

REGRESSION CALIBRATION USING RESPONSE VARIABLES IN LINEAR MODELS

Steve Y. H. Huang

Tamkang University

Abstract: Regression calibration is an easy way to improve estimation in errors-in-variables models. This method replaces missing covariates with estimates that are more accurate than surrogates. One might expect better estimation using response variables together with surrogates to estimate or predict missing values. However, the introducing of response variables generates bias in the estimating function. In this article, we use response variables to calibrate the missing covariates and provide an estimation method for the regression parameters in linear models. When errors in variables are small, we show that regression calibration using response variables outperforms the conventional regression calibration. A small simulation study comparing the performances of these methods in finite sample is provided.

Key words and phrases: Errors in variables, missing data, regression calibration, response variables.

1. Introduction

In applied problems, researchers are often interested in the relationships between response variables and covariates. In many situations, however, the true covariates are expensive to collect, hard to measure, and available only in limited supply. If a surrogate W of the true covariate is collected for each subject, then we can partition the sample into two subsamples, one with true covariates (the validation data set) and the other with surrogates only (the primary data set)—that is, the subset in which the true covariates are missing. If one ignores the missing values and proceeds, it is called the complete data analysis. Under certain conditions, the factor causing data to be missing can be ignored (Rubin (1976)), and a complete case analysis is a valid one. Clearly such an analysis wastes information contained in the primary data set. An easy way to make use of the information in a data set with missing values is Regression Calibration (RC)—a kind of imputation which estimates the missing covariates first, replaces missing values with their estimates, then proceed with analysis as if there were no missing values. Estimations with such “replacements” have been widely used; see, for example, Carroll and Stefanski (1990) for quasi-likelihood estimation, Liang and Liu (1991) for generalized linear models, and Lee and Sepanski (1995)

for linear and nonlinear errors-in-variables models. The method of RC is simple and potentially applicable to any regression model, provided the approximation is sufficiently accurate (Carroll, Rupper and Stefanski (1995, Chap.3)).

In this article, we consider the simple errors-in-variables model

$$Y = \beta_0 + \beta_1 X + \epsilon^*, \quad W = X + \delta^*,$$

where W is the surrogate of X , δ^* is the error in measuring X , and ϵ^* is the error in measuring $\beta_0 + \beta_1 X$. For any missing X , the conventional RC replaces X with $E(X | W)$ or an estimate of it. This procedure can be justified by observing that $E(Y | W) = \beta_0 + \beta_1 E(X | W)$, which is also a regression function with the same parameters β_0 and β_1 . Doing so, however, requires distribution assumptions to find the form of $E(X | W)$, and it is unclear how much additional efficiency is gained. In later sections, we propose using response variables together with surrogates to improve estimation of missing values, and we develop an approximation method based on small errors to calculate the efficiency of different estimators.

In Section 2, we motivate and propose the estimating functions. Section 3 contains a theoretical comparison of the proposed method with conventional RC and complete case analysis. A simulation study is provided in Section 4. In Section 5, we discuss the possible extension of RC with response variables to more general models. The appendix sketches proofs of the theorems in Section 2.

2. Regression Calibration using Response Variables

Rewrite the model in the previous section by letting $\epsilon^* = \sigma_1 \epsilon$ and $\delta^* = \sigma_2 \delta$, where σ_1 and σ_2 are such that $E(\epsilon^2) = 1$ and $E(\delta^2) = 1$. Then

$$Y = \beta_0 + \beta_1 X + \sigma_1 \epsilon, \quad W = X + \sigma_2 \delta. \quad (2.1)$$

Assume that X has mean μ and variance σ_x^2 , that X, ϵ and δ are independently distributed with finite fourth moments, and that both ϵ and δ have mean 0. Normality is not needed here, but $E(\epsilon^3) = 0$ is assumed for technical reasons.

We divide the observations into two sets, V and P . Set V consists of observations (Y, X, W) and set P contains (Y, W) only. Thus, (Y_i, W_i, X_i) is observed if $i \in V$, and (Y_i, W_i) is observed if $i \in P$. Let N_P denote the size of P and N_V the size of V ; the total sample size is then $N = N_P + N_V$. The probability that an observation is missing is assumed constant.

The conventional RC consists of two steps. First, the validation data is used to estimate the regression function of X on W , which may be accomplished by solving

$$\sum_{i \in V} (X_i - \gamma_0 - \gamma_1 W_i) \begin{pmatrix} 1 \\ W_i \end{pmatrix} = \mathbf{0}$$

for γ_0 and γ_1 . Let $\hat{\gamma}_0$ and $\hat{\gamma}_1$ be the solutions and let $m_i = \hat{\gamma}_0 + \hat{\gamma}_1 W_i$. Second, replace the missing X_i with m_i and proceed as if there were no measurement errors. That is, solve

$$\sum_{i \in V} (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \sum_{i \in P} (Y_i - \beta_0 - \beta_1 m_i) \begin{pmatrix} 1 \\ m_i \end{pmatrix} = 0$$

for β_0 and β_1 , if least squares estimation is adopted.

2.1. The estimating procedure

Our procedure uses the best linear predictor H_i of X_i , where $H_i = a + bW_i + cY_i$ minimizes $E(X_i - H_i)^2$. It is easy to show that (a, b, c) satisfies the equation

$$E \begin{bmatrix} 1 & W & Y \\ W & W^2 & WY \\ Y & WY & Y^2 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = E \begin{pmatrix} X \\ WX \\ XY \end{pmatrix}. \tag{2.2}$$

When replacing the unobserved X_i with H_i , we are led to the estimating equation

$$\sum_{i \in V} (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \sum_{i \in P} (Y_i - \beta_0 - \beta_1 H_i) \begin{pmatrix} 1 \\ H_i \end{pmatrix} = 0 \tag{2.3}$$

for β_0 and β_1 , provided (a, b, c) is known. However, the left-hand side of (2.3) is a biased estimating function because H_i is correlated with $Y_i - \beta_0 - \beta_1 H_i$. To remove the bias, we observe that

$$\begin{aligned} E \left[\sum_{i \in V} (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \sum_{i \in P} (Y_i - \beta_0 - \beta_1 H_i) \begin{pmatrix} 1 \\ H_i \end{pmatrix} \right] \\ = N_P \beta_1 E \left[(X_i - H_i) \begin{pmatrix} 1 \\ H_i \end{pmatrix} \right] + N_P E \left[\epsilon_i \begin{pmatrix} 1 \\ H_i \end{pmatrix} \right]. \end{aligned}$$

The first term on the right-hand side is $\mathbf{0}$ since $(X_i - H_i)$ is orthogonal to H_i , and the last term is $N_P(0, c\sigma_1^2)'$. We then subtract it from the original estimating function to obtain

$$\sum_{i \in V} (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \sum_{i \in P} \left[(Y_i - \beta_0 - \beta_1 H_i) \begin{pmatrix} 1 \\ H_i \end{pmatrix} - \begin{pmatrix} 0 \\ c\sigma_1^2 \end{pmatrix} \right]. \tag{2.4}$$

An algorithm that iteratively estimates (a, b, c) and σ_1^2 and utilizes (2.4) to estimate β_0 and β_1 is described as follows.

1. Use the validation data set to estimate the regression coefficients of X on (W, Y) . Obtain estimates of the best linear predictor H_i for each missing X_i .
2. Use the validation data to compute the least squares estimates of β_0, β_1 and σ_1^2 as the initial estimates.

3. Use the current estimate of σ_1^2 to solve (2.4) with respect to β_0 and β_1 ; derive new estimates of β_0 and β_1 ; then compute the residuals and update the estimate of σ_1^2 .
4. Repeat Step 3 until the estimates of σ_1^2 converge.

This algorithm is equivalent to solving

$$\sum_{i \in V} [(Y_i - \beta_0 - \beta_1 X_i)^2 - \sigma_1^2] = 0$$

$$\sum_{i \in V} (X_i - a - bW_i - cY_i) \begin{pmatrix} 1 \\ W_i \\ Y_i \end{pmatrix} = \mathbf{0} \tag{2.5}$$

$$\sum_{i \in V} (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \sum_{i \in P} [(Y_i - \beta_0 - \beta_1 H_i) \begin{pmatrix} 1 \\ H_i \end{pmatrix} - \begin{pmatrix} 0 \\ c\sigma_1^2 \end{pmatrix}] = \mathbf{0}$$

for estimates of $\boldsymbol{\eta} = (\sigma_1^2, a, b, c, \beta_0, \beta_1)$, which is a one-to-one transformation of the original parameters $\boldsymbol{\theta} = (\mu, \sigma_x^2, \sigma_1^2, \sigma_2^2, \beta_0, \beta_1)$. Next, define

$$A_i = (Y_i - \beta_0 - \beta_1 X_i)^2 - \sigma_1^2, \quad B_i = (X_i - a - bW_i - cY_i) \begin{pmatrix} 1 \\ W_i \\ Y_i \end{pmatrix},$$

$$C_i = (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix}, \quad D_i = (Y_i - \beta_0 - \beta_1 H_i) \begin{pmatrix} 1 \\ H_i \end{pmatrix} - \begin{pmatrix} 0 \\ c\sigma_1^2 \end{pmatrix},$$

and $\rho = N_V/N$. The asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$, the solution of (2.5), equals

$$\begin{pmatrix} \rho E \frac{\partial A}{\partial \boldsymbol{\eta}} \\ \rho E \frac{\partial B}{\partial \boldsymbol{\eta}} \\ \rho E \frac{\partial C}{\partial \boldsymbol{\eta}} + (1 - \rho) E \frac{\partial D}{\partial \boldsymbol{\eta}} \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \rho E A A' & \rho E A B' & \rho E A C' \\ \rho E B' A & \rho E B B' & \rho E B C' \\ \rho E C' A & \rho E C' B & \rho E C C' + (1 - \rho) E D D' \end{pmatrix} \begin{pmatrix} \rho E \frac{\partial A}{\partial \boldsymbol{\eta}} \\ \rho E \frac{\partial B}{\partial \boldsymbol{\eta}} \\ \rho E \frac{\partial C}{\partial \boldsymbol{\eta}} + (1 - \rho) E \frac{\partial D}{\partial \boldsymbol{\eta}} \end{pmatrix}'^{-1} \tag{2.6}$$

The matrix (2.6) is rather complicated, and hence an approximation is considered.

2.2. Small errors approximation

Here we assume that the ratio of the two error variances σ_2^2/σ_1^2 , denoted by k , remains fixed as σ_1^2 tends to 0. If $\mu = 0$, then (2.6) is equal to $T^{-1} M T^{-1'} + \mathbf{O}(\sigma_1^3)$,

where

$$T = - \begin{pmatrix} \rho & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & \rho\beta_0 & 0 & 0 \\ 0 & 0 & \rho\sigma_x^2 + \sigma_1^2 & \rho\beta_1\sigma_x^2 & 0 & 0 \\ 0 & \rho\beta_0 & \rho\beta_1\sigma_x^2 & \rho(\beta_0^2 + \beta_1^2\sigma_x^2 + \sigma_1^2) & 0 & 0 \\ 0 & (1-\rho)\beta_1 & 0 & (1-\rho)\beta_0\beta_1 & 1 & 0 \\ (1-\rho)c & 0 & (1-\rho)\beta_1\sigma_x^2 & (1-\rho)\beta_1^2\sigma_x^2 & 0 & \sigma_x^2 - \frac{(1-\rho)k\sigma_1^2}{1+\beta_1^2k} \end{pmatrix},$$

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho\frac{k}{1+\beta_1^2k} & 0 & \rho\beta_0\frac{k}{1+\beta_1^2k} & -\rho c & 0 \\ 0 & 0 & \rho\sigma_x^2\frac{k}{1+\beta_1^2k} & \rho\beta_1\sigma_x^2\frac{k}{1+\beta_1^2k} & 0 & -\rho c\sigma_x^2 \\ 0 & \frac{\rho\beta_0k}{1+\beta_1^2k} & \frac{\rho\beta_1\sigma_x^2k}{1+\beta_1^2k} & \rho(\beta_0^2 + \beta_1^2\sigma_x^2)\frac{k}{1+\beta_1^2k} & -\rho\beta_0c & -\rho\beta_1c\sigma_x^2 \\ 0 & -\rho c & 0 & -\rho\beta_0c & \rho + \frac{1-\rho}{1+\beta_1^2k} & 0 \\ 0 & 0 & -\rho c\sigma_x^2 & -\rho c\beta_1\sigma_x^2 & 0 & \rho\sigma_x^2 + \frac{1-\rho}{1+\beta_1^2k}\sigma_x^2 \end{pmatrix} \sigma_1^2,$$

Hence, $T^{-1}MT^{-1'}$ is an approximation of (2.6) when errors are “small.” The lower-right part of $T^{-1}MT^{-1'}$ corresponds to the covariance matrix of the estimated regression coefficients, denoted by $\hat{\beta}_0$ and $\hat{\beta}_1$, as stated in Theorem 1.

Theorem 1. *Under (2.1) and the finite fourth moments assumption for (X, ϵ, δ) , the solution $\hat{\eta}$ obtained by solving (2.5) is consistent. Moreover, if σ_2^2/σ_1^2 (denoted by k) remains fixed as $\sigma_1^2 \rightarrow 0$ and $E(\epsilon^3) = 0$, then $\hat{\eta}$ has asymptotic covariance $T^{-1}MT^{-1'} + \mathbf{O}(\sigma_1^3)$. In particular, the asymptotic covariance matrix of $\hat{\beta}_0$ and $\hat{\beta}_1$ is*

$$\begin{pmatrix} 1 & -\mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\rho+\beta_1^2k}{\rho+\rho\beta_1^2k} & 0 \\ 0 & \frac{1}{\sigma_x^2} \frac{\rho+\beta_1^2k}{\rho+\rho\beta_1^2k} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3).$$

To compare the proposed method with conventional RC, a similar analysis is applied to find the asymptotic covariance matrix of the estimators defined as the solutions to

$$\sum_{i \in V} (X_i - \gamma_0 - \gamma_1 W_i) \begin{pmatrix} 1 \\ W_i \end{pmatrix} = \mathbf{0},$$

$$\sum_{i \in V} (Y_i - \beta_0 - \beta_1 X_i) \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \sum_{i \in P} (Y_i - \beta_0 - \beta_1(\gamma_0 + \gamma_1 W_i)) \begin{pmatrix} 1 \\ \gamma_0 + \gamma_1 W_i \end{pmatrix} = \mathbf{0}. \tag{2.7}$$

Theorem 2. *Under the same conditions as in Theorem 1, the estimators of β_0 and β_1 obtained by solving (2.7) have asymptotic covariance matrix*

$$\begin{pmatrix} 1 & -\mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{(1-\rho)\beta_1^2k+\rho}{\rho} & 0 \\ 0 & \frac{(1-\rho)\beta_1^2k+\rho}{\rho\sigma_x^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3).$$

3. Efficiency Comparison

Three estimators are compared here: $\hat{\beta}$, $\hat{\beta}_{rc}$ and $\hat{\beta}_c$, where $\hat{\beta}$ is our estimator, $\hat{\beta}_{rc}$ is the conventional RC estimator, and $\hat{\beta}_c$ is the least squares estimator of $(\beta_0, \beta_1)'$ from the complete case analysis. In Section 2, it was shown that

$$NE(\hat{\beta}_{rc} - \beta)(\hat{\beta}_{rc} - \beta)' = \begin{pmatrix} 1 & -\mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{(1-\rho)\beta_1^2 k + \rho}{\rho} & 0 \\ 0 & \frac{(1-\rho)\beta_1^2 k + \rho}{\rho\sigma_x^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3),$$

$$NE(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \begin{pmatrix} 1 & -\mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\rho + \beta_1^2 k}{\rho + \rho\beta_1^2 k} & 0 \\ 0 & \frac{1}{\sigma_x^2} \frac{\rho + \beta_1^2 k}{\rho + \rho\beta_1^2 k} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3),$$

where $\beta = (\beta_0, \beta_1)'$. It was also shown that

$$N_V E(\hat{\beta}_c - \beta)(\hat{\beta}_c - \beta)' = \begin{pmatrix} 1 & -\mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix} \sigma_1^2 + \mathbf{o}(\sigma_1^2).$$

In order to compare these covariance matrices, we set $\mu = 0$, ignore terms smaller than $O(\sigma_1^2)$, and standardize by N_V . We find that leading terms for the asymptotic covariance matrices of $\hat{\beta}_c$, and $\hat{\beta}_{rc}$ and $\hat{\beta}$ are

$$\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{pmatrix} \sigma_1^2, \quad [(1 - \rho)\beta_1^2 k + \rho] \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{pmatrix} \sigma_1^2, \quad \frac{\rho + \beta_1^2 k}{1 + \beta_1^2 k} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma_x^2} \end{pmatrix} \sigma_1^2. \quad (3.1)$$

When σ_1 is small and N is sufficiently large, we observe the following.

1. The traditional RC estimator outperforms the least squares estimators when $[(1 - \rho)\beta_1^2 k + \rho] < 1$. This condition is likely to hold when $|\beta_1|$ or the variance of the measurement error is small.
2. The proposed estimator outperforms the least squares estimator.
3. The proposed estimator outperforms the traditional RC estimator.

We also note that when ρ is close to 1, the three estimation methods differ by little.

4. Simulation Studies

To assess the performance of these estimation methods in finite samples, computer simulations were conducted. We set β_0 at -1 and β_1 at 0.5 or 1.5 , fixed σ_2^2 and σ_1^2 at 0.25 , and drew X_i from both the standard normal and the standardized uniform distributions. Based on samples of size 300 , the algorithm in Section 2 was iterated three times. Tables 1 and 2 report the mean square errors and estimated variances averaged over $1,000$ replications.

Table 1. Comparison results when $X \sim N(0, 1)$.

$\rho = 0.2, \beta_1 = 0.5$			$\rho = 0.5, \beta_1 = 0.5$		
	MSE*10 ³	Estimated variance *10 ³ (*10 ³)		MSE*10 ³	Estimated variance *10 ³ (*10 ³)
$\hat{\beta}_c$	(3.95, 4.52)*	4.13 ^o (0.76), 4.16 ^o (0.88)	$\hat{\beta}_c$	(1.65,1.50)	1.66(0.19), 1.68(0.28)
$\hat{\beta}_{rc}$	(1.54,1.86)	1.68(0.67), 1.66(0.63)	$\hat{\beta}_{rc}$	(1.10,1.05)	1.04(0.13), 1.05(0.15)
$\hat{\beta}$	(1.42,1.85)	1.44(0.48), 1.44(0.46)	$\hat{\beta}$	(1.05,1.01)	0.99(0.11), 1.00(0.14)
$\rho = 0.2, \beta_1 = 1.5$			$\rho = 0.5, \beta_1 = 1.5$		
	MSE*10 ³	Estimated variance *10 ³ (*10 ³)		MSE*10 ³	Estimated variance *10 ³ (*10 ³)
$\hat{\beta}_c$	(3.95,4.52)	4.13(0.76), 4.16(0.88)	$\hat{\beta}_c$	(1.65,1.50)	1.66(0.19), 1.68(0.28)
$\hat{\beta}_{rc}$	(6.95,8.07)	8.41(5.70), 8.26(5.44)	$\hat{\beta}_{rc}$	(2.56,2.39)	2.70(0.83), 2.68(0.70)
$\hat{\beta}$	(2.92,3.26)	3.11(5.69), 3.13(6.11)	$\hat{\beta}$	(1.41,1.28)	1.37(0.20), 1.38(0.22)

Table 2. Comparison results when $X \sim Uni(-0.5, 0.5) * 3.4641$.

$\rho = 0.2, \beta_1 = 0.5$			$\rho = 0.5, \beta_1 = 0.5$		
	MSE*10 ³	Estimated variance *10 ³ (*10 ³)		MSE*10 ³	Estimated variance *10 ³ (*10 ³)
$\hat{\beta}_c$	(3.86, 4.10)	4.10(0.77), 4.11(0.81)	$\hat{\beta}_c$	(1.64,1.57)	1.66(0.19), 1.67(0.23)
$\hat{\beta}_{rc}$	(1.61,1.73)	1.68(0.50), 1.67(0.48)	$\hat{\beta}_{rc}$	(1.03,1.03)	1.04(0.13), 1.04(0.13)
$\hat{\beta}$	(1.50,1.63)	1.48(0.34), 1.47(0.33)	$\hat{\beta}$	(1.00,0.96)	1.00(0.11), 1.00(0.12)
$\rho = 0.2, \beta_1 = 1.5$			$\rho = 0.5, \beta_1 = 1.5$		
	MSE*10 ³	Estimated variance *10 ³ (*10 ³)		MSE*10 ³	Estimated variance *10 ³ (*10 ³)
$\hat{\beta}_c$	(3.86,4.10)	4.10(0.77), 4.11(0.81)	$\hat{\beta}_c$	(1.64,1.57)	1.66(0.19), 1.67(0.23)
$\hat{\beta}_{rc}$	(7.20,7.26)	8.54(4.13), 8.44(3.95)	$\hat{\beta}_{rc}$	(2.42,2.23)	2.72(0.66), 2.71(0.59)
$\hat{\beta}$	(3.09,2.91)	3.04(3.18), 3.03(3.42)	$\hat{\beta}$	(1.39,1.22)	1.39(0.16), 1.40(0.17)

Note: The vector “(*)” represents the mean square error in estimation of β_0 and β_1 . The symbols “o” and “o” indicate the average values of variance estimates from (3.1) of estimators of β_0 and β_1 , respectively. The sample standard deviation of these variance estimates are in parentheses.

There is not much difference between Tables 1 and 2, indicating that normality is not necessary for the proposed method. We also note that when $\beta_1 = 0.5$, $\hat{\beta}_{rc}$ is more efficient than $\hat{\beta}_c$. But, when $\beta_1 = 1.5$, the reverse is true. In all cases, $\hat{\beta}$ demonstrates superior performance, as expected. However, the variance estimates should be used with caution when the amount of validation data is relatively small, and the parameter β_1 is not close to 0.

5. Discussion

Response variable regression calibration can be extended to multiple regression. The model (2.1) can be rewritten as

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \sigma_1 \epsilon_i$$

with $\beta_2 = 0$, where the covariates X_i and W_i are correlated. Compare this equation with a multiple regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_m X_{mi} + \sigma_1 \epsilon_i,$$

with a portion of the X_{1i} missing completely at random and with surrogates W_{1i} available for all individuals. There is not much difference between these two equations, except that the number of variables correlated with X_{1i} may go up to m in multiple regression models. A best linear predictor of X_{1i} using response variables is still available, and the estimation procedure is similar to the univariate case.

It is also possible to extend the idea to nonlinear models. For example, consider the nonlinear model $Y = g(\mathbf{X}, \beta) + \epsilon$ described in Lee and Sepanski (1995). They projected $g(\mathbf{X}, \beta)$ onto the subspace consisting of linear functions of surrogates and used a nonlinear least squares method to derive estimates. The method we propose is equivalent to replacing the mean function $\beta_0 + \beta_1 X$ by its projection $\beta_0 + \beta_1 H$ onto the subspace of linear functions of W and Y . In the same spirit, we can project the function $g(\mathbf{X}, \beta)$ onto the subspace of linear functions of surrogates and response variables, though explicit form of the projection may be hard to derive. The situation can be complicated, and we believe its exploration is an interesting area for future research.

Acknowledgements

The author thanks the Editor and referees for helpful comments, and wishes to specially thank an associate editor for valuable assistance in correcting this paper. This research was supported by the National Science Council of Taiwan.

Appendix

We only prove Theorem 1. Theorem 2 can be shown in a similar way. First, we define notation and establish two lemmas. When $\mu = 0$, one can express H_i

as $X_i - \Delta X_i - \eta_i$, where $\Delta = (1 - b - c\beta_1)$ and $\eta_i = -(b\sigma_2\delta_i + c\sigma_1\epsilon_i)$. Let $e_i = X_i - H_i = \Delta X_i + \eta_i$ denote the difference between X_i and its predictor H_i .

Lemma 3. *Let V and V_0 denote the asymptotic covariance matrix of estimators of β_0 and β_1 obtained by solving (2.5) when $\mu = 0$ and when $\mu \neq 0$. Then,*

$$V = \begin{pmatrix} 1 & -\mu \\ 0 & 1 \end{pmatrix} V_0 \begin{pmatrix} 1 & 0 \\ -\mu & 1 \end{pmatrix}.$$

Proof. Note that the equations in (2.5) can be written as

$$\begin{aligned} \sum_{i \in V} [(Y_i - (\beta_0 + \beta_1\mu) - \beta_1 X_i^*)^2 - \sigma_1^2] &= 0, \\ \sum_{i \in V} (X_i^* - a^* - bW_i^* - cY_i) \begin{pmatrix} 1 \\ W_i^* \\ Y \end{pmatrix} &= \mathbf{0}, \\ \sum_{i \in V} (Y_i - (\beta_0 + \beta_1\mu) - \beta_1 X_i^*) \begin{pmatrix} 1 \\ X_i^* \end{pmatrix} \\ + \sum_{i \in P} [(Y_i - (\beta_0 + \beta_1\mu) - \beta_1 H_i^*) \begin{pmatrix} 1 \\ H_i^* \end{pmatrix} - \begin{pmatrix} 0 \\ c\sigma_1^2 \end{pmatrix}] &= \mathbf{0}, \end{aligned}$$

where $X_i^* = X_i - \mu$, $W_i^* = W_i - \mu$, $H_i^* = a^* + bW_i^* + cY_i$ and $a^* = a - (1 - b)\mu$. The estimators $(b_0 + b_1\mu)$ and b_1 of $(\beta_0 + \beta_1\mu)$ and β_1 have asymptotic covariance V_0 . Since

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} 1 - \mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_0 + b_1\mu \\ b_1 \end{pmatrix},$$

the conclusion follows easily.

Lemma 4. *Assume that X, ϵ and δ all have finite fourth moments with $\mu = 0$ and that $\sigma_2^2/\sigma_1^2 = k$ remains fixed as σ_1^2 approaches 0. Then we have*

$$\begin{aligned} a &= \frac{-\beta_0\beta_1k}{1 + \beta_1^2k} + O(\sigma_1^2), \quad b = \frac{1}{1 + \beta_1^2k} + O(\sigma_1^2), \quad c = \frac{\beta_1k}{1 + \beta_1^2k} + O(\sigma_1^2), \\ E(H) &= 0, \quad E(H^2) = \sigma_x^2 - \frac{k\sigma_1^2}{1 + \beta_1^2k} + O(\sigma_1^4), \\ E(e) &= E(eH) = E(eW) = E(eY) = 0, \quad E(eX) = \Delta\sigma_x^2 = O(\sigma_1^2), \\ E(e^2) &= \frac{k}{1 + \beta_1^2k}\sigma_1^2 + O(\sigma_1^3), \quad E(e^2W) = O(\sigma_1^3), \\ E(e^2Y) &= \frac{\beta_0k}{1 + \beta_1^2k}\sigma_1^2 + O(\sigma_1^3), \quad E(e^2X) = O(\sigma_1^3), \quad E(e^2H) = O(\sigma_1^3), \\ E(e^2W^2) &= \frac{k\sigma_x^2}{1 + \beta_1^2k}\sigma_1^2 + O(\sigma_1^3), \quad E(e^2Y^2) = (\beta_0^2 + \beta_1^2\sigma_x^2)\frac{k}{1 + \beta_1^2k}\sigma_1^2 + O(\sigma_1^3), \end{aligned}$$

$$\begin{aligned}
E(e^2WY) &= \frac{\beta_1 k \sigma_x^2}{1 + \beta_1^2 k} \sigma_1^2 + O(\sigma_1^3), \quad E(e^2H^2) = \frac{k \sigma_x^2}{1 + \beta_1^2 k} \sigma_1^2 + O(\sigma_1^3). \\
E(\epsilon) &= 0, \quad E(\epsilon W) = 0, \quad E(\epsilon Y) = \sigma_1, \quad E(\epsilon H) = c \sigma_1, \\
E(\epsilon^2 H) &= 0, \quad E(\epsilon^2 H^2) = (1 - 2\Delta) \sigma_x^2 + O(\sigma_1^2), \\
E(e\epsilon) &= -c \sigma_1, \quad E(e\epsilon H) = 0, \quad E(e\epsilon H^2) = -c \sigma_1 \sigma_x^2 + O(\sigma_1^3), \\
E(e\epsilon W) &= 0, \quad E(e\epsilon X) = 0, \quad E(e\epsilon Y) = -\beta_0 c \sigma_1, \\
E(e\epsilon XW) &= -c \sigma_1 \sigma_x^2, \quad E(e\epsilon XY) = -c \sigma_1 \beta_1 \sigma_x^2 + O(\sigma_1^3).
\end{aligned}$$

Proof. We only establish the forms for a, b, c and $E(e^2WY)$. From (2.2) and some straightforward calculations, it follows that

$$\begin{aligned}
\begin{pmatrix} a \\ b \\ c \end{pmatrix} &= \left(E \begin{bmatrix} 1 & W & Y \\ W & W^2 & WY \\ Y & WY & Y^2 \end{bmatrix} \right)^{-1} E \begin{pmatrix} X \\ WX \\ XY \end{pmatrix} \\
&= \begin{bmatrix} 1 & 0 & \beta_0 \\ 0 & \sigma_x^2 + \sigma_2^2 & \beta_1 \sigma_x^2 \\ \beta_0 & \beta_1 \sigma_x^2 & \beta_0^2 + \beta_1^2 \sigma_x^2 + \sigma_1^2 \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ \sigma_x^2 \\ \beta_1 \sigma_x^2 \end{pmatrix} \\
&= \frac{1}{1 + \beta_1^2 k} \begin{pmatrix} -\beta_0 \beta_1 k \\ 1 \\ \beta_1 k \end{pmatrix} + \mathbf{O}(\sigma_1^2).
\end{aligned}$$

Recall that $e = X - H = \Delta X + \eta$, where $\eta = -b\sigma_2\delta - c\sigma_1\epsilon = O_p(\sigma_1)$ and $\Delta = 1 - b - c\beta_1$. It is easy to show that $\Delta = O(\sigma_1^2)$ and $E(\eta^2) = (k/(1 + \beta_1^2 k))\sigma_1^2 + O(\sigma_1^2)$. It follows that

$$\begin{aligned}
E(e^2WY) &= E[(\Delta^2 X^2 + \eta^2 + 2\Delta X\eta)(X + \sigma_2\delta)(\beta_0 + \beta_1 X + \sigma_1\epsilon)] \\
&= E[(\eta^2(X)(\beta_0 + \beta_1 X)] + O(\sigma_1^3) = \beta_1(b^2\sigma_2^2 + c^2\sigma_1^2)\sigma_x^2 + O(\sigma_1^3) \\
&= \beta_1 \left[\left(\frac{1}{1 + \beta_1^2 k} + O(\sigma_1^2) \right)^2 k \sigma_1^2 + \left(\frac{\beta_1 k}{1 + \beta_1^2 k} + O(\sigma_1^2) \right)^2 \sigma_1^2 \right] \sigma_x^2 + O(\sigma_1^3) \\
&= \frac{\beta_1 k \sigma_x^2}{1 + \beta_1^2 k} \sigma_1^2 + O(\sigma_1^3).
\end{aligned}$$

Proof of Theorem 1. We assume that $E(X) = 0$ to derive a formula and then apply Lemma 3 for the case $E(X) \neq 0$.

Since the expectations of the equations on the left-hand sides of (2.5) are $\mathbf{0}$ (if and only if they are evaluated at the true parameter $\boldsymbol{\eta}$), the consistency property is established. To compute the matrices in (2.6), we see that

$$E \frac{\partial A}{\partial \boldsymbol{\eta}} = \left(\frac{\partial A}{\partial \sigma_1^2}, \frac{\partial A}{\partial a}, \frac{\partial A}{\partial b}, \frac{\partial A}{\partial c}, \frac{\partial A}{\partial \beta_0}, \frac{\partial A}{\partial \beta_1} \right) = (-1, 0, 0, 0, 0, 0),$$

$$E \frac{\partial B}{\partial \boldsymbol{\eta}} = \begin{pmatrix} 0 & -1 & 0 & -\beta_0 & 0 & 0 \\ 0 & 0 & -(\sigma_x^2 + \sigma_2^2) & -\beta_1 \sigma_x^2 & 0 & 0 \\ 0 & -\beta_0 & -\beta_1 \sigma_x^2 & -(\beta_0^2 + \beta_1^2 \sigma_x^2 + \sigma_1^2) & 0 & 0 \end{pmatrix}, \text{ and}$$

$$E \frac{\partial C}{\partial \boldsymbol{\eta}} = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sigma_x^2 \end{pmatrix}.$$

The term $E(\partial D/\partial \boldsymbol{\eta})$ equals $E\partial \left(\begin{matrix} Y - \beta_0 - \beta_1 H \\ (\beta_1 e + \sigma_1 \epsilon)H - c\sigma_1^2 \end{matrix} \right) / \partial \boldsymbol{\eta}$ and consists of

$$\frac{\partial D}{\partial \sigma_1^2} = \begin{pmatrix} 0 \\ -c \end{pmatrix}, \quad \frac{\partial D}{\partial a} = \begin{pmatrix} -\beta_1 \frac{\partial H}{\partial a} \\ (\beta_1 e + \sigma_1 \epsilon) \frac{\partial H}{\partial a} - \beta_1 H \frac{\partial H}{\partial a} \end{pmatrix},$$

$$\frac{\partial D}{\partial b} = \begin{pmatrix} -\beta_1 \frac{\partial H}{\partial b} \\ (\beta_1 e + \sigma_1 \epsilon) \frac{\partial H}{\partial b} - \beta_1 H \frac{\partial H}{\partial b} \end{pmatrix}, \quad \frac{\partial D}{\partial c} = \begin{pmatrix} -\beta_1 \frac{\partial H}{\partial c} \\ (\beta_1 e + \sigma_1 \epsilon) \frac{\partial H}{\partial c} - \beta_1 H \frac{\partial H}{\partial c} - \sigma_1^2 \end{pmatrix},$$

$$\frac{\partial D}{\partial \beta_0} = \begin{pmatrix} -1 \\ -H \end{pmatrix} \text{ and } \frac{\partial D}{\partial \beta_1} = \begin{pmatrix} -H \\ -H^2 \end{pmatrix}.$$

Replacing $(\partial H/\partial a)$, $(\partial H/\partial b)$ and $(\partial H/\partial c)$ by 1, W , and Y and H by $(1-\Delta)X-\eta$ and using Lemma 4, we have

$$E \frac{\partial D}{\partial \boldsymbol{\eta}} = \begin{pmatrix} 0 & -\beta_1 & 0 & -\beta_0 \beta_1 & -1 & 0 \\ -c & 0 & -\beta_1 \sigma_x^2 & -\beta_1^2 \sigma_x^2 & 0 & -\sigma_x^2 + \frac{k\sigma_1^2}{1+\beta_1^2 k} \end{pmatrix} + \mathbf{O}(\sigma_1^4).$$

Replacing c by $[(\beta_1 k \sigma_x^2)/(\sigma_x^2(1 + \beta_1^2 k) + k\sigma_1^2)]$, it follows that

$$\begin{pmatrix} \rho E \frac{\partial A}{\partial \boldsymbol{\eta}} \\ \rho E \frac{\partial B}{\partial \boldsymbol{\eta}} \\ \rho E \frac{\partial C}{\partial \boldsymbol{\eta}} + (1-\rho)E \frac{\partial D}{\partial \boldsymbol{\eta}} \end{pmatrix} \tag{A.1}$$

$$= - \begin{pmatrix} \rho & 0 & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & \rho \beta_0 & 0 & 0 \\ 0 & 0 & \rho(\sigma_x^2 + \sigma_2^2) & \rho \beta_1 \sigma_x^2 & 0 & 0 \\ 0 & \rho \beta_0 & \rho \beta_1 \sigma_x^2 & \rho(\beta_0^2 + \beta_1^2 \sigma_x^2 + \sigma_1^2) & 0 & 0 \\ 0 & (1-\rho)\beta_1 & 0 & (1-\rho)\beta_0 \beta_1 & 1 & 0 \\ \frac{(1-\rho)\beta_1 k \sigma_x^2}{\sigma_x^2(1+\beta_1^2 k)+k\sigma_1^2} & 0 & (1-\rho)\beta_1 \sigma_x^2 & (1-\rho)\beta_1^2 \sigma_x^2 & 0 & \sigma_x^2 - \frac{(1-\rho)k\sigma_1^2}{1+\beta_1^2 k} \end{pmatrix}$$

$$+ \mathbf{O}(\sigma_1^4).$$

From Lemma 4 and straightforward computations, one can show that

$$E(AA') = O(\sigma_1^4), E(BB') = \begin{pmatrix} 1 & 0 & \beta_0 \\ 0 & \sigma_x^2 & \beta_1 \sigma_x^2 \\ \beta_0 & \beta_1 \sigma_x^2 & (\beta_0^2 + \beta_1^2 \sigma_x^2) \end{pmatrix} \frac{k\sigma_1^2}{1 + \beta_1^2 k} + \mathbf{O}(\sigma_1^3).$$

$$E(CC') = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \end{pmatrix} \sigma_1^2, E(DD') = \begin{pmatrix} \frac{1}{1+\beta_1^2 k} & 0 \\ 0 & \frac{\sigma_x^2}{1+\beta_1^2 k} \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3),$$

$$E(AB') = (0, 0, 0) + O(\sigma_1^3), E(AC') = (0, 0), E(BC') = -c \begin{pmatrix} 1 & 0 \\ 0 & \sigma_x^2 \\ \beta_0 & \beta_1 \sigma_x^2 \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3).$$

We also note that $E(AD') = \mathbf{0}$, $E(BD') = \mathbf{0}$ and $E(CD') = \mathbf{0}$, because the individuals in the two sets are independent. In conclusion, the middle factor in (2.6) is

$$\begin{pmatrix} \rho EAA' & \rho EAB' & \rho EAC' \\ \rho EB'A & \rho EBB' & \rho EBC' \\ \rho EC'A & \rho EC'B & \rho ECC' + (1-\rho)EDD' \end{pmatrix} \quad (\text{A.2})$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\rho k}{1+\beta_1^2 k} & 0 & \frac{\rho \beta_0 k}{1+\beta_1^2 k} & -\rho c & 0 \\ 0 & 0 & \frac{\rho k \sigma_x^2}{1+\beta_1^2 k} & \frac{\rho \beta_1 k \sigma_x^2}{1+\beta_1^2 k} & 0 & -\rho c \sigma_x^2 \\ 0 & \frac{\rho \beta_0 k}{1+\beta_1^2 k} & \frac{\rho \beta_1 k \sigma_x^2}{1+\beta_1^2 k} & \frac{\rho(\beta_0^2 + \beta_1^2 \sigma_x^2)k}{1+\beta_1^2 k} & -\rho \beta_0 c & -\rho \beta_1 c \sigma_x^2 \\ 0 & -\rho c & 0 & -\rho c \beta_0 & \frac{1+\rho \beta_1^2 k}{1+\beta_1^2 k} & 0 \\ 0 & 0 & -\rho c \sigma_x^2 & -\rho c \beta_1 \sigma_x^2 & 0 & \frac{1+\rho \beta_1^2 k}{1+\beta_1^2 k} \sigma_x^2 \end{pmatrix} \sigma_1^2 + \mathbf{O}(\sigma_1^3).$$

The determinant of (A.1) is of order $O(\sigma_1^2)$. If terms of $O(\sigma_1^4)$ in (A.1) and terms of $O(\sigma_1^3)$ in (A.2) are ignored, they become the matrices T and M in (2.6). Multiplying by the matrix $T^{-1}MT^{-1'}$ completes the proof of Theorem 1.

References

- Carroll, R. J., Rupper, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, London.
- Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasilielihood estimation in models with surrogate predictor. *J. Amer. Statist. Assoc.* **85**, 652-663.
- Lee, L. and Sepanski, J. H. (1995). Estimation of linear and nonlinear errors-in-variables models using validation data. *J. Amer. Statist. Assoc.* **90**, 130-140.
- Liang, K. Y. and Liu, X. H. (1991). Estimating equations in generalized linear models with measurement error. In *Estimating Functions* (Edited by V. P. Godambe), 47-63.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Department of Mathematics, Tamkang University, Tamsui, Taipei County, Taiwan.

E-mail: huang@math.tku.edu.tw

(Received December 2001; accepted September 2004)