Statistica Sinica **24** (2014), 357-374
doi:http://dx.doi.org/10.5705/ss.2012.181

CORRECTED SCORE WITH SIZABLE COVARIATE MEASUREMENT ERROR: PATHOLOGY AND REMEDY

Yijian Huang

Emory University

Abstract: Corrected score (Nakamura (1990); Stefanski (1989)) is an important consistent functional modeling method for covariate measurement error in nonlinear regression. Although its pathological behaviors are known to exacerbate with increasing error contamination, neither their nature nor severity is well understood. We conduct a detailed investigation with the loglinear model for count data in the presence of sizable measurement error. Our study reveals that multiple roots, estimate-finding failure, and skewness in distribution are common and may persist even when the sample size is large. These pathological behaviors are attributed to a surprising fact that the desirable trend of the corrected score always goes astray as the parameter space approaches extremes. A novel remedy is proposed to constrain the derivatives with additional estimating functions. The resulting trend-constrained corrected score may also substantially improve estimation efficiency. These findings and the estimation strategy shed light on the developments for other nonlinear models and for the nonparametric correction method.

Key words and phrases: Empirical likelihood, functional modeling, loglinear model, method of moments, multiple roots, nonlinear model, Poisson regression, random effects Poisson regression, trend-constrained corrected score.

1. Introduction

Covariates in regression analyses may not always be accurately measured. Medical examples of error-contaminated covariates include CD4 lymphocyte count and HIV viral load in HIV/AIDS studies, and fat intake in nutritional epidemiology. Such errors are not necessarily small, and can be comparable to the true underlying covariates or even larger as in the case of HIV viral load. Naively treating mismeasured covariates as the truth can result in substantial estimation bias.

Statistical methods to accommodate covariate measurement error have been well developed for linear regression (Fuller (1987)). Current research efforts are mostly devoted to nonlinear models (see Carroll et al. (2006)) and we focus on such methods. These techniques are broadly classified as structural and functional modeling; the former imposes distributional assumptions on the true covariates and the latter spares them. Consistent functional modeling methods are particularly appealing for their robustness. Corrected score is one such method for a number of nonlinear models with additive measure error whose distribution is either known or consistently estimable. Enlisting a reference estimating function or objective function with the true covariates, the method constructs a so-called corrected function with error-contaminated covariates to share the same limit, by the method of moments. Provided that the reference admits only consistent estimates, so does the corrected function in a compact parameter space. Nakamura (1990) and Stefanski (1989) originally proposed the method for regression models whose likelihood scores are correction amenable, with Poisson regression as the prime application. Nakamura (1992) later extended it to Cox regression; see also Kong and Gu (1999) and Huang and Wang (2000). Huang and Wang (2001) developed a corrected score method for logistic regression, after devising new references since the likelihood score is not correction amenable (Stefanski (1989)).

Alternative functional modeling methods are available, most notably conditional score (Stefanski and Carroll (1987)) and locally efficient score (Tsiatis and Ma (2004)); the latter may be viewed as an extension of the former (Ma and Tsiatis (2006)). In comparison, corrected score has a number of distinct and appealing characteristics. First, the approach has a stronger consistency property since conditional score and locally efficient score admit not only consistent but also inconsistent roots. In fact, corrected score has been used for consistent root identification of conditional score (Huang and Wang (2001)). Second, corrected score can accommodate a general error distribution whereas conditional score is typically restricted to normal error. Further, corrected score is directly tied to a reference function rather than the model. For that reason, as an example, the corrected score for the Poisson regression model remains valid under the loglinear supermodel, while this is not true for the conditional score.

Nonetheless, corrected score has not often been adopted in practice, largely due to its finite-sample pathological behaviors. These behaviors were noticed by, for example, Nakamura (1990, 1992), Huang and Wang (2000, 2001), and Song and Huang (2005). But neither the nature nor the severity has been well understood. In the estimating function literature, the issue of multiple roots has received much attention and research; see a review by Small, Wang, and Yang (2000). However, the pathological behaviors of corrected score are much more complicated and challenging. Another but perhaps less important reason is its limited efficiency. Mean squared error often favors regression calibration (e.g., Prentice (1982); Carroll and Stefanski (1990); Gleser (1990)), an approximate but generally inconsistent method. Also, Stefanski (1989) demonstrated that the corrected score can be substantially less efficient than the conditional score for Poisson regression with normal measurement error. We conduct a detailed investigation on the pathological behaviors with the loglinear model and provide a comprehensive remedy that can also achieve remarkable efficiency gain. The findings are more generally applicable and the remedy may motivate similar developments for other nonlinear models.

2. Corrected Score for the Loglinear Model and Pathological Behaviors

Write Y as a count response variable and $p\times 1$ vector $\mathbf{X}$ as the true covariates. The loglinear model postulates

$$E(Y \mid \mathbf{X}) = \exp\left(\alpha + \boldsymbol{\beta}^\top \mathbf{X}\right), \tag{2.1}$$

where α is the intercept and $\boldsymbol{\beta}$ the slope vector. The distribution of Y given $\mathbf{X}$ is not modeled beyond the mean, and thus the Poisson regression model and random effects Poisson regression model are submodels. With an iid sample $\{(Y_i, \mathbf{X}_i) : i = 1, \ldots, n\}$ of size n, the standard estimation procedure is to maximize objective function,

$$n^{-1} \sum_{i=1}^{n} \left\{ Y_i \left(a + \mathbf{b}^\top \mathbf{X}_i\right) - \exp\left(a + \mathbf{b}^\top \mathbf{X}_i\right) \right\}, \tag{2.2}$$

or equivalently to find the zero-crossing of the estimating function

$$n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \exp\left(a + \mathbf{b}^\top \mathbf{X}_i\right) \right\} \begin{pmatrix} 1 \\ \mathbf{X}_i \end{pmatrix}; \tag{2.3}$$

these are the normalized log likelihood and likelihood score, respectively, for the Poisson regression submodel. By profiling out a in (2.3), one may also work with the profile score for $\boldsymbol{\beta}$,

$$\frac{\sum_{i=1}^{n} Y_i \mathbf{X}_i}{\sum_{i=1}^{n} Y_i} - \frac{\sum_{i=1}^{n} \mathbf{X}_i \exp\left(\mathbf{b}^\top \mathbf{X}_i\right)}{\sum_{i=1}^{n} \exp\left(\mathbf{b}^\top \mathbf{X}_i\right)}. \tag{2.4}$$

In practice, α is typically not of as much interest as $\boldsymbol{\beta}$.

2.1. Corrected score

In the presence of covariate measurement error, $\mathbf{X}$ is not directly observed but through its surrogate $\mathbf{W}$. Adopt the classical additive measurement error model

$$\mathbf{W} = \mathbf{X} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \perp\!\!\!\perp \{\mathbf{X}, Y\}, \tag{2.5}$$

where $\boldsymbol{\varepsilon}$ is the error vector and $\perp\!\!\!\perp$ denotes statistical independence. The elements of $\boldsymbol{\varepsilon}$ corresponding to accurately measured covariates, if any, are zeros. To focus on the main issues, we suppose that the distribution of $\boldsymbol{\varepsilon}$ is completely known.

Consider now an iid sample, $\{(Y_i, \mathbf{W}_i) : i = 1, \ldots, n\}$. The corrected score approach utilizes (2.2) and (2.3) as references to construct corrected functions based on the available data. Write the cumulant-generating function of $\boldsymbol{\varepsilon}$ as $\Omega(\mathbf{b}) \equiv \ln E\{\exp(\mathbf{b}^\top \boldsymbol{\varepsilon})\}$ and its derivative as $\dot{\Omega}(\mathbf{b}) \equiv \partial\Omega(\mathbf{b})/\partial\mathbf{b}$; in the normal error case, $\Omega(\mathbf{b}) = \mathbf{b}^\top E(\boldsymbol{\varepsilon}) + \mathbf{b}^\top \mathrm{var}(\boldsymbol{\varepsilon})\mathbf{b}/2$ and $\dot{\Omega}(\mathbf{b}) = E(\boldsymbol{\varepsilon}) + \mathrm{var}(\boldsymbol{\varepsilon})\mathbf{b}$. The corrected objective function (Nakamura (1990)) is

$$n^{-1} \sum_{i=1}^{n} \left(Y_i \left[a + \mathbf{b}^\top \left\{ \mathbf{W}_i - \dot{\Omega}(\mathbf{0}) \right\} \right] - \exp\left\{ a + \mathbf{b}^\top \mathbf{W}_i - \Omega(\mathbf{b}) \right\} \right), \tag{2.6}$$

which has the same expectation as (2.2). The estimation is then to maximize (2.6) or to find a zero-crossing of its derivative, the corrected score

$$\boldsymbol{\eta}(a, \mathbf{b}) \equiv n^{-1} \sum_{i=1}^{n} \boldsymbol{\eta}_i(a, \mathbf{b}), \tag{2.7}$$

where

$$\boldsymbol{\eta}_i(a, \mathbf{b}) \equiv Y_i \left\{ \begin{array}{c} 1 \\ \mathbf{W}_i - \dot{\Omega}(\mathbf{0}) \end{array} \right\} - \exp\left\{ a + \mathbf{b}^\top \mathbf{W}_i - \Omega(\mathbf{b}) \right\} \left\{ \begin{array}{c} 1 \\ \mathbf{W}_i - \dot{\Omega}(\mathbf{b}) \end{array} \right\}.$$

The corrected score may also be constructed directly from (2.3) to have the same expectation. Inherited from the reference, the corrected score almost surely has a unique and consistent root in a compact parameter space containing $(\alpha, \boldsymbol{\beta}^\top)^\top$. The estimator is asymptotically normal.

Note that a is uniquely determined for given $\mathbf{b}$ from the first element of (2.7). The corrected profile score for $\boldsymbol{\beta}$ is

$$\boldsymbol{\xi}(\mathbf{b}) = \frac{\sum_{i=1}^{n} Y_i \mathbf{W}_i}{\sum_{i=1}^{n} Y_i} - \frac{\sum_{i=1}^{n} \mathbf{W}_i \exp(\mathbf{b}^\top \mathbf{W}_i)}{\sum_{i=1}^{n} \exp(\mathbf{b}^\top \mathbf{W}_i)} + \dot{\Omega}(\mathbf{b}) - \dot{\Omega}(\mathbf{0}), \tag{2.8}$$

which has the same limit as profile score (2.4) in a compact parameter space containing $\boldsymbol{\beta}$ (cf., Huang and Wang (2006)). The difference between the corrected profile score and (2.7) is nothing but algebraic.

2.2. Pathological behaviors

The asymptotic justification of the corrected score requires regularity conditions including compactness of the parameter space. Such conditions are fairly standard and, for many statistical problems, the asymptotic results typically provide a good approximation for practical purposes. It is indeed the case when the measurement error is small. But pathological behaviors soon emerge as the error increases. They are complex and can be prevalent enough to cause serious practical concerns.

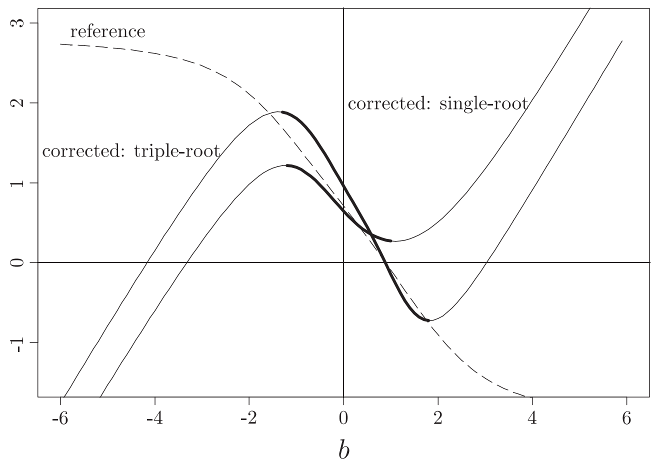

Figure 1. Two observed root patterns of corrected profile score $\xi(b)$. Although the reference is decreasing, the corrected function is so only in the thickened portion.

Consider a single-covariate model with normal error contamination, $\varepsilon \sim \text{Normal}(0, \Lambda)$. For ease of visualization, we work with (2.8), where $\dot{\Omega}(b) - \dot{\Omega}(0) = b\Lambda$. Note that its reference (2.4) is monotonically decreasing and has a unique root, provided that X_i, $i = 1, \ldots, n$, are not all equal to each other. Then, by asymptotic arguments, $\xi(b)$ is approximately decreasing in a compact parameter space when the sample size is sufficiently large. Despite this, rather surprisingly, the overall trend as dominated by $\dot{\Omega}(b) - \dot{\Omega}(0)$ turns out to be the opposite regardless of the sample size,

$$\xi(b) = \begin{cases} -\infty, & b = -\infty, \\ \infty, & b = \infty. \end{cases} \tag{2.9}$$

Upon ignoring the trivial local extremum case that $\xi(b)$ touches 0 but does not cross, $\xi(b)$ has an odd number of roots, with the minimum being 1. Furthermore, the increasing roots–those with increasing local trend–outnumber decreasing ones by 1. In our numerical studies, we have observed only single- and triple-root patterns, however, and they are illustrated along with the reference in Figure 1.

Only decreasing roots of $\xi(b)$ may be appropriate as estimates, since the increasing roots correspond to saddle points, rather than local maximizers, of (2.6). Then, estimate-finding failure arises with the single-root pattern since the only root is increasing. More concerning is its high prevalence, particularly when the measurement error is sizable. We conducted a simulation study under the Poisson regression submodel with $\alpha = 0$, $\beta > 0$, and X having mean 0. Sample sizes ranged from 100 to 800, and several X distributions were studied. The

Table 1. Behaviors of the corrected profile score under $Y \sim$ Poisson $\{\exp(\boldsymbol{\beta} X)\}$ with $E(X) = 0$, $\boldsymbol{\beta} > 0$, and $\varepsilon \sim$ Normal with mean 0, based on 1,000 simulation iterations.

	distribution of X											
	Normal		Uniform		$\chi^2(1)^\dagger$		Normal		Uniform		$\chi^2(1)^\dagger$	
size	SR	SK	SR	SK	SR	SK	SR	SK	SR	SK	SR	SK
	$\text{var}(\varepsilon)/\text{var}(X) = 0.5$											
	$\boldsymbol{\beta}^2\text{var}(X) = 0.5$						$\boldsymbol{\beta}^2\text{var}(X) = 1$					
100	2.1	2.74	4.4	2.56	0.6	2.81	15.0	2.66	22.2	2.00	16.4	2.64
200	0.0	3.72	0.2	2.64	0.1	2.08	4.3	2.98	8.4	2.49	12.7	3.65
400	0.0	3.12	0.0	1.44	0.1	2.87	0.8	2.82	3.0	2.84	9.4	2.72
800	0.0	2.53	0.0	1.74	0.1	3.31	0.1	2.57	0.0	1.91	8.4	3.17
	$\text{var}(\varepsilon)/\text{var}(X) = 1$											
	$\boldsymbol{\beta}^2\text{var}(X) = 0.5$						$\boldsymbol{\beta}^2\text{var}(X) = 1$					
100	24.8	1.60	30.5	1.77	9.4	2.00	48.8	1.19	53.6	1.36	57.8	1.53
200	11.9	2.15	15.5	2.25	1.0	3.17	36.0	1.74	39.4	1.33	46.2	2.00
400	2.7	2.73	8.4	2.78	0.8	2.20	21.9	2.22	32.1	1.53	35.9	2.23
800	0.4	2.60	1.7	2.22	0.4	2.27	12.0	3.06	19.3	2.54	27.8	3.02

SR: prevalence (%) of the single-root pattern; SK: skewness measure (2.10) of the corrected estimate.

† shifted and rescaled.

results based on 1,000 iterations are reported in Table 1. Consider the case $\text{var}(X) = \text{var}(\varepsilon) = \beta = 1$. With sample size 100, the failure could be as frequent as around 50%. Although its occurrence decreased with increasing sample size, it was still over 10% when the sample size was 800. The failure rate declined with decreasing $\text{var}(\varepsilon)$ and β, but it is clear that the estimate-finding failure would not be negligible in many practical situations.

The issue of multiple roots with the triple-root pattern can be readily resolved since there is only one decreasing root, designated by Nakamura (1990) as the corrected estimate. Nevertheless, this corrected estimate can be highly skewed. Table 1 also reports a crude skewness measure,

$$\frac{\max \widehat{\beta} - \beta}{\beta - \min \widehat{\beta}}, \tag{2.10}$$

where $\widehat{\beta}$ is the corrected estimate. The sub-distributions were clearly skewed as shown. Like estimate-finding failure, the skewness may also be attributed to the misbehaved overall trend but as a less severe manifestation, which might explain that the skewness did not always improve with increasing sample size.

The results in Table 1 are invariant to the values of $\text{var}(X)$, $\text{var}(\varepsilon)$, and β once $\text{var}(\varepsilon)/\text{var}(X)$ and $\beta^2\text{var}(X)$ are fixed. Also, not only $\text{var}(\varepsilon)/\text{var}(X)$ but also $\beta^2\mathrm{var}(X)$ are relevant in the discussion of pathological behaviors. When $\mathrm{var}(\varepsilon)$ and $\mathrm{var}(X)$ are comparable, the pathological behaviors of the corrected score could nonetheless be negligible if $\beta^2\mathrm{var}(X)$ is sufficiently small. Still, they could be of enormous concern even with a small $\mathrm{var}(\varepsilon)$ relative to $\mathrm{var}(X)$ if $\beta^2\mathrm{var}(X)$ is very large. Throughout, our discussion is understood to be in the context that both the variance and coefficients of the true covariates are roughly unity unless otherwise specified.

These pathological behaviors of $\xi(b)$ are not unique to normal error. Trend (2.9) holds whenever the error distribution is not bounded at both extremes, because $\dot\Omega(b) - \dot\Omega(0)$ is then unbounded. In the case of a bounded measurement error distribution, one might expect the pathological behaviors to be less serious, but not completely eliminated. When there are more covariates and one or more of them are contaminated, the pathological behaviors may be more complex but should be similar in nature.

3. Improving Corrected Score via Trend Constraining

Multiple roots and estimate-finding failure are the two most prominent pathological behaviors. The corrected score is too liberal in adopting any root, yet too stringent in using a root only, as estimate. The local trend needs to be taken into consideration, and an approximate root should be admissible. We propose a remedy by imposing trend constraints through additional estimating functions and defining an estimate via empirical likelihood. This remedy mitigates the skewness and improves estimation efficiency as well in the mean time.

3.1. A class of estimating functions

Local trend of the corrected score is quantified by the derivatives. Write $\ddot\Omega(\mathbf{b}) \equiv \partial\dot\Omega(\mathbf{b})/\partial\mathbf{b}^\top$. The first derivative $\partial\boldsymbol{\eta}(a,\mathbf{b})/\partial(a,\mathbf{b}^\top)$ is

$$\dot{\boldsymbol{\eta}}(a,\mathbf{b}) = -n^{-1}\sum_{i=1}^{n} \exp\left\{a + \mathbf{b}^\top\mathbf{W}_i - \Omega(\mathbf{b})\right\} \times \left[\left\{\begin{matrix} 1 \\ \mathbf{W}_i - \dot\Omega(\mathbf{b}) \end{matrix}\right\}^{\otimes 2} - \mathrm{diag}\left\{0, \ddot\Omega(\mathbf{b})\right\}\right], \qquad (3.1)$$

where $\mathbf{v}^{\otimes 2} \equiv \mathbf{v}\mathbf{v}^\top$. Section 2 suggests that inappropriate roots of $\boldsymbol{\eta}(a,\mathbf{b})$ can be eliminated should $\dot{\boldsymbol{\eta}}(a,\mathbf{b})$ be constrained. For this purpose, we establish an identity.

Proposition 1. *Suppose that the loglinear model* (2.1) *and the classical additive measurement error model* (2.5) *hold. Writing* $\mathbf{b} \equiv (b_1,\ldots,b_p)^\top$*, if the moment-*

generating functions of $\mathbf{X}$ *and* $\boldsymbol{\varepsilon}$ *exist, then*

$$E\left[Y\,\frac{\partial^{k_1+\cdots+k_p}\exp\left\{\mathbf{b}^\top\mathbf{W}-\Omega(\mathbf{b})\right\}}{\partial b_1^{k_1}\cdots\partial b_p^{k_p}}\bigg|_{\mathbf{b}=\mathbf{0}} - \frac{\partial^{k_1+\cdots+k_p}\exp\left\{\alpha+\mathbf{b}^\top\mathbf{W}-\Omega(\mathbf{b})\right\}}{\partial b_1^{k_1}\cdots\partial b_p^{k_p}}\bigg|_{\mathbf{b}=\boldsymbol{\beta}}\right]=0, \tag{3.2}$$

for $k_q \geq 0$*,* $q=1,\ldots,p$*.*

Proof. Given $E[\exp\{a+\mathbf{b}^\top\mathbf{W}-\Omega(\mathbf{b})\} \mid \mathbf{X}] = \exp(a+\mathbf{b}^\top\mathbf{X})$ under (2.5), one obtains directly that

$$E\left[\frac{\partial^{k_1+\cdots+k_p}\exp\left\{a+\mathbf{b}^\top\mathbf{W}-\Omega(\mathbf{b})\right\}}{\partial b_1^{k_1}\cdots\partial b_p^{k_p}}\;\middle|\;\mathbf{X}\right] = \exp\left(a+\mathbf{b}^\top\mathbf{X}\right)\prod_{q=1}^{p}X_q^{k_q}.$$

Then,

$$E\left[\frac{\partial^{k_1+\cdots+k_p}\exp\left\{\mathbf{b}^\top\mathbf{W}-\Omega(\mathbf{b})\right\}}{\partial b_1^{k_1}\cdots\partial b_p^{k_p}}\bigg|_{\mathbf{b}=\mathbf{0}}\;\middle|\;\mathbf{X}\right] = \prod_{q=1}^{p}X_q^{k_q}$$

follows. These two equations, coupled with (2.1), imply (3.2).

Equation (3.2) leads to a class of estimating functions for $(\alpha,\boldsymbol{\beta}^\top)^\top$. The corrected score $\boldsymbol{\eta}(a,\mathbf{b})$, as a special case, corresponds to $\sum_{q=1}^{p}k_q = 0$ and 1. Similarly, the estimating function, as a symmetric matrix,

$$n^{-1}\sum_{i=1}^{n}\left(Y_i\left[\left\{\mathbf{W}_i-\dot{\Omega}(\mathbf{0})\right\}^{\otimes 2}-\ddot{\Omega}(\mathbf{0})\right] - \exp\left\{a+\mathbf{b}^\top\mathbf{W}_i-\Omega(\mathbf{b})\right\}\left[\left\{\mathbf{W}_i-\dot{\Omega}(\mathbf{b})\right\}^{\otimes 2}-\ddot{\Omega}(\mathbf{b})\right]\right), \tag{3.3}$$

corresponds to $\sum_{q=1}^{p}k_q = 2$. In light of (3.1), having both $\boldsymbol{\eta}(a,\mathbf{b})$ and (3.3) equal to zero leads to the constraint

$$\dot{\boldsymbol{\eta}}(a,\mathbf{b}) = -n^{-1}\sum_{i=1}^{n}Y_i\left[\begin{Bmatrix}1\\ \mathbf{W}_i-\dot{\Omega}(\mathbf{0})\end{Bmatrix}^{\otimes 2}-\operatorname{diag}\left\{0,\ddot{\Omega}(\mathbf{0})\right\}\right].$$

The right-hand side is independent of $(a,\mathbf{b}^\top)^\top$, and almost surely negative definitive since

$$E\left[\begin{Bmatrix}1\\ \mathbf{W}-\dot{\Omega}(\mathbf{0})\end{Bmatrix}^{\otimes 2}-\operatorname{diag}\{0,\ddot{\Omega}(\mathbf{0})\}\;\middle|\;\mathbf{X}\right] = \begin{pmatrix}1\\ \mathbf{X}\end{pmatrix}^{\otimes 2}$$

under (2.5). Thus, when incorporating (3.3), the first derivative of $\boldsymbol{\eta}(a, \mathbf{b})$ at an estimate is required to be almost surely negative definitive. In the single covariate case, this means that $\xi(b)$ needs to have a negative derivative almost surely at the estimate.

Along the same line, estimating functions corresponding to $\sum_{q=1}^{p} k_q > 2$, when additionally incorporated, would effectively impose constraints on higher-order derivatives of $\boldsymbol{\eta}(a, \mathbf{b})$ at an estimate.

3.2. Estimation and inference

After constraining the local trend, the number of estimating functions exceeds that of the parameters. We employ empirical likelihood and adopt the maximum empirical likelihood estimator (Owen (2001); Qin and Lawless (1994)). Write a generic estimating function as $\boldsymbol{\varphi}(a, \mathbf{b}) \equiv n^{-1} \sum_{i=1}^{n} \boldsymbol{\varphi}_i(a, \mathbf{b})$. The estimator is the global maximizer of the empirical likelihood ratio function

$$\mathcal{L}(a, \mathbf{b}) = \max \left\{ \prod_{i=1}^{n} n w_i : \sum_{i=1}^{n} w_i \boldsymbol{\varphi}_i(a, \mathbf{b}) = \mathbf{0},\ \sum_{i=1}^{n} w_i = 1,\ w_i \geq 0 \, \forall \, i = 1, \ldots, n \right\}; \tag{3.4}$$

algorithms for the computation can be found in Owen (2001). For our problem, $\boldsymbol{\varphi}(a, \mathbf{b})$ consists of $\boldsymbol{\eta}(a, \mathbf{b})$ and the additional estimating functions for trend constraining. We call this approach the trend-constrained corrected score, and the special case the first derivative-constrained corrected score if (3.3) is the only additional estimating function.

While best recognized as a means of dealing with more estimating functions than parameters, empirical likelihood also provides a natural way to define an approximate root as estimate even when the numbers of estimating functions and parameters are the same. This leads to a redefined corrected score by adopting the empirical likelihood with $\boldsymbol{\varphi}(a, \mathbf{b}) = \boldsymbol{\eta}(a, \mathbf{b})$. Write $\widehat{w}_i$ as the value of w_i for the maximization in (3.4) with given $(a, \mathbf{b}^\top)^\top$. The estimate is chosen as a local maximizer of $\mathcal{L}(a, \mathbf{b})$ such that $\sum_{i=1}^{n} \widehat{w}_i \dot{\boldsymbol{\eta}}_i(a, \mathbf{b})$ is negative definite. To illustrate, Figure 2 shows the empirical likelihood ratio functions, upon profiling out a, for the data sets having single- and triple-root patterns in Figure 1. The redefined corrected score can identify a local maximizer as estimate when no root deems appropriate in the single-root pattern, but the global maximizer with the first derivative-constrained corrected score offers a more comprehensive remedy and has better computational properties. We view the redefined corrected score as an *ad hoc* remedy to multiple roots and estimate-finding failure.

According to Qin and Lawless (1994), the maximum empirical likelihood estimator is consistent and asymptotically normal with variance

$$n^{-1} \left[E\dot{\boldsymbol{\varphi}}_1(\alpha, \boldsymbol{\beta})^\top E \left\{ \boldsymbol{\varphi}_1(\alpha, \boldsymbol{\beta})^{\otimes 2} \right\}^{-1} E\dot{\boldsymbol{\varphi}}_1(\alpha, \boldsymbol{\beta}) \right]^{-1}, \tag{3.5}$$

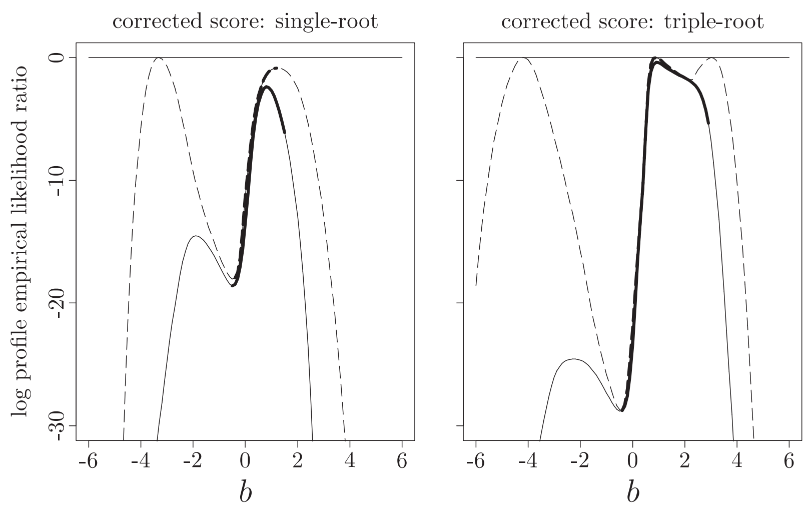

Figure 2. Log profile empirical likelihood ratio functions for the two data sets in Figure 1. Dashed and solid lines correspond to the corrected score and first derivative-constrained corrected score, respectively. Thickened portion indicates the region of b where the qualitative trend constraint is satisfied.

under mild moment conditions. The estimator reaches the asymptotic efficiency bound among those based on linear combinations of elements in $\boldsymbol{\varphi}(a, \mathbf{b})$. Thus, the trend-constrained corrected score estimator is asymptotically no less efficient than the corrected score estimator. To provide a sense of the efficiency improvement, we analytically compute the asymptotic efficiency under a single-covariate Poisson regression submodel. Both true covariate X and measurement error ε are normal with mean 0, and the intercept α is 0. Table 2 reports the asymptotic efficiency of the first derivative-constrained corrected score estimator relative to the corrected score estimator at various values of $\mathrm{var}(\varepsilon)/\mathrm{var}(X)$ and $\beta^2\mathrm{var}(X)$. When $\mathrm{var}(\varepsilon)/\mathrm{var}(X)$ or $\beta^2\mathrm{var}(X)$ approaches 0, the asymptotic relative efficiency is close to 1 as expected. But it can be well over 3 when $\mathrm{var}(\varepsilon) = \mathrm{var}(X) = 1$ and $|\beta| = 1$.

For interval estimation, one standard approach is to construct Wald-type confidence intervals by adopting a plug-in estimate of the asymptotic variance (3.5); see Qin and Lawless (1994, p.306). An alternative is to use the empirical likelihood ratio on the basis of Wilks' theorem; see Qin and Lawless (1994, Thm. 2 and Coro. 5).

4. Simulation Studies

Simulation studies were conducted to assess the performance of the trend-constrained corrected score and to identify specific trend-constrained corrected

Table 2. Asymptotic efficiency of the first derivative-constrained corrected score estimator relative to the corrected score estimator under Poisson regression submodel with a single and error-contaminated covariate: $X \sim$ Normal with mean 0, $\varepsilon \sim$ Normal with mean 0, and $\alpha = 0$.

	$\beta^2\text{var}(X)$							
$\text{var}(\varepsilon)/\text{var}(X)$	0.25		0.5		0.75		1	
0.2	1.00	1.00	1.01	1.06	1.07	1.24	1.27	1.60
0.4	1.00	1.02	1.05	1.18	1.24	1.56	1.72	2.13
0.6	1.00	1.04	1.10	1.32	1.46	1.85	2.25	2.53
0.8	1.01	1.08	1.18	1.46	1.73	2.12	2.84	2.88
1	1.02	1.11	1.27	1.61	2.05	2.37	3.46	3.20

The two columns correspond to estimators of α and β.

scores that best improve the estimation under practical sample sizes. We report the results from single- and double-covariate models with substantial normal covariate measurement error.

Three single-covariate models were considered. The true covariate X had mean 0 and variance 1, and error ε was standard normal. The regression coefficient $(\alpha, \beta)^\top$ was set to $(0, 1)^\top$.

Case A. Y followed a Poisson regression model given X, and X was standard normal.

Case B. Y followed a Poisson regression model given X, and X was a modified chi-square with 1 degree of freedom: first truncated at 5, then location-shifted to mean 0, and rescaled to variance 1.

Case C. Y was Poisson with mean $\exp(\alpha + \beta X + Z)$ given X and Z, X was standard normal, and unobserved Z was normal with mean -0.2 and variance .4. Thus, Y followed a random effects Poisson regression model given X.

For the double-covariate model, we adopted a Poisson regression model where the true covariate $\mathbf{X}$ was standard bivariate normal with correlation coefficient .5. The first covariate was subject to contamination by standard normal error, whereas the second covariate was accurately measured. The regression coefficient $(\alpha, \beta_1, \beta_2)^\top$ was set to $(0, 1, -1)^\top$.

Sample sizes 100, 200, 400, and 800 were investigated. For each scenario, 1,000 samples were simulated. We report the results on point and interval estimation separately.

4.1. Point estimation

The naive approach, regression calibration, conditional score, and redefined corrected score were included for comparison. The regression calibration used

(2.3) with $\mathbf{X}_i$ replaced by its best linear approximate given $\mathbf{W}_i$,

$$\overline{\mathbf{W}} - \dot{\Omega}(\mathbf{0}) + \{\widehat{\mathbf{\Sigma}} - \ddot{\Omega}(\mathbf{0})\}\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{W}_i - \overline{\mathbf{W}}),$$

where $\overline{\mathbf{W}}$ and $\widehat{\mathbf{\Sigma}}$ are the sample mean and variance of $\mathbf{W}$ (cf., Carroll et al. (2006)). Despite its inconsistency in general, numerical stability, moderate bias, and reasonable mean squared error in many small to moderate error contamination situations make regression calibration the default choice in practice. The conditional score was given in Stefanski and Carroll (1987, p.710) with $t(\cdot)$ linear as commonly used, and the root closest to the naive estimate was adopted. We concentrate on estimators of $\boldsymbol{\beta}$ because α is typically of less interest.

Table 3 and Figure 3 summarize the simulation results for the estimators in the single-covariate models. For the trend-constrained corrected score, three sets of additional estimating functions were considered to effectively impose constraints on the first k derivatives of the corrected score, $k = 1, 2, 3$. As expected, the naive estimator had large bias under all three cases. The regression calibration estimator of β is consistent under Cases A and C, which explains the small bias observed; in Case B, however, it was even more biased than the naive estimator. The conditional score estimator showed little bias and high efficiency under Cases A and B, but it is no longer consistent and incurred substantial bias under Case C. Both the redefined corrected score and trend-constrained corrected score estimators are consistent under all three cases, and they showed decreasing bias with increasing sample size. But the bias reduction of the former appeared slower and the efficiency not nearly as good. The quantile-quantile plots show that the redefined corrected score estimator deviated from normality considerably even when the sample size was 800. Notably, the first derivative-constrained corrected score estimator quickly behaved more like a normal with increasing sample size. In comparison with higher-order derivatives-constrained corrected scores, the first derivative-constrained one seemed favorable in bias and mean squared error despite the fact that the asymptotic efficiency would suggest otherwise.

Table 4 reports the simulation results for the double-covariate model. The measurement error in general had impact not only on the coefficient estimation of the error-prone covariate but also on that of the accurately measured. The relative performance of these estimators largely followed a pattern similar to that for the single-covariate models. In consideration of trend constraints for the trend-constrained corrected score, the first derivative now involves three elements, those in the upper triangle of matrix (3.3). We investigated subsets of these three as constraints; higher-order derivatives-constrained corrected scores were not considered due to unfavorable performance in the single-covariate models. In comparison, the first derivative-constrained corrected score estimator had the smallest mean squared error albeit not necessarily the smallest bias.

Table 3. Simulation summary statistics with the single-covariate models: naive (NV), regression calibration (RC), conditional score (ConS), redefined corrected score (CS), and first k derivatives-constrained corrected score (TC:k), $k = 1, 2, 3$.

		NV		RC		ConS			CS		TC:1		TC:2		TC:3	
size		B	E	B	E	F	B	E	B	E	B	E	B	E	B	E
								Case A								
100	α	241	277	244	297	0	-9	184	-87	304	101	236	148	289	215	316
	β	-504	514	63	335		38	245	132	330	-93	301	-179	577	-291	551
200	α	243	261	245	268	0	-5	125	-122	284	58	156	81	183	162	244
	β	-503	508	23	184		18	164	156	345	-44	186	-35	282	-166	353
400	α	249	258	250	262	0	3	85	-118	270	33	113	37	125	103	173
	β	-502	505	7	128		5	107	145	333	-23	147	6	163	-72	202
800	α	250	254	250	255	0.1	-1	61	-101	231	17	75	11	81	53	111
	β	-500	502	5	90		5	78	120	282	-15	100	23	118	-19	123
								Case B								
100	α	195	264	218	322	0	1	168	-293	510	127	238	146	333	196	340
	β	-285	301	570	795		23	167	259	419	-5	251	-143	632	-320	696
200	α	207	239	219	262	0	2	114	-262	471	76	168	73	182	135	237
	β	-293	300	472	545		7	104	193	357	16	150	33	300	-69	387
400	α	219	235	224	246	0	5	79	-177	382	49	113	34	119	83	168
	β	-299	302	428	462		2	71	123	273	6	99	69	150	41	207
800	α	223	231	226	236	0	1	53	-93	244	26	75	8	70	24	88
	β	-301	303	410	428		2	52	64	176	0	61	48	99	72	136
								Case C								
100	α	224	279	227	297	0.6	-227	370	-103	313	67	241	88	262	148	306
	β	-504	519	55	368		359	660	115	339	-87	293	-121	465	-257	540
200	α	240	265	242	272	0.2	-230	299	-126	303	44	177	59	189	129	233
	β	-504	513	16	226		374	496	148	352	-40	231	-21	250	-137	302
400	α	245	259	244	260	1.1	-220	266	-137	295	24	124	24	136	85	169
	β	-502	507	12	164		349	427	162	356	-19	156	11	178	-73	194
800	α	246	253	245	253	1.2	-214	243	-105	249	14	89	9	90	43	113
	β	-501	503	4	114		324	372	116	300	-16	108	10	114	-25	119

F: estimate-finding failure (%); B: bias ($\times 1000$); E: root mean squared error ($\times 1000$). Only conditional score was observed to experience estimate-finding failure and its associated summary statistics are based on successful iterations only.

From these simulations, we found that the first derivative-constrained corrected score compares favorably with those having other choices of trend constraints; much larger sample sizes might be needed for higher-order derivatives-constrained corrected scores to outperform. Relative to existing methods, the first derivative-constrained corrected score does well given substantial error con-

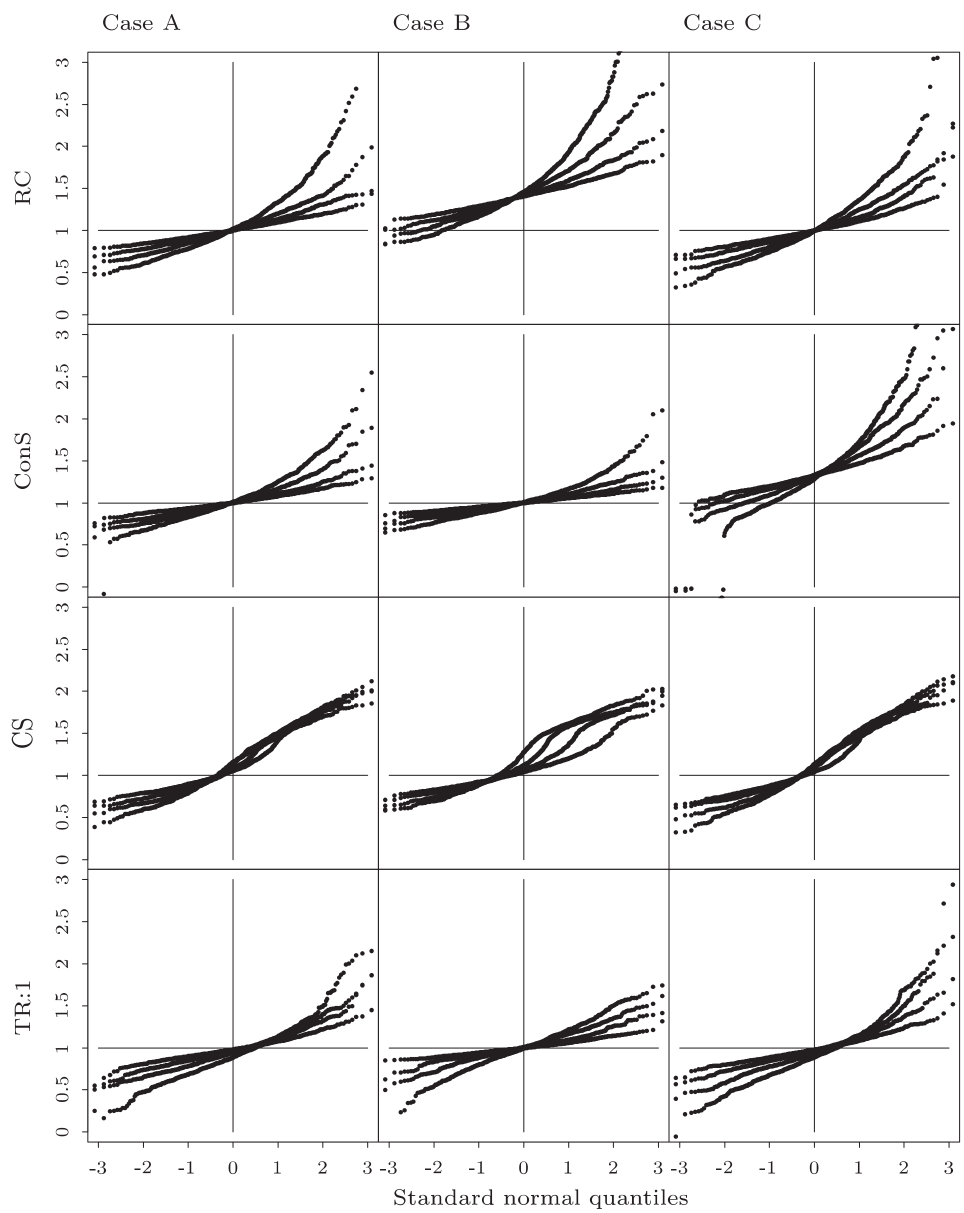

Figure 3. Quantile-quantile plots for the slope estimators in the single-covariate models, where $\beta = 1$. Plots in the order of decreasing slopes correspond to sample sizes 100, 200, 400, and 800. Conditional score is based on successful estimate-finding iterations only.

tamination under moderate sample size, and it had an efficiency comparable to the regression calibration in the estimation of $\boldsymbol{\beta}$ when the latter was consistent.

Table 4. Simulation summary statistics with the double-covariate model: naive (NV), regression calibration (RC), conditional score (ConS), redefined corrected score (CS), and trend-constrained corrected scores (TC:1/3, TC:2/3, TC:1).

		NV		RC		ConS		CS			TC:1/3		TC:2/3		TC:1	
size		B	E	B	E	B	E	F	B	E	B	E	B	E	B	E
100	α	200	239	199	262	-18	241	0.1	-55	254	1	339	76	272	140	226
	β_1	-578	587	123	515	41	403		85	299	-42	403	-134	402	-211	359
	β_2	293	323	-54	315	-21	249		-97	309	-66	410	13	345	95	253
200	α	205	226	206	235	-6	126	0	-102	244	-38	298	37	197	89	172
	β_1	-574	578	52	258	25	202		149	335	40	323	-43	260	-106	237
	β_2	292	309	-22	175	-9	155		-109	286	-90	356	-18	249	46	172
400	α	210	220	211	225	-2	88	0	-118	240	-48	218	14	129	54	113
	β_1	-573	575	24	160	13	131		168	335	55	260	-15	183	-62	165
	β_2	288	297	-8	116	-5	106		-111	257	-76	257	-21	180	26	122
800	α	212	217	212	219	-1	61	0	-103	212	-51	180	-1	90	26	80
	β_1	-571	572	12	107	6	88		146	308	64	224	2	130	-27	118
	β_2	285	289	-5	79	-3	73		-99	225	-72	216	-24	121	9	87

TC:1/3 and TC:2/3 correspond to using the (1,1)-th element and the first row of matrix (3.3) as the additional estimating function, respectively. TC:1 is the first derivative-constrained corrected score. F: estimate-finding failure (%); B: bias ($\times 1000$); E: root mean squared error ($\times 1000$). Only redefined corrected score was observed to experience estimate-finding failure and its associated summary statistics are based on successful iterations only.

Although it could be less efficient than the conditional score, the comparison is not completely fair since the conditional score requires the Poisson regression submodel and exploits the extra modeling assumptions; note that the conditional score is not a viable approach for the Case C single-covariate model. Additional simulations, not shown, revealed that, in the absence of measurement error, the first derivative-constrained corrected score performed about the same as the standard regression despite that the trend constraint is not informative at all with the Poisson regression submodel.

4.2. Interval estimation

Three confidence intervals were constructed for the first derivative-constrained corrected score. In addition to the Wald-type, the others were obtained by inverting the empirical likelihood ratio test, where the critical value was based on the asymptotic chi-square distribution or the bootstrap calibration (cf., Owen (2001, Chap. 3.3)) with bootstrap size 199.

Table 5 reports the empirical coverage of the 95% confidence intervals in the single- and double-covariate models. The Wald-type confidence interval and the

Table 5. Coverage of 95% Wald-type (W) and empirical likelihood-based (EL, ELC) confidence intervals for first derivative-constrained corrected score.

size	100			200			400			800		
	W	EL	ELC	W	EL	ELC	W	EL	ELC	W	EL	ELC
	Single-covariate model: Case A											
α	79.8	79.8	95.4	85.1	86.0	95.4	89.3	89.0	93.9	93.1	93.1	93.7
β	81.0	89.5	93.5	86.8	90.5	94.1	89.7	91.6	94.0	93.0	94.4	93.7
	Single-covariate model: Case B											
α	83.3	79.4	94.1	85.4	85.2	96.0	89.2	88.6	95.4	91.3	91.2	95.4
β	87.9	84.9	90.9	91.4	87.8	93.4	93.1	90.8	93.9	94.3	93.9	94.7
	Single-covariate model: Case C											
α	83.7	84.7	95.3	88.0	89.2	94.5	91.3	90.7	93.3	91.9	91.4	94.6
β	81.9	88.4	93.7	87.0	91.9	95.0	92.0	93.3	94.5	92.3	94.5	93.6
	Double-covariate model											
α	72.0	73.8	96.7	75.8	76.9	95.2	84.3	85.7	95.1	90.1	90.4	95.0
β_1	70.3	78.2	93.4	79.4	85.1	95.0	88.0	89.8	96.0	91.3	93.3	95.4
β_2	80.3	86.3	91.9	84.5	90.5	94.3	89.3	90.9	93.9	93.2	93.6	93.6

EL and ELC are empirical likelihood-based confidence intervals with critical value determined by asymptotic chi-square distribution and bootstrap calibration, respectively.

empirical likelihood-based one using the asymptotic critical value both had under-coverage, although improved with increasing sample size. In comparison, the empirical likelihood-based confidence interval using bootstrap-calibrated critical value achieved much more accurate coverage for all sample sizes considered.

5. Final Remarks

Commonly adopted approximate methods, including regression calibration and simulation extrapolation (Cook and Stefanski (1994); Stefanski and Cook (1995)), and existing functional modeling methods may accommodate small or, at best, moderate covariate error contamination in a nonlinear regression model. In the presence of substantial error, they are all deficient even with large sample size. We have developed the trend-constrained corrected score as a functional modeling method for the loglinear model, which enjoys considerable advantages over existing methods. In our development, we have assumed that the error distribution is known. Of course, this is rarely realistic in practice where parameters in a parametric error distribution often need to be estimated. The trend-constrained corrected score can be extended in a straightforward fashion by including additional estimating functions for the parameter estimation of the error.

Huang and Wang (2006) discovered that the corrected profile score (2.8) for the loglinear model has an intimate relationship with those for logistic and Cox regression. Therefore, the pathological behaviors and their cause for the latter models are expected to be similar. This might also explain the limited success in previous remedy attempts for Cox regression (Nakamura (1992); Huang and Wang (2001)) by reducing or eliminating the asymptotically negligible bias of the corrected partial score. In fact, this bias is similar in nature to that of the corrected profile score (2.8); however, the corresponding corrected score (2.7) is nonetheless unbiased. Our proposed strategy of trend constraining warrants further investigation for logistic and Cox regression.

With additional data available on the measurement error, Huang and Wang (1999, 2000, 2001, 2006) generalized the corrected score and developed the nonparametric correction method which further spares distributional assumptions on the error. Not surprisingly, nonparametric correction estimating functions suffer from pathological behaviors as well. While the nature of these pathological behaviors might not be exactly the same, the results on the corrected score nonetheless shed light on the investigation, which is the focus of our current research.

Acknowledgement

This research was partially supported by funds AI078835, AI050409, and HL113451 from the U.S. National Institutes of Health. The author thanks the reviewers for their helpful comments which have led to an improved presentation.

References

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition. Chapman and Hall, Boca Raton.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictors. *J. Amer. Statist. Assoc.* **85**, 652-663.

Cook, J. and Stefanski, L. A. (1994). A simulation extrapolation method for parametric measurement error models. *J. Amer. Statist. Assoc.* **89**, 1314-1328.

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

Gleser, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In *Statistical Analysis of Measurement Errors Models and Application* (Edited by P. J. Brown and W. A. Fuller). American Mathematics Society, Providence.

Huang, Y. and Wang, C. Y. (1999). Nonparametric correction to errors in covariates. Technical Report, Fred Hutchinson Cancer Research Center, Seattle.

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

Huang, Y. and Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *J. Amer. Statist. Assoc.* **96**, 1469-1482.

Huang, Y. and Wang, C. Y. (2006). Errors-in-covariates effect on estimating functions: Additivity in limit and nonparametric correction. *Statist. Sinica* **16**, 861-881.

Kong, F. H. and Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statist. Sinica* **9**, 953-969.

Ma, Y. and Tsiatis, A. A. (2006). Closed form semiparametric estimators for measurement error models. *Statist. Sinica* **16**, 183-193.

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

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**, 829-838.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman and Hall/CRC, Boca Raton.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331-342.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Small, C. G., Wang, J. and Yang, Z. (2000). Eliminating multiple root problems in estimation (with discussion). *Statist. Sci.* **15**, 313-341.

Song, X. and Huang, Y. (2005). On corrected score approach for proportional hazards model with covariates measured with error. *Biometrics* **61**, 702-714.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. Statist. Ser. 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.

Stefanski, L. A. and Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. *J. Amer. Statist. Assoc.* **90**, 1247-1256.

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

Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, 1518 Clifton Rd. NE, Atlanta, GA 30322, U.S.A.

E-mail: yhuang5@emory.edu

(Received June 2012; accepted December 2012)