II, the remaining difficulty is with estimating $\mathcal{E}\{\exp(\mathbf{b}^T\varepsilon)\}$ for known constant $\mathbf{b}$, as can be seen from the expression of α_*. Nevertheless, a fully parametric approach may not be necessary. For example, if the distribution of ε is symmetric, $\widehat{\mathcal{E}}^{1/2}[\exp\{\mathbf{b}^T(\mathbf{W}^{(1)} - \mathbf{W}^{(2)})\}]\exp\{\mathbf{b}^T\mathcal{E}(\varepsilon)\}$ is a consistent estimator of $\mathcal{E}\{\exp(\mathbf{b}^T\varepsilon)\}$.

Example 1. (Continued). Normal regression model.

The nonparametrically corrected estimating function is

$$\widehat{\mathbf{\Phi}}(\alpha,\boldsymbol{\beta}) = \widehat{\mathcal{E}}\left\{(Y - \alpha - \boldsymbol{\beta}_e^T\mathbf{W}^{(1)} - \boldsymbol{\beta}_a^T\mathbf{X}_a)\begin{pmatrix} 1 \\ \mathbf{W}^{(2)} \\ \mathbf{X}_a \end{pmatrix}\right\}.$$

It yields the nonparametric-correction estimator, given as

$$\left[\widehat{\mathcal{E}}\left\{\begin{pmatrix} 1 \\ \mathbf{W}^{(1)} \\ \mathbf{X}_a \end{pmatrix}\begin{pmatrix} 1 \\ \mathbf{W}^{(2)} \\ \mathbf{X}_a \end{pmatrix}^T\right\}\right]^{-1}\widehat{\mathcal{E}}\left\{Y\begin{pmatrix} 1 \\ \mathbf{W}^{(1)} \\ \mathbf{X}_a \end{pmatrix}\right\},$$

if the inverse exists. The estimator is consistent for $\{\alpha_0 - \boldsymbol{\beta}_{e0}^T\mathcal{E}(\varepsilon), \boldsymbol{\beta}_0^T\}^T$. Notice that

$$\widehat{\mathcal{E}}\left\{\begin{pmatrix} 1 \\ \mathbf{W}^{(1)} \\ \mathbf{X}_a \end{pmatrix}\begin{pmatrix} 1 \\ \mathbf{W}^{(2)} \\ \mathbf{X}_a \end{pmatrix}^T\right\} = \widehat{\mathcal{E}}\left\{\begin{pmatrix} 1 \\ \overline{\mathbf{W}} \\ \mathbf{X}_a \end{pmatrix}^{\otimes 2} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \overline{\mathbf{W}}^{\otimes 2} - \mathbf{W}^{(1)}\mathbf{W}^{(2)T} & 0 \\ 0 & 0 & 0 \end{pmatrix}\right\},$$

where $\overline{\mathbf{W}}$ is the average of the replicate surrogates and $\mathbf{v}^{\otimes 2} = \mathbf{v}\mathbf{v}^T$ for vector $\mathbf{v}$. Further, $\widehat{\mathcal{E}}\{\overline{\mathbf{W}}^{\otimes 2} - \mathbf{W}^{(1)}\mathbf{W}^{(2)T}\}$ is a consistent variance estimate of $\overline{\mathbf{W}}$; in case of $R_m = 2$ for all $m = 1,\cdots,M$, it reduces to $\widehat{\mathcal{E}}\{(\mathbf{W}^{(1)} - \mathbf{W}^{(2)})^{\otimes 2}\}/4$. As such, if $\mathcal{E}(\varepsilon) = \mathbf{0}$, the nonparametric-correction estimator coincides with the corrected-score estimator given by Nakamura (1990) using $\overline{\mathbf{W}}$ and the aforementioned variance estimate, as well as that in Fuller (1987, p. 5) among others. The fact indicates that the normality assumption in the corrected-score method with the Normal regression model does not play a role in the consistency of the estimator. However, it is well known that this phenomenon does not extend to nonlinear models in general.

Example 2. (Continued). Poisson regression model.

The nonparametrically corrected estimating function is

$$\widehat{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta}) = \widehat{\mathcal{E}} \left[\{ Y - \exp(\alpha + \boldsymbol{\beta}_e^T \mathbf{W}^{(1)} + \boldsymbol{\beta}_a^T \mathbf{X}_a) \} \begin{pmatrix} 1 \\ \mathbf{W}^{(2)} \\ \mathbf{X}_a \end{pmatrix} \right]$$

and the resulting estimator is consistent for $(\alpha_0 - \log[\mathcal{E}\{\exp(\boldsymbol{\beta}_{e0}^T \boldsymbol{\varepsilon})\}], \boldsymbol{\beta}_0^T)^T$.

4. NUMERICAL RESULTS FOR POISSON REGRESSION

In this section, we present numerical results for Poisson regression, which is widely used in the analysis of count data. This investigation also serves as an illustration of our methodology.

One issue of interest is on efficiency. Particularly, how are the nonparametric- and parametric-correction estimators compared to each other when the distributional assumptions on measurement error required for the latter indeed hold? We consider the Poisson regression model given in Example 2 with a single and error-prone covariate. Suppose that the true covariate X and error ε are both normal with mean 0, and with variances σ_X^2 and σ_ε^2, respectively. Further, two replicates of the surrogate $W = X + \varepsilon$ are available. One may use their average $\overline{W}$ to apply the parametric correction of Nakamura (1990). While the parametrically corrected score function involves σ_ε^2, we consider two scenarios, namely, with and without the knowledge of σ_ε^2. For the latter, we use the variance estimate $\widehat{\mathcal{E}}\{(W^{(1)} - W^{(2)})^2\}/2$ of the error ε instead. The asymptotic variances for the estimators of slope β_0 can be evaluated analytically; Table 1 shows the asymptotic efficiency of the nonparametric-correction estimator relative to the parametric-correction ones. Note that the efficiency is a function of α_0, $\beta_0 \sigma_X$, and $\sigma_\varepsilon / \sigma_X$. As expected, the nonparametric-correction, the parametric-correction with σ_ε^2 estimated, and the parametric-correction with σ_ε^2 known are in the order of increasing asymptotic efficiency. Our results also show that the relative efficiency of the nonparametric-correction estimator decreases with the increase of $\sigma_\varepsilon / \sigma_X$. Nevertheless, much surprisingly, the efficiency difference is remarkably

small, if any, under the settings of consideration.

To examine the finite-sample performance of our proposed estimator, simulation studies were conducted. The setting is the same as the earlier efficiency study unless otherwise specified. We set $\alpha_0 = 0$ and $\beta_0 = 1$ and chose the true covariate X to be uniformly distributed with mean 0 and variance 1. Both normal and non-normal distributions for the error ε were considered. Under the latter, the parametric-correction estimators are in general inconsistent; we present here a study with location-shifted and scale-changed Beta(3,1) distribution. For both the normal and Beta distributions, we specified zero mean but considered various standard deviations. Sample sizes of 200 and 400 were chosen. We studied five estimators for β_0: naive, regression calibration, parametric-correction with σ_ε^2 known, parametric-correction with σ_ε^2 estimated, and our proposed nonparametric-correction. The naive approach uses $\overline{W}$ in place of the true covariate in the standard estimation procedure, i.e. in the score function. As widely applied, the regression calibration approach reduces but in general does not eliminate bias from measurement error; we used the formulas given in Carroll et al. (1995, Section 3.4.2). Under each setting, 500 simulated data sets were generated and the empirical biases and variances of these estimators are presented in Table 2. Both the naive and regression calibration estimators suffer from bias though it is less serious for the latter. Not surprisingly, we also observe biases of the parametric-correction estimators when ε has the Beta distribution; they can be even worse than the regression calibration estimator when σ_ε is large. In comparison, the biases of the nonparametric-correction estimator are reasonably small for both the normal and Beta error distributions. Additionally, they in general decrease as the sample size increases. When the error distribution is normal, the empirical variances of our proposed estimator are close to those of the parametric-correction ones. This finding is in accordance with our earlier result from the asymptotic relative efficiency study.

Of note, for a finite sample, parametrically and nonparametrically corrected estimat-

ing functions may not have unique zero-crossings. In fact, there may exist multiple roots or none. In our study, we employed the Newton-Raphson algorithm for root finding, using the regression calibration estimator as the initial trial value. In case of no root, one may adopt the convention to use the zero-crossing of the tangent plane of the estimating function at the initial trial value. When the sample size is small and the measurement error is large, our numerical experience suggests that the occurrence rates of no root are similar for the nonparametrically corrected estimating function and the parametrically corrected ones if the measurement error is normally distributed. Nevertheless, such pathological cases were rare in our simulations under moderate sample size and measurement error, which is consistent with the findings of Stefanski (1989) and Nakamura (1990) for the parametric-correction estimators.

5. Remarks and extensions

The estimator based on a corrected estimating function converges to the same limit as that based on the original estimating function (in the absence of covariate measurement error), which holds regardless of whether the regression model is misspecified or not. Carroll et al. (1995, Section 6.8) pointed out that for the parametric-correction estimator this distributional robustness property, however, depends critically on the valid distributional assumptions of the measurement error. In this regard, the proposed nonparametric-correction estimator is more favorable.

In this article, we have proposed the nonparametric correction technique for the original functions of form (5) under the additive measurement error model (6). In fact, this methodology can be more broadly applied, given that that the essence is to exploit the independence structure of replicated surrogate covariates. See a companion paper, Huang and Wang (2000), for its application to Cox regression, where the partial-score function is not of form (5). Also it can accommodate multiplicative measurement error in normal regression as considered in Nakamura (1990, Section 4.2), or more generally,

polynomial regression in a recent paper by Iturria, Carroll and Firth (1999). Solely due to identifiability, the means of the multiplicative errors, however, need to be known.

As indicated in Section 1, Buzas (1997) also investigated the setting of replicated mismeasured covariate data. Despite that the idea of modifying the original estimating function is common, his and our approaches are different in the assumption requirement on the errors and in the modification. In fact, Buzas (1997) required the covariate measurement errors to be zero-unbiased and symmetric in distribution: In case of Poisson regression, this requirement is needed for not only the consistency of the intercept estimator but also that of the slope one. In contrast, our proposal has completely spared any error assumption for the slope estimation. More fundamentally, Buzas (1997) pursued the notion of *unbiased estimating function*, whereas our proposal aims directly at the root-consistency by taking advantage of the fact that the original estimating function possesses such a property. Although it is not an issue in the case of Poisson regression, further justification and/or assumption are needed in general for the approach of Buzas (1997) to achieve the root-consistency, as desired.

Indeed, Buzas (1997) considered the more general setting of instrumental variable, of which a replicate measurement is a special case; see also Carroll et al. (1995, Chapter 5) and the references therein. As will be seen, our proposal can be extended to instrumental variable estimation as well, with less assumptions, if any. To be an instrumental variable of $\mathbf{X}_e$, variable $\mathbf{T}$ satisfies three basic requirements: (i) correlated with $\mathbf{X}_e$; (ii) independent of $\boldsymbol{\varepsilon} = \mathbf{W} - \mathbf{X}_e$; and (iii) independent of Y given $\mathbf{X}$. Note that any function of $\mathbf{T}$ and $\mathbf{X}_a$, say $\mathbf{f}(\mathbf{T}, \mathbf{X}_a)$ of the same length as $\mathbf{X}_e$, is also an instrumental variable if the function meets the three requirements. The data typically consist of independent realizations of $(Y, \mathbf{X}_a, \mathbf{W}, \mathbf{T})$. Taking the Poisson regression model as an example, we construct the following estimating function,

$$\widehat{\mathcal{E}} \left[\{Y - \exp(\alpha + \boldsymbol{\beta}_e^T \mathbf{W} + \boldsymbol{\beta}_a^T \mathbf{X}_a)\} \begin{pmatrix} 1 \\ \mathbf{f}(\mathbf{T}, \mathbf{X}_a) \\ \mathbf{X}_a \end{pmatrix} \right],$$

which has the limit

$$\mathcal{E}\left[\{Y - \exp(\alpha + \boldsymbol{\beta}_e^T\mathbf{W} + \boldsymbol{\beta}_a^T\mathbf{X}_a)\}\begin{pmatrix} 1 \\ \mathcal{E}\{\mathbf{f}(\mathbf{T},\mathbf{X}_a)|\mathbf{X}\} \\ \mathbf{X}_a \end{pmatrix}\right].$$

As seen, the limit is zero at $(\alpha_*, \boldsymbol{\beta}_0^T)^T$. Therefore, the estimating function is root-consistent for $\boldsymbol{\beta}_0$ if the zero-crossing uniqueness of the limit can be further established. For instance, the scenario that $\mathcal{E}\{\mathbf{f}(\mathbf{T},\mathbf{X}_a)|\mathbf{X}\}$ is a linear function of $\mathbf{X}$ is often considered. In this case, write $\mathcal{E}[\{1, \mathbf{f}(\mathbf{T},\mathbf{X}_a)^T, \mathbf{X}_a^T\}^T|\mathbf{X}] = \mathbf{A}(1, \mathbf{X}^T)^T$ and suppose that matrix $\mathbf{A}$ is of full rank. Then the limit becomes

$$\mathbf{A}\,\mathcal{E}\left[\{Y - \exp(\alpha + \boldsymbol{\beta}_e^T\mathbf{W} + \boldsymbol{\beta}_a^T\mathbf{X}_a)\}\begin{pmatrix} 1 \\ \mathbf{X} \end{pmatrix}\right],$$

which has a unique zero-crossing, as is inherited from the original estimating function. Note that the linearity of $\mathcal{E}\{\mathbf{f}(\mathbf{T},\mathbf{X}_a)|\mathbf{X}\}$ in $\mathbf{X}$ is one sufficient condition for the root-consistency; generalizing the condition is a current research topic.

In our setting of replication data, it is assumed that $R_m \geq K + 1$, $m = 1, \cdots, M$. In some applications, however, due to practical constraints, this requirement may only be met on a subset of the sample. For example, in case of the Poisson regression model, duplicated surrogates may be available only on a subset and, for the rest of the sample, $R_m = 1$. While it is valid to apply the nonparametric-correction approach to the replication subset, it would be desirable to take advantage of the whole data set. For this purpose, one may rewrite the original estimating function for the Poisson regression model,

$$\widetilde{\boldsymbol{\Phi}}(\alpha, \boldsymbol{\beta}) = \widehat{\mathcal{E}}\left\{Y\begin{pmatrix} 1 \\ \mathbf{X}_e \\ \mathbf{X}_a \end{pmatrix}\right\} - \widehat{\mathcal{E}}\left\{\exp(\alpha + \boldsymbol{\beta}_e^T\mathbf{X}_e + \boldsymbol{\beta}_a^T\mathbf{X}_a)\begin{pmatrix} 1 \\ \mathbf{X}_e \\ \mathbf{X}_a \end{pmatrix}\right\}.$$

Since the first item only involves a single appearance of $\mathbf{X}_e$, one could use the whole data set for its correction if the chance with a single surrogate is completely at random. The potential efficiency gain, nonetheless, merits further investigation.

For brevity, we have largely focused our discussion on point estimation. To make inference, one can construct sandwich variance estimate (cf. Carroll et al., 1995, Ap-

pendix A) or use bootstrap (Efron and Tibshirani, 1993). These statistical procedures are now fairly standard.

References

Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, **10**, 1100–1120.

Buonaccorsi, J. P. (1996) A modified estimating equation approach to correcting for measurement error in regression. *Biometrika*, **83**, 433–440.

Buzas, J. S. (1997) Instrumental variable estimation in nonlinear measurement error models. *Commun. Statist.* A, **26**, 2861–2877.

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

Crowder, M. (1986) On consistency and inconsistency of estimating functions. *Econometric Theory*, **2**, 305–330.

Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap.* New York: Chapman and Hall.

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

Huang, Y. and Wang, C. Y. (2000) Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *J. Am. Statist. Assoc.*, **95**, 1209–1219.

Iturria, S. J., Carroll, R. J. and Firth, D. (1999) Polynomial regression and estimating functions in the presence of multiplicative measurement error. *J. R. Statist. Soc.* B, **61**, 547–561.

Nakamura, T. (1990) Corrected score function for errors in variables models: methodology and application to generalized linear models. *Biometrika*, **77**, 127–137.

Stefanski, L. A. (1989) Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun. Statist.* A, **18**, 4335–4358.

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

Table 1: Asymptotic efficiency of the nonparametric-correction slope estimator in the Poisson regression model relative to the parametric-correction one with σ_ε^2 (a) known or (b) estimated

| | | $\sigma_\varepsilon/\sigma_X$ | | | | | | | |
| | | 1/4 | | 1/2 | | 3/4 | | 1 | |
α_0	$\beta_0\sigma_X$	a	b	a	b	a	b	a	b
-1	1/3	1.000	1.000	0.999	1.000	0.995	1.000	0.986	0.999
	2/3	1.000	1.000	0.994	0.999	0.976	0.997	0.942	0.990
	1	0.999	1.000	0.984	0.995	0.949	0.979	0.901	0.947
0	1/3	1.000	1.000	0.997	1.000	0.986	1.000	0.964	0.999
	2/3	0.999	1.000	0.985	0.998	0.948	0.993	0.897	0.982
	1	0.997	0.999	0.972	0.991	0.931	0.972	0.884	0.938
1	1/3	0.999	1.000	0.991	1.000	0.965	0.999	0.920	0.997
	2/3	0.997	1.000	0.969	0.996	0.912	0.988	0.855	0.974
	1	0.994	0.998	0.962	0.988	0.921	0.968	0.877	0.934

Table 2: Empirical biases (×1000) and variances (×1000) of slope estimators: naive, regression calibration (RC), parametric-correction with σ_ε^2 known (PC$_a$), parametric-correction with σ_ε^2 estimated (PC$_b$), and nonparametric-correction (NPC)

size	σ_ε	Naive		RC		PC$_a$		PC$_b$		NPC	
		bias	var	bias	var	bias	var	bias	var	bias	var
						$\varepsilon \sim$ Normal					
200	1/4	-47.4	5.6	-16.8	6.0	2.4	6.8	3.2	6.8	2.8	6.8
	1/2	-158.4	4.8	-52.9	6.5	12.2	10.4	11.4	11.0	11.2	11.3
	3/4	-287.4	3.8	-83.1	8.3	38.6	25.7	44.3	31.1	41.6	30.1
400	1/4	-47.1	3.2	-17.1	3.4	2.1	3.8	2.3	3.8	2.1	3.8
	1/2	-161.2	1.9	-55.5	2.6	4.7	4.0	5.7	4.3	5.7	4.4
	3/4	-289.9	1.7	-89.1	3.9	17.1	9.2	19.7	11.9	19.1	11.0
						ε: Beta(3,1)					
200	1/4	-45.6	5.1	-15.4	5.4	4.4	6.1	4.4	6.2	1.7	6.2
	1/2	-151.9	4.7	-44.8	6.2	26.3	9.7	27.2	10.3	1.9	9.0
	3/4	-264.9	4.3	-54.0	8.7	131.3	31.1	130.9	34.0	24.1	17.5
400	1/4	-48.0	2.7	-18.0	2.9	1.3	3.2	1.3	3.3	-1.6	3.2
	1/2	-149.0	2.3	-41.9	3.1	29.2	4.9	29.1	5.2	5.4	4.6
	3/4	-267.0	2.0	-58.7	3.9	108.6	12.2	114.6	16.4	12.7	7.8