Statistica Sinica **13**(2003), 1201-1210

ESTIMATION OF NONLINEAR BERKSON-TYPE MEASUREMENT ERROR MODELS

Liqun Wang

University of Manitoba

Abstract: This paper studies a minimum distance moment estimator for general nonlinear regression models with Berkson-type measurement errors in predictor variables. The estimator is based on the first two conditional moments of the response variable given the observed predictor variable. It is shown that under general regularity conditions the proposed estimator is consistent and asymptotically normally distributed.

Key words and phrases: Asymptotic normality, consistency, errors-in-variables, least squares method, minimum distance estimator, moment estimation.

1. Introduction

Many scientific studies involve fitting a nonlinear relationship of the form

$$y = g(x;\theta) + \varepsilon, \tag{1}$$

where $y \in I\!R$ is the response variable, $x \in I\!R$ is the predictor variable, $\theta \in I\!R^p$ is the unknown regression parameter and ε is the random error. Often the predictor variable x cannot be measured directly, or it is measured with substantial random error. A special type of measurement error is described by Berkson (1950): a controlled variable z is observed, related to the true predictor variable through

$$x = z + \delta, \tag{2}$$

where δ is the unobserved random measurement error.

For example, in a study of the relationship between the temperature used to dry a sample for chemical analysis and the resulting concentration of a volatile constituent, an oven is used to prepare samples. The temperature z is set at $300, 350, 400, 450$ and 500 degrees Fahrenheit, respectively. The true temperature x inside the oven, however, may vary randomly around the pre-set values. In such a situation (2) is a reasonable model for the measurement error.

The distinguishing stochastic feature of Berkson's measurement error model (2) is that, in this model, the measurement error δ is independent of the observed predictor z but is dependent on the unobserved true variable x. This is fundamentally different from the classical errors-in-variables model, where the measurement error is independent of x, but dependent on z. This difference in the stochastic structure leads to completely different procedures in parameter estimation and inference about the models.

Estimation of linear regression models with Berkson-type measurement errors has been discussed in Fuller (1987) and Cheng and Van Ness (1999). For nonlinear models, an approximate estimation procedure called regression calibration has been investigated in Carroll, Ruppert and Stefanski (1995). Recently, Huwang and Huang (2000) studied a polynomial model where $g(x;\theta)$ is a polynomial in x for which the order is known. In particular, they showed that the model can be identified using the first two conditional moments of y given z and model parameters can be estimated consistently by the least squares method.

In many applications, however, the underlying theory suggests a specific nonlinear structure of the regression relationship between y and x. From a technical point of view, genuine nonlinear models are parsimonious and computationally more efficient than polynomial models, which sometimes suffer from such problems as multicollinearity.

In this paper, we propose a minimum distance estimator for general nonlinear models (1) and (2), using the first two conditional moments of y given z. Our method generalizes the results of Huwang and Huang (2000). Furthermore, we derive the consistency and asymptotic normality of this estimator.

Throughout the paper we assume that ε and δ are normally distributed with zero means and variances σ_ε^2 and σ_δ^2, respectively. In addition, we assume that z, δ and ε are independent and y has finite second moment. We also adopt the common assumption that the measurement error is "non-differential" in the sense that the conditional expectation of y given x and z is the same as the conditional expectation of y given x. Although in this paper z is assumed to be a random variable, it is easy to see that all results continue to hold if the observations $z_1, z_2, \ldots, z_n$ are treated as fixed constants such that $\lim_{n\to\infty} \sum_{i=1}^n z_i/n$ and $\lim_{n\to\infty} \sum_{i=1}^n z_i^2/n$ are finite.

In Section 2, we give two examples to motivate our estimation method. In Section 3, we formally define the minimum distance estimator and derive its consistency and asymptotic normality under some general regularity conditions. Conclusions and a discussion are given in Section 4. Finally, proofs of the theorems are given in Section 5.

2. Motivation

To motivate our estimation procedure, we consider some simple examples.

Example 1. Take $g(x;\theta) = \theta_1 \exp(\theta_2 x)$ and $\theta_1\theta_2 \neq 0$. For this model the first two conditional moments can be written as

$$E(y|z) = \theta_1 e^{\theta_2 z} E\left(e^{\theta_2 \delta}\right) = \varphi_1 e^{\varphi_2 z}, \tag{3}$$

where $\varphi_1 = \theta_1 \exp(\theta_2^2 \sigma_\delta^2/2)$ and $\varphi_2 = \theta_2$; and

$$E(y^2|z) = \theta_1^2 e^{2\theta_2 z} E\left(e^{2\theta_2 \delta}\right) + E(\varepsilon^2) = \psi_1 e^{2\varphi_2 z} + \psi_2, \tag{4}$$

where $\psi_1 = \theta_1^2 \exp(2\theta_2^2\sigma_\delta^2)$ and $\psi_2 = \sigma_\varepsilon^2$. Since (3) and (4) are the usual nonlinear regression equations and both y and z are observable, it is easy to see that (φ_i, ψ_j) are identified by these equations and, therefore, can be consistently estimated by the nonlinear least squares method. Furthermore, the original parameters $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2)$ are identified because the mapping $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2) \mapsto (\varphi_i, \psi_j)$ is bijective. Indeed, straightforward calculation shows that $\theta_1 = \varphi_1^2/\sqrt{\psi_1}$, $\theta_2 = \varphi_2$, $\sigma_\delta^2 = \log(\psi_1/\varphi_1^2)/\varphi_2^2$ and $\sigma_\varepsilon^2 = \psi_2$.

Example 2. Now let $g(x;\theta) = \theta_0 + \theta_1 x + \theta_2 x^2$, where $\theta_2 \neq 0$. For this model the first two conditional moments are respectively

$$\begin{aligned} E(y|z) &= \theta_0 + \theta_1 E[(z+\delta)|z] + \theta_2 E[(z+\delta)^2|z] \\ &= \theta_0 + \theta_1 z + \theta_2 z^2 + \theta_2\sigma_\delta^2 = \varphi_1 + \varphi_2 z + \varphi_3 z^2, \end{aligned} \tag{5}$$

where $\varphi_1 = \theta_0 + \theta_2\sigma_\delta^2$, $\varphi_2 = \theta_1$ and $\varphi_3 = \theta_2$; and

$$\begin{aligned} E(y^2|z) &= E[(\theta_0 + \theta_1 x + \theta_2 x^2)^2|z] + E(\varepsilon^2|z) \\ &= \psi_1 + \psi_2 z + \psi_3 z^2 + 2\varphi_2\varphi_3 z^3 + \varphi_3^2 z^4, \end{aligned} \tag{6}$$

where $\psi_1 = \theta_0^2 + (2\theta_0\theta_2 + \theta_1^2)\sigma_\delta^2 + 3\theta_2^2\sigma_\delta^4 + \sigma_\varepsilon^2$, $\psi_2 = 2\theta_1(\theta_0 + 3\theta_2\sigma_\delta^2)$ and $\psi_3 = 2\theta_0\theta_2 + \theta_1^2 + 6\theta_2^2\sigma_\delta^2$. Again, parameters (φ_i, ψ_j) are identified by the usual nonlinear regression equations (5) and (6), and the original parameters $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2)$ are identified because the mapping $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2) \mapsto (\varphi_i, \psi_j)$ is bijective. In fact, we find easily that $\theta_0 = \varphi_1 - \varphi_3\sigma_\delta^2$, $\theta_1 = \varphi_2$, $\theta_2 = \varphi_3$, $\sigma_\delta^2 = (\psi_3 - 2\varphi_1\varphi_3 - \varphi_2^2)/(4\varphi_3^2)$ and $\sigma_\varepsilon^2 = \psi_1 - \varphi_1^2 - \varphi_2^2\sigma_\delta^2 - 2\varphi_3^2\sigma_\delta^4$.

The above examples suggest that in many situations, parameters in nonlinear models can be identified and, therefore, consistently estimated using the first two conditional moments of y given z. Indeed, Huwang and Huang (2000) have shown that a polynomial model of any given order can be identified and estimated in this way. It is worthwhile noting that in general the mapping $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2) \mapsto (\varphi_i, \psi_j)$ needs not be bijective. In the case where more than one set of values of $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2)$ correspond to the same value of (φ_i, ψ_j), restrictions on $(\theta_i, \sigma_\delta^2, \sigma_\varepsilon^2)$ are needed to ensure their identifiability. In the next section, we develop a minimum distance estimator for the general nonlinear model (1) and (2) based on the first two conditional moments.

3. Minimum Distance Estimator

Under the assumptions for model (1) and (2), it is easy to see that the first two conditional moments are given by

$$E(y|z) = E[g(z+\delta;\theta)|z] + E(\epsilon|z) = \int g(z+t;\theta) f_\delta(t)dt, \tag{7}$$

$$E(y^2|z) = E[g^2(z+\delta;\theta)|z] + E(\varepsilon^2|z) = \int g^2(z+t;\theta) f_\delta(t)dt + \sigma_\varepsilon^2, \tag{8}$$

where $f_\delta(t) = (2\pi\sigma_\delta^2)^{1/2} \exp(-t^2/(2\sigma_\delta^2))$.

Let $\gamma = (\theta', \sigma_\delta, \sigma_\varepsilon)'$ denote the vector of model parameters and $\Gamma = \Theta \times \Sigma_\delta \times \Sigma_\varepsilon \subset I\!R^{p+2}$ the parameter space. The true parameter value of model (1) and (2) is denoted by $\gamma_0 \in \Gamma$. For every $\gamma \in \Gamma$, define

$$m_1(z;\gamma) = \int g(z+t;\theta) f_\delta(t)dt, \tag{9}$$

$$m_2(z;\gamma) = \int g^2(z+t;\theta) f_\delta(t)dt + \sigma_\varepsilon^2. \tag{10}$$

Note that $m_1(z;\gamma_0) = E(y|z)$ and $m_2(z;\gamma_0) = E(y^2|z)$.

Now, suppose $(y_i, z_i)', i = 1, \ldots, n$ is an i.i.d. random sample. Then the minimum distance estimator (MDE) $\hat{\gamma}_n$ for γ can be found by minimizing the objective function

$$Q_n(\gamma) = \sum_{i=1}^{n} [(y_i - m_1(z_i;\gamma))^2 + (y_i^2 - m_2(z_i;\gamma))^2], \tag{11}$$

given that γ is identified by (7) and (8). Regularity conditions under which $\hat{\gamma}_n$ is identified, consistent and asymptotically normally distributed are well-known, and often varied, in nonlinear regression literature, see e.g., Amemiya (1985), Gallant (1987) and Seber and Wild (1989).

In the rest of the paper we give regularity conditions in terms of the regression function $g(x;\theta)$ and measurement error distribution $f_\delta(t)$. We adopt the set-up of Amemiya (1985). Let μ denote the Lebesgue measure on $I\!R$ and $\|\cdot\|$ the Euclidean norm in $I\!R^p$. We assume the following.

A1 Parameter spaces $\Theta \subset I\!R^p$, $\Sigma_\delta \subset I\!R$ and $\Sigma_\varepsilon \subset I\!R$ are compact.

A2 $g(x;\theta)$ is continuous in θ for μ-almost all x.

A3 $E \int \sup_{\Theta} g^4(z+t;\theta) \sup_{\Sigma_\delta} f_\delta(t)dt < \infty$.

A4 $E[m_1(z;\gamma) - m_1(z;\gamma_0)]^2 + E[m_2(z;\gamma) - m_2(z;\gamma_0)]^2 = 0$ if and only if $\gamma = \gamma_0$.

These conditions are common in the literature. In particular, the compactness of space Θ for the regression parameter θ is often used. From a practical point of view, assumption A1 is not as restrictive as it seems to be, because for any given problem one usually has some information about the possible range of the parameters. In some situations, the error variance σ^2 can be estimated using the corresponding residuals. In such cases the compactness of Σ is not needed. In our approach, however, the variances σ_ε^2 and σ_δ^2 are simultaneously estimated together with θ. The advantage is that the joint distribution of estimators $\hat{\theta}$, $\hat{\sigma}_\delta$ and $\hat{\sigma}_\varepsilon$ can be obtained. Assumption A2 is usually used to ensure that the objective function $Q_n(\gamma)$ is continuous in γ. Assumption A3 is a moment condition which implies the uniform convergence of $Q_n(\gamma)$. In view of (7) and (8), this assumption is equivalent to y and ε having finite fourth moment. Finally, assumption A4 is the usual condition for identifiability of parameters, which means that the true parameter γ_0 is the unique minimizer of the objective function $Q_n(\gamma)$ for large n.

Theorem 1. *Under* A1−A4, *the MDE* $\hat{\gamma}_n \xrightarrow{P} \gamma_0$, *as* $n \to \infty$.

Proof. See the appendix.

For asymptotic normality, we assume further regularity as follows.

A5 The true parameter θ_0 is contained in an open subset Θ_0 of Θ. For μ-almost all x, the function $g(x;\theta)$ has continuous first and second order partial derivatives with respect to $\theta \in \Theta_0$.

A6 The first two derivatives of $g(x;\theta)$ satisfy

$$E\int \sup_{\Theta_0}\left\|\frac{\partial g(z+t;\theta)}{\partial\theta}\right\|^2 \sup_{\Sigma_\delta} f_\delta(t)dt < \infty,$$
$$E\int \sup_{\Theta_0}\left\|\frac{\partial^2 g(z+t;\theta)}{\partial\theta\partial\theta'}g(z+t;\theta)\right\| \sup_{\Sigma_\delta} f_\delta(t)dt < \infty,$$
$$E\int \sup_{\Theta_0}\left\|\frac{\partial^2 g(z+t;\theta)}{\partial\theta\partial\theta'}\right\| \sup_{\Sigma_\delta} f_\delta(t)dt < \infty.$$

A7 The matrix $A = 2E[(\partial m_1(z;\gamma_0)/\partial\gamma)(\partial m_1(z;\gamma_0)/\partial\gamma') + (\partial m_2(z;\gamma_0)/\partial\gamma)(\partial m_2(z;\gamma_0)/\partial\gamma')]$ is nonsingular, where $\partial m_1(z;\gamma)/\partial\gamma$ is the column vector with elements

$$\frac{\partial m_1(z;\gamma)}{\partial\theta} = \int \frac{\partial g(z+t;\theta)}{\partial\theta} f_\delta(t)dt,$$
$$\frac{\partial m_1(z;\gamma)}{\partial\sigma_\delta^2} = \int g(z+t;\theta)\frac{t^2-\sigma_\delta^2}{2\sigma_\delta^4} f_\delta(t)dt,$$
$$\frac{\partial m_1(z;\gamma)}{\partial\sigma_\varepsilon^2} = 0,$$

and $\partial m_2(z;\gamma)/\partial\gamma$ is the column vector with elements

$$\frac{\partial m_2(z;\gamma)}{\partial\theta} = 2\int \frac{\partial g(z+t;\theta)}{\partial\theta} g(z+t;\theta) f_\delta(t)dt,$$
$$\frac{\partial m_2(z;\gamma)}{\partial\sigma_\delta^2} = \int g^2(z+t;\theta)\frac{t^2-\sigma_\delta^2}{2\sigma_\delta^4} f_\delta(t)dt,$$
$$\frac{\partial m_2(z;\gamma)}{\partial\sigma_\varepsilon^2} = 1.$$

Again, A5−A7 are commonly found as sufficient for the asymptotic normality of the estimators: A5 ensures that the first derivative of $Q_n(\gamma)$ admits a first-order Taylor expansion; A6 implies the uniform convergence of the second derivative of $Q_n(\gamma)$; A7 implies that the second derivative of $Q_n(\gamma)$ has a non-singular limiting matrix.

Theorem 2. *Under* A1−A7, *as* $n\to\infty$, $\sqrt{n}(\hat\gamma_n-\gamma_0)\xrightarrow{d} N(0,A^{-1}BA^{-1})$, *where* $B=\lim_{n\to\infty} E[(1/n)(\partial Q_n(\gamma_0)/\partial\gamma)(\partial Q_n(\gamma_0)/\partial\gamma')]$ and $(\partial Q_n(\gamma)/\partial\gamma) = -2\sum_{i=1}^n [(y_i-m_1(z_i;\gamma))(\partial m_1(z_i;\gamma)/\partial\gamma) + (y_i^2-m_2(z_i;\gamma))(\partial m_2(z_i;\gamma)/\partial\gamma)]$.

Proof. See the appendix.

4. Conclusion and Discussion

We propose a minimum distance moment estimator for general nonlinear regression models with Berkson-type measurement errors in predictor variables that is based on the first two conditional moments of the response variable given the observed predictor variable. It is consistent and asymptotically normally distributed under fairly simple and general regularity conditions. The results obtained generalize those of Huwang and Huang (2000) who deal with polynomial models.

The proposed estimator is obtained by minimizing the objective function $Q_n(\gamma)$ in (11), which can be done using the usual procedures of numerical computation. For some regression functions $g(x;\theta)$, explicit forms of the integrals in (9) and (10) may be hard or impossible to derive. In this case, numerical integration techniques can be used. Alternatively, the method of simulated moments (MSM) of McFadden (1989) or Pakes and Pollard (1989) can be applied. In some special cases the numerical minimization can be simplified. For example, for the polynomial models of Huwang and Huang (2000), a consistent estimator is obtained by using the least squares regression of y on z and then solving a one-dimensional minimization problem for σ_δ^2.

Appendix

We restate some theorems of Amemiya (1985) which are used in the proofs. For this purpose, let $w=(w_1,\ldots,w_n)$ be an i.i.d. random sample and γ a vector of unknown parameters. Further, let $H(w_1,\gamma)$ and $S_n(w,\gamma)$ be measurable functions for any $\gamma \in \Gamma$, and be continuous in γ for any given w. In addition, the parameter space $\Gamma \subset I\!R^k$ is compact. Using these notations, Theorems 4.1.1, 4.2.1 and 4.1.5 of Amemiya (1985) are as follows.

Lemma 3. *Suppose* $H(w_1,\gamma)$ *satisfies* $EH(w_1,\gamma) = 0$ *and* $E\sup_{\gamma\in\Gamma}|H(w_1,\gamma)| < \infty$. *Then* $n^{-1}\sum_{i=1}^n H(w_i,\gamma)$ *converges in probability to zero uniformly in* $\gamma \in \Gamma$.

Lemma 4. *Suppose, as* $n \to \infty$, $S_n(w,\gamma)$ *converges in probability to a non-stochastic function* $S(\gamma)$ *uniformly in* $\gamma \in \Gamma$, *and* $S(\gamma)$ *attains a unique minimum at* γ_0. *Then the estimator* $\hat\gamma_n$ *satisfying* $S_n(w,\hat\gamma_n) = \max_{\gamma\in\Gamma} S_n(w,\gamma)$ *converges in probability to* γ_0.

Lemma 5. *Suppose, as* $n \to \infty$, $S_n(w,\gamma)$ *converges in probability to a non-stochastic function* $S(\gamma)$ *uniformly in* γ *in an open neighborhood of* γ_0, *and* $S(\gamma)$ *is continuous at* γ_0. *Then* $\operatorname{plim}\hat\gamma_n = \gamma_0$ *implies* $\operatorname{plim} S_n(w,\hat\gamma_n) = S(\gamma_0)$.

Proof of Theorem 1. We show that A1-A4 are sufficient for all conditions of Lemma 4. First, by A3 and Hölder's inequality we have $E\int\sup_\Theta|g(z+t;\theta)|^j\sup_{\Sigma_\delta}f_\delta(t)dt < \infty$ for $j = 1,2,3$. It follows from A2 and the Dominated Convergence Theorem, that $Q_n(\gamma)$ is continuous in γ. Further, let $Q(\gamma) = E[y_1 - m_1(z_1;\gamma)]^2 + E[y_1^2 - m_2(z_1;\gamma)]^2$ and $H(y_i,z_i,\gamma) = [y_i - m_1(z_i;\gamma)]^2 + [y_i^2 - m_2(z_i;\gamma)]^2 - Q(\gamma)$. Then $EH(y_i,z_i,\gamma) = 0$. Again, by A3 and Hölder's inequality we have

$$\begin{aligned} E\sup_\Gamma[y_i - m_1(z_i;\gamma)]^2 &\le 2Ey_i^2 + 2E\sup_\Gamma m_1^2(z_i;\gamma) \\ &\le 2Ey_i^2 + 2E\int\sup_\Theta g^2(z+t;\theta)\sup_{\Sigma_\delta}f_\delta(t)dt < \infty \end{aligned}$$

and $E\sup_\Gamma[y_i^2 - m_2(z_i;\gamma)]^2 \le 2Ey_i^4 + 2E\int\sup_\Theta g^4(z+t;\theta)\sup_{\Sigma_\delta}f_\delta(t)dt < \infty$. It follows that $E\sup_{\gamma\in\Gamma}|H(y_i,z_i,\gamma)| < \infty$ and, therefore by Lemma 3, that

$$\sup_\Gamma\left|\frac{1}{n}Q_n(\gamma) - Q(\gamma)\right| = \sup_\Gamma|\frac{1}{n}\sum_{i=1}^n H(y_i,z_i,\gamma)| = o_p(1).$$

Further, since $E[y_1 - m_1(z_1;\gamma)]^2 = E[y_1 - m_1(z_1;\gamma_0)]^2 + E[m_1(z_1;\gamma) - m_1(z_1;\gamma_0)]^2$ and $E[y_1^2 - m_2(z_1;\gamma)]^2 = E[y_1^2 - m_2(z_1;\gamma_0)]^2 + E[m_2(z_1;\gamma) - m_2(z_1;\gamma_0)]^2$, it follows that $Q(\gamma) \ge Q(\gamma_0)$ and, by A4, equality holds if and only if $\gamma = \gamma_0$. Thus all conditions of Lemma 4 hold and $\hat\gamma_n \xrightarrow{P} \gamma_0$ follows.

Proof of Theorem 2. By A5, the first derivative $\partial Q_n(\gamma)/\partial\gamma$ exists and has a first order Taylor expansion in a neighborhood of γ_0. Since $\partial Q_n(\hat\gamma_n)/\partial\gamma = 0$ and

$\hat{\gamma}_n \xrightarrow{P} \gamma_0$, for sufficiently large n, we have

$$0 = \frac{\partial Q_n(\gamma_0)}{\partial \gamma} + \frac{\partial^2 Q_n(\tilde{\gamma})}{\partial \gamma \partial \gamma'}(\hat{\gamma}_n - \gamma_0), \tag{12}$$

where $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma}_n - \gamma_0\|$. The first derivative in (12) is given by

$$\frac{\partial Q_n(\gamma)}{\partial \gamma} = -2\sum_{i=1}^{n}\left[(y_i - m_1(z_i;\gamma))\frac{\partial m_1(z_i;\gamma)}{\partial \gamma} + (y_i^2 - m_2(z_i;\gamma))\frac{\partial m_2(z_i;\gamma)}{\partial \gamma}\right],$$

where $\partial m_1(z_i;\gamma)/\partial\gamma$ is the column vector with elements

$$\frac{\partial m_1(z_i;\gamma)}{\partial \theta} = \int \frac{\partial g(z+t;\theta)}{\partial \theta} f_\delta(t)dt,$$

$$\frac{\partial m_1(z_i;\gamma)}{\partial \sigma_\delta^2} = \int g(z+t;\theta)\frac{t^2 - \sigma_\delta^2}{2\sigma_\delta^4} f_\delta(t)dt,$$

$$\frac{\partial m_1(z_i;\gamma)}{\partial \sigma_\varepsilon^2} = 0,$$

and $\partial m_2(z_i;\gamma)/\partial\gamma$ is the column vector with elements

$$\frac{\partial m_2(z_i;\gamma)}{\partial \theta} = 2\int \frac{\partial g(z+t;\theta)}{\partial \theta} g(z+t;\theta) f_\delta(t)dt,$$

$$\frac{\partial m_2(z_i;\gamma)}{\partial \sigma_\delta^2} = \int g^2(z+t;\theta)\frac{t^2 - \sigma_\delta^2}{2\sigma_\delta^4} f_\delta(t)dt,$$

$$\frac{\partial m_2(z_i;\gamma)}{\partial \sigma_\varepsilon^2} = 1.$$

The second derivative in (12) is given by

$$\begin{aligned}\frac{\partial^2 Q_n(\gamma)}{\partial \gamma \partial \gamma'} &= 2\sum_{i=1}^{n}\left[\frac{\partial m_1(z_i;\gamma)}{\partial \gamma}\frac{\partial m_1(z_i;\gamma)}{\partial \gamma'} + \frac{\partial m_2(z_i;\gamma)}{\partial \gamma}\frac{\partial m_2(z_i;\gamma)}{\partial \gamma'}\right] \\ &\quad - 2\sum_{i=1}^{n}\left[(y_i - m_1(z_i;\gamma))\frac{\partial^2 m_1(z_i;\gamma)}{\partial \gamma \partial \gamma'} + (y_i^2 - m_2(z_i;\gamma))\frac{\partial^2 m_2(z_i;\gamma)}{\partial \gamma \partial \gamma'}\right],\end{aligned}$$

where the non-zero elements in $\partial^2 m_1(z_i;\gamma)/\partial\gamma\partial\gamma'$ are

$$\frac{\partial^2 m_1(z_i;\gamma)}{\partial \theta \partial \theta'} = \int \frac{\partial^2 g(z+t;\theta)}{\partial \theta \partial \theta'} f_\delta(t)dt,$$

$$\frac{\partial^2 m_1(z_i;\gamma)}{\partial \theta \partial \sigma_\delta^2} = \int \frac{\partial g(z+t;\theta)}{\partial \theta}\frac{t^2 - \sigma_\delta^2}{2\sigma_\delta^4} f_\delta(t)dt,$$

$$\frac{\partial^2 m_1(z_i;\gamma)}{(\partial\sigma_\delta^2)^2} = \int g(z+t;\theta)\frac{t^4 - 6\sigma_\delta^2 t^2 + 3\sigma_\delta^4}{4\sigma_\delta^8} f_\delta(t)dt,$$

and the non-zero elements in $\partial^2 m_2(z_i;\gamma)/\partial\gamma\partial\gamma'$ are

$$\begin{aligned}
&\frac{\partial^2 m_2(z_i;\gamma)}{\partial\theta\partial\theta'} \\
&= 2\int \frac{\partial^2 g(z+t;\theta)}{\partial\theta\partial\theta'} g(z+t;\theta) f_\delta(t)dt + 2\int \frac{\partial g(z+t;\theta)}{\partial\theta}\frac{\partial g(z+t;\theta)}{\partial\theta'} f_\delta(t)dt,
\end{aligned}$$

$$\frac{\partial^2 m_2(z_i;\gamma)}{\partial\theta\partial\sigma_\delta^2} = 2\int \frac{\partial g(z+t;\theta)}{\partial\theta} g(z+t;\theta)\frac{t^2-\sigma_\delta^2}{2\sigma_\delta^4} f_\delta(t)dt,$$

$$\frac{\partial^2 m_2(z_i;\gamma)}{(\partial\sigma_\delta^2)^2} = \int g^2(z+t;\theta)\frac{t^4 - 6\sigma_\delta^2 t^2 + 3\sigma_\delta^4}{4\sigma_\delta^8} f_\delta(t)dt.$$

Analogous to the proof of Theorem 1, we can verify by A5 and A6 that $(1/n)\partial^2 Q_n(\gamma)/\partial\gamma\partial\gamma'$ converges in probability to $\partial^2 Q(\gamma)/\partial\gamma\partial\gamma'$ uniformly in a neighborhood of γ_0. It follows from Lemma 5 that

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2 Q_n(\tilde\gamma)}{\partial\gamma\partial\gamma'} \xrightarrow{P} \frac{\partial^2 Q(\gamma_0)}{\partial\gamma\partial\gamma'} \qquad (13)\\
&= 2E\left[\frac{\partial m_1(z;\gamma_0)}{\partial\gamma}\frac{\partial m_1(z;\gamma_0)}{\partial\gamma'} + \frac{\partial m_2(z;\gamma_0)}{\partial\gamma}\frac{\partial m_2(z;\gamma_0)}{\partial\gamma'}\right]\\
&- 2E\left[(y - m_1(z;\gamma_0))\frac{\partial^2 m_1(z;\gamma_0)}{\partial\gamma\partial\gamma'} + (y^2 - m_2(z;\gamma_0))\frac{\partial^2 m_2(z;\gamma_0)}{\partial\gamma\partial\gamma'}\right] = A,
\end{aligned}$$

which is non-singular by A7. Furthermore, by the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\frac{\partial Q_n(\gamma_0)}{\partial\gamma} \xrightarrow{L} N(0,B). \qquad (14)$$

The theorem follows then from (12)−(14) and Slutsky's Theorem (Amemiya (1985)).

Acknowledgement

I am grateful to Professor Brian Macpherson, an associate editor and two anonymous referees for their useful comments and suggestions. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

Amemiya, T. (1985). *Advanced Econometrics.* Basil Blackwell Ltd., Oxford, UK.

Berkson, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45**, 164-180.

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

Cheng, C. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*, Arnold, London.

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

Gallant, A. R. (1987). *Nonlinear Statistical Model*, Wiley, New York.

Huwang, L. and Huang, Y. H. S. (2000). On errors-in-variables in polynomial regression - Berkson case. *Staist. Sinica* **10**, 923-936.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica* **57**, 995-1026.

Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica* **57**, 1027-1057.

Seber, G. A. and Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

Department of Statistics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada.

E-mail: liqun_wang@umanitoba.ca

(Received October 2002; accepted March 2003)