Statistical Papers 38, 1–25 (1997)

Statistical Papers

Estimation of nonlinear errors-in-variables models: an approximate solution *

C. Hsiao, L. Wang, Q. Wang

Received: January 4, 1995; revised version: November 10, 1995

We propose an easy to derive and simple to compute approximate least squares or maximum likelihood estimator for nonlinear errors-in-variables models that does not require the knowledge of the conditional density of the latent variables given the observables. Specific examples and Monte Carlo studies demonstrate that the bias of this approximate estimator is small even when the magnitude of the variance of measurement errors to the variance of measured covariates is large.

Keywords: Measurement error; Nonlinear models; Approximate least squares; Bias adjusted estimators

*Cheng Hsiao and Qing Wang's work was supported in part by National Science Foundation grant SES91-22481 and SBR94-09540. Liqun Wang gratefully acknowledges the financial support from Swiss National Science Foundation. We wish to thank Professor H. Schneeweiss and a referee for helpful comments and suggestions.

1. Introduction

We consider nonlinear errors-in-variables models of the form

$$E(y \mid x) = g\left(x; \theta^0\right) \tag{1.1}$$

and

$$z = x + \epsilon, \tag{1.2}$$

where the $K \times 1$ vector of explanatory variables x are unobservable and only their proxies z are observed. The measurement errors ϵ are assumed to be independent of x and are independently, identically distributed with mean 0 and covariance matrix V. The $p \times 1$ vector of parameters θ^0 are assumed to lie in the interior of a convex compact set $\Theta \subset R^p$, with R^p denoting a p-dimensional Euclidean space. We assume that $g(x;\theta)$ is nonlinear in x and is differentiable with respect to x and θ. Let u denote the difference between the observed y and its conditional expectation (1.1), then

$$y = g\left(x; \theta^0\right) + u \tag{1.3}$$

and u is assumed to be independent of ϵ.

When variables are subject to measurement errors, it was shown by Y. Amemiya (1985) and Hsiao (1989) that the mere existence of instruments that are correlated with the latent variables, x, but are uncorrelated with the errors ϵ is not sufficient to identify nor will it provide a consistent estimator of θ^0 if $g(x;\theta)$ is nonlinear in x variables. To obtain consistent estimators of θ^0, Amemiya and Fuller (1988), Stefanski and Carroll (1985), Wolter and Fuller (1982a, b) etc. have relied on the assumption of increasing sample size and decreasing error variances. Alternatively, a structural errors-in-variables approach which assumes a known conditional distribution of x given z (or conditional distribution of ϵ given z), $f\left(x \mid z; \delta^0\right)$ can be used, where δ^0 denotes the $q \times 1$ vector of parameters. Combining $f\left(x \mid z; \delta^0\right)$ with the conditional distribution of y given x, $f\left(y \mid x; \lambda^0\right)$, where λ^0 is the parameter vector including θ^0 as a subvector, we can derive the conditional distribution of y given z,

$$f\left(y \mid z; \lambda^0, \delta^0\right) = \int f\left(y \mid x; \lambda^0\right) f\left(x \mid z; \delta^0\right) dx, \tag{1.4}$$

or the expected value of y given z,

$$\begin{aligned} E\left(y \mid z; \gamma^0\right) &= \int g\left(x; \theta^0\right) f\left(x \mid z; \delta^0\right) dx \\ &= G\left(z; \gamma^0\right), \end{aligned}$$

where $\gamma^0 = (\theta^{0\prime}, \delta^{0\prime})'$. Under fairly general conditions, it can be shown that the maximum likelihood estimator (MLE) that maximizes the conditional likelihood function (1.4) or the minimum distance estimator (MDE) that minimizes some distance measure of $[y - G(z;\gamma)]$ is consistent and asymptotically normally distributed (e.g. Hsiao (1989, 1991)). However, to implement the MLE or MDE, not only the error distribution needs to be known *a priori*, also computationally it can be unwieldy because of the need to take multiple integrations. Therefore, Carroll and Stefanski (1990), Stefanski (1985), Whittemore and Keller (1988), etc. have suggested alternative computationally simpler bias adjusted or approximate maximum likelihood estimators which do not need to take multiple integrations and only require the knowledge of the first two moments of measurement errors given the observed covariates. That is,

$$E(\epsilon \mid z) = \alpha(z) = c\tilde{\alpha}(z) + o(c), \tag{1.5}$$

$$E(\epsilon\epsilon' \mid z) = \Omega(z) = c\tilde{\Omega}(z) + o(c), \tag{1.6}$$

are assumed known or estimable, and

$$E\left(\|\epsilon\|^3 \mid z\right) = o(c), \tag{1.7}$$

where c is a positive scalar that reflects the magnitude of the variance of ϵ relative to that of z, $\tilde{\alpha}$ and $\tilde{\Omega}$ are independent of c. They demonstrate that the bias of their bias adjusted or approximate MLE estimators is of order $o(c)$. In this paper we shall also assume (1.5) - (1.7) and propose alternative approximate MDE or MLE, which is straightforward to implement and appears to perform better.

There are many practical situations where assumptions (1.5) and (1.6) are applicable, in particular in situations when replicated measurements are available. Some of these situations were described by Berkson (1950), Carroll and Stefanski (1990), Whittemore and Keller (1988), etc. Applications of some special nonlinear errors-in-variables models in medicine and epidemiology may be found in Rosner, Willett and Spiegelman (1989) and Tosteson, Stefanski and Schafer (1989). A survey of approaches to estimation is given by Carroll (1989).

In section 2 we introduce the alternative approximation to MDE or MLE and discuss how this method may be implemented. In section 3 we provide some examples that evaluate the bias of various approximate estimators analytically and discuss how these methods may be implemented. In section 4 we provide some Monte Carlo evaluations. Conclusions are in section 5.

2. Approximate Nonlinear Least Squares Estimator

Suppose the conditional distribution of x given z $f(x \mid z; \delta^0)$ is known. Then one can compute the conditional mean of y given z

$$G\left(z_i; \gamma^0\right) = E\left(y_i \mid z_i\right) = E\left[g(x_i; \theta^0) \mid z_i\right]. \tag{2.1}$$

It was shown by Hsiao (1989) that under fairly general conditions the nonlinear least squares estimator (NLS) that minimizes $\sum_{i=1}^{n} \left[y_i - G\left(z_i; \gamma\right)\right]^2$ is consistent and asymptotically normally distributed. However, the computation of the nonlinear least squares estimator involves multiple integrations in the form of (2.1) at each iterative step, which can be quite complicated. Moreover, the approach is sensitive to the specification of $f\left(x \mid z; \delta^0\right)$ (Shafer (1987)). Thus, various computationally feasible procedures that avoid multiple integrations have been proposed to obtain bias adjusted or approximate MLE or NLS (e.g. Carroll and Stefanski (1990), Stefanski (1985), Whittemore and Keller (1988)) for the case when measurement error variances are small relative to the variance of observables z.[1] In this paper we propose another method of obtaining approximate NLS which is easy to derive and simple to implement and does not require the knowledge of $f(x \mid z; \delta^0)$. The estimator also appears to possess good properties even when the measurement error variance is not small.

There are two approaches to derive the approximate NLS. One approach is to take a Taylor series expansion of the NLS around $c = 0$ (e.g. Whittemore and Keller (1988)). The other approach is to obtain an approximation of $G(z; \gamma)$, say $A(z; \gamma)$, then solve for the approximate NLS by minimizing $\sum_{i=1}^{n} \left[y_i - A\left(z_i; \gamma\right)\right]^2$ (Carroll and Stefanski (1990), Rudemo, Ruppert and Streibig (1989)). In this paper, we shall follow the latter approach.

There are many ways to obtain the approximation of $E(y \mid z)$. Chesher (1991) has suggested an approximation of the conditional density of y given z, $f(y \mid z)$, which has approximation error $o(c)$ and avoids the computation of multiple integrations. The approximation of the conditional mean of y given z can then be obtained by using this approximate density function. Here we suggest an alternative approximation that is straightforward to derive and the approximation error is also of $o(c)$.

Let $\phi(z)$, which may also depend on c, be a predictor of unobserved

[1] Strictly speaking, the procedures they propose are for obtaining approximate maximum likelihood estimator. However, they can be easily adapted to the present setting.

x. Taking a Taylor series expansion of $g(x;\theta)$ around $\phi(z)$, we have

$$\begin{aligned} g(x;\theta) &= g(\phi(z),\theta) + \frac{\partial g(\phi(z),\theta)}{\partial x'}(x-\phi(z)) \\ &+\frac{1}{2}(x-\phi(z))'\frac{\partial^2 g(\phi(z),\theta)}{\partial x\partial x'}(x-\phi(z)) \\ &+e(x,\phi(z),\theta). \end{aligned} \tag{2.2}$$

Then $G(z;\gamma)$ can be approximated by taking expectation of (2.2) without the fourth term on the right hand side of the equality, conditional on z:

$$\begin{aligned} A(z;\gamma) &= g(\phi(z);\theta) + \frac{\partial g(\phi(z);\theta)}{\partial x'}E(x-\phi(z)\mid z) \\ &+\frac{1}{2}\mathrm{tr}\left\{\frac{\partial^2 g(\phi(z);\theta)}{\partial x\partial x'}\,E\left[(x-\phi(z))(x-\phi(z))'\mid z\right]\right\}. \end{aligned} \tag{2.3}$$

The approximation error is given by $E[e(x,\phi(z),\theta)\mid z]$. For instance, if $\phi(z)=z$, then

$$A(z;\gamma) = g(z;\theta) - \frac{\partial g(z;\theta)}{\partial x}\alpha(z) + \frac{1}{2}\mathrm{tr}\left[\frac{\partial^2 g(z;\theta)}{\partial x\partial x'}\Omega(z)\right] \tag{2.4}$$

and the order of approximation error $E[e(x,\phi(z),\theta)\mid z]$ is $o(c)$ under (1.5)-(1.7), given that the third derivative of the function $g(x;\theta)$ with respect to x is bounded. Note that (2.4) does not require the knowledge of $f(x\mid z)$. It only needs the knowledge of $E(x\mid z)$ and Cov $(x\mid z)$. So is our proposed approximation formula (2.7) below. Therefore, without causing confusion, we shall use $A(z;\theta)$ to denote $A(z;\gamma)$ conditioning on $\alpha(z)$ and $\Omega(z)$ and use $G(z;\theta)$ to denote $G(z;\gamma)$ conditioning on δ.

Following the standard nonlinear least squares framework, we assume that

A.1 $(y_i,z_i')'$ are independently, identically distributed (i.i.d.) and $Ey^2<\infty$.

A.2 $A(z;\theta)$ is continuous in θ and $E\sup_{\theta\in\Theta}|A(z;\theta)|^2<\infty$.

A.3 $Q(\theta)=E[G(z;\theta^0)-A(z;\theta)]^2$ attains the unique minimum at an interior point $\tilde{\theta}\in\Theta$.

Let $S_n(\theta)=\sum_{i=1}^n[y_i-A(z_i;\theta)]^2$ and $\hat{\theta}_A$ be the solution to $\min_{\theta\in\Theta}S_n(\theta)$.

Theorem 2.1. *Under assumptions A.1 - A.3, the approximate NLS for model (1.1) - (1.2), $\hat{\theta}_A$, converges in probability to $\tilde{\theta}$.*

Proof: Denote $G_i(\theta) = G(z_i;\theta)$ and $A_i(\theta) = A(z_i;\theta)$. Then we can write

$$\begin{aligned}\frac{1}{n}S_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left[y_i - G_i(\theta^0)\right]^2 \\ &+ \frac{2}{n}\sum_{i=1}^{n}\left[y_i - G_i(\theta^0)\right]\left[G_i(\theta^0) - A_i(\theta)\right] \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left[G_i(\theta^0) - A_i(\theta)\right]^2 .\end{aligned} \tag{2.5}$$

By A.1 and the Kolmogorov law of large numbers, the first term on the right-hand side of (2.5) converges in probability to the constant $E\left[\mathrm{Var}(y_i \mid z_i)\right]$. Under A.1 and A.2, the second term converges to 0 in probability uniformly in $\theta \in \Theta$ by Theorem 4.2.1 of T. Amemiya (1985). Similarly, the third term on the right-hand side of (2.5) converges to $Q(\theta)$ in probability uniformly in $\theta \in \Theta$. The result then follows from A.3 and Theorem 4.1.1 of T. Amemiya (1985). □

In general $\tilde{\theta}$ is different from the true parameter θ^0. The difference $\tilde{\theta} - \theta^0$ gives the asymptotic bias of $\hat{\theta}_A$. It is well-known that the naive estimator $\hat{\theta}(0)$ that ignores the measurement error issue has the bias of order $O(c)$. For the approximate NLS, $\hat{\theta}_A$, we have the following general result:

Theorem 2.2. *In addition to A.1 - A.3, we assume*

A.4 *$A(z;\theta)$ is continuously differentiable with respect to θ and $E\sup_{\theta\in\Theta}\|\partial A(z;\theta)/\partial\theta\|^2 < \infty$.*

A.5 *$B(\theta) = (1/2)\,\partial^2 Q(\theta)/\partial\theta\partial\theta'$ is nonsingular for all θ on a straight line between θ^0 and $\tilde{\theta}$.*

Then $\operatorname{plim}_{n\to\infty}\hat{\theta}_A = \tilde{\theta} = \theta^0 + B(\theta^*)^{-1}H$, where θ^* lies between θ^0 and $\tilde{\theta}$, and

$$H = E\left[\left(G\left(z;\theta^0\right) - A\left(z;\theta^0\right)\right)\frac{\partial A\left(z;\theta^0\right)}{\partial\theta}\right].$$

Therefore, if $Q\left(\theta^0\right) = E\left[G\left(z;\theta^0\right) - A\left(z;\theta^0\right)\right]^2 = o(c)$, then $\operatorname{plim}_{n\to\infty}\hat{\theta}_A = \theta^0 + o(c)$.

Proof: By A.5 and a Taylor expansion of $\partial Q(\theta)/\partial\theta$ around θ^0, we have

$$\tilde{\theta} = \theta^0 - \left[\frac{\partial^2 Q\left(\theta^*\right)}{\partial\theta\partial\theta'}\right]^{-1}\frac{\partial Q\left(\theta^0\right)}{\partial\theta},$$

where θ^* lies between θ^0 and $\tilde{\theta}$. By definition, A.2 and A.4, $\partial Q/\partial\theta\,|_{\theta^0}= -2H$. The second part of the theorem follows from the Cauchy-Schwarz inequality. □

Furthermore, following the proofs of Hsiao (1989) or T. Amemiya (1985), we can establish the following result.

Theorem 2.3. *Suppose A.1 - A.5 and*

A.6 *$A(z;\theta)$ has continuous second order derivative with respect to θ and*

$$E\sup_{\theta\in\Theta}\left\|[y-A(z;\theta)]\frac{\partial^2 A(z;\theta)}{\partial\theta\partial\theta'}\right\|<\infty.$$

Then

$$\sqrt{n}\left(\hat{\theta}_A-\tilde{\theta}\right)\xrightarrow{d}N\left(0,B^{-1}CB^{-1}\right),\tag{2.6}$$

where

$$B=E\left\{\frac{\partial A\left(z;\tilde{\theta}\right)}{\partial\theta}\frac{\partial A\left(z;\tilde{\theta}\right)}{\partial\theta'}-\left[G\left(z;\theta^0\right)-A\left(z;\tilde{\theta}\right)\right]\frac{\partial^2 A\left(z;\tilde{\theta}\right)}{\partial\theta\partial\theta'}\right\}$$

and

$$C=E\left\{\left[Var\left(y\mid z\right)+\left(G\left(z;\theta^0\right)-A\left(z;\tilde{\theta}\right)\right)^2\right]\frac{\partial A\left(z;\tilde{\theta}\right)}{\partial\theta}\frac{\partial A\left(z;\tilde{\theta}\right)}{\partial\theta'}\right\}.$$

The approximate least squares estimator is obtained by minimizing $S_n(\theta)$. Since, as Theorem 2.2 has demonstrated, the bias of the approximate NLS, in general, is of the same order as the order of approximation error $E\left[G\left(z;\theta^0\right)-A\left(z;\theta^0\right)\right]^2$, a less biased and simple to compute estimator can be derived by finding a good, simple approximation of $E(y\mid z)$.

Because the bias of the approximate NLS in general is of the same order as the order of the approximation error $E\left[e\left(x,\phi(z),\theta\right)\mid z\right]$, it would be desirable if one can choose ϕ such that the approximation error is minimized. Unfortunately, an optimal predictor $\phi(z)$, even if it exists, depends on $E(y\mid z)$, hence brings up the same computational difficulty as the NLS. However, a good choice of $\phi(z)$ appears to be the conditional mean of x given z, $E(x\mid z)$, since it is the solution of $\min_\phi E\left(|x-\phi(z)|^h\mid z\right)$ for $h=1,2,3,\ldots$ as long as the conditional distribution of x given z is symmetric around its mean.

Motivated by the above intuition, we let $\phi(z_i) = \mu_i = E(x_i \mid z_i)$ and $\Omega(z_i) = \Omega_i = \mathrm{Cov}(x_i \mid z_i)$, and propose

$$A(z_i;\theta) = g(\mu_i;\theta) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial^2 g(\mu_i;\theta)}{\partial x \partial x'}\Omega_i\right) \tag{2.7}$$

as an approximation of $G(z_i;\theta)$. Equation (2.7), just like (2.4), does not require the knowledge of $f(x \mid z)$, but does require the knowledge of $E(x \mid z)$ and Cov $(x \mid z)$. For this approximation, it is straightforward to show by Cauchy-Schwarz inequality that assumption A.2 is implied by

$$E\left\|\Omega(z)\right\|^2 < \infty \tag{2.8}$$

and

$$E\sup_{\theta\in\Theta} g(\mu(z);\theta)^2 < \infty, \quad E\sup_{\theta\in\Theta}\left\|\frac{\partial^2 g(\mu(z);\theta)}{\partial x \partial x'}\right\|^2 < \infty; \tag{2.9}$$

assumption A.4 is implied by (2.8) and

$$E\sup_{\theta\in\Theta}\left\|\frac{\partial g(\mu(z);\theta)}{\partial \theta}\right\|^2 < \infty, \quad E\sup_{\theta\in\Theta}\left|\frac{\partial^3 g(\mu(z);\theta)}{\partial x_j \partial x_k \partial \theta_l}\right|^2 < \infty, \quad \forall\, j,k,l; \tag{2.10}$$

whereas assumption A.6 follows from (2.8) - (2.9) and

$$E\sup_{\theta\in\Theta}\left\|\frac{\partial^2 g(\mu(z);\theta)}{\partial \theta \partial \theta'}\right\|^2 < \infty, \quad E\sup_{\theta\in\Theta}\left|\frac{\partial^4 g(\mu(z);\theta)}{\partial x_j \partial x_k \partial \theta_l \partial \theta_m}\right|^2 < \infty, \quad \forall\, j,k,l,m. \tag{2.11}$$

It is easier to check conditions (2.8)- (2.11) than A.2, A.4 and A.6 because these conditions are directly expressed in terms of the function $g(x;\theta)$. In the rest of this paper we denote $\hat{\theta}_A$ as the approximate NLS of θ^0 corresponding to $A(z_i;\theta)$ in (2.7). Thus we have the following result.

Theorem 2.4. *Suppose A1, A3, A5, (2.8) - (2.11) hold and that $g(x;\theta^0)$ has a bounded third order derivative with respect to x. Then all results of Theorems 2.1 - 2.3 hold for the estimator $\hat{\theta}_A$ corresponding to $A(z_i;\theta)$ in (2.7).*

Remark 2.1. The approximation of $E(y \mid z)$ by (2.4) or (2.7) is usually simple to implement because the function $g(x;\theta)$ is often specified in simple closed form. All we need to do is to calculate its first and/or second order partial derivatives. Equation (2.4) uses z and (2.7) uses

$E(x \mid z)$ to predict x. Equation (2.4) is used by Carroll and Stefanski (1990), Whittemore and Keller (1988), etc. However, $E(x \mid z)$ is a better predictor of x than z. When c is small, there is little difference between using $E(x \mid z)$ or z to predict the unobserved covariates x. When c is large, the difference can be significant. Thus, even though $\hat{\theta}_A$ may have the same order of bias as other approximate estimators, it is likely to be more robust relative to the magnitude of measurement error variance than those estimators that rely on z to predict x.

Remark 2.2. If $\alpha(z)$ and $\Omega(z)$ are unknown, there are many ways to obtain their approximations. For instance, Carroll and Stefanski (1990) gives the general form of (1.5) and (1.6) in their Lemma A.1 for the case when the joint density of z and x and the first two conditional moments of ϵ given z are three times differentiable with respect to c. Under the assumption of independence between x and ϵ, they become

$$E(\epsilon \mid z) = -V\frac{\partial \log f_x(z)}{\partial x} + o(c) \tag{2.12}$$

and

$$Cov(\epsilon \mid z) = V + o(c), \tag{2.13}$$

where V is the covariance matrix of ϵ and $f_x(z)$ is the density function of x evaluated at z. In Whittemore and Keller (1988) conditional mean and variance of measurement errors (1.5) and (1.6) are assumed known. In the case they are unknown, but the validation data are available (e.g., Carroll and Stefanski (1990), Lee and Sepanski (1995)), they can be estimated directly. If no validation data are available, but there are repeated measurements, (say r replications for each $i = 1, 2, ..., n$), we can obtain an estimate of V by $\sum_{i=1}^{n}\sum_{t=1}^{r}(z_{it} - \overline{z}_i)(z_{it} - \overline{z}_i)'/n(r-1)$ and Cov(z) by $\sum_{i=1}^{n}\sum_{t=1}^{r}(z_{it} - \overline{z})(z_{it} - \overline{z})'/(nr-1)$, where $\overline{z}_i = \sum_{t=1}^{r} z_{it}/r$ and $\overline{z} = \sum_{i=1}^{n}\overline{z}_i/n$. When x and ϵ are normally distributed, using the well-known conditional mean and conditional variance formulae, we can again estimate (1.5) and (1.6) directly. In the case that x and ϵ are not normally distributed but if the distribution of x is known and the first two conditional moments of ϵ given z are three times differentiable with respect to c, one can use (2.12) and (2.13) to approximate $E(\epsilon \mid z)$ and $Cov(\epsilon \mid z)$. The approximation errors for μ_i and Ω_i are of orders $O\left(r^{-1/2}\right)$ and $O\left((nr)^{-1}\right)$ respectively. Under the additional assumptions that $\partial g(x;\theta)/\partial x$ and $\partial^2 g(x;\theta)/\partial x \partial x'$ are uniformly bounded, substituting $\widehat{\mu}_i$ and $\widehat{\Omega}_i$ for μ_i and Ω_i in (2.4) or (2.7) introduce an approximation error of $O\left(r^{-1/2}\right)$. Using Lemma 1 of Whittemore and Keller

(1988) and Theorem 2.2, it can be shown that the bias of $\hat{\theta}_A$ remains to be $o(c)$ as long as $\|\hat{\mu}_i - \mu_i\| = o(c)$, $\left\|\hat{\Omega}_i - \Omega_i\right\| = o(c)$ and $\partial g(x;\theta)/\partial x$, $\partial^2 g(x;\theta)/\partial x \partial x'$ are uniformly bounded. Furthermore $\hat{\theta}_A$ converges to $\tilde{\theta}$ in probability. However, the asymptotic variance-covariance matrix will be different. Suppose that the conditional mean and variance-covariance matrix of x given z depends on the $q \times 1$ parameter δ^0 and $\hat{\delta}_n$ is a consistent estimator of δ^0, then the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}_A - \tilde{\theta})$ is equal to (Hsiao 1989)

$$B^{-1}\left[C + S_{\theta\delta}\,\mathrm{Var}\left(\sqrt{n}\hat{\delta}_n\right) S_{\delta\theta} + \Sigma S_{\delta\theta} + S_{\theta\delta}\Sigma'\right] B^{-1}$$

where

$$S_{\theta\delta} = \lim_{n\to\infty} E\left(\frac{1}{n}\frac{\partial^2 \tilde{S}_n\left(\tilde{\theta},\delta^0\right)}{\partial\theta\partial\delta'}\right)$$

and

$$\Sigma = \lim_{n\to\infty} E\left[\frac{1}{\sqrt{n}}\frac{\partial \tilde{S}_n\left(\tilde{\theta},\delta^0\right)}{\partial\theta}\cdot\sqrt{n}\left(\hat{\delta}_n - \delta^0\right)'\right].$$

Remark 2.3. The proposed approach can be easily adapted to obtain an approximate maximum likelihood estimator (MLE). Let $f(y \mid x;\lambda^0)$ denote the conditional density of y given x. Then the conditional density of y given z can be approximated by

$$A(y_i,\mu_i;\lambda,c) = f(y_i \mid \mu_i;\lambda) + \frac{1}{2}\,tr\left[\frac{\partial^2 f(y_i \mid \mu_i,\lambda)}{\partial x\partial x'}\Omega(z_i)\right].$$

An approximate MLE of λ can be defined as a solution of the approximate score equations:

$$\sum_{i=1}^{n}\frac{1}{A(y_i,\mu_i;\lambda,c)}\frac{\partial A(y_i,\mu_i,;\lambda,c)}{\partial\lambda} = 0.$$

Remark 2.4. Similarly, the approach can be adapted to obtain maximum quasi-likelihood estimator (MQLE). The MQLE as defined by Carroll and Stefanski (1990) is the solution of the equation

$$Q_n(\theta) = \sum_{i=1}^{n}\frac{y_i - G_i(\theta)}{\sigma_i^2}\frac{\partial G_i(\theta)}{\partial\theta} = 0.$$

where

$$\begin{aligned}\sigma_i^2(\theta) &= \mathrm{Var}(y_i \mid z_i)\\ &= \mathrm{Var}(u_i) + E\left[g^2(x_i;\theta) \mid z_i\right] - \left\{E\left[g(x_i;\theta) \mid z_i\right]\right\}^2.\end{aligned}$$

The $G_i(\theta) = E[g(x_i;\theta) \mid z_i]$ can be approximated by (2.7). Using a Taylor expansion of $g^2(x_i;\theta)$ for x_i around $\mu_i = E(x_i \mid z_i)$, $E[g^2(x_i;\theta) \mid z_i]$ may be approximated by

$$g^2(\mu_i;\theta) + tr\left\{\Omega_i\left[g(\mu_i;\theta)\frac{\partial^2 g(\mu_i;\theta)}{\partial x \partial x'} + \frac{\partial g(\mu_i;\theta)}{\partial x}\frac{\partial g(\mu_i;\theta)}{\partial x'}\right]\right\}.$$

Hence, $\sigma_i^2(\theta)$ may be approximated by

$$\tilde{\sigma}_i^2(\theta) = \ Var(u_i) + tr\left[\Omega_i \frac{\partial g(\mu_i;\theta)}{\partial x}\frac{\partial g(\mu_i;\theta)}{\partial x'}\right] - \frac{1}{4} tr\left[\Omega_i \frac{\partial^2 g(\mu_i;\theta)}{\partial x \partial x'}\right]^2$$

and an approximate MQLE, $\hat{\theta}_{AQL}$, may be defined as the solution to

$$\sum_{i=1}^{n} \frac{y_i - A_i(\theta)}{\tilde{\sigma}_i^2}\frac{\partial A_i(\theta)}{\partial \theta} = 0.$$

Under the additional assumption that

$$E\left[\frac{G_i(\theta^0) - A_i(\theta)}{\tilde{\sigma}_i^2}\frac{\partial A_i(\theta)}{\partial \theta}\right] = 0$$

has a unique solution at an interior point $\tilde{\theta}^* \in \Theta$,

$$\sqrt{n}\left(\hat{\theta}_{AQL} - \tilde{\theta}^*\right) \xrightarrow{d} N\left(0, \tilde{B}^{-1}\tilde{C}\tilde{B}^{-1}\right),$$

where

$$\tilde{B} = E\left[\frac{1}{\tilde{\sigma}_i^2\left(\tilde{\theta}^*\right)}\frac{\partial A_i\left(\tilde{\theta}^*\right)}{\partial \theta}\frac{\partial A_i\left(\tilde{\theta}^*\right)}{\partial \theta'} - \frac{G_i\left(\theta^0\right) - A_i\left(\tilde{\theta}^*\right)}{\tilde{\sigma}_i^2(\tilde{\theta}^*)}\frac{\partial^2 A_i\left(\tilde{\theta}^*\right)}{\partial \theta \partial \theta'}\right],$$

and

$$\tilde{C} = E\left[\left(\frac{y_i - A_i\left(\tilde{\theta}^*\right)}{\tilde{\sigma}_i^2\left(\tilde{\theta}^*\right)}\right)^2 \frac{\partial A_i\left(\tilde{\theta}^*\right)}{\partial \theta}\frac{\partial A_i\left(\tilde{\theta}^*\right)}{\partial \theta'}\right].$$

3. Some Special Cases

There are a number of bias-adjusted or approximate estimators suggested in the literature. Carroll and Stefanski (1990) have provided three other approximations to $G(z;\theta)$. The first one is to use (2.4), which can be rewritten in terms of c as

$$A_1(z;\theta) = g(z;\theta) - c\frac{\partial g(z;\theta)}{\partial x'}\tilde{\alpha}(z) + \frac{c}{2}\mathrm{tr}\left[\frac{\partial^2 g(z;\theta)}{\partial x \partial x'}\tilde{\Omega}(z)\right]. \tag{3.1}$$

The second approximation is

$$A_2(z;\theta) = g\left(z - c\tilde{\alpha}(z);\theta\right). \tag{3.2}$$

The third approximation is a modification of $A_2(z;\theta)$,

$$A_3(z;\theta) = g\left(z - c\beta(z);\theta\right), \tag{3.3}$$

where $\beta(z)$ is given by

$$\left[\frac{\partial g\,(z;\theta)}{\partial x'}\frac{\partial g\,(z;\theta)}{\partial x}\right]^{-1}\left\{\frac{\partial g\,(z;\theta)}{\partial x'}\tilde{\alpha}(z) - \frac{1}{2}\mathrm{tr}\left[\frac{\partial^2 g(z;\theta)}{\partial x \partial x'}\tilde{\Omega}(z)\right]\right\}\frac{\partial g(z;\theta)}{\partial x}.$$

In addition, Stefanski (1985) has proposed a bias adjusted estimator that corrects the bias of the naive estimator,

$$\hat{\theta}_s = \hat{\theta}(0) - \frac{c}{2}\left.\frac{\partial^2 \hat{\theta}(c)}{\partial c^2}\right|_{c=0},$$

where $\hat{\theta}(c)$ denotes the MLE or MDE and $\hat{\theta}(0)$ denotes the naive estimator that treats z as if it were x. Whittemore and Keller (1988) have proposed an approximate MLE or MDE estimator,

$$\hat{\theta}_w = \hat{\theta}(0) + c\left.\frac{\partial \hat{\theta}(c)}{\partial c}\right|_{c=0}.$$

All these estimators have bias of order $o(c)$ except for the estimator that uses (3.2) which is $O(c)$ because (3.2) differs from $G(z;\theta)$ by $O(c)$.

It is difficult to compare the exact bias of various estimators in its general form. In the following we consider some special models for which more concrete results concerning the bias of these estimators can be derived. To facilitate comparison, we shall assume that the conditional mean and covariance of measurement errors ϵ given z are known, i.e., $E(\epsilon \mid z) = \alpha(z)$ and $E(\epsilon\epsilon' \mid z) = \Omega(z)$.

3.1. Linear Model

First we consider the linear model $g(x;\theta) = x'\theta$. Then $G(z_i;\theta) = \mu_i'\theta$. Since now $\partial g(x;\theta)/\partial x = \theta$ and $\partial^2 g(x;\theta)/\partial x \partial x' = 0$, the approximation (2.7) is $A(z_i;\theta) = \mu_i'\theta = G(z_i;\theta)$. So is the approximation (2.4). Therefore, both $\hat{\theta}_A$ and the Carroll and Stefanski (1990) estimator $\hat{\theta}_c$ based on (2.4) are consistent.[2] However, the naive estimator $\hat{\theta}(0)$, the Stefanski (1985) estimator $\hat{\theta}_s$ and the Whittemore and Keller (1988) estimator $\hat{\theta}_w$ are in general inconsistent. To see this, consider the example that $g(x;\theta) = \theta^0 x$ with $x \sim N(a,1)$ and $\epsilon \sim N(0,c)$. A straightforward calculation shows that $\hat{\theta}(0)$, $\hat{\theta}_s$ and $\hat{\theta}_w$ have asymptotic bias $-c\theta^0/\left(1+c+a^2\right)$, $-c^2\theta^0/\left(1+c+a^2\right)^2$ and $2c^2a^2\theta^0/\left(1+c+a^2\right)^2$ respectively. The Whittemore and Keller (1988) estimator, $\hat{\theta}_w$, is consistent only when $Ex = a = 0$.

3.2. Polynomial Regression Model

Consider the model

$$g(x;\theta) = \sum_{j=1}^{p} \theta_j x^{j-1} = \tilde{x}'\theta,$$

where $x \in R$, $\tilde{x} = \left(1, x, x^2, \ldots, x^{p-1}\right)'$ and $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$. To avoid the trivial case we assume $p \geq 3$. Then

$$G(z_i;\theta) = \sum_{j=1}^{p} \theta_j E\left(x_i^{j-1} \mid z_i\right)$$

and the approximation (2.7) is

$$A(z_i;\theta) = \left(\tilde{\mu}_i + \frac{1}{2}\Omega_i \tilde{\mu}_i^{(2)}\right)'\theta,$$

where $\tilde{\mu}_i = \left(1, \mu_i, \mu_i^2, \ldots, \mu_i^{p-1}\right)'$ and $\tilde{\mu}_i^{(2)} = \partial^2 \tilde{\mu}_i / \partial \mu_i^2$. Now since

$$\frac{\partial A(z_i;\theta)}{\partial \theta} = \tilde{\mu}_i + \frac{1}{2}\Omega_i \tilde{\mu}_i^{(2)}$$

and

$$\frac{\partial^2 A(z_i;\theta)}{\partial \theta \partial \theta'} = 0,$$

[2]Here we assume that $\alpha(z)$ and $\Omega(z)$ are known. If one were to use $o(c)$ approximation (3.1) - (3.3), then the Carroll and Stefanski estimator will have an asymptotic bias $(1-c)c^2\theta^0/[(1+c)(1-c)^2 + a^2]$.

we have

$$G(z_i;\theta) - A(z_i;\theta) = \begin{cases} 0, & p \leq 3 \\ \sum_{j=3}^{p-1} \theta_{j+1} \sum_{k=3}^{j} \binom{j}{k} \mu_i^{j-k} E\left[(x_i - \mu_i)^k \mid z_i\right], & p > 3. \end{cases}$$

This implies immediately the following results.

Corollary 3.1. *Under the conditions of Theorem 2.1, the estimator* $\hat{\theta}_A$ *is consistent if*

(1) $p \leq 3$*; or*

(2) $p = 4$ *and* $E\left[(x_i - \mu_i)^3 \mid z_i\right] = 0$.

The results in Corollary 3.1 are quite natural because the approximation $A(z_i;\theta)$ is based on a second order Taylor expansion. It is easily seen that the accuracy of the approximation will increase along with the order of the Taylor expansion. Indeed, the results of Corollary 3.1 may be made more general as follows.

Corollary 3.2. *Suppose the first two moments* $\mu_i \neq 0$, $\Omega_i \neq 0$. *Then under the conditions of Theorem 2.1,*

(1) The estimator $\hat{\theta}_A$ *is consistent if and only if* $A(z_i,\theta)$ *is based on a* $(p-1)$*-th order Taylor expansion;*

(2) If p *is even and the conditional distribution* $f(x \mid z)$ *is symmetric, then* $\hat{\theta}_A$ *is consistent if and only if* $A(z_i;\theta)$ *is based on a* $(p-2)$*-th order Taylor expansion.*

Finally we note that even under the conditions of Corollary 3.1 or 3.2 the other estimators are still inconsistent. To see this, we consider again a univariate model $g(x;\theta) = \theta x^2$ with $x \sim N(0,1)$ and $\epsilon \sim N(0,c)$. Then the asymptotic biases of the estimators $\hat{\theta}(0)$, $\hat{\theta}_c$, $\hat{\theta}_s$ and $\hat{\theta}_w$ are $-(5c+3c^2)\theta^0/3(1+c)^2$, $(11c^2-8c^3-12c^4)\theta^0/(3-4c-13c^2+8c^3+12c^4)$, $-(7c^2+3c^3)\theta^0/3(1+c)^3$ and $(7c^2+c^3)\theta^0/9(1+c)^3$ respectively. Another example for the model $g(x;\theta) = \theta x^3$ shows similar results.

The results of Corollary 3.1 and 3.2 are useful in view of the fact that every smooth (or piecewise smooth) function can be approximated by a polynomial function with arbitrarily given precision.

3.3. Exponential Model

Now let us consider the model

$$g(x,\theta) = \exp(-x'\theta).$$

For this model we have

$$\begin{aligned} G(z_i,\theta) &= E\left[\exp\left(-x_i'\theta\right) \mid z_i\right] \\ &= \exp\left(-z_i'\theta\right) E\left[\exp\left(\epsilon_i'\theta\right) \mid z_i\right]. \end{aligned}$$

Under certain conditions, e.g., all moments $E\left(\|\epsilon\|^j \mid z\right) = o\left(c^j\right), j = 1, 2, ...,$ we have by a Taylor expansion

$$\begin{aligned} G(z_i,\theta) &= \exp\left(-z_i'\theta\right)\left[1 + E\left(\epsilon_i'\theta \mid z_i\right) + E\left(\left(\epsilon_i'\theta\right)^2 \mid z_i\right) + \cdots\right] \\ &= \exp\left(-z_i'\theta\right)\left[1 + \theta'\alpha(z_i) + \tfrac{1}{2}\theta'\Omega(z_i)\theta + o(c)\right]. \end{aligned}$$

Since now

$$\frac{\partial g(x_i,\theta)}{\partial x} = -\exp\left(-x_i'\theta\right)\theta$$

and

$$\frac{\partial^2 g(x_i,\theta)}{\partial x \partial x'} = \exp\left(-x_i'\theta\right)\theta\theta',$$

it is easy to calculate

$$\begin{aligned} A(z_i,\theta) &= \exp\left(-\mu_i'\theta\right)\left[1 + \tfrac{1}{2}\theta'\Omega_i\theta\right] \\ &= \exp\left(-z_i'\theta\right)\exp\left[\theta' E\left(\epsilon_i \mid z_i\right)\right]\left[1 + \tfrac{1}{2}\theta'\Omega_i\theta\right] \\ &= \exp\left(-z_i'\theta\right)\left[1 + \theta'\alpha(z_i) + o(c)\right]\left[1 + \tfrac{1}{2}\theta'\Omega(z_i)\theta\right] \\ &= \exp\left(-z_i'\theta\right)\left[1 + \theta'\alpha(z_i) + \tfrac{1}{2}\theta'\Omega(z_i)\theta + o(c)\right] \end{aligned}$$

which differs from $G(z_i,\theta)$ by $o(c)$. Analogously from (3.1) and (3.2) we have

$$\begin{aligned} A_1(z_i,\theta) &= \exp\left(-z_i'\theta\right)\left[1 + c\theta'\widetilde{\alpha}(z_i) + \tfrac{c}{2}\theta'\widetilde{\Omega}(z_i)\theta\right] \\ &= \exp\left(-z_i'\theta\right)\left[1 + \theta'\alpha(z_i) + \tfrac{1}{2}\theta'\Omega(z_i)\theta + o(c)\right] \end{aligned}$$

and

$$\begin{aligned} A_2(z_i,\theta) &= \exp\left(-z_i'\theta\right)\exp\left[c\theta'\widetilde{\alpha}(z_i)\right] \\ &= \exp\left(-z_i'\theta\right)\left[1 + \theta'\alpha(z_i) + o(c)\right]. \end{aligned}$$

Thus, $A_1(z_i,\theta)$ has an approximation error of $o(c)$, whereas $A_2(z_i,\theta)$ has $O(c)$.

4. A Monte Carlo Study

In this section we provide numerical evaluation of the performance of various approximate estimators. We consider three models. The first model is a polynomial model of the form $y = \theta x^4 + u$. The second model is an exponential model of the form $y = \exp(\theta x) + u$. The third model assumes that the conditional density of y given x is $\lambda(x)\exp(-\lambda(x)y)$, hence $E(y \mid x) = \lambda(x)^{-1}$ and $\text{Var}(y \mid x) = \lambda(x)^{-2}$. We assume that x is unobservable, only its proxy $z = x + \epsilon$ is observed. In conformity with previous Monte Carlo studies in the literature (e.g., Whittemore and Keller (1988)), we assume that the conditional mean and variance of ϵ given z are known and let c take various values, in order to examine the performances and sensitivity of various bias adjusted and approximate estimators.

We generate ϵ from $N(0, c^*)$. For the polynomial model, we let $\theta = 5$, u be generated from $N(0,1)$, x be generated from $N(3,1)$. For the second and third model, we generate x from $N(0.5, 1)$. The second model assumes $\theta = -1$ and $u \sim N(0,1)$. The third model assumes $\lambda(x) = \exp(x)$, hence $\text{Var}(y \mid x) = \exp(-2x)$. We consider $c^* = .1, .25, .5, .75$ and 1. In other words, we are examining the performance and sensitivity of various bias adjusted or approximate estimators when the order of the variance of measurement error in relation to the variance of observable is $c = c^*/(1 + c^*)$. Five hundred replications are conducted for sample size $n = 50$, 100 and 500 respectively. We compare the bias, standard deviation (SD) and root mean squared errors (RMSE) of the approximate NLS, $\hat{\theta}_A$, the Whittemore and Keller (1988) estimator, $\hat{\theta}_w$, the Stefanski (1985) estimator, $\hat{\theta}_s$, the Carroll and Stefanski (1990) estimator, $\hat{\theta}_c$ (which uses (3.1) to approximate $G(z;\theta)$), the naive estimator $\hat{\theta}(0)$ and the Gleser (1989) estimator, $\hat{\theta}_e$, that treats z or $E(x \mid z)$ as if it were x respectively and apply the standard least squares method. The first four estimators have bias of order $o(c)$. The last two estimators have bias of order $O(c)$. The results for the polynomial model and two variants of exponential model are summarized in Tables 1, 2 and 3 respectively.

As one can see from these Tables, the naive procedure of using z in place of x, $\hat{\theta}(0)$, or substituting the missing x by the conditional mean of x given z, $\hat{\theta}_e$, yields estimators which are severely biased and have large RMSE even when the measurement error variance c^* is small. All the adjusted estimators have more or less similar magnitude of bias and RMSE when c^* is small, say $c^* = 0.1$. When c^* is of moderate magnitude $(0.25 < c^* < 0.5)$, the proposed estimator, $\hat{\theta}_A$, and the Whittemore and Keller (1988) estimator, $\hat{\theta}_w$, are the best. Both the Stefanski (1985) estimator, $\hat{\theta}_s$, and the Carroll and Stefanski (1990) estimator, $\hat{\theta}_c$, are sensitive to the magnitude of the variance of measurement errors c^*. When c^* becomes large $(c^* > 0.75)$, the proposed estimator dominates other approximate estimators in terms of bias and RMSE. Furthermore, the performance of $\hat{\theta}_A$ improves when sample size increases.

5. Conclusions

In this paper we propose an alternative approximate least squares or maximum likelihood estimators for nonlinear errors-in-variables models. The approximation formula is simple to derive and easy to compute. It only requires the knowledge of first two moments of measurement errors given measured covariates. There is no need for the knowledge of the conditional distribution of the measurement errors given measured covariates. The specific examples and Monte Carlo studies demonstrate that the performance of this approximate estimator is quite good and robust to the magnitude of the variance of measurement errors to the variance of measured covariates while the performance of other approximate estimators are sensitive to this magnitude.

Table 1. The Performance of Various Estimators for the Model
$y = 5x^4 + u,\ u \sim N(0,1)$

	$\hat{\theta}(0)$	$\hat{\theta}_w$	$\hat{\theta}_s$	$\hat{\theta}_c$	$\hat{\theta}_e$	$\hat{\theta}_A$
$n = 50,\ c^* = 0.1$						
Bias	-0.49	0.00	0.00	0.05	0.11	-0.03
SD	0.42	0.44	0.49	0.44	0.45	0.43
RMSE	0.65	0.44	0.49	0.44	0.46	0.43
$n = 50,\ c^* = 0.25$						
Bias	-1.07	0.14	-0.06	0.50	0.30	-0.04
SD	0.58	0.65	0.81	0.69	0.68	0.62
RMSE	1.22	0.67	0.81	0.85	0.74	0.62
$n = 50,\ c^* = 0.5$						
Bias	-1.79	0.46	-0.29	1.53	0.58	-0.03
SD	0.66	0.89	1.14	1.35	0.89	0.77
RMSE	1.91	1.01	1.18	2.04	1.06	0.77
$n = 50,\ c^* = 0.75$						
Bias	-2.32	0.80	-0.58	-3.67	0.82	-0.01
SD	0.68	1.09	1.35	2.78	1.02	0.85
RMSE	2.41	1.35	1.47	4.61	1.31	0.85
$n = 50,\ c^* = 1.0$						
Bias	-2.71	1.09	-0.88	-6.06	1.02	0.00
SD	0.66	1.27	1.48	1.20	1.11	0.89
RMSE	2.79	1.67	1.73	6.18	1.51	0.89
$n = 100,\ c^* = 0.1$						
Bias	-0.48	0.03	0.00	0.07	0.14	-0.01
SD	0.32	0.34	0.38	0.34	0.35	0.33
RMSE	0.58	0.34	0.38	0.35	0.37	0.33
$n = 100,\ c^* = 0.25$						
Bias	-1.08	0.17	-0.10	0.54	0.33	-0.01
SD	0.44	0.50	0.61	0.53	0.52	0.48
RMSE	1.16	0.53	0.62	0.76	0.61	0.48

Table 1. Continued

$n = 100$, $c^* = 0.5$						
Bias	-1.82	0.50	-0.39	1.54	0.60	-0.01
SD	0.50	0.68	0.85	0.88	0.68	0.59
RMSE	1.89	0.84	0.94	1.78	0.90	0.59
$n = 100$, $c^* = 0.75$						
Bias	-2.37	0.82	-0.74	-4.23	0.82	-0.01
SD	0.50	0.83	0.98	2.19	0.77	0.64
RMSE	2.42	1.17	1.23	4.77	1.13	0.64
$n = 100$, $c^* = 1.0$						
Bias	-2.77	1.10	-1.08	-6.22	1.01	0.00
SD	0.48	0.97	1.05	0.77	0.83	0.66
RMSE	2.82	1.46	1.51	6.27	1.31	0.66
$n = 500$, $c^* = 0.1$						
Bias	-0.49	0.05	-0.02	0.09	0.15	0.01
SD	0.15	0.15	0.18	0.15	0.16	0.15
RMSE	0.51	0.16	0.18	0.18	0.22	0.15
$n = 500$, $c^* = 0.25$						
Bias	-1.10	0.20	-0.16	0.58	0.34	0.01
SD	0.21	0.23	0.30	0.24	0.24	0.22
RMSE	1.12	0.31	0.34	0.63	0.42	0.22
$n = 500$, $c^* = 0.50$						
Bias	-1.87	0.54	-0.52	1.56	0.61	0.01
SD	0.24	0.31	0.41	0.48	0.31	0.26
RMSE	1.88	0.62	0.66	1.63	0.68	0.26
$n = 500$, $c^* = 0.75$						
Bias	-2.43	0.87	-0.93	-4.86	0.84	0.01
SD	0.24	0.38	0.48	1.06	0.34	0.28
RMSE	2.44	0.95	1.04	4.97	0.91	0.28
$n = 500$, $c^* = 1.0$						
Bias	-2.85	1.13	-1.32	-6.24	1.03	0.02
SD	0.23	0.45	0.51	0.29	0.37	0.29
RMSE	2.86	1.22	1.41	6.25	1.09	0.29

Table 2. The Performance of Various Estimators for the Model
$y = \exp(-x) + u,\ u \sim N(0,1)$

	$\hat{\theta}(0)$	$\hat{\theta}_w$	$\hat{\theta}_s$	$\hat{\theta}_c$	$\hat{\theta}_e$	$\hat{\theta}_A$
$n = 50,\ c^* = 0.1$						
Bias	0.06	0.00	-0.01	-0.03	-0.01	0.01
SD	0.06	0.06	0.07	0.07	0.07	0.06
RMSE	0.09	0.06	0.08	0.07	0.07	0.06
$n = 50,\ c^* = 0.25$						
Bias	0.13	-0.01	-0.02	-0.26	-0.03	0.01
SD	0.09	0.09	0.13	0.34	0.10	0.09
RMSE	0.16	0.09	0.13	0.42	0.11	0.09
$n = 50,\ c^* = 0.5$						
Bias	0.22	-0.04	-0.01	0.08	-0.07	0.01
SD	0.10	0.12	0.18	0.64	0.13	0.11
RMSE	0.25	0.13	0.18	0.64	0.15	0.11
$n = 50,\ c^* = 0.75$						
Bias	0.29	-0.09	0.02	0.60	-0.10	0.01
SD	0.11	0.14	0.22	0.25	0.16	0.12
RMSE	0.31	0.17	0.22	0.65	0.18	0.13
$n = 50,\ c^* = 1.0$						
Bias	0.35	-0.14	0.05	0.83	-0.13	0.01
SD	0.11	0.16	0.24	0.10	0.17	0.13
RMSE	0.36	0.21	0.25	0.83	0.21	0.13
$n = 100,\ c^* = 0.1$						
Bias	0.06	0.00	-0.01	-0.04	-0.02	0.00
SD	0.05	0.05	0.06	0.06	0.06	0.05
RMSE	0.08	0.05	0.06	0.07	0.06	0.05
$n = 100,\ c^* = 0.25$						
Bias	0.13	-0.01	-0.01	-0.25	-0.04	0.00
SD	0.07	0.08	0.10	0.21	0.08	0.08
RMSE	0.15	0.08	0.10	0.33	0.09	0.08

Table 2. Continued

$n = 100, c^* = 0.5$						
Bias	0.22	-0.05	0.01	0.23	-0.07	0.00
SD	0.08	0.10	0.14	0.37	0.11	0.10
RMSE	0.24	0.11	0.14	0.44	0.13	0.10
$n = 100, c^* = 0.75$						
Bias	0.29	-0.10	0.04	0.65	-0.10	0.00
SD	0.09	0.12	0.17	0.11	0.13	0.10
RMSE	0.31	0.15	0.17	0.66	0.17	0.10
$n = 100, c^* = 1.0$						
Bias	0.35	-0.15	0.08	0.85	-0.13	0.00
SD	0.09	0.13	0.18	0.06	0.14	0.11
RMSE	0.36	0.20	0.20	0.85	0.19	0.11
$n = 500, c^* = 0.1$						
Bias	0.06	0.00	0.00	-0.04	-0.02	0.00
SD	0.03	0.03	0.03	0.03	0.03	0.03
RMSE	0.06	0.03	0.03	0.05	0.04	0.03
$n = 500, c^* = 0.25$						
Bias	0.13	-0.02	0.01	-0.22	-0.04	0.00
SD	0.04	0.04	0.06	0.09	0.05	0.04
RMSE	0.14	0.05	0.06	0.24	0.06	0.04
$n = 500, c^* = 0.5$						
Bias	0.23	-0.06	0.04	0.36	-0.07	0.00
SD	0.05	0.06	0.08	0.10	0.06	0.05
RMSE	0.23	0.08	0.09	0.38	0.10	0.05
$n = 500, c^* = 0.75$						
Bias	0.30	-0.10	0.09	0.69	-0.10	-0.01
SD	0.05	0.06	0.09	0.05	0.07	0.06
RMSE	0.30	0.12	0.12	0.69	0.13	0.06
$n = 500, c^* = 1.0$						
Bias	0.36	-0.15	0.13	0.87	-0.13	-0.01
SD	0.05	0.07	0.09	0.03	0.08	0.06
RMSE	0.36	0.17	0.16	0.87	0.15	0.06

Table 3. The Performance of Various Estimators for the Model
$f(y \mid x) = \lambda(x)\exp[-\lambda(x)y]$, $\lambda(x) = \exp(x)$

	$\hat{\theta}(0)$	$\hat{\theta}_w$	$\hat{\theta}_s$	$\hat{\theta}_c$	$\hat{\theta}_e$	$\hat{\theta}_A$
$n = 50$, $c^* = 0.1$						
Bias	0.11	0.05	0.04	0.02	0.04	0.06
SD	0.18	0.19	0.21	0.20	0.20	0.19
RMSE	0.21	0.20	0.21	0.20	0.20	0.20
$n = 50$, $c^* = 0.25$						
Bias	0.18	0.05	0.04	-0.21	0.02	0.06
SD	0.18	0.20	0.24	0.48	0.21	0.20
RMSE	0.25	0.20	0.24	0.52	0.21	0.20
$n = 50$, $c^* = 0.5$						
Bias	0.27	0.01	0.05	0.07	-0.01	0.06
SD	0.18	0.22	0.27	0.62	0.24	0.20
RMSE	0.32	0.22	0.28	0.62	0.24	0.21
$n = 50$, $c^* = 0.75$						
Bias	0.33	-0.03	0.08	0.58	-0.04	0.06
SD	0.17	0.23	0.30	0.22	0.26	0.21
RMSE	0.38	0.24	0.31	0.62	0.26	0.22
$n = 50$, $c^* = 1.0$						
Bias	0.39	-0.08	0.12	0.81	-0.06	0.05
SD	0.17	0.25	0.32	0.11	0.27	0.22
RMSE	0.42	0.27	0.34	0.81	0.28	0.22
$n = 100$, $c^* = 0.1$						
Bias	0.09	0.03	0.03	0.00	0.02	0.04
SD	0.15	0.15	0.16	0.16	0.16	0.15
RMSE	0.17	0.16	0.17	0.16	0.16	0.16
$n = 100$, $c^* = 0.25$						
Bias	0.16	0.02	0.03	-0.22	0.00	0.04
SD	0.14	0.16	0.19	0.35	0.17	0.16
RMSE	0.22	0.16	0.19	0.42	0.17	0.16

Table 3. Continued

$n = 100$, $c^* = 0.5$						
Bias	0.25	-0.02	0.05	0.23	-0.04	0.04
SD	0.14	0.17	0.21	0.33	0.19	0.16
RMSE	0.29	0.17	0.21	0.40	0.19	0.17
$n = 100$, $c^* = 0.75$						
Bias	0.32	-0.06	0.09	0.63	-0.06	0.03
SD	0.13	0.18	0.22	0.13	0.20	0.16
RMSE	0.35	0.19	0.23	0.64	0.21	0.17
$n = 100$, $c^* = 1.0$						
Bias	0.37	-0.11	0.12	0.83	-0.09	0.03
SD	0.13	0.20	0.22	0.08	0.21	0.17
RMSE	0.40	0.23	0.26	0.84	0.23	0.17
$n = 500$, $c^* = 0.1$						
Bias	0.07	0.01	0.01	-0.02	0.00	0.01
SD	0.09	0.09	0.10	0.10	0.10	0.09
RMSE	0.12	0.09	0.10	0.10	0.10	0.09
$n = 500$, $c^* = 0.25$						
Bias	0.15	0.00	0.02	-0.22	-0.03	0.01
SD	0.09	0.10	0.11	0.11	0.10	0.09
RMSE	0.17	0.10	0.12	0.25	0.11	0.10
$n = 500$, $c^* = 0.5$						
Bias	0.24	-0.04	0.06	0.35	-0.06	0.01
SD	0.08	0.10	0.13	0.11	0.11	0.10
RMSE	0.25	0.11	0.14	0.37	0.13	0.10
$n = 500$, $c^* = 0.75$						
Bias	0.31	-0.09	0.10	0.68	-0.09	0.01
SD	0.08	0.11	0.13	0.06	0.12	0.10
RMSE	0.32	0.14	0.17	0.68	0.15	0.10
$n = 500$, $c^* = 1.0$						
Bias	0.37	-0.14	0.14	-0.14	-0.11	0.00
SD	0.08	0.12	0.14	0.03	0.12	0.09
RMSE	0.37	0.18	0.20	0.86	0.17	0.09

References

[1] Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

[2] Amemiya, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, 28, 273-290.

[3] Amemiya, Y. and W.A. Fuller (1988). Estimation for the Nonlinear Functional Relationship. *Annals of Statistics*, 16, 147-160.

[4] Berkson, J. (1950). Are There Two Regressions? *Journal of the American Statistical Association*, 45, 164-180.

[5] Carroll, R.J. (1989). Covariance Analysis in Generalized Linear Measurement Error Models. *Statistics in Medicine*, 8, 1075-1093.

[6] Carroll, R.J. and L.A. Stefanski (1990). Approximate Quasi-likelihood Estimation in Models with Surrogate Predictors. *Journal of the American Statistical Association*, 85, 652-663.

[7] Chesher, A. (1991). The Effects of Measurement Error. *Biometrika*, 78, 451-462.

[8] Gleser, L.J. (1989). Improvement of the Naive Approach to Estimation in Nonlinear Errors-in-Variables Regression Models. Reprint.

[9] Hsiao, C. (1989). Consistent Estimation for Some Nonlinear Errors-in-Variables Models. *Journal of Econometrics*, 41, 159-185.

[10] Hsiao, C. (1991). Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data. *Review of Economic Studies*, 58, 717-731.

[11] Lee, L.F. and J.H. Sepanski (1995). Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data. *Journal of the American Statistical Association*, 90, 130-140.

[12] Rosner, B., Willett, W.C. and Spiegelman D. (1989). Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-Person Measurement Error, *Statistics in Medicine*, 8, 1051-1070.

[13] Rudemo, M., D. Ruppert and J.C. Streibig (1989). Random Effect Models in Nonlinear Regression with Applications to Bioassay. *Biometrics*, 45, 349-362.

[14] Shafer, D. (1987). Covariate Measurement Error in Generalized Linear Models. *Biometrika*, 74, 385-391.

[15] Stefanski, L.A. (1985). The Effects of Measurement Error on Parameter Estimation. *Biometrika*, 72, 583-592.

[16] Stefanski, L.A. and R.J. Carroll (1985). Covariate Measurement Error in Logistic Regression. *Annals of Statistics*, 13, 1335-1351.

[17] Tosteson, T., Stefanski, L.A. and Schafter, D.W. (1989). A Measurement Error Model for Binary and Ordinal Regression, *Statistics in Medicine*, 8, 1139-1147.

[18] Whittemore, A.S. and J.B. Keller (1988). Approximations for Regression With Covariate Measurement Error. *Journal of American Statistical Association*, 83, 1057-1066.

[19] Wolter, K.M. and W.A. Fuller (1982a). Estimation of Nonlinear Errors-in-Variables Models. *Annals of Statistics*, 10, 539-548.

[20] Wolter, K.M. and W.A. Fuller (1982b). Estimation of the quadratic errors-in-variables models. *Biometrika*, 69, 175-182.

Cheng Hsiao
Department of Economics
University of Southern California
Los Angeles
CA 90089-0253, USA

Qing Wang
Graduate School of Business
University of Chicago
Chicago
IL 60637, USA

Liqun Wang
Institute for Statistics and Econometrics
University of Basel
Holbeinstrasse 12
CH-4051 Basel
Switzerland