Corrected score with sizable covariate measurement error: pathology and remedy

Yijian Huang, Emory University

Journal Title: Statistica Sinica
Publisher: Academia Sinica, Institute of Statistical Science | 2014
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.5705/ss.2012.181
Permanent URL: https://pid.emory.edu/ark:/25593/pfsqv

Final published version: http://dx.doi.org/10.5705/ss.2012.181

Accessed November 2, 2017 9:43 PM EDT
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 (2000). 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 \( X \) as the true covariates. The loglinear model postulates

\[
E(Y \mid X) = \exp \left( \alpha + \beta^\top X \right),
\]

(2.1)

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

\[
n^{-1} \sum_{i=1}^{n} \left\{ Y_i \left( a + b^\top X_i \right) - \exp \left( a + b^\top X_i \right) \right\},
\]

(2.2)

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

\[
n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \exp \left( a + b^\top X_i \right) \right\} \left( \frac{1}{X_i} \right); \quad (2.3)
\]

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 \( \beta \),

\[
\frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} Y_i} - \frac{\sum_{i=1}^{n} X_i \exp \left( b^\top X_i \right)}{\sum_{i=1}^{n} \exp \left( b^\top X_i \right)}. \quad (2.4)
\]

In practice, \( \alpha \) is typically not of as much interest as \( \beta \).

2.1. Corrected score

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

\[
W = X + \epsilon, \quad \epsilon \perp \!
\!
\!\perp \{X, Y\},
\]

(2.5)

where \( \epsilon \) is the error vector and \( \perp \!
\!
\!\perp \) denotes statistical independence. The elements of \( \epsilon \) corresponding to accurately measured covariates, if any, are zeros. To focus on the main issues, we suppose that the distribution of \( \epsilon \) is completely known.
Consider now an iid sample, \{(Y_i, 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 \( \varepsilon \) as \( \Omega(b) \equiv \ln E\{\exp(b^\top \varepsilon)\} \) and its derivative as \( \dot{\Omega}(b) \equiv \partial\Omega(b)/\partial b \); in the normal error case, \( \Omega(b) = b^\top E(\varepsilon) + b^\top \text{var}(\varepsilon) b/2 \) and \( \dot{\Omega}(b) = E(\varepsilon) + \text{var}(\varepsilon) b \). The corrected objective function (Nakamura (1990)) is

\[
n^{-1} \sum_{i=1}^{n} \left( Y_i \left[ a + \left. b^\top \left\{ W_i - \dot{\Omega}(0) \right\} \right. \right] - \exp \left\{ a + b^\top W_i - \Omega(b) \right\} \right), \quad (2.6)
\]

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

\[
\eta(a, b) \equiv n^{-1} \sum_{i=1}^{n} \eta_i(a, b), \quad (2.7)
\]

where

\[
\eta_i(a, b) \equiv Y_i \left\{ W_i - \dot{\Omega}(0) \right\} - \exp \left\{ a + b^\top W_i - \Omega(b) \right\} \left\{ \frac{1}{W_i} - \dot{\Omega}(b) \right\}.
\]

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 \((\alpha, \beta)^\top\). The estimator is asymptotically normal.

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

\[
\xi(b) = \frac{\sum_{i=1}^{n} Y_i W_i}{\sum_{i=1}^{n} Y_i} - \frac{\sum_{i=1}^{n} W_i \exp(b^\top W_i) - \Omega(b)}{\sum_{i=1}^{n} \exp(b^\top W_i)} + \dot{\Omega}(b) - \dot{\Omega}(0), \quad (2.8)
\]

which has the same limit as profile score (2.4) in a compact parameter space containing \( \beta \) (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 $\hat{\Omega}(b) - \hat{\Omega}(0) = b\Lambda$. 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 $\hat{\Omega}(b) - \hat{\Omega}(0)$ turns out to be the opposite regardless of the sample size,

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

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 $Y \sim \text{Poisson } \{\exp(\beta X)\}$ with $E(X) = 0$, $\beta > 0$, and $\varepsilon \sim \text{Normal with mean } 0$, based on 1,000 simulation iterations.

<table>
<thead>
<tr>
<th>size</th>
<th>$\var(\varepsilon)/\var(X) = 0.5$</th>
<th>$\var(\varepsilon)/\var(X) = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta^2 \var(X) = 0.5$</td>
<td>$\beta^2 \var(X) = 1$</td>
</tr>
<tr>
<td></td>
<td>$\beta^2 \var(X) = 1$</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>$2.1$ 2.74 4.4 2.56 0.6 2.81</td>
<td>$15.0$ 2.66 22.2 2.00 16.4 2.64</td>
</tr>
<tr>
<td>200</td>
<td>$0.0$ 3.72 0.2 2.64 0.1 2.08</td>
<td>$4.3$ 2.98 8.4 2.49 12.7 3.65</td>
</tr>
<tr>
<td>400</td>
<td>$0.0$ 3.12 0.0 1.44 0.1 2.87</td>
<td>$0.8$ 2.82 3.0 2.84 9.4 2.72</td>
</tr>
<tr>
<td>800</td>
<td>$0.0$ 2.53 0.0 1.74 0.1 3.31</td>
<td>$0.1$ 2.57 0.0 1.91 8.4 3.17</td>
</tr>
</tbody>
</table>

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

\[\max \hat{\beta} - \beta \over \beta - \min \hat{\beta},\]  

(2.10)

where $\hat{\beta}$ 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 $\var(X)$, $\var(\varepsilon)$, and $\beta$ once $\var(\varepsilon)/\var(X)$ and $\beta^2 \var(X)$ are fixed. Also, not only $\var(\varepsilon)/\var(X)$ but
also $\beta^2 \text{var}(X)$ are relevant in the discussion of pathological behaviors. When \( \text{var}(\varepsilon) \) and \( \text{var}(X) \) are comparable, the pathological behaviors of the corrected score could nonetheless be negligible if $\beta^2 \text{var}(X)$ is sufficiently small. Still, they could be of enormous concern even with a small \( \text{var}(\varepsilon) \) relative to \( \text{var}(X) \) if $\beta^2 \text{var}(X)$ 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 (2.21) holds whenever the error distribution is not bounded at both extremes, because $\hat{\Omega}(b) - \hat{\Omega}(0)$ 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 $\hat{\Omega}(b) \equiv \partial \hat{\Omega}(b)/\partial b^T$. The first derivative $\partial \hat{\eta}(a, b)/\partial (a, b^T)$ is

$$
\hat{\eta}(a, b) = -n^{-1} \sum_{i=1}^{n} \exp\left\{ a + b^T W_i - \Omega(b) \right\} \\
\times \left[ \left\{ \frac{1}{W_i - \hat{\Omega}(b)} \right\}^\otimes 2 \right] - \text{diag} \left\{ 0, \hat{\Omega}(b) \right\},
$$

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

**Proposition 1.** Suppose that the loglinear model (2.11) and the classical additive measurement error model (2.5) hold. Writing $b \equiv (b_1, \ldots, b_p)^T$, if the moment-
generating functions of $X$ and $\varepsilon$ exist, then
\[
E \left[ \begin{array}{c}
Y \\
\frac{\partial^{k_1+\ldots+k_p} \exp \left\{ \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b}) \right\}}{\partial b_1^{k_1} \ldots \partial b_p^{k_p}} \\
- \frac{\partial^{k_1+\ldots+k_p} \exp \left\{ \alpha + \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b}) \right\}}{\partial b_1^{k_1} \ldots \partial b_p^{k_p}}
\end{array} \bigg|_{\mathbf{b}=\theta} \right] = 0,
\]
for $k_q \geq 0, q = 1, \ldots, p$.

**Proof.** Given $E[\exp\{a + \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b})\} \mid X] = \exp(a + \mathbf{b}^\top X)$ under (2.3), one obtains directly that
\[
E \left[ \begin{array}{c}
\frac{\partial^{k_1+\ldots+k_p} \exp \left\{ a + \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b}) \right\}}{\partial b_1^{k_1} \ldots \partial b_p^{k_p}} \\
\frac{\partial^{k_1+\ldots+k_p} \exp \left\{ \alpha + \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b}) \right\}}{\partial b_1^{k_1} \ldots \partial b_p^{k_p}}
\end{array} \bigg|_{\mathbf{b}=\theta} \right] = \begin{pmatrix} X \\ X \end{pmatrix} = \prod_{q=1}^p X_q^{k_q}.
\]
Then,
\[
E \left[ \begin{array}{c}
\frac{\partial^{k_1+\ldots+k_p} \exp \left\{ a + \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b}) \right\}}{\partial b_1^{k_1} \ldots \partial b_p^{k_p}} \\
\frac{\partial^{k_1+\ldots+k_p} \exp \left\{ \alpha + \mathbf{b}^\top \mathbf{W} - \Omega(\mathbf{b}) \right\}}{\partial b_1^{k_1} \ldots \partial b_p^{k_p}}
\end{array} \bigg|_{\mathbf{b}=\theta} \right] = \prod_{q=1}^p X_q^{k_q}
\]
follows. These two equations, coupled with (2.1), imply (3.2).

Equation (3.2) leads to a class of estimating functions for $(\alpha, \beta^\top)^\top$. The corrected score $\eta(a, b)$, as a special case, corresponds to $\sum_{q=1}^p k_q = 0$ and 1. Similarly, the estimating function, as a symmetric matrix,
\[
n^{-1} \sum_{i=1}^n \left( Y_i \left[ \left\{ \mathbf{W}_i - \hat{\Omega}(0) \right\}^\otimes 2 - \tilde{\Omega}(0) \right] - \exp \left\{ a + \mathbf{b}^\top \mathbf{W}_i - \Omega(\mathbf{b}) \right\} \left[ \left\{ \mathbf{W}_i - \hat{\Omega}(0) \right\}^\otimes 2 - \tilde{\Omega}(0) \right] \right),
\]
corresponds to $\sum_{q=1}^p k_q = 2$. In light of (3.1), having both $\eta(a, b)$ and (3.3) equal to zero leads to the constraint
\[
\dot{\eta}(a, b) = -n^{-1} \sum_{i=1}^n Y_i \left[ \left\{ \mathbf{W}_i - \hat{\Omega}(0) \right\}^\otimes 2 - \text{diag} \left\{ 0, \tilde{\Omega}(0) \right\} \right].
\]
The right-hand side is independent of $(a, \beta^\top)^\top$, and almost surely negative definitive since
\[
E \left[ \left\{ \mathbf{W} - \hat{\Omega}(0) \right\}^\otimes 2 - \text{diag} \{0, \tilde{\Omega}(0)\} \mid X \right] = \left( \frac{1}{X} \right)^\otimes 2.
\]
under (2.3). Thus, when incorporating (3.3), the first derivative of $\eta(a, b)$ 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 $\sum_{q=1}^{p} k_q \tau_q > 2$, when additionally incorporated, would effectively impose constraints on higher-order derivatives of $\eta(a, b)$ 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 $\varphi(a, b) = n^{-1} \sum_{i=1}^{n} \varphi_i(a, b)$. The estimator is the global maximizer of the empirical likelihood ratio function

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

(3.4)

algorithms for the computation can be found in Owen (2001). For our problem, $\varphi(a, b)$ consists of $\eta(a, b)$ 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 $\varphi(a, b) = \eta(a, b)$. Write $\hat{w}_i$ as the value of $w_i$ for the maximization in (3.4) with given $a, b^\top$. The estimate is chosen as a local maximizer of $\mathcal{L}(a, b)$ such that $\sum_{i=1}^{n} \hat{w}_i \eta_i(a, b)$ 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

$$
n^{-1} \left[ E\varphi_1(a, \beta)^\top E \{ \varphi_1(a, \beta)^\otimes 2 \}^{-1} E\varphi_1(a, \beta) \right]^{-1},
$$

(3.5)
under mild moment conditions. The estimator reaches the asymptotic efficiency bound among those based on linear combinations of elements in $\varphi(a, b)$. 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 $\varepsilon$ are normal with mean 0, and the intercept $\alpha$ 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 $\text{var}(\varepsilon)/\text{var}(X)$ and $\beta^2\text{var}(X)$. When $\text{var}(\varepsilon)/\text{var}(X)$ or $\beta^2\text{var}(X)$ approaches 0, the asymptotic relative efficiency is close to 1 as expected. But it can be well over 3 when $\text{var}(\varepsilon) = \text{var}(X) = 1$ 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; 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 $\alpha = 0$.

<table>
<thead>
<tr>
<th>$\text{var}(\varepsilon)/\text{var}(X)$</th>
<th>$\beta^2 \text{var}(X)$</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>1.00</td>
<td>1.01</td>
<td>1.06</td>
<td>1.07</td>
<td>1.24</td>
</tr>
<tr>
<td>0.4</td>
<td>1.00</td>
<td>1.02</td>
<td>1.05</td>
<td>1.18</td>
<td>1.24</td>
</tr>
<tr>
<td>0.6</td>
<td>1.00</td>
<td>1.04</td>
<td>1.10</td>
<td>1.32</td>
<td>1.46</td>
</tr>
<tr>
<td>0.8</td>
<td>1.01</td>
<td>1.08</td>
<td>1.18</td>
<td>1.46</td>
<td>1.73</td>
</tr>
<tr>
<td>1</td>
<td>1.02</td>
<td>1.11</td>
<td>1.27</td>
<td>1.61</td>
<td>2.05</td>
</tr>
</tbody>
</table>

The two columns correspond to estimators of $\alpha$ and $\beta$.

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 $\varepsilon$ 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 $0.4$. 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 $X$ was standard bivariate normal with correlation coefficient $0.5$. 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
with \( \mathbf{X}_i \) replaced by its best linear approximate given \( \mathbf{W}_i \),

\[
\mathbf{W} - \hat{\Omega}(0) + (\hat{\Sigma} - \bar{\Omega}(0))\hat{\Sigma}^{-1}(\mathbf{W}_i - \mathbf{W}),
\]

where \( \mathbf{W} \) and \( \hat{\Sigma} \) 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 \( \beta \) because \( \alpha \) 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 \( \beta \) 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:$k$), $k = 1, 2, 3$.

<table>
<thead>
<tr>
<th>size</th>
<th>NV B</th>
<th>E</th>
<th>RC B</th>
<th>E</th>
<th>ConS F</th>
<th>B</th>
<th>E</th>
<th>CS B</th>
<th>E</th>
<th>TC:1 B</th>
<th>E</th>
<th>TC:2 B</th>
<th>E</th>
<th>TC:3 B</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>100α</td>
<td>241</td>
<td>277</td>
<td>244</td>
<td>297</td>
<td>0</td>
<td>-9</td>
<td>184</td>
<td>-87</td>
<td>304</td>
<td>101</td>
<td>236</td>
<td>148</td>
<td>289</td>
<td>215</td>
<td>316</td>
</tr>
<tr>
<td>β</td>
<td>-504</td>
<td>514</td>
<td>63</td>
<td>335</td>
<td>38</td>
<td>245</td>
<td>132</td>
<td>330</td>
<td>-93</td>
<td>301</td>
<td>-179</td>
<td>577</td>
<td>-291</td>
<td>551</td>
<td></td>
</tr>
<tr>
<td>200α</td>
<td>243</td>
<td>261</td>
<td>245</td>
<td>268</td>
<td>0</td>
<td>-5</td>
<td>125</td>
<td>-122</td>
<td>284</td>
<td>58</td>
<td>156</td>
<td>81</td>
<td>183</td>
<td>162</td>
<td>244</td>
</tr>
<tr>
<td>β</td>
<td>-503</td>
<td>508</td>
<td>23</td>
<td>184</td>
<td>18</td>
<td>164</td>
<td>156</td>
<td>345</td>
<td>-44</td>
<td>186</td>
<td>-35</td>
<td>282</td>
<td>-166</td>
<td>353</td>
<td></td>
</tr>
<tr>
<td>400α</td>
<td>249</td>
<td>258</td>
<td>250</td>
<td>262</td>
<td>0</td>
<td>3</td>
<td>85</td>
<td>-118</td>
<td>270</td>
<td>33</td>
<td>113</td>
<td>37</td>
<td>125</td>
<td>103</td>
<td>173</td>
</tr>
<tr>
<td>β</td>
<td>-502</td>
<td>505</td>
<td>7</td>
<td>128</td>
<td>5</td>
<td>107</td>
<td>145</td>
<td>333</td>
<td>-23</td>
<td>147</td>
<td>6</td>
<td>163</td>
<td>-72</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>800α</td>
<td>250</td>
<td>254</td>
<td>250</td>
<td>255</td>
<td>0.1</td>
<td>-1</td>
<td>61</td>
<td>-101</td>
<td>231</td>
<td>17</td>
<td>75</td>
<td>11</td>
<td>81</td>
<td>53</td>
<td>111</td>
</tr>
<tr>
<td>β</td>
<td>-500</td>
<td>502</td>
<td>5</td>
<td>90</td>
<td>5</td>
<td>78</td>
<td>120</td>
<td>282</td>
<td>-15</td>
<td>100</td>
<td>23</td>
<td>118</td>
<td>-19</td>
<td>123</td>
<td></td>
</tr>
</tbody>
</table>

Case B

| size | NV α | 195 | 264 | 218 | 322 | 0.6 | -227 | 370 | -103 | 313 | 67  | 241 | 88  | 262 | 148 | 306 |
| β    | -285 | 301 | 570 | 795 | 23 | 167 | 259 | 419 | -5   | 251 | -143 | 632 | -320 | 696 |
| 200α | 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 |
| 400α | 219  | 235 | 224 | 246 | 0  | 5   | 79  | -177 | 382 | 49   | 113 | 34   | 119 | 83  | 168 |
| β    | -299 | 302 | 428 | 462 | 2  | 7   | 123 | 273 | 6    | 99  | 69   | 150 | 41   | 207 |
| 800α | 223  | 231 | 226 | 236 | 0  | 1   | 53  | -93  | 244 | 26   | 75  | 8    | 70  | 24  | 88  |
| β    | -301 | 303 | 410 | 428 | 2  | 5   | 64  | 176 | 0    | 61  | 48   | 99  | 72  | 136 |

Case C

| size | NV 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 |
| 200α | 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 |
| 400α | 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 |
| 800α | 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 $\beta$ 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).

<table>
<thead>
<tr>
<th>size</th>
<th>NV</th>
<th>RC</th>
<th>ConS</th>
<th>CS</th>
<th>TC:1/3</th>
<th>TC:2/3</th>
<th>TC:1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>E</td>
<td>B</td>
<td>E</td>
<td>B</td>
<td>E</td>
<td>B</td>
</tr>
<tr>
<td>100</td>
<td>-18 241</td>
<td>0.1</td>
<td>-55 254</td>
<td>1 339</td>
<td>76 272</td>
<td>140 226</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>-102 244</td>
<td>-38 298</td>
<td>37 197</td>
<td>89 172</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-118 240</td>
<td>-48 218</td>
<td>14 129</td>
<td>54 113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>168 335</td>
<td>55 260</td>
<td>15 183</td>
<td>62 165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-111 257</td>
<td>-76 257</td>
<td>21 180</td>
<td>26 122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>146 308</td>
<td>64 224</td>
<td>2 130</td>
<td>27 118</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-99 225</td>
<td>-72 216</td>
<td>24 121</td>
<td>9 87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TC:1/3 and TC:2/3 correspond to using the (1,1)-th element and the first row of matrix $(\mathbf{S}_2)$ 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.

<table>
<thead>
<tr>
<th>size</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>79.8</td>
<td>79.8</td>
<td>95.4</td>
<td>85.1</td>
<td>86.0</td>
<td>95.4</td>
<td>89.3</td>
<td>89.0</td>
<td>93.9</td>
<td>93.1</td>
<td>93.1</td>
<td>93.7</td>
</tr>
<tr>
<td>200</td>
<td>81.0</td>
<td>89.5</td>
<td>93.5</td>
<td>86.8</td>
<td>90.5</td>
<td>94.1</td>
<td>89.7</td>
<td>91.6</td>
<td>94.0</td>
<td>93.0</td>
<td>94.4</td>
<td>93.7</td>
</tr>
<tr>
<td>400</td>
<td>83.3</td>
<td>79.4</td>
<td>94.1</td>
<td>85.4</td>
<td>85.2</td>
<td>96.0</td>
<td>89.2</td>
<td>88.6</td>
<td>95.4</td>
<td>91.3</td>
<td>91.2</td>
<td>95.4</td>
</tr>
<tr>
<td>800</td>
<td>87.9</td>
<td>84.9</td>
<td>90.9</td>
<td>91.4</td>
<td>87.8</td>
<td>93.4</td>
<td>93.1</td>
<td>90.8</td>
<td>93.9</td>
<td>94.3</td>
<td>93.9</td>
<td>94.7</td>
</tr>
</tbody>
</table>

Single-covariate model: Case A

<table>
<thead>
<tr>
<th>size</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>83.7</td>
<td>84.7</td>
<td>95.3</td>
<td>88.0</td>
<td>89.2</td>
<td>94.5</td>
<td>91.3</td>
<td>90.7</td>
<td>93.3</td>
<td>91.9</td>
<td>91.4</td>
<td>94.6</td>
</tr>
<tr>
<td>200</td>
<td>81.9</td>
<td>88.4</td>
<td>93.7</td>
<td>87.0</td>
<td>91.9</td>
<td>95.0</td>
<td>92.0</td>
<td>93.3</td>
<td>94.5</td>
<td>92.3</td>
<td>94.5</td>
<td>93.6</td>
</tr>
</tbody>
</table>

Single-covariate model: Case B

<table>
<thead>
<tr>
<th>size</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>72.0</td>
<td>73.8</td>
<td>96.7</td>
<td>75.8</td>
<td>76.9</td>
<td>95.2</td>
<td>84.3</td>
<td>85.7</td>
<td>95.1</td>
<td>90.1</td>
<td>90.4</td>
<td>95.0</td>
</tr>
<tr>
<td>200</td>
<td>70.3</td>
<td>78.2</td>
<td>93.4</td>
<td>79.4</td>
<td>85.1</td>
<td>95.0</td>
<td>88.0</td>
<td>89.8</td>
<td>96.0</td>
<td>91.3</td>
<td>93.3</td>
<td>95.4</td>
</tr>
<tr>
<td>400</td>
<td>80.3</td>
<td>86.3</td>
<td>91.9</td>
<td>84.5</td>
<td>90.5</td>
<td>94.3</td>
<td>89.3</td>
<td>90.9</td>
<td>93.9</td>
<td>93.2</td>
<td>93.6</td>
<td>93.6</td>
</tr>
</tbody>
</table>

Double-covariate model

<table>
<thead>
<tr>
<th>size</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
<th>W</th>
<th>EL</th>
<th>ELC</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>79.8</td>
<td>79.8</td>
<td>95.4</td>
<td>85.1</td>
<td>86.0</td>
<td>95.4</td>
<td>89.3</td>
<td>89.0</td>
<td>93.9</td>
<td>93.1</td>
<td>93.1</td>
<td>93.7</td>
</tr>
<tr>
<td>200</td>
<td>81.0</td>
<td>89.5</td>
<td>93.5</td>
<td>86.8</td>
<td>90.5</td>
<td>94.1</td>
<td>89.7</td>
<td>91.6</td>
<td>94.0</td>
<td>93.0</td>
<td>94.4</td>
<td>93.7</td>
</tr>
<tr>
<td>400</td>
<td>83.3</td>
<td>79.4</td>
<td>94.1</td>
<td>85.4</td>
<td>85.2</td>
<td>96.0</td>
<td>89.2</td>
<td>88.6</td>
<td>95.4</td>
<td>91.3</td>
<td>91.2</td>
<td>95.4</td>
</tr>
<tr>
<td>800</td>
<td>87.9</td>
<td>84.9</td>
<td>90.9</td>
<td>91.4</td>
<td>87.8</td>
<td>93.4</td>
<td>93.1</td>
<td>90.8</td>
<td>93.9</td>
<td>94.3</td>
<td>93.9</td>
<td>94.7</td>
</tr>
</tbody>
</table>

Single-covariate model: Case C

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

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


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)