On the Conditional Score and Corrected Score Estimation in Nonlinear Measurement Error Models

By CHI-LUN CHENG

Institute of Statistical Science, Academia Sinica

Abstract. This paper reviews the conditional score and corrected score estimation of the unknown parameters in nonlinear measurement error (errors-in-variables) models. This includes the functional and structural models. The connection among these methodologies and total least squares (TLS) is also examined. A compendium of existing results as well as some possible extensions are discussed.

1. Introduction

The ordinary regression models assume that the independent variables are measured without error. However, in many situations, the independent variables cannot measured precisely. When the measurement error is too large to ignore, the estimators for the regression parameters are biased and inconsistent. Measurement error (ME) models are important alternatives for ordinary regression models, in which we assume that relation between the independent variable y and independent variable $\boldsymbol{\xi}$ is known but one cannot observe $\boldsymbol{\xi}$ directly. Instead, one observes $\mathbf{x} = \boldsymbol{\xi} + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is the measurement error with mean zero and is independents of $\boldsymbol{\xi}$. If the $\boldsymbol{\xi}_i$ are unknown constants, the model is known as a functional model; whereas, if the $\boldsymbol{\xi}_i$ are independent identically distributed random variables, the model is known as a *structural model*. We also assume that $\boldsymbol{\xi}$ is independent of the measurement error $\boldsymbol{\delta}$ in the structural model.

More general terms suggested by Carroll, Ruppert and Stefanski (1995) seem to be useful. The term *functional modelling* is referred to the situation in which the true (latent) variable $\boldsymbol{\xi}$ is either fixed or random, but if it is random, then no, or at least minimal, distributional assumptions are imposed on $\boldsymbol{\xi}$. On the other hand, *structural modelling* refers to the situation in which $\boldsymbol{\xi}$ is random, and distributional assumptions, usually parametric, are made on the true variable $\boldsymbol{\xi}$.

There are several general methodologies proposed in the literature to estimate the regression parameters in nonlinear measurement error models. We are interested in two of them. The first one is the *conditional score* method that was proposed by Stefanski and Carroll (1987). The second one is called *corrected score* method, which was proposed by Stefanski (1989) and Nakamura (1990) independently.

Another important issue is the types of measurement error. We will follow the definition made by Carroll *et al.* (1995) and define the nondifferential measurement error meaning that $\mathbf{x}$ has no information of y other than what is available in $\boldsymbol{\xi}$. That is, the conditional distribution of y given $(\boldsymbol{\xi}, \mathbf{x})$ depends only on $\boldsymbol{\xi}$ and hence $\boldsymbol{\delta}$ has nothing to do with the distribution of y. In this case, $\mathbf{x}$ is called a surrogate. Putting in symbol, the nondifferential measurement error means that $f(y \mid \mathbf{x}, \boldsymbol{\xi}) = f(y \mid \boldsymbol{\xi})$, which leads to the following two other equivalent expressions (Liang and Liu, 1991), namely,

$$f(y, \mathbf{x} \mid \xi) = f(y \mid \boldsymbol{\xi}) f(\mathbf{x} \mid \boldsymbol{\xi}),$$

or

$$f(\mathbf{x} \mid \boldsymbol{\xi}, y) = f(\mathbf{x} \mid \boldsymbol{\xi}),$$

where f denotes the appropriate density (mass) function. The measurement is differential otherwise.

In this paper, we will review these two methods. In our view, they have some fundamental difference in their assumptions that has been neglected in the literature. We will also bring some recent developments to attention and some possible extensions are discussed. Finally the connection between the conditional score method and **TLS (total least squares)** is addressed.

1.1. Generalized linear models

In this subsection, we briefly discuss the generalized linear models that will be the focus on the conditional and corrected score estimation methods. It should be pointed that the conditional and corrected score estimation methods are not necessarily restricted to generalized linear models. We will return to this point later.

An important class of nonlinear ME models is the **generalized linear models (GLM's)** (McCullagh and Nelder, 1989). These models have received considerable attention lately, partly because they are very useful in many disciplines such as biostatistics. These are important models and deserves special attention. Assuming we observed $(y_i, \mathbf{x}_i')$ and $\mathbf{x}_i$ are subject to measurement error:

$$\mathbf{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i,$$

for $i = 1, \ldots, n$ and y is a scalar and $\mathbf{x}$ is a $p \times 1$ vector. The generalized linear models considered in the present article are with "canonical links", so that y has density or mass function

$$f(y \mid \boldsymbol{\xi}, \boldsymbol{\tau}) = \exp\{\frac{y\boldsymbol{\tau} - c(\boldsymbol{\tau})}{\phi} + h(y, \phi)\}, \tag{1}$$

where $\tau = \boldsymbol{\beta}'(1, \boldsymbol{\xi}')' = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi}$ is called the **natural (canonical) parameter**. The functions $c(\cdot)$ and $h(\cdot)$ are assumed to be known and ϕ is called **dispersion** parameter. The **regression parameter** $\boldsymbol{\theta}' =$

$(\beta_0, \boldsymbol{\beta}')$, associated with the dispersion parameter ϕ is the unknown parameter to be estimated. Note that in some situations the dispersion parameter ϕ is known.

$$E(\partial L/\partial \tau) = 0,$$

$$E(\partial^2 L/\partial \tau^2) + E(\partial L/\partial \tau)^2 = 0,$$

which give

$$E(y \mid \xi) = c'(\tau) = \mu,$$

and

$$var(y \mid \xi) = \phi c''(\tau) = \phi V(\mu),$$

2. Conditional Score Estimating Functions

In this section, we will investigate *conditional score* estimation method, which was introduced Stefanski and Carroll (1987). Essentially, the conditional score method is applied to a class of generalized linear measurement models with canonical form when the independent variable is measurement with error. The basic assumptions are, the measurement error $\boldsymbol{\delta}$ is normally distributed and its variance $\boldsymbol{\Sigma}_{\delta\delta} = \phi\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is known. Note that $\boldsymbol{\Sigma}_{\delta\delta}$ could be less than full rank, which means that some of the components of $\boldsymbol{\delta}$ are measured without error. The latent variable $\boldsymbol{\xi}$ is fixed, that is, the functional model is considered here.

The following discussions are based on the work by Stefanski and Carroll (1987) and Carroll, Ruppert and Stefanski (1995, Chapter 6).

Consider the generalized linear model defined by (1). The parameter of interest is $\boldsymbol{\Psi} = (\boldsymbol{\theta}', \phi)$, which can be estimated by solving the unbiased estimating equations (recall that $c'(\tau)$ and $\phi c''(\tau)$ are the mean and variance, respectively)

$$\sum \{y_i - c'(\tau_i)\} \begin{pmatrix} 1 \\ \boldsymbol{\xi}_i \end{pmatrix} = 0 \tag{2}$$

$$\sum \left[\phi - \frac{\{y_i - c'(\tau_i)\}^2}{c''(\tau_i)} \right] = 0, \tag{3}$$

if $\boldsymbol{\xi}_i$ were known.

If we regard $\boldsymbol{\xi}$ as an unknown parameter and all other parameters are assumed to be known, Stefanski and Carroll (1987) show that $\boldsymbol{\pi} = \mathbf{x} + y\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/\phi$ is a complete sufficient statistics for $\boldsymbol{\xi}$. Moreover, the conditional distribution of y given $\boldsymbol{\pi}$ has the form of generalized linear model (1) with $\boldsymbol{\xi}$ replaced by $\boldsymbol{\pi}$ and the following substitutions

$$\tau_* = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\pi};$$

$$h_*(y, \phi, \boldsymbol{\beta}'\Sigma_{\delta\delta}\boldsymbol{\beta}) = h(y, \phi) - (1/2)(y/\phi)^2 \boldsymbol{\beta}'\Sigma_{\delta\delta}\boldsymbol{\beta};$$

$$c_*(\tau_*, \phi, \boldsymbol{\beta}'\Sigma_{\delta\delta}\boldsymbol{\beta}) = \phi \; log \left[\int exp\{y\tau_*/\phi + h_*(y, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})\} d\mu(y) \right],$$

where the integral is a sum if y is discrete and an integral otherwise. In other words, the conditional density or mass function is

$$f(y \mid \boldsymbol{\pi}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}_{\delta\delta}) = exp\{\frac{y\tau_* - c_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})}{\phi} + h_*(y, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})\}$$

To find the unbiased estimating functions for the parameter Ψ, we can use (2)-(3) above with the change of the mean m_* and variance v_* by

$$m_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = {c_*}'(\tau_*) = \frac{\partial}{\partial \tau_*} c_*,$$

$$\phi v_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = \phi c_*''(\tau_*) = \phi \frac{\partial}{\partial {\tau_*}^2} c_*. \tag{4}$$

To be more precise, the estimators of $\mathbf{\Psi}' = (\beta_0, \boldsymbol{\beta}', \phi)$ can be obtained by solving

$$\sum \{y_i - m_*(\tau_*, \phi, \boldsymbol{\beta}'\mathbf{\Sigma}_{\delta\delta}\boldsymbol{\beta})\} \begin{pmatrix} 1 \\ \boldsymbol{\pi}_i \end{pmatrix} = 0 \tag{5}$$

$$\sum \left[\phi - \frac{\{y_i - m_*(\tau_*, \phi, \boldsymbol{\beta}'\mathbf{\Sigma}_{\delta\delta}\boldsymbol{\beta})\}^2}{v_*(\tau_*, \phi, \boldsymbol{\beta}'\mathbf{\Sigma}_{\delta\delta}\boldsymbol{\beta})} \right] = 0, \tag{6}$$

It is conceivable that solving (5)-(6) is much more difficulty than solving (2)-(3).

Example 1. (Stefanski and Carroll, 1987)
Linear functional model. If y is normally distributed with mean $\beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi}$ and variance $\phi = \sigma_\varepsilon^2$, then

$$m_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = \frac{\tau_*}{1 + \phi^{-1}\boldsymbol{\beta}'\Sigma_{\delta\delta}\boldsymbol{\beta}};$$

$$v_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = \frac{1}{1 + \phi^{-1}\boldsymbol{\beta}'\Sigma_{\delta\delta}\boldsymbol{\beta}}.$$

This example is basically for illustration purpose because the estimation in linear model has been discussed widely and one does not need to use this complicated method to find consistent estimator.

For simple functional model, it is easy to see that $\pi = x + \beta_1 y/\lambda$, where $\lambda = \sigma_\varepsilon^2/\sigma_\delta^2$, the ratio of error variances. Moreover, $m_* = \lambda\tau_*/(\lambda + \beta_1^2)$ and $v_* = \lambda/(\lambda + \beta_1^2)$. Then the estimating equations become

$$\sum (y_i - \beta_0 - \beta_1 x_i) = 0,$$

$$\sum (y_i - \beta_0 - \beta_1 x_i)(\lambda x_i + \beta_1 y_i) = 0,$$

$$\sum \left[\sigma_\varepsilon^2 - \frac{\lambda (y_i - \beta_0 - \beta_1 x_i)^2}{\lambda + \beta_1^2} \right] = 0.$$

The estimators of β_0 and σ_ε^2 are easily obtained and they are the usual method of moments estimators, namely,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\sigma}_\varepsilon^2 = n^{-1} \sum \frac{\lambda (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\lambda + \hat{\beta}_1^2}.$$

To estimate the slope β_1, the middle estimating equation is a quadratic form of β_1 and it yields the two solutions

$$\hat{\beta}_1 = \frac{s_{yy} - \lambda s_{xx} \pm \{(s_{yy} - \lambda s_{xx})^2 + 4\lambda s_{xy}^2\}^{1/2}}{2 s_{xy}}.$$

Unfortunately, the conditional estimation method does not indicate which root is appropriate, unlike the maximum likelihood and/or generalized least squares method. We will return to this point later in discussion section.

Example 2. (Stefanski and Carroll, 1987)
Logistic measurement-error model. Note that $\phi = 1$ in logistic regression model. The conditional mean and variance are

$$m_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = H(\tau_* - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/2);$$

$$v_*(\tau_*, \phi, \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = H^{(1)}(\tau_* - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/2),$$

where $H^{(1)} = H(1-H)$ is the logistic density function. Then the unbiased estimating equation becomes

$$\sum \{y_i - H(\beta_0 - \boldsymbol{\beta}'\boldsymbol{\pi}_i - 0.5\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})\} \begin{pmatrix} 1 \\ \boldsymbol{\pi}_i \end{pmatrix} = 0.$$

Again, there are multiple solutions to the equation above and they have to solved numerically.

Generally speaking, the conditional score unbiased estimating equation depends on the first two moments, namely equation (4), of y given $\boldsymbol{\pi}$ and there are no closed forms of them except in some rare cases such as linear and logistic measurement-error models. Therefore, numerical integration or summation is needed to find these moments. Next example is a typical case of such occasion.

Example 3. (Stefanski and Carroll, 1987)
Poisson measurement-error model with $\phi = 1$. Then

$$c_*(\tau_*, \phi, \boldsymbol{\beta}'\Sigma_{\delta\delta}\boldsymbol{\beta}) = log\left\{\sum_{y=0}^{\infty}(y!)^{-1}exp(y\tau_* - y^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/2)\right\},$$

Carroll et al. (1995, p. 129) gave the formula to find the mean and variance, which are the first and second derivative of c_* with respect to τ_*, respectively. They are $m_* = s_1$ and $v_* = s_2 - s_1^2$, where

$$s_j = E(y^j \mid \boldsymbol{\pi}) = \frac{\sum_{y=0}^{\infty} y^j (y!)^{-1} exp(y\tau_* - y^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/2)}{\sum_{y=0}^{\infty}(y!)^{-1}exp(y\tau_* - y^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/2)}.$$

2.1. Discussion

The key feature of conditional score methodology is to observe that a complete sufficient statistic for $\boldsymbol{\xi}$, namely, $\boldsymbol{\pi} = \mathbf{x} + y\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}/\phi$, when all other parameters are fixed. Then the distribution of y given $\boldsymbol{\pi}$ is independent of $\boldsymbol{\xi}$. This method applied to the logistic, Poisson, and gamma measurement-error models with normal measurement error. The normality assumption is not crucial, at least in theory, but this has not been shown in the literature. However, the existence of the a complete sufficient statistic for $\boldsymbol{\xi}$ is crucial.

3. Corrected Score Estimating Functions

The *corrected score* estimation method was introduced independently by Stefanski (1989) and Nakamura (1990). The basic assumptions are, the measurement error $\boldsymbol{\delta}$ is normally distributed and its variance $\boldsymbol{\Sigma}_{\delta\delta}$ is known. Note that $\boldsymbol{\Sigma}_{\delta\delta}$ could be less than full rank, which means that some of the components of $\boldsymbol{\delta}$ are measured without error. The latent variable $\boldsymbol{\xi}$ could be fixed or random and no distributional assumption is imposed in the latter case. In other words, this estimation method is a typical functional modelling methodology.

The following is based on the work by Stefanski (1989), Nakamura (1990) and Buzas and Stefanski (1996). Suppose that there exists an unbiased estimating function such that

$$E\boldsymbol{\psi}(y_i, \boldsymbol{\xi}_i; \boldsymbol{\theta}) = 0. \tag{7}$$

for all i. Further, suppose one can find a $\tilde{\boldsymbol{\psi}}$ such that

$$E\{\tilde{\boldsymbol{\psi}}(y_i, \mathbf{x}_i; \boldsymbol{\theta} \mid y_i, \boldsymbol{\xi}_i)\} = \boldsymbol{\psi}(y_i, \boldsymbol{\xi}_i; \boldsymbol{\theta}) \tag{8}$$

for all $(y_i, \boldsymbol{\xi}_i')$. Note that the conditional expectation in (8) is only respect to the distribution of the measurement error $\boldsymbol{\delta}$. Because $\mathbf{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i$ and taking expectation with respect to $\boldsymbol{\xi}_i$ and y_i again, one has

$$E\{\tilde{\boldsymbol{\psi}}(y_i, \mathbf{x}_i; \boldsymbol{\theta})\} = 0. \tag{9}$$

Equation (9) leads to solve the unbiased estimating function

$$\sum \tilde{\boldsymbol{\psi}}(y_i, \mathbf{x}_i; \boldsymbol{\theta}) = 0, \tag{10}$$

and the resulting estimator is known as the corrected score estimator.

To obtain (7) is quite easy because one can use the maximum likelihood score function or least squares in ordinary regression pretending that $\boldsymbol{\xi}_i$ are known. We will use the subscript ml, mm, and ls for $\boldsymbol{\psi}$ and $\tilde{\boldsymbol{\psi}}$ to denote the source, that is, the maximum likelihood induced, the method-of-moments, and the least squares induced, respectively, of the estimator. The main difficult is how to find $\tilde{\boldsymbol{\psi}}$ that satisfies (8). Such $\tilde{\boldsymbol{\psi}}$ might not exist at all and it is not easy to find even it exists. On the other hand, the corrected score estimation method provides a general methodology for functional modelling. We now illustrate the corrected score estimation method via following examples. By using the normal generating function, one can show that the following three identities, namely (recall that $\boldsymbol{\theta}' = (\beta, \boldsymbol{\beta}'))$

$$E\{exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x}) \mid \boldsymbol{\xi}\} = exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}),$$

$$E\{\mathbf{x}exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x}) \mid \boldsymbol{\xi}\} = (\boldsymbol{\xi} + \boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})exp(\boldsymbol{\beta}'\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}),$$

$$E\{\mathbf{x}\mathbf{x}'exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x}) \mid \boldsymbol{\xi}\} = \{\boldsymbol{\Sigma}_{\delta\delta} + (\boldsymbol{\xi} + \boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})(\boldsymbol{\xi} + \boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})'\}exp(\boldsymbol{\beta}'\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}),$$

hold true when $\boldsymbol{\delta}$ is normally distributed with mean zero and variance $\boldsymbol{\Sigma}_{\delta\delta}$.

Example 4

Linear measurement error model that is described as follows.

$$\mathbf{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \qquad y_i = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi}_i + \varepsilon_i,$$

with $i = 1, \ldots, n$. The well-known "method-of-moment" estimator in linear ME model can be treated as a corrected score estimator. Using least squares to regress y on $\boldsymbol{\xi}$ will result in the unbiased estimating function

$$\boldsymbol{\psi}_{ls}(y, \boldsymbol{\xi}; \boldsymbol{\theta}) = (y - \beta_0 - \boldsymbol{\beta}'\boldsymbol{\xi})(1, \boldsymbol{\xi}')'.$$

It is easy to verify that

$$\tilde{\boldsymbol{\psi}}_{ls}(y, \mathbf{x}; \boldsymbol{\theta}) = (y - \beta_0 - \boldsymbol{\beta}'\mathbf{x})(1, \mathbf{x})' + \boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}(0, 1)'$$

is the corrected score function satisfying (8). In this case, the estimator can be treated as either a least squares induced or method-of-moments estimator. Note that no distributional assumption needs to be imposed on the measurement error $\boldsymbol{\delta}$.

The example above is interesting because Nakamura (1990, p. 131) used normally distributed $\boldsymbol{\delta}$ to obtain the same estimator, although Nakamura pointed it out that the normality assumption of $\boldsymbol{\delta}$ is not necessary. Buzas and Stefanski (1996, p. 2) directly pointed that (9) is the corrected score function for the naive estimating function (7), which implicitly suggests that least squares, not likelihood being used.

Example 5 (Nakamura 1990)
Poisson regression with canonical link. Assuming $\boldsymbol{\xi}$ is fixed, the log-likelihood based on $(y, \boldsymbol{\xi}')$ is

$$l(\boldsymbol{\theta}) = \sum \{-exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi}_i) + y_i(\beta_0, \boldsymbol{\beta})'(1, \boldsymbol{\xi}_i) - log\ y_i!\},$$

Because $E(y \mid \boldsymbol{\xi}) = exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi})$ and the maximum likelihood score derived from $(y, \boldsymbol{\xi})$ is

$$\boldsymbol{\psi}_{ml}(y, \boldsymbol{\xi}; \boldsymbol{\theta}) = \{y - exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi})\}(1, \boldsymbol{\xi}')'.$$

Using the first two identities proved by Nakamura (1990), the corrected score function becomes

$$\tilde{\boldsymbol{\psi}}_{ml}(y, \mathbf{x}; \beta_0) = y - exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x} - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})$$

$$\tilde{\boldsymbol{\psi}}_{ml}(y, \mathbf{x}; \boldsymbol{\beta}) = y\mathbf{x} - (\mathbf{x} - \boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x} - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})$$

Finding the corrected score function $\tilde{\boldsymbol{\psi}}$ is not difficult if the original score function $\boldsymbol{\psi}$ depends on $\boldsymbol{\xi}$ only through $\boldsymbol{\xi}$, $exp(\boldsymbol{\beta}'\boldsymbol{\xi})$ and $\boldsymbol{\xi}exp(\boldsymbol{\beta}'\boldsymbol{\xi})$ by invoking the three identities above. There is more to say about this. Shklyar and Schneeweiss (2005) showed that

$$E\{f(\mathbf{x})\lambda(\mathbf{x},\boldsymbol{\theta}) \mid \boldsymbol{\xi}\} = exp(\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})E\{f(\mathbf{x}+\boldsymbol{\Sigma}_{\delta\delta}) \mid \boldsymbol{\xi}\}, \qquad (11)$$

where $\lambda(\mathbf{x},\boldsymbol{\theta}) = exp(\beta_0+\boldsymbol{\beta}'\mathbf{x})$, and f is an arbitrary function for which the expectation above exists. Therefore, if the ψ function is a function involved with λ function, then there is a good chance to find the corresponding corrected function $\tilde{\psi}$. The reason for assuming normal measurement error is quite obvious via equation (11).

Example 6 (Cheng and Schneeweiss, 1998)
Polynomial measurement error model. The model can be described as follows.

$$x_i = \xi_i + \delta_i, \qquad y_i = \beta_0 + \beta_1\xi_i + \beta_2\xi_i^2 + \ldots + \beta_k\xi_i^k + \varepsilon_i,$$

where $i = 1, \ldots, n$ and k is a positive integer. We assume that ξ_i are fixed. If not, we can condition on ξ_i. If ξ_i, $i = 1, \ldots, n$, were observable, one can simply use ordinary least squares and minimize

$$\sum \left(y_i - \beta_0 - \beta_1\xi_i^2 - \ldots - \beta_k\xi_i^k\right)^2.$$

Taking derivatives with respect to the β's leads to the unbiased estimating equations

$$\sum(y_i\xi_i^j - \beta_0\xi_i^j - \beta_1\xi_i^{j+1} - \beta_2\xi_i^{j+2} - \ldots - \beta_k\xi_i^{j+k}) = 0,$$

for $j = 0, 1, \ldots, k$. The equations above are the ψ function in (7). The question is to find the corresponding $\tilde{\psi}$. The problem is to find t_{ir} such that

$E(t_{ir}) = \xi_i^r$ for $r = 0, 1, 2, \ldots, 2k$. The solution for finding t_{ir} was first suggested by Chan and Mak (1985) and then more completely by Cheng and Schneeweiss (1998). The resulting estimator is called adjusted least squares (ALS) estimator in Cheng and Schneeweiss (1998). However, it can be viewed as a corrected score induced by least squares.

Stefanski (1989, p. 4338) proved that a corrected score function does not always exist. The necessary condition for the existence of a corrected score function is that the underlying estimation function has to be an entire function in the complex plane and it does not allow any singularities in the complex numbers. Stefanski (1989, p. 4350) also showed that the corrected score function induced by the maximum likelihood of the logistic regression model with probability mass function $p(y = 1 \mid \boldsymbol{\xi}) = \{1 + exp(-\beta_0 - \boldsymbol{\beta}'\boldsymbol{\xi})\}^{-1}$ possessing singularities. Consequently, the corrected score function induced by (corrected) log-likelihood for the logistic model does not exist.

3.1 Discussion and Further Development

In this subsection, we will briefly discuss the extensions of corrected score estimators, based on the work of Buzas and Stefanski (1996), see also Carroll *et al.* (1995, Chapter 6).

The first extension of corrected score estimation is to the class called *mean and variance* models, which also called *quasi-likelihood and variance function* (QVF) models described by Carroll and Ruppert (1988) and McCullagh and Nelder (1989). Because it is mathematically complicated, we omit it and refer the reader to Buzas and Stefanski (1996).

Buzas and Stefanski (1996) also suggested a potentially useful extension of the method to nonnormal, additive measurement error models. Suppose that $\mathbf{x} = \boldsymbol{\xi} + \boldsymbol{\delta}$, and its moment of generating function of $\boldsymbol{\delta}$, $m_\delta = E\{exp(t'\boldsymbol{\delta})\}$, exists for some t and is known. For normal errors it has shown that the corrected score is a function of terms of the form $exp(j\beta_0 + j\boldsymbol{\beta}'\mathbf{x} - \frac{1}{2}j^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})$, for nonnegative integer j. Note that for normally distributed $\boldsymbol{\delta}$, $m_\delta(t) = E\{exp(\frac{1}{2}t'\boldsymbol{\Sigma}_{\delta\delta}t)\}$ and thus

$$exp(j\beta_0 + j\boldsymbol{\beta}'\mathbf{x} - \frac{1}{2}j^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})$$

$$= exp(j\beta_0 + j\boldsymbol{\beta}'\mathbf{x})/exp(\frac{1}{2}j^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}) = exp(j\beta_0 + j\boldsymbol{\beta}'\mathbf{x})/m_\delta(j\beta_0 + j\boldsymbol{\beta}).$$

The extension for general error distributions is to replace all terms of the form $exp(j\beta_0 + j\boldsymbol{\beta}'\mathbf{x} - \frac{1}{2}j^2\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta})$ by $exp(j\beta_0 + j\boldsymbol{\beta}'\mathbf{x})/m_\delta(j\beta_0 + j\boldsymbol{\beta})$. The key issue is that one has to know the moment generating function of $\boldsymbol{\delta}$. Augustin (2004) found the exact corrected log-likelihood function for Cox's proportional hazards model using this assumption. However, the extension proposed by Buzas and Stefanski (1996) has not been fully explored yet.

Recently, Novick and Stefanski (2002) proposed a new method of computing corrected score function and the resulting unbiased estimating function is called Monte Carlo corrected score function. In principle, as long as the estimating ψ function is an entire function in the complex plane and suitably smooth, the Monte Carlo corrected score approach works well. The theory needs some knowledge of complex variable analysis that is not familiar with many statisticians. Moreover, the Monte Carlo corrected score method has some robust property. When the measurement error is not normal, it provides bias-reduced estimator.

Nakamura (1990) introduced the corrected score estimation by considering the log-likelihood function of a given ME model. Suppose that the log-likelihood function is denoted by $l(y, \mathbf{x}; \boldsymbol{\theta})$, and if there exists a function $\tilde{l}(y, \mathbf{x}; \boldsymbol{\theta})$, called a corrected log-likelihood, and it satisfies

$$E\{\tilde{l}(y_i, \mathbf{x}_i; \boldsymbol{\theta} \mid y_i, \boldsymbol{\xi}_i)\} = l(y_i, \boldsymbol{\xi}_i; \boldsymbol{\theta}), \tag{12}$$

for all $\boldsymbol{\theta}$ in the parameter space. If the expectation and partial derivative with respect to $\boldsymbol{\theta}$ are interchangeable, then it gives

$$E\{\frac{\partial \tilde{l}}{\partial \boldsymbol{\theta}}(y_i, \mathbf{x}_i; \boldsymbol{\theta} \mid y_i, \boldsymbol{\xi}_i)\} = \frac{\partial l}{\partial \boldsymbol{\theta}}(y_i, \boldsymbol{\xi}_i; \boldsymbol{\theta}). \tag{13}$$

Comparing the equation above with (8), the corrected score function is just $\partial \tilde{l}/\partial \boldsymbol{\theta}$. Therefore, there are two ways to define the corrected score estimate, one is the maximizer of the corrected log-likelihood, the other is the solution to the corrected score function. Nakamura (1990, p. 128) chose the latter one. Note that there are some differences between these two definitions. First of all, the maximizer of the log-likelihood might not exist. Secondly, even the maximizer exists, usually there are some restrictions on the data, as it happens frequently in the maximum likelihood estimator in general parametric models. On the other hand, solution to the corrected score function does not require such restrictions. From the viewpoint of large sample properties, if the maximizer exists, then the two definitions are the same because the restrictions are satisfied when the sample is large enough and the model is correct.

It should also be noted that, no matter which definitions one uses, the corrected score estimator induced by likelihood is not the same as the true maximum likelihood estimator in general. The corrected score estimator induced by corrected log-likelihood only requires to specify some distribution assumption, such as normality on the measurement error $\boldsymbol{\delta}$. Moreover, the latent variable $\boldsymbol{\xi}$ could be either functional or structural because of conditioning on y and $\boldsymbol{\xi}$. In the latter (structural) case, no distributional assumption is imposed on the latent variable either. Therefore, it is clear that the corrected score estimator is quite different from the maximum likelihood estimator in general.

Example 7 (Example 4 continued)

Both Stefanski (1989) and Nakamura (1990) used corrected score estimation in linear ME (functional) models. The corrected log-likelihood (when $\boldsymbol{\xi}$ is fixed) is, omitting the constants,

$$\sum \tilde{l}(y_i, \mathbf{x}_i; \boldsymbol{\theta}) = -\frac{1}{2\sigma_\varepsilon^2} \sum \{(y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{x}_i)^2 - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\beta}\}.$$

The resulting estimator is the same as that in Example 4, which is induced by least squares. It is also the maximum likelihood estimator (with some restrictions) when $\boldsymbol{\xi}$ is also normally distributed (Fuller, 1987, Cheng and Van Ness, 1999). Note that these restrictions are automatically satisfied when the sample is large and the ME modelling is correct.

It is worth noting that Nakamura (1990) (implicitly) assumed the functional model, that is, the latent variable $\boldsymbol{\xi}$ is non-stochastic, in finding the corrected log-likelihood function. In other words, the log-likelihood function $l(y, \mathbf{x}; \boldsymbol{\theta})$ assumes that $\boldsymbol{\xi}$ is being fixed in the first place. If not, the original log-likelihood function l will be quite different as it depends on the distribution of $\boldsymbol{\xi}$. For instance, the $\boldsymbol{\xi}$ is normal in the example above, then the corrected log-likelihood function is not the same as that in Example 7. If the latent variable $\boldsymbol{\xi}$ is random, it will greatly complicate the search for corrected log-likelihood function. However, one can modify the definition of corrected log-likelihood to be the one that satisfied (16) with the log-likelihood function $l(y, \mathbf{x}; \boldsymbol{\theta})$ replacing by the conditional log-likelihood function $l(y, \mathbf{x}; \boldsymbol{\theta} \mid \boldsymbol{\xi})$. Then the problem vanishes.

Corrected score estimation method was treated in a broader manner by Stefanski (1989). That is, the corrected score estimator is the solution to the unbiased estimating function (10). It is not necessarily induced by the score function, which comes from likelihood function. Certainly likelihood function is an important and natural source but by no means the only source. Nonlinear least squares is a reasonable source, for example. Nakamura (1990, p. 128 and 129) twice pointed it out that corrected score function is an unbiased score (estimating) function but the converse is not necessarily true but he did not explore the other sources. In Nakamura's paper, corrected score function is just the derivative of corrected log-likelihood function and is an unbiased estimating function. It is somewhat different from the broad meaning recognized by Carroll *et al.* (1995, Section 6.5) and Buzas and Stefanski (1996). We use the term corrected score function to stand for unbiased estimating function satisfying (7) and (8).

Finally, it is worth noting that corrected score estimation is not restricted to any particular distribution of the measurement error $\boldsymbol{\delta}$. The normality assumption is just for convenience. On the other hand, there is no results available for any other distributed $\boldsymbol{\delta}$ except the suggestion by Buzas and Stefanski (1996), as discussed earlier. It is because, in order to construct the corrected score function, one needs to evaluate the conditional expectation (8) and it seems to be inevitable that one must specify the distribution of $\boldsymbol{\delta}$, at least to some certain degree.. It would be of some interest to explore corrected estimation to certain family of distributions, such as exponential family. This work has not yet been done.

4. Concluding Remarks

Both conditional score and corrected score are functional methods, that is, the latent variables $\boldsymbol{\xi}_i$ are either fixed or random, but there is no distributional assumption imposed on it in the latter case. However, they differ in many aspects. First, conditional score method needs the existence of a complete sufficient statistics. Although the normality assumption of the measurement $\boldsymbol{\delta}$ does not seem to be crucial, it is not easy to find a complete sufficient statistics in general.

On the other hand, corrected score method seems more variable because it can be induced by least squares, method of moments, log-likelihood function, etc. Moreover, the Monte Carlo corrected score method greatly expands the applications of this methodology.

Generally speaking, conditional score and corrected score apply to different models with very few exceptions, such as Poisson regression model. If one compares these two estimation methods on Poisson regression model, one will find the corrected score has a closed form expression whereas the conditional score has infinite series that requires more heavy computations. Stefanski (1989) found that conditional score estimator is more efficient for Poisson model in some practical cases.

When the latent variable $\boldsymbol{\xi}$ does have a specific distribution, such as normal, then it will be interesting to compare the functional method with other methods that use the knowledge of the distribution of the latent variable. Some work has been done this aspect. Kukush, Schneeweiss and Wolf (2004) compared the corrected score estimator and structural quasi-score estimator in the Poisson model. The latter is based in conditional mean-variance, taking the distribution of the latent covariate into account. They found that the corrected score is preferred, unless the error variance is large, because the corrected score method is insensitive to the latent variable assumption.

The link between the conditional score method and total least squares can be seen via the (simple) linear ME model. The conditional score estimator coincide with the TLS estimator, see Example 1. However, such connection is not clear for nonlinear model. It needs some further investigation.