Statistica Sinica **10**(2000), 923-936

ON ERRORS-IN-VARIABLES IN POLYNOMIAL REGRESSION-BERKSON CASE

Longcheen Huwang and Y. H. Steve Huang

University of Michigan and National Tsing Hua University

Abstract: A nonlinear Berkson model, particularly the polynomial Berkson model, is considered in this work. It is shown that without making any identifiability assumptions, all coefficient parameters in this model can be estimated consistently. In particular, the model is shown to be identifiable. However, unlike the linear Berkson model where one can estimate the coefficient parameters by ignoring the measurement error, in the polynomial Berkson model we must take into account the measurement error. An iterative reweighted least squares approach is taken to estimate the parameters in the model. The resulting estimates are found to be the solution of a set of estimating equations. Simulation results of the three methods discussed for large samples in a quadratic Berkson model are compared.

Key words and phrases: Berkson model, controlled variable, identifiability assumption, polynomial model, weighted least squares.

1. Introduction

In a classical linear measurement error (errors-in-variables) model, the true explanatory variable ξ is not observed. The observed variable x equals ξ plus a measurement error δ where ξ and δ are independent. This is fundamentally different from the model described by Berkson (1950), where x is controlled by the experimenter and ξ and δ are dependent. The controlled explanatory variable x is called a controlled variable in the Berkson model.

More precisely the linear Berkson model is defined as follows:

$$y_i = \beta_0 + \beta_1\xi_i + \epsilon_i, \quad \xi_i = x_i - \delta_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where the variables x_i (either stochastic or nonstochastic) are controlled by the experimenter, and the random errors (ϵ_i, δ_i) are two i.i.d. sequences of random variables. In this case ξ_i and δ_i are dependent, which is the main difference from the classical errors-in-variables model. Model (1.1) can be written as

$$y_i = \beta_0 + \beta_1 x_i + (\epsilon_i - \beta_1\delta_i), \quad i = 1, \ldots, n,$$

where x_i is uncorrelated with $\epsilon_i - \beta_1\delta_i$. In this situation, Berkson (1950) pointed out that ordinary least squares should be applied and one can estimate the coefficient parameters by simply ignoring the measurement error δ_i. Furthermore, there is no need to supply extra information if the goal is to estimate the coefficient parameters only. On the contrary, if ξ_i and δ_i are independent (x_i's are not controlled) in (1.1), the ordinary least squares method is not appropriate for the estimation of the same parameters because x_i is correlated with $\epsilon_i - \beta_1\delta_i$. Now the question is whether we can estimate the coefficient parameters in a polynomial Berkson model by ignoring the measurement error δ_i? A straightforward derivation shows that the answer is no. However, by using a different approach, we show that all parameters in the polynomial Berkson model can still be estimated consistently. In particular, no identifiability assumption is required. In other words, the polynomial Berkson model is intrinsically identifiable. Theoretically, one can use a higher order regression to obtain a set of consistent estimates of the parameters in the model. However this is not the most efficient procedure due to error variance heteroscedasticity. Hence we propose an iterative reweighted least squares method which takes heteroscedasticity into account. These iterative reweighted least squares estimates are found as the solution of a set of estimating equations.

The difference between the classical linear measurement error model and the linear Berkson model was first observed by Berkson (1950). Geary (1953) investigated a cubic Berkson model with replicated data. Fedorov (1974) generalized Berkson's model and estimated the parameters by using a certain approximation technique and the iterative weighted least squares method. Fuller (1987, p.81) discussed general results in the linear Berkson model and pointed out the inadequacy of using ordinary least squares if the response is quadratic. He also referred to Box (1961) who investigated the effects of measurement errors in experiments with nonlinear response. Burr (1988) considered the binary Berkson probit model and Rudemo, Ruppert and Streibig (1989) used transformation and weighting techniques to investigate nonlinear Berkson models. Carroll and Stefanski (1990) studied the general theory of common measurement error models (including the Berkson model) with quasi-likelihood estimation.

Throughout the paper, we consider the model

$$y_i = \beta_0 + \beta_1\xi_i + \cdots + \beta_p\xi_i^p + \epsilon_i, \xi_i = x_i - \delta_i, \quad p > 1, i = 1, \ldots, n, \tag{1.2}$$

where ϵ_i are i.i.d. $N(0, \sigma_\epsilon^2)$, δ_i are i.i.d. $N(0, \sigma_\delta^2)$, and the ϵ_i are independent of δ_j for all i and j. Under certain mild conditions all results in the paper hold when the controlled varibles x_i are stochastic, but we will focus the study on the case of nonstochastic x_i. It is assumed that the x_i are fixed and

$$\frac{1}{n}\sum_{i=1}^{n} x_i^l \to \mu_l, \quad l = 0, \ldots, 4p, \tag{1.3}$$

$$\mathbf{M}_n \triangleq \frac{1}{n}\sum (x_i - \mu_1, \ldots, x_i^{2p} - \mu_{2p})'(x_i - \mu_1, \ldots, x_i^{2p} - \mu_{2p}) \to \mathbf{M}, \tag{1.4}$$

and

$$\frac{1}{n^2}\sum_{i=1}^{n} x_i^{2l} \to 0, \quad l = 0, \ldots, 4p, \tag{1.5}$$

where μ_l are real constants and $\mathbf{M}$ is a positive definite matrix.

The rest of the paper is organized as follows. In Section 2 we prove that the model is identifiable. Explanation is also given as to why higher order regression is not used. Section 3 utilizes a numerical method to derive a consistent estimate of σ_δ^2 and hence the rest of parameters. These estimates are called initial estimates. In Section 4, by using the constructed initial estimates, we propose an iterative reweighted least squares method to estimate the parameters. These estimates are shown to be the solution of a set of estimating equations. Section 5 uses simulation results to compare the reweighted least squares estimates with those mentioned in Sections 2 and 3 in a quadratic Berkson model. Proofs are in the Appendix.

2. Consistent Estimates without Identifiability Assumptions

From (1.2) it follows that the conditional distribution of ξ_i given x_i is $N(x_i, \sigma_\delta^2)$. As a consequence, the rth moment of ξ_i given x_i equals

$$E(\xi_i^r \mid x_i) = E[(x_i + \sigma_\delta z)^r \mid x_i] = \sum_{l=0}^{r} \binom{r}{l} x_i^{r-l}\sigma_\delta^l m_l, \tag{2.1}$$

where the random variable z has a standard normal distribution and m_l is the lth moment of the distribution, $l = 0, \ldots, r$. By (2.1),

$$E(y_i \mid x_i) = \beta_0^* + \beta_1^* x_i + \cdots + \beta_p^* x_i^p, \tag{2.2}$$

where $d_{jk} = \binom{j}{k} m_k$, $j = 1, \ldots, p$, $k = 0, \ldots, j$, and

$$(\beta_0^*, \ldots, \beta_p^*) = (\beta_0, \beta_1, \ldots, \beta_p) \begin{bmatrix} 1 & 0 & 0 & \cdot\ \cdot & 0 \\ d_{11}\sigma_\delta & 1 & 0 & \cdot\ \cdot & 0 \\ d_{22}\sigma_\delta^2 & d_{21}\sigma_\delta & 1 & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ d_{pp}\sigma_\delta^p & d_{p\,p-1}\sigma_\delta^{p-1} & d_{p\,p-2}\sigma_\delta^{p-2} & \cdot\ \cdot & 1 \end{bmatrix}. \tag{2.3}$$

Similarly,

$$\begin{aligned} E(y_i^2 \mid x_i) &= E(a_0 + a_1\xi_i + \cdots + a_{2p}\xi_i^{2p} \mid x_i) + E(\epsilon_i^2 \mid x_i) \\ &= a_0^* + a_1^* x_i + \cdots + a_{2p}^* x_i^{2p}, \end{aligned}$$

where

$$a_s = \sum_{\{0\le k,l\le p,k+l=s\}} \beta_k\beta_l, \quad s = 0, \ldots, 2p \tag{2.4}$$

and

$$(a_0^* - \sigma_\epsilon^2, a_1^*, \ldots, a_{2p}^*) = (a_0, \ldots, a_{2p}) \begin{bmatrix} 1 & 0 & 0 & \cdot\ \cdot & 0 \\ d_{11}\sigma_\delta & 1 & 0 & \cdot\ \cdot & 0 \\ d_{22}\sigma_\delta^2 & d_{21}\sigma_\delta & 1 & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ d_{2p\,2p}\sigma_\delta^{2p} & d_{2p\,2p-1}\sigma_\delta^{2p-1} & d_{2p\,2p-2}\sigma_\delta^{2p-2} & \cdot\ \cdot & 1 \end{bmatrix}. \tag{2.5}$$

Since all odd moments of $N(0,1)$ are zero, by (2.3) and (2.5) we have

$$\beta_p^* = \beta_p,\ \beta_{p-1}^* = \beta_{p-1},\ \beta_{p-2}^* = \beta_{p-2} + \beta_p \sigma_\delta^2 \frac{p(p-1)}{2} \tag{2.6}$$

and

$$a_{2p-2}^* = a_{2p-2} + a_{2p}\sigma_\delta^2 \frac{2p(2p-1)}{2} = (2\beta_p\beta_{p-2} + \beta_{p-1}^2) + \beta_p^2\sigma_\delta^2 p(2p-1). \tag{2.7}$$

Substituting (2.6) into (2.7) and simplifying the result, we obtain

$$a_{2p-2}^* = 2\beta_{p-2}^*\beta_p^* + \beta_{p-1}^{*2} + p^2\beta_p^{*2}\sigma_\delta^2. \tag{2.8}$$

Now if we regress y^2 and y on x, we can obtain the consistent estimates $\hat{a}_0^*, \ldots, \hat{a}_{2p}^*$ of $a_0^*, \ldots, a_{2p}^*$ and $\hat{\beta}_0^*, \ldots, \hat{\beta}_p^*$ of $\beta_0^*, \ldots, \beta_p^*$. Then from (2.8), a consistent estimate $\hat{\sigma}_\delta^2$ of σ_δ^2 can be constructed. Consequently, the consistent estimates $\hat{\beta}_0, \ldots, \hat{\beta}_p$ of $\beta_0, \ldots, \beta_p$ can be derived by substituting $\hat{\beta}_0^*, \ldots, \hat{\beta}_p^*$ and $\hat{\sigma}_\delta^2$ for $\beta_0^*, \ldots, \beta_p^*$ and σ_δ^2 in (2.3), respectively. Also, a consistent estimates $\hat{\sigma}_\epsilon^2$ of σ_ϵ^2 can be obtained from (2.5) since we already established the consistent estimates of $a_0^*, \ldots, a_{2p}^*, \sigma_\delta^2$, and $a_0, \ldots, a_{2p}$ (which are functions of $\beta_0, \ldots, \beta_p$). In conclusion, all parameters in the model can be estimated consistently and hence the model is identifiable.

Although we have a way to estimate all parameters in the model, it may not be adequate because the high order regression of y^2 on x is adopted. A common understanding in the literature of regression analysis (see, for example, Seber (1976)) indicates that high order polynomial regression may create ill-conditioned matrix to be inverted in applying the ordinary least squares method. Furthermore, there is a possibility that an equation different from (2.8) may be derived and hence a set of different consistent estimates of σ_δ^2 and other parameters could be obtained. This leads to the question of which consistent estimates should be used.

3. Alternative Consistent Estimates of σ_δ^2 and Other Parameters

In this section we use a different approach to construct a set of consistent estimates for the parameters. First of all, by (2.3), it is easy to observe that the parameters $\boldsymbol{\theta} = (\beta_0, \ldots, \beta_p, \sigma_\delta^2, \sigma_\epsilon^2)$ and $\boldsymbol{\theta}^* = (\beta_0^*, \ldots, \beta_p^*, \sigma_\delta^2, \sigma_\epsilon^2)$ are one-to-one correspondent. Define $\boldsymbol{\psi} = (\beta_0^*, \ldots, \beta_p^*)$ and $\hat{\boldsymbol{\psi}} = (\hat{\beta}_0^*, \ldots, \hat{\beta}_p^*)$, where $\hat{\beta}_0^*, \ldots, \hat{\beta}_p^*$ are the ordinary least squares estimates of $\beta_0^*, \ldots, \beta_p^*$ obtained by regressing y on x. Treating $\beta_0, \ldots, \beta_p$ in (2.3) as functions of $\boldsymbol{\psi}$ and σ_δ^2 (denoted by $\tilde{\beta}_0(\sigma^2, \boldsymbol{\psi}), \ldots, \tilde{\beta}_p(\sigma^2, \boldsymbol{\psi})$ in (3.1)), we can rewrite (2.3) as

$$(\tilde{\beta}_0(\sigma^2, \boldsymbol{\psi}), \ldots, \tilde{\beta}_p(\sigma^2, \boldsymbol{\psi})) = (\beta_0^*, \ldots, \beta_p^*) \begin{bmatrix} 1 & 0 & 0 & \cdot\ \cdot & 0 \\ d_{11}\sigma & 1 & 0 & \cdot\ \cdot & 0 \\ d_{22}\sigma^2 & d_{21}\sigma & 1 & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ d_{pp}\sigma^p & d_{p\,p-1}\sigma^{p-1} & d_{p\,p-2}\sigma^{p-2} & \cdot\ \cdot & 1 \end{bmatrix}^{-1}, \tag{3.1}$$

where the parameter σ^2 is a dummy of σ_δ^2, $0 \le \sigma^2 < \infty$. Similarly, from (2.5), we have

$$(\tilde{a}_0^*(\sigma^2, \boldsymbol{\psi}, \sigma_\epsilon^2) - \sigma_\epsilon^2, \tilde{a}_1^*(\sigma^2, \boldsymbol{\psi}), \ldots, \tilde{a}_{2p}^*(\sigma^2, \boldsymbol{\psi})) = (\tilde{a}_0(\sigma^2, \boldsymbol{\psi}), \ldots, \tilde{a}_{2p}(\sigma^2, \boldsymbol{\psi}))$$
$$\begin{bmatrix} 1 & 0 & 0 & \cdot\ \cdot & 0 \\ d_{11}\sigma & 1 & 0 & \cdot\ \cdot & 0 \\ d_{22}\sigma^2 & d_{21}\sigma & 1 & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot \\ d_{2p\,2p}\sigma^{2p} & d_{2p\,2p-1}\sigma^{2p-1} & d_{2p\,2p-2}\sigma^{2p-2} & \cdot\ \cdot & 1 \end{bmatrix},$$

where

$$\tilde{a}_s(\sigma^2, \boldsymbol{\psi}) = \sum_{\{0 \le k, l \le p,\, k+l=s\}} \tilde{\beta}_k(\sigma^2, \boldsymbol{\psi}) \tilde{\beta}_l(\sigma^2, \boldsymbol{\psi}), \;\; s = 0, \ldots, 2p.$$

From the above relation and $\hat{\boldsymbol{\psi}} \xrightarrow{P} \boldsymbol{\psi}$, it follows that

$$\tilde{\beta}_j(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) \xrightarrow{P} \beta_j, \;\; \tilde{a}_h^*(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) \xrightarrow{P} a_h^*, \;\; j = 0, \ldots, p, \;\; h = 1, \ldots, 2p.$$

Now, we will mimic the regression of y^2 on x to construct a consistent estimate of σ_δ^2. Observe that the sums of squares $\sum_1^n (y_i^2 - a_0^* - a_1^* x_i - \cdots - a_{2p}^* x_i^{2p})^2$ and $\sum_1^n (y_i^2 - \overline{y^2} - a_1^*(x_i - \bar{x}) - \cdots - a_{2p}^*(x_i^{2p} - \overline{x^{2p}}))^2$, $\overline{y^2} = n^{-1} \sum_1^n y_i^2$ and $\overline{x^k} = n^{-1} \sum_1^n x_i^k$, $k = 1, \ldots, 2p$, have the same minimum value on the space of $(a_0^*, \ldots, a_{2p}^*)$ and $(a_1^*, \ldots, a_{2p}^*)$, respectively. In order to get rid of parameter a_0^* (which involves σ_ϵ^2), we use the second form to define the average sum of squares

$$S_n(\sigma^2, \boldsymbol{\psi}) = \frac{1}{n} \sum_1^n [y_i^2 - \overline{y^2} - \tilde{a}_1^*(\sigma^2, \boldsymbol{\psi})(x_i - \bar{x}) - \cdots - \tilde{a}_{2p}^*(\sigma^2, \boldsymbol{\psi})(x_i^{2p} - \overline{x^{2p}})]^2$$

$$= \frac{1}{n}\sum_1^n [y_i^{2c} - \tilde{a}_1^*(\sigma^2, \boldsymbol{\psi})x_i^{1c} - \cdots - \tilde{a}_{2p}^*(\sigma^2, \boldsymbol{\psi})x_i^{2pc}]^2,$$

where $y_i^{2c} = y_i^2 - \overline{y^2}$ and $x_i^{kc} = x_i^k - \overline{x^k}$, $k = 1, \ldots, 2p$. Note that $S_n(\sigma^2, \hat{\boldsymbol{\psi}}) = S_n(\sigma^2, \boldsymbol{\psi}) \mid_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}$ is a function of data and the parameter σ^2 only. Suppose that $\hat{\sigma}_\delta^2$ is the value of σ^2, $0 \le \sigma^2 < \infty$, where $S_n(\sigma^2, \hat{\boldsymbol{\psi}})$ achieves its minimum, i.e.,

$$S_n(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}) = \underset{0\le\sigma^2<\infty}{inf} S_n(\sigma^2, \hat{\boldsymbol{\psi}}). \tag{3.2}$$

It can be seen (see the Remark after Lemma 2) that $S_n(\sigma^2, \hat{\boldsymbol{\psi}})$ always achieves its minimum value at certain finite number. It is also worth noting that only the regression of y on x (to obtain $\hat{\boldsymbol{\psi}}$) and some numerical minimization technique are needed to obtain $\hat{\sigma}_\delta^2$ in this approach. The higher order regression of y^2 on x has been avoided and the problem of inverting ill-conditioned matrix that might arise during the regression will not be encountered. Now, we prove the consistency of $\hat{\sigma}_\delta^2$ in (3.2).

Theorem. *Assume that Model* (1.2) *and Conditions* (1.3) − (1.5) *hold with* $p \ge 2$ *and* $\beta_0, \ldots, \beta_p, \sigma_\delta^2$ *and* σ_ϵ^2 *are unknown parameters. Then* $\hat{\sigma}_\delta^2 \xrightarrow{P} \sigma_\delta^2$, *where* $\hat{\sigma}_\delta^2$ *is as defined in* (3.2).

Proof. See the Appendix.

After constructing the consistent estimates $\hat{\boldsymbol{\psi}}$ and $\hat{\sigma}_\delta^2$, by (3.1) we obtain the consistent estimates $(\hat{\beta}_0, \ldots, \hat{\beta}_p) = (\tilde{\beta}_0(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}), \ldots, \tilde{\beta}_p(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}))$. For estimating σ_ϵ^2, note that $\mathrm{Var}\,(y_i \mid x_i) = (\beta_1, \ldots, \beta_p)\boldsymbol{\Sigma}_i(\beta_1, \ldots, \beta_p)' + \sigma_\epsilon^2$, where $\boldsymbol{\Sigma}_i$ is the covariance matrix of the conditional distribution of $(\xi_i, \xi_i^2, \ldots, \xi_i^p)$ given x_i. Here $\boldsymbol{\Sigma}_i$ can be computed by (2.1) and estimated by $\hat{\boldsymbol{\Sigma}}_i$ which is obtained with $\hat{\sigma}_\delta^2$ in the place of σ_δ^2. Now, define $e_i = y_i - \hat{\beta}_0^* - \cdots - \hat{\beta}_p^* x_i^p$ where $\hat{\beta}_0^*, \ldots, \hat{\beta}_p^*$ are the least squares estimates of $\beta_0^*, \ldots, \beta_p^*$. Then it is easy to show that

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}[\sum_1^n e_i^2 - \sum_1^n \hat{\boldsymbol{\beta}}_1' \hat{\boldsymbol{\Sigma}}_i \hat{\boldsymbol{\beta}}_1] \bigvee 0, \tag{3.3}$$

where $\hat{\boldsymbol{\beta}}_1 = (\tilde{\beta}_1(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}), \ldots, \tilde{\beta}_p(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}))'$ converges to σ_ϵ^2 in probability.

4. Improved Estimates

The procedure to derive the consistent estimates of $\beta_0, \ldots, \beta_p, \sigma_\delta^2$ and σ_ϵ^2 in the last section is not efficient. The least squares estimates $\hat{\boldsymbol{\psi}} = (\hat{\beta}_0^*, \ldots, \hat{\beta}_p^*)$ used in formulating $S_n(\sigma^2, \hat{\boldsymbol{\psi}})$ in (3.2) are not efficient estimates because the conditional variance of y given x depends on x. Also, in minimizing $S_n(\sigma^2, \hat{\boldsymbol{\psi}})$

over $[0, \infty)$, we mimic the regression of y^2 on x. However, we do not take into account that the conditional variance of y^2 given x depends on x as well.

In regression, the usual way to resolve the heteroscedasticity is to apply a weighted least squares approach. Denote the estimates $\hat{\boldsymbol{\psi}}, \hat{\sigma}_\delta^2$, and $\hat{\sigma}_\epsilon^2$ obtained in the last section as the initial estimates $\hat{\boldsymbol{\psi}}_0$ $(=(\hat{\beta}_{00}^*, \ldots, \hat{\beta}_{p0}^*))$, $\hat{\sigma}_{\delta 0}^2$, and $\hat{\sigma}_{\epsilon 0}^2$. We seek weighted least squares estimates of $\boldsymbol{\psi}$ in (2.2), i.e., we find the minimizer of

$$\sum_1^n \hat{W}_i^{-1}(y_i - \beta_0^* - \cdots - \beta_p^* x_i^p)^2, \tag{4.1}$$

where $\hat{W}_i$ is a consistent estimate of $W_i = \mathrm{Var}\,(y_i \mid x_i)$ defined below. We also find the minimizer of

$$\sum_1^n \hat{U}_i^{-1}[y_i^{2c} - \tilde{a}_1^*(\sigma^2, \hat{\boldsymbol{\psi}})x_i^{1c} - \cdots - \tilde{a}_{2p}^*(\sigma^2, \hat{\boldsymbol{\psi}})x_i^{2pc}]^2, \tag{4.2}$$

where $\hat{U}_i$ is a consistent estimate of $U_i = Var(y_i^2 \mid x_i)$ (which is approximately $E[y_i^{2c} - \tilde{a}_1^*(\sigma_\delta^2, \hat{\boldsymbol{\psi}}_0)x_i^{1c} - \cdots - \tilde{a}_{2p}^*(\sigma_\delta^2, \hat{\boldsymbol{\psi}}_0)x_i^{2pc}]^2$ for large n). To obtain $\hat{W}_i$, one only needs to replace $\beta_0, \ldots, \beta_p, \sigma_\delta^2$, and σ_ϵ^2 with $\hat{\beta}_{00}, \ldots, \hat{\beta}_{p0}$ (can be obtained from $\hat{\boldsymbol{\psi}}_0$ and $\hat{\sigma}_{\delta 0}^2$), $\hat{\sigma}_{\delta 0}^2$, and $\hat{\sigma}_{\epsilon 0}^2$ in $W_i = (\beta_1, \ldots, \beta_p)\boldsymbol{\Sigma}_i(\beta_1, \ldots, \beta_p)' + \sigma_\epsilon^2$, respectively. Hence, $\hat{W}_i = W_i(\hat{\boldsymbol{\psi}}_0, \hat{\sigma}_{\delta 0}^2, \hat{\sigma}_{\epsilon 0}^2)$. Also, observe that

$$\begin{aligned} U_i = {} & var(a_0 + a_1\xi_i + \cdots + a_{2p}\xi_i^{2p} + \epsilon_i^2 \mid x_i) + var[2(\beta_0 + \cdots + \beta_p\xi_i^p)\epsilon_i \mid x_i] \\ & +2cov\{[a_0 + a_1\xi_i + \cdots + a_{2p}\xi_i^{2p} + \epsilon_i^2, 2(\beta_0 + \cdots + \beta_p\xi_i^p)\epsilon_i] \mid x_i\}, \end{aligned} \tag{4.3}$$

where $a_0, \ldots, a_{2p}$ are given in (2.4). The first term on the right of (4.3) equals

$$(a_1, \ldots, a_{2p})\boldsymbol{\Sigma}_i^*(a_1, \ldots, a_{2p})' + 2\sigma_\epsilon^4, \tag{4.4}$$

where $\boldsymbol{\Sigma}_i^*$ is the covariance matrix of the conditional distribution of $(\xi_i, \ldots, \xi_i^{2p})$ given x_i. Here $\boldsymbol{\Sigma}_i^*$ can be computed by (2.1) and estimated by $\hat{\boldsymbol{\Sigma}}_i^* = \boldsymbol{\Sigma}_i^*(\hat{\boldsymbol{\psi}}_0, \hat{\sigma}_{\delta 0}^2, \hat{\sigma}_{\epsilon 0}^2)$. In addition, $a_k, k = 1, \ldots, 2p$, and σ_ϵ^2 in (4.4) can be estimated by $\hat{a}_k = \tilde{a}_k(\hat{\sigma}_{\delta 0}^2, \hat{\boldsymbol{\psi}}_0)$ and $\hat{\sigma}_{\epsilon 0}^2$, respectively. The second term of (4.3) is

$$4\sigma_\epsilon^2[a_0 + a_1 E(\xi_i \mid x_i) + \cdots + a_{2p}E(\xi_i^{2p} \mid x_i)],$$

which again can be computed by (2.1) and estimated consistently. Finally, the last term of (4.3) equals zero since $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ and ϵ_i is independent of ξ_i. Putting all results together, we have $\hat{U}_i = U_i(\hat{\boldsymbol{\psi}}_0, \hat{\sigma}_{\delta 0}^2, \hat{\sigma}_{\epsilon 0}^2)$.

Applying the weighted least squares approach above, we propose an iterative algorithm to produce more efficient estimates of the parameters in the model.

- 1. Use the initial estimates $\hat{\boldsymbol{\psi}}_0, \hat{\sigma}^2_{\delta 0}$, and $\hat{\sigma}^2_{\epsilon 0}$ (established in Section 3) to compute the weights $\hat{W}_i$ and $\hat{U}_i$.
- 2. Obtain the weighted least squares estimates $\hat{\beta}_0^{*^w}, \ldots, \hat{\beta}_p^{*^w}$ from (4.1) and update $\hat{\boldsymbol{\psi}} = (\hat{\beta}_0^{*^w}, \ldots, \hat{\beta}_p^{*^w})$.
- 3. Generate $\tilde{a}_1^*(\sigma^2, \hat{\boldsymbol{\psi}}), \ldots, \tilde{a}_{2p}^*(\sigma^2, \hat{\boldsymbol{\psi}})$ through the updated $\hat{\boldsymbol{\psi}}$ and find the value of $\sigma^2, \hat{\sigma}_\delta^{2^w}$, minimizing (4.2). Update $\hat{\sigma}^2_\delta$ by letting $\hat{\sigma}^2_\delta = \hat{\sigma}_\delta^{2^w}$.
- 4. Update $\hat{\sigma}^2_\epsilon$ in (3.3) by using the latest $\hat{\boldsymbol{\psi}}$ and $\hat{\sigma}^2_\delta$.
- 5. Treating $\hat{\boldsymbol{\psi}}, \hat{\sigma}^2_\delta$, and $\hat{\sigma}^2_\epsilon$ as the initial estimates, repeat the previous steps K more times, K decided by the user. Alternatively, stop the iteration if the change in $\hat{\sigma}^2_\delta$ is small.

If $\hat{\boldsymbol{\psi}}$ and $\hat{\sigma}^2_\delta$ are the estimates obtained in the final stage of the above algorithm, we can compute the estimates $\hat{\beta}_0, \ldots, \hat{\beta}_p$ from (3.1). Note that the full iteration of the above algorithm corresponds to setting $K = \infty$, which is equivalent to solving the following equations

$$\begin{aligned}
&\sum_1^n W_i^{-1}(y_i - \beta_0^* - \beta_1^* x_i - \cdots - \beta_p^* x_i^p)(1, x_i, \ldots, x_i^p)' = \mathbf{0}, \\
&\sum_1^n U_i^{-1} \frac{\partial}{\partial \sigma^2_\delta}(y_i^{2c} - a_1^* x_i^{1c} - \cdots - a_{2p}^* x_i^{2pc}) = 0, \qquad (4.5)\\
&\sum_1^n \{n^{-1}[\sum_1^n e_i^2 - \sum_1^n (\beta_1, \ldots, \beta_p)\boldsymbol{\Sigma}_i(\beta_1, \ldots, \beta_p)'] \bigvee 0 - \sigma^2_\epsilon\} = 0.
\end{aligned}$$

The parameters $\beta_0^*, \ldots, \beta_p^*, a_1^*, \ldots, a_{2p}^*$ in the equations of (4.5) are functions of $\beta_0, \ldots, \beta_p$, and σ^2_δ. We omit the relation in notation for simplicity. Therefore, the equations in (4.5) involve the parameters $\beta_0, \ldots, \beta_p, \sigma^2_\delta$, and σ^2_ϵ only, and are a set of estimating equations of them.

To obtain the estimates of the parameters in the model by directly solving (4.5) can be cumbersome due to the complexity of the equations and hence the iterative reweighted least squares method is recommended. For a more detailed discussion of estimating equations in this topic, see Carroll, Ruppert and Stefanski (1995).

5. Simulation

In this section, we use statistical simulation to compare the performance of the reweighted least squares method with that of the other two methods mentioned in Sections 2 and 3. Since the iteration number does not affect the asymptotic distributions of the reweighted least squares estimates (Carroll and Ruppert (1988)), it is chosen here to be two in the procedure of computing the estimates. In Tables 1-3, we assume that

$$y_i = 3 + 2\xi_i + \xi_i^2 + \epsilon_i,\ \xi_i = x_i - \delta_i,$$

where $(\epsilon_i, \delta_i) \overset{i.i.d.}{\sim} N(0, diag(\sigma_\epsilon^2, \sigma_\delta^2))$, $\sigma_\epsilon^2 = \sigma_\delta^2 = 0.5, 1$, $n = 200, 1000$. For the controlled variables x_i, in order to satisfy (1.3)-(1.5) they are generated as the random samples from N(0,1), Uniform(0, 3.464), and $\chi_1^2/1.414$ (each distribution has variance one). To treat as an initial guess of σ_δ^2 in minimizing $S_n(\sigma^2, \hat{\psi})$ of (3.2), we draw a random number from Uniform$(0, 2S_x^2)$, where $S_x^2 = n^{-1}\sum_1^n (x_i - \bar{x})^2$. Note that the estimates of β_1 and β_2 computed by the method in Section 2 are the same as those found by the method in Section 3, because $d_{11} = d_{21} = 0$ in (2.3). Also it is sensible to define the estimates of σ_ϵ^2 computed by the methods in Sections 2 and 3 to be zero when they are negative. Since the results for $\sigma_\epsilon^2 = 1$ are not appreciably different from those for $\sigma_\epsilon^2 = 0.5$, only the latter cases are reported. Entries in Tables 1-3 represent the average of 300 repeated estimates, with standard deviations included in parentheses.

From the tables, we conclude that the reweighted least squares method is the best among the three methods. When the x_i are drawn from $N(0, 1)$, the reweighted least squares method gives results very close to the desired values. Between the methods in Sections 2 and 3, the one in Section 3 is better than the other in estimating β_0, σ_δ^2, and σ_ϵ^2. When the x_i are from uniform and chi-squared distributions, the reweighted least squares method still provides results close to the desired values. At the same time, the method in Section 3 provides reasonable estimates as well, except for estimating σ_ϵ^2. On the other hand, the method in Section 2 gives erroneous results in estimating β_0 and σ_δ^2 (the estimate of β_0 even presents the opposite sign). Although the estimate of σ_ϵ^2 in this case seems to be the best among the three methods, negative estimates of σ_ϵ^2 have been masked by truncating them to zero. However, the reweighted least squares estimate of σ_ϵ^2 is still comparable with this estimate for large samples. Note that

Table 1. Simulation results for $x_i \sim N(0, 1)$.
(replication=300)

	$\beta_0 = 3$	$\beta_1 = 2$	$\beta_2 = 1$	$\sigma_\delta^2 = 0.5$	$\sigma_\epsilon^2 = 0.5$
n=200	*3.022	*2.022	*0.980	*0.527	*0.469
	(0.191)	(0.188)	(0.137)	(0.148)	(0.506)
	Δ3.052	Δ 2.007	Δ 0.998	Δ 0.457	Δ 1.327
	(0.375)	(0.215)	(0.164)	(0.240)	(1.358)
	☐ 2.363	☐ 2.007	☐ 0.998	☐ 1.261	☐ 2.765
	(1.393)	(0.215)	(0.164)	(1.754)	(3.540)
n=1000	*3.004	*1.997	*0.994	*0.509	*0.475
	(0.082)	(0.081)	(0.054)	(0.070)	(0.326)
	Δ3.014	Δ2.002	Δ 0.998	Δ 0.489	Δ 0.891
	(0.190)	(0.096)	(0.071)	(0.139)	(0.879)
	☐ 2.758	☐ 2.002	☐ 0.998	☐ 0.768	☐ 1.899
	(0.757)	(0.096)	(0.071)	(0.815)	(2.375)

Table 2. Simulation results for $x_i \sim$ Uniform(0, 3.464). (replication=300)

	$\beta_0 = 3$	$\beta_1 = 2$	$\beta_2 = 1$	$\sigma_\delta^2 = 0.5$	$\sigma_\epsilon^2 = 0.5$
n=200	*2.985	*2.091	*0.968	*0.575	*1.227
	(0.521)	(0.934)	(0.308)	(0.262)	(1.957)
	Δ3.054	Δ 1.946	Δ 1.014	Δ 0.555	Δ 1.536
	(0.579)	(1.120)	(0.363)	(0.295)	(2.136)
	□-20.390	□ 1.946	□ 1.014	□ 33.250	□ 0.330
	(43.800)	(1.120)	(0.363)	(85.540)	(1.632)
n=1000	*2.984	*2.009	*1.001	*0.508	*0.770
	(0.225)	(0.413)	(0.135)	(0.083)	(1.000)
	Δ2.993	Δ 1.994	Δ 1.006	Δ 0.505	Δ 0.960
	(0.252)	(0.488)	(0.158)	(0.097)	(1.212)
	□ -6.269	□ 1.994	□ 1.006	□ 9.922	□ 0.372
	(13.330)	(0.488)	(0.158)	(14.160)	(1.247)

Table 3. Simulation results for $x_i \sim \chi_1^2/1.414$. (replication=300)

	$\beta_0 = 3$	$\beta_1 = 2$	$\beta_2 = 1$	$\sigma_\delta^2 = 0.5$	$\sigma_\epsilon^2 = 0.5$
n=200	*2.942	*2.148	*0.971	*0.565	*1.112
	(0.376)	(0.704)	(0.247)	(0.345)	(1.530)
	Δ3.095	Δ 1.988	Δ 1.017	Δ 0.403	Δ 2.836
	(0.334)	(0.928)	(0.302)	(0.339)	(2.891)
	□-17.170	□ 1.988	□ 1.017	□ 26.700	□ 0.275
	(33.710)	(0.928)	(0.302)	(98.780)	(1.010)
n=1000	*2.959	*2.063	*0.989	*0.534	*0.874
	(0.234)	(0.286)	(0.104)	(0.216)	(1.108)
	Δ3.025	Δ 2.016	Δ 1.001	Δ 0.453	Δ 2.691
	(0.335)	(0.452)	(0.144)	(0.337)	(2.988)
	□ -8.737	□ 2.016	□ 1.001	□ 11.930	□ 0.326
	(18.560)	(0.452)	(0.144)	(18.090)	(1.024)

*: Reweighted least squares estimates

Δ: Estimates proposed in Section 3

□: Estimates proposed in Section 2

Number in the parenthesis is standard deviation

the convergence of $\hat{\sigma}_\epsilon^2$ to σ_ϵ^2 for the reweighted least squares method, except x_i from normal distribution, is very slow and does need very large sample size to have the desired accuracy. For the same parameter setting of Table 2, the reweighted least squares method is simulated to obtain $\hat{\sigma}_\epsilon^2 = .675$ (standard deviation = .766) for $n = 2000$ and .650 (.584) for $n = 5000$ (for chi-squared distribution, the results are similar).

Acknowledgements

The authors thank Prof. C. F. Jeff Wu, the Referee, and the Associate editor for their valuable comments and suggestions on this work.

Appendix

Throughout the Appendix, define $\tilde{\mathbf{a}}_1^*(\sigma^2,\boldsymbol{\psi}) = (\tilde{a}_1^*(\sigma^2,\boldsymbol{\psi}),\ldots,\tilde{a}_{2p}^*(\sigma^2,\boldsymbol{\psi}))'$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_{2p})'$, $\boldsymbol{x}_i = (x_i,\ldots,x_i^{2p})'$, and $\boldsymbol{x}_i^c = (x_i^{1c},\ldots,x_i^{2pc})' = (x_i-\bar{x}, x_i^2-\overline{x^2},\ldots,x_i^{2p}-\overline{x^{2p}})'$. By a direct expansion,

$$\frac{1}{n}\sum_1^n y_i^2 = \sum_{0\le h,l,h+l\le 2p}\frac{1}{n}\sum_{i=1}^n c_{h,l}x_i^h\delta_i^l + \sum_{0\le m,k,m+k\le p}\frac{1}{n}\sum_{i=1}^n f_{m,k}x_i^m\delta_i^k\epsilon_i + \frac{1}{n}\sum_1^n\epsilon_i^2,$$

where $c_{h,l}$ and $f_{m,k}$ are sums of coefficients of $x_i^h\delta_i^l$ and $x_i^m\delta_i^k\epsilon_i$, respectively. Since $n^{-1}\sum_1^n x_i^h\delta_i^l$ has mean $n^{-1}\sum_1^n x_i^h E\delta_i^l$ and variance $n^{-2}\sum_1^n x_i^{2h}var(\delta_i^l)$ (which converges to 0 by (1.5)), as a result $n^{-1}\sum_1^n x_i^h\delta_i^l \xrightarrow{P} \mu_h E\delta_i^l$ by Kolmogrov's theorem and (1.3). Similarly, $n^{-1}\sum_1^n x_i^m\delta_i^k\epsilon_i \xrightarrow{P} 0$. In conclusion,

$$\frac{1}{n}\sum_1^n y_i^2 \xrightarrow{P} \sum_{0\le h,l,h+l\le 2p} c_{h,l}\mu_h E\delta_i^l + \sigma_\epsilon^2 \triangleq \mu_{y^2}.$$

To prove the main theorem, we need the following three lemmas.

Lemma 1. *Suppose that Model* (1.2) *and Conditions* (1.3) – (1.5) *hold with* $p\ge 2$ *and that* $\beta_0,\ldots,\beta_p,\sigma_\delta^2$, *and* σ_ϵ^2 *are unknown parameters. Define* $\underline{S}_i(\sigma^2) = E[y_i^2-\mu_{y_i^2}-\tilde{a}_1^*(\sigma^2,\boldsymbol{\psi})(x_i-\mu_1)-\cdots-\tilde{a}_{2p}^*(\sigma^2,\boldsymbol{\psi})(x_i^{2p}-\mu_{2p})]^2$, *where* $\mu_{y_i^2} = Ey_i^2$ *and* μ_l, $1\le l\le 2p$, *are given in* (1.3). *For any* $\tau>0$, *there exists a* $n_0(\tau)>0$ *and* $\rho(\tau)>0$ *such that* $n^{-1}\sum_1^n\underline{S}_i(\sigma^2) - n^{-1}\sum_1^n\underline{S}_i(\sigma_\delta^2)\ge\rho(\tau)$ *if* $n>n_0(\tau)$ *and* $|\sigma^2-\sigma_\delta^2|>\tau$.

Proof. Since $a_j^* = \tilde{a}_j^*(\sigma_\delta^2,\boldsymbol{\psi})$, $1\le j\le 2p$, it follows that $E(y_i^2\mid x_i) = a_0^* + \tilde{\mathbf{a}}_1^*(\sigma_\delta^2,\boldsymbol{\psi})'\boldsymbol{x}_i$. By a straightforward calculation,

$$\begin{aligned}\frac{1}{n}\sum_1^n\underline{S}_i(\sigma^2) = \frac{1}{n}\sum_1^n\underline{S}_i(\sigma_\delta^2) &+ \frac{2}{n}\sum_1^n E\{[y_i^2-\mu_{y_i^2}-\tilde{\mathbf{a}}_1^*(\sigma_\delta^2,\boldsymbol{\psi})'(\boldsymbol{x}_i-\boldsymbol{\mu})][\tilde{\mathbf{a}}_1^*(\sigma_\delta^2,\boldsymbol{\psi})\\ &-\tilde{\mathbf{a}}_1^*(\sigma^2,\boldsymbol{\psi})]'(\boldsymbol{x}_i-\boldsymbol{\mu})\} + [\tilde{\mathbf{a}}_1^*(\sigma_\delta^2,\boldsymbol{\psi})-\tilde{\mathbf{a}}_1^*(\sigma^2,\boldsymbol{\psi})]'\boldsymbol{M}_n[\tilde{\mathbf{a}}_1^*(\sigma_\delta^2,\boldsymbol{\psi})\\ &-\tilde{\mathbf{a}}_1^*(\sigma^2,\boldsymbol{\psi})],\end{aligned}\tag{A.1}$$

where $\mathbf{M}_n = n^{-1}\sum_1^n(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'$. By (1.3) the second term on the right of (A.1) converges to 0. Also, by (1.4) $\lambda_{min}(\boldsymbol{M}_n)\to\lambda_{min}(\boldsymbol{M})>0$, where $\lambda_{min}(\mathbf{M}_n)$ and $\lambda_{min}(\mathbf{M})$ are minimum eigenvalues of $\mathbf{M}_n$ and $\mathbf{M}$, respectively. Then the result follows easily.

Lemma 2. *Let the assumptions of Lemma 1 hold. If $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} = O_p(n^{-\frac{1}{2}})$, there exists a number $b > \sigma_\delta^2$ such that $\lim_{n\to\infty} P(\hat{\sigma_\delta^2} > b) = 0$.*

Proof. First we will show that for any $c > 0$, there exists a number $b_1 > 0$ such that

$$\lim_{n\to\infty} P[\inf_{\sigma^2 > b_1} S_n(\sigma^2, \hat{\boldsymbol{\psi}}) > c] = 1. \tag{A.2}$$

Observe that

$$S_n(\sigma^2, \hat{\boldsymbol{\psi}}) \geq \frac{1}{n}\sum_1^n (g_i - \bar{g})^2 + (\frac{1}{n}\sum_1^n q_i^2)^{\frac{1}{2}}[(\frac{1}{n}\sum_1^n q_i^2)^{\frac{1}{2}} - 2(\frac{1}{n}\sum_1^n (g_i - \bar{g})^2)^{\frac{1}{2}}], \tag{A.3}$$

where $g_i = y_i^2 - \tilde{\mathbf{a}}_1^*(\sigma_\delta^2, \boldsymbol{\psi})' \boldsymbol{x}_i$ and $q_i = [\tilde{\mathbf{a}}_1^*(\sigma_\delta^2, \boldsymbol{\psi}) - \tilde{\mathbf{a}}_1^*(\sigma^2, \hat{\boldsymbol{\psi}})]' \boldsymbol{x}_i^c$. By a direct expansion of $n^{-1}\sum_1^n (g_i - \bar{g})^2$, (1.3), (1.4), and Kolmogrov's theorem, it follows that $n^{-1}\sum_1^n (g_i - \bar{g})^2 = \sigma_g^2 + o_p(1)$, where σ_g^2 is a constant. Also observe that

$$\frac{1}{n}\sum_1^n q_i^2 \geq \mid \tilde{a}_{2p-2}^*(\sigma_\delta^2, \boldsymbol{\psi}) - \tilde{a}_{2p-2}^*(\sigma^2, \hat{\boldsymbol{\psi}}) \mid^2 \lambda_{min}(\mathbf{M}_n^*), \tag{A.4}$$

where $\mathbf{M}_n^* = n^{-1}\sum_1^n \boldsymbol{x}_i^c \boldsymbol{x}_i^{c'}$ and $\lambda_{min}(\boldsymbol{M}_n^*)$ is the minimum eigenvalue of $\mathbf{M}_n^*$. The right hand side of (A.4), by (2.8), equals

$$(2\beta_{p-2}^*\beta_p^* + \beta_{p-1}^{*2} + p^2\beta_p^{*2}\sigma_\delta^2 - 2\hat{\beta}_{p-2}^*\hat{\beta}_p^* - \hat{\beta}_{p-1}^{*2} - p^2\hat{\beta}_p^{*2}\sigma^2)^2 \lambda_{min}(\mathbf{M}_n^*). \tag{A.5}$$

Since $\hat{\beta}_p^*, \hat{\beta}_{p-1}^*$, and $\hat{\beta}_{p-2}^*$ converge to $\beta_p^*(\neq 0), \beta_{p-1}^*$, and β_{p-2}^* in probability, respectively and $\lambda_{min}(\mathbf{M}_n^*) \to \lambda_{min}(\mathbf{M})(> 0)$, consequently, for any number c^* we can choose b_1^* large enough that $\lim_{n\to\infty} P[\inf_{\sigma^2 > b_1^*}(\frac{1}{n}\sum_1^n q_i^2)^{\frac{1}{2}} > c^*] = 1$. Applying this result and $n^{-1}\sum_1^n (g_i - \bar{g})^2 = \sigma_g^2 + o_p(1)$ to (A.3), we obtain (A.2) holds. Placing $\sigma^2 = \sigma_\delta^2$ in $n^{-1}\sum_1^n q_i^2$ of (A.3), and using that $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} = O_p(n^{-\frac{1}{2}})$ and $n^{-1}\sum_1^n (g_i - \bar{g})^2 = \sigma_g^2 + o_p(1)$, we conclude $S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) \geq \sigma_g^2 + o_p(1)$. Similarly, $S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) \leq \sigma_g^2 + o_p(1)$. In conclusion, $S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) = \sigma_g^2 + o_p(1)$. For $2\sigma_g^2 > 0$, by (A.2) there exists a number $b > 0$ such that

$$\lim_{n\to\infty} P[\inf_{\sigma^2 > b} S_n(\sigma^2, \hat{\boldsymbol{\psi}}) > 2\sigma_g^2] = 1. \tag{A.6}$$

The event $\{\hat{\sigma}_\delta^2 > b\}$ implies that $S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) \geq S_n(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}) = \inf_{\sigma^2 > b} S_n(\sigma^2, \hat{\boldsymbol{\psi}})$, and hence $P(\hat{\sigma}_\delta^2 > b) \leq P[S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) \geq \inf_{\sigma^2 > b} S_n(\sigma^2, \hat{\boldsymbol{\psi}})]$. From $S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) = \sigma_g^2 + o_p(1)$ and (A.6), the desired result follows.

Remark. From (A.5) it follows that $P[\lim_{\sigma^2\to\infty}$ (A.5)$= \infty] = 1$, and hence $P[\lim_{\sigma^2\to\infty} S_n(\sigma^2, \hat{\boldsymbol{\psi}}) = \infty] = 1$. This guarantees that $S_n(\sigma^2, \hat{\boldsymbol{\psi}})$ always achieves its minimum value at a finite positive number.

Lemma 3. *Under the assumptions of Lemma* 1, *for any number* $\eta > 0$ *and* $\nu > 0$,

$$\lim_{n\to\infty} P[\sup_{0\le\sigma^2\le\eta} \mid S_n(\sigma^2, \hat{\boldsymbol{\psi}}) - \frac{1}{n}\sum_1^n \underline{S}_i(\sigma^2) \mid > \nu] = 0,$$

where $\underline{S}_i(\sigma^2)$ *are defined as in Lemma* 1.

Proof. Because $\tilde{\beta}_j(\sigma^2, \boldsymbol{\psi})$ is a continuous function of $\boldsymbol{\psi}$ for fixed σ^2, and of σ^2 on $[0, \eta]$ for fixed $\boldsymbol{\psi}$, based on $\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} = O_p(n^{-\frac{1}{2}})$ that

$$\sup_{0\le\sigma^2\le\eta} \mid \tilde{a}_j^*(\sigma^2, \hat{\boldsymbol{\psi}}) - \tilde{a}_j^*(\sigma^2, \boldsymbol{\psi}) \mid = O_p(n^{-\frac{1}{2}}),\ j = 1, \ldots, 2p. \tag{A.7}$$

Recall that

$$\begin{aligned} S_n(\sigma^2, \hat{\boldsymbol{\psi}}) - S_n(\sigma^2, \boldsymbol{\psi}) = {} & [\tilde{\mathbf{a}}_1^*(\sigma^2, \boldsymbol{\psi}) - \tilde{\mathbf{a}}_1^*(\sigma^2, \hat{\boldsymbol{\psi}})]'(\frac{2}{n}\sum_1^n \boldsymbol{x}_i^c y_i^{2c}) + [\tilde{\mathbf{a}}_1^*(\sigma^2, \hat{\boldsymbol{\psi}}) \\ & + \tilde{\mathbf{a}}_1^*(\sigma^2, \boldsymbol{\psi})]'(\frac{1}{n}\sum_1^n \mathbf{x}_i^c \boldsymbol{x}_i^{c'})[\tilde{\mathbf{a}}_1^*(\sigma^2, \hat{\boldsymbol{\psi}}) - \tilde{\mathbf{a}}_1^*(\sigma^2, \boldsymbol{\psi})], \end{aligned}$$

where $n^{-1}\sum_1^n \boldsymbol{x}_i^c \boldsymbol{x}_i^{c'} \to \boldsymbol{M}$ and $n^{-1}\sum_1^n \boldsymbol{x}_i^c y_i^{2c} \xrightarrow{P}$ some constant. Since $\tilde{a}_j^*(\sigma^2, \boldsymbol{\psi})$, $1 \le j \le 2p$, are bounded on $[0, \eta]$ for fixed $\boldsymbol{\psi}$, we have from (A.7) that

$$\sup_{0\le\sigma^2\le\eta} \mid S_n(\sigma^2, \hat{\boldsymbol{\psi}}) - S_n(\sigma^2, \boldsymbol{\psi}) \mid \xrightarrow{P} 0. \tag{A.8}$$

By a straightforward calculation,

$$\begin{aligned} & S_n(\sigma^2, \boldsymbol{\psi}) - \frac{1}{n}\sum_1^n \underline{S}_i(\sigma^2) \\ & = \frac{1}{n}\sum_1^n [(y_i^{2c})^2 - E(y_i^2 - \mu_{y_i^2})^2] + \sum_{j=1}^{2p} \tilde{a}_j^{*2}(\sigma^2, \boldsymbol{\psi})[\frac{1}{n}\sum_{i=1}^n (x_i^j - \overline{x^j})^2 - \frac{1}{n}\sum_{i=1}^n (x_i^j - \mu_j)^2] \\ & \quad -2\sum_{j=1}^{2p} \tilde{a}_j^*(\sigma^2, \boldsymbol{\psi})[\frac{1}{n}\sum_{i=1}^n (x_i^j - \overline{x^j}) y_i^{2c} - \frac{1}{n}\sum_{i=1}^n (x_i^j - \mu_j) E(y_i^2 - \mu_{y_i^2})] \\ & \quad + \sum_{1\le l\ne k\le 2p} \tilde{a}_l^*(\sigma^2, \boldsymbol{\psi}) \tilde{a}_k^*(\sigma^2, \boldsymbol{\psi})[\frac{1}{n}\sum_{i=1}^n (x_i^l - \overline{x^l})(x_i^k - \overline{x^k}) - \frac{1}{n}\sum_{i=1}^n (x_i^l - \mu_l)(x_i^k - \mu_k)]. \end{aligned}$$

Using the argument of Lemma 2, (1.3), and the boundeoness of $\tilde{a}_j^*(\sigma^2, \boldsymbol{\psi})$ on $[0, \eta]$ for fixed $\boldsymbol{\psi}$, $j = 1, \ldots, 2p$, we conclude that

$$\sup_{0\le\sigma^2\le\eta} \mid S_n(\sigma^2, \boldsymbol{\psi}) - \frac{1}{n}\sum_1^n \underline{S}_i(\sigma^2) \mid \xrightarrow{P} 0. \tag{A.9}$$

Combining (A.8) and (A.9), we establish the result.

Proof of Theorem. Let

$$E_1 = \{0 \leq \hat{\sigma}_\delta^2 \leq b\},\ E_2 = \{\mid \hat{\sigma}_\delta^2 - \sigma_\delta^2 \mid > \tau\},$$
$$E_3 = \{ \sup_{0 \leq \sigma^2 \leq b} \mid S_n(\sigma^2, \hat{\boldsymbol{\psi}}) - n^{-1} \sum_1^n \underline{S}_i(\sigma^2) \mid < \frac{\rho}{3}\},$$

where $\underline{S}_i(\sigma^2), \tau, \rho$, and b are defined in Lemmas 1 and 2. By Lemma 1, there exists a n_0 such that $E_2 \subset \{n^{-1} \sum_1^n \underline{S}_i(\hat{\sigma}_\delta^2) - n^{-1} \sum_1^n \underline{S}_i(\sigma_\delta^2) \geq \rho\}$ if $n \geq n_0$. As a consequence, for $n \geq n_0$, $E_1 \cap E_2 \cap E_3 \subset \{S_n(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}) - S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) > \rho/3\}$ and $P(E_1 \cap E_2 \cap E_3) \leq P\{S_n(\hat{\sigma}_\delta^2, \hat{\boldsymbol{\psi}}) - S_n(\sigma_\delta^2, \hat{\boldsymbol{\psi}}) > \rho/3\} = 0$, where the last equality holds because $S_n(\sigma^2, \hat{\boldsymbol{\psi}})$, $0 \leq \sigma^2 < \infty$, achieves its minimum at $\sigma^2 = \hat{\sigma}_\delta^2$. Since by Lemmas 2 and 3 $\lim_{n\to\infty} P(E_1) = 1 = \lim_{n\to\infty} P(E_3)$, we have $\lim_{n\to\infty} P(E_2) = 0$, i.e., $\hat{\sigma}_\delta^2 \xrightarrow{P} \sigma_\delta^2$.

References

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

Box, G. E. P. (1961). The effects of errors in factor levels and experimental design. *Bull. Int. Statist. Inst.* **38**, 339-355.

Burr, D. (1988). On errors-in-variables in binary regression - Berkson case. *J. Amer. Statist. Assoc.* **83**, 739-743.

Carroll, R. J. and Rupport, D. (1988). *Transformation and Weighting in Regression.* Chapman & Hall, London.

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

Fedorov, W. A. (1974). Regression problems with controllable variables subject to error. *Biometrika* **61**, 49-56.

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

Geary, R. C. (1953). Non-linear functional relationship between two variables when one variable is controlled. *J. Amer. Statist. Assoc.* **48**, 94-103.

Rudemo, M., Ruppert, D. and Streibig, J. C. (1989). Random- effect models in nonlinear regression with applications to bioassay. *Biometrics* **45**, 349-362.

Seber, G. A. F. (1977). *Linear Regression Analysis.* Wiley, New York.

Department of Statistics, University of Michigan, Ann Arbor, U.S.A.

E-mail: huwang@stat.nthu.edu.tw

Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan.

(Received February 1999; accepted September 1999)