

Comparisons Between Some Estimators  
in Functional Errors-in-Variables Regression Models

by

Raymond J. Carroll\*

Paul P. Gallo\*\*

Key words and phrases: Errors-in-variables, analysis of covariance, least squares, functional model, Monte-Carlo, robustness, measurement error.

\* Department of Statistics, University of North Carolina at Chapel Hill. Research supported by the Air Force Office of Scientific Research under Grant AFOSR-80-0080, and partially completed while the author was at the National Heart, Lung and Blood Institute.

\*\*Statistical Design and Analysis Department, Lederle Laboratories, Pearl River, N.Y. Research supported in part by a National Science Foundation Graduate Fellowship and by the Air Force Office of Scientific Research under Grant AFOSR-80-0080.

## ABSTRACT

We study the functional errors-in-variables regression model. In the case of no equation error (all randomness due to measurement errors), the maximum likelihood estimator computed assuming normality is asymptotically better than the usual moments estimator, even if the errors are not normally distributed. For certain statistical problems such as randomized two group analysis of covariance, the least squares estimate is shown to be better than the aforementioned errors-in-variables methods for estimating certain important contrasts.

## 1. Introduction

The problem of estimating linear regression parameters when the variables are subject to measurement or observation error has a long history and has recently been the focus of considerable attention. Reilly and Patino-Leal (1981) list a number of recent publications concerning situations in which the problem arises; see Wu, Ware and Feinlieb (1980) for a simple but particularly interesting example in a biomedical context. Blomquist (1977), Nussbaum (1980), Fuller (1980) and Gleser (1981) have recently addressed various theoretical aspects of the problem.

The purposes of this paper are three. First, by exploiting a particular representation of estimators we unify and extend some of the asymptotic results for the normal theory maximum likelihood estimator (normality-MLE) and the "method of moments" estimators developed by Fuller (1980). Second, having obtained the asymptotic distributions of the method of moments estimators and the normality-MLE, we are in a position to compare the two via limiting variances. In a particular important special case, we are able to show that the normality-MLE is better than the method of moments estimator in the sense of having an asymptotic normal distribution centered about the true regression parameter and with smaller asymptotic variance. This is perhaps not too surprising at the normal model, but it in fact holds even if assumptions of normality are violated. Our Monte-Carlo study confirms this result, but we also discuss reasons why one would want to use the method of moments estimator in practice, especially when using Fuller's small sample modification.

The third major purpose of this paper is to study the least squares estimator (LSE), computed as if the variables were observed exactly. The LSE is generally inconsistent for regression parameters, and thus has not been considered much in the literature. This is unfortunate because, as has not been generally recognized, there are important statistical problems in which the LSE is consistent; one example is two-group analysis of covariance for a

randomized study, where the LSE consistently estimates the treatment effect difference even when there are errors in the variables. A heuristic asymptotic result suggests that when the LSE is consistent for a particular contrast, it will be better than the normality-MLE in large samples. The conjecture is explicitly confirmed for two group analysis of covariance. Our small Monte-Carlo study is illuminating here.

There are two other features of the paper which are important. First, to the best of our knowledge the Monte-Carlo results are among the first of their kind for the errors-in-variables problems we consider, although Wolter and Fuller (1982 a,b) have Monte-Carlo as well. Second, the Monte-Carlo study includes recently introduced generalizations of M-estimates (Carroll and Gallo (1982)), which we show to work quite well.

## 2. Models, Assumptions and Estimates

We consider a general errors-in-variables (EIV) regression model in which some subset of the variables is subject to error, while some are observed exactly; the response is replicated  $s$  times and the predictor variables subject to measurement error are replicated  $r$  times. Specifically,

$$\begin{aligned} Y_i &= X_1 \beta_1 + X_2 \beta_2 + \epsilon_i \\ \epsilon_i &= \delta + V_i & i = 1, \dots, s \\ C_j &= X_2 + U_j & j = 1, \dots, r. \end{aligned}$$

Here,  $\beta_1$  is a  $(p_1 \times 1)$  vector and  $\beta_2$  is  $(p_2 \times 1)$ ,  $p = p_1 + p_2$ . The vectors  $Y_i$ ,  $\delta$ , and  $V_i$  are of dimension  $(N \times 1)$ , where  $N$  is the sample size in the study.  $X_1$  and  $X_2$  are constant matrices of order  $(N \times p_1)$  and  $(N \times p_2)$ , respectively.  $X_1$  is observable, however, because of measurement error  $U_j$ ,  $X_2$  is not observable but rather the  $(N \times p_2)$  matrices  $C_j$  are observed. The  $(N \times 1)$  random vector  $\delta$  is called the equation error, while the  $\{V_i\}$  are the measurement errors in the response. The assumption that  $X_1$  and  $X_2$  are constant puts us in the functional EIV model. In some cases we will assume

no equation error ( $\delta = 0$ ), in which case we have the classical linear functional relationship. The concept of equation error was introduced by Fuller (1980).

We assume that the  $\{V_i\}$  are mutually independent and independent of  $\delta$  as are the  $\{U_j\}$ . The elements of  $\delta$  and those of each  $V_i$  are i.i.d. with zero mean and finite variances  $\sigma_\delta^2$  and  $\sigma_v^2$  respectively, while the rows of each  $U_j$  are i.i.d. with mean zero and non-singular covariance matrix  $\Sigma_u$ . We define

$$\sigma^2 = \sigma_\delta^2 + \sigma_v^2$$

$$X = [X_1 \ X_2]$$

and assume that  $X$  is of full column rank such that

$$\Delta = \lim_{N \rightarrow \infty} N^{-1} X' X \text{ is positive definite.} \quad (2.1)$$

We further define

$$U_{j*} = \begin{bmatrix} 0 & U_j \\ N \times p_1 \end{bmatrix} \quad \beta' = [\beta'_1 \ \beta'_2] \quad C_{j*} = [X_1 \ C_j]$$

$$\Sigma_{u*} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_u \end{bmatrix} \quad \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}$$

Where  $\Delta_{11}$  and the upper left-hand submatrix of zeroes in  $\Sigma_{u*}$  are  $(p_1 \times p_1)$ .

We will discuss a number of special cases of our EIV model and define an estimator of  $\beta$  in each.

Case No. 1 (No replication) Gleser (1981) and the majority of researchers assume no replication is available ( $r = s = 1$ ). Here, we suppress the subscripts referring to the replicates and write  $Y$ ,  $\epsilon$ ,  $C$ , etc. We assume that the rows of  $[U \ \epsilon]$  have finite fourth moments and that a matrix  $\Sigma_0$  is known such that

$$\Sigma = \sigma^2 \Sigma_0 = \sigma^2 \begin{bmatrix} \Sigma_{u0} & \Sigma_{\epsilon u0} \\ \Sigma'_{\epsilon u0} & 1 \end{bmatrix}; \quad (2.2)$$

we define  $\Sigma_{0*}$ ,  $\Sigma_{u0*}$ ,  $\Sigma_{\epsilon u0*}$  correspondingly.

With  $I_N$  the identity matrix of order  $N$ , write

$$R = I_N - X_1 (X_1' X_1)^{-1} X_1'$$

$$W = [C \ Y]' R [C \ Y].$$

Let  $\theta$  be the smallest eigenvalue of  $\Sigma_0^{-1}W$ . If  $C_*' C_* - \theta \Sigma_{0*}$  is non-singular (Gallo (1982b) has shown that this holds a.s. if the error distribution is absolutely continuous), we define

$$\hat{\beta}_M = (C_*' C_* - \theta \Sigma_{0*})^{-1} (C_*' Y - \theta \Sigma_{\epsilon u0*}) \quad (2.3)$$

This estimate is the maximum likelihood estimate for jointly normally distributed errors (note: if we omit assumption (2.2), the supremum of the likelihood is infinite). The estimate was derived in a more general framework by Healy (1975) and was shown by Gleser (1981) to be equivalent to a generalized weighted least squares estimate. We emphasize that we will study (2.3) and the other estimates of  $\beta$  without assuming normality.

Case No. 2 (Equal replication) Here we let  $s = r > 1$ . The equal replication is convenient since it admits simpler notation, but it is by no means necessary. It does arise in practical circumstances. For example, if one predictor is baseline diastolic blood pressure the response is change in diastolic blood pressure, as in the Framingham Heart Study of the National Heart, Lung and Blood Institute and in other studies, a common practice is to take one replicate, i.e.,  $r = s = 2$ . A method of moments estimator motivated by the work of Fuller (1980) is

$$\hat{\beta}_R = \left( \sum_{j=1}^r \sum_{k=1}^r \Sigma_{j \neq k} C_{j*}' C_{k*} \right)^{-1} \sum_{j=1}^r \sum_{k=1}^r \Sigma_{j \neq k} C_{j*}' Y_k \quad (2.4)$$

We assume that the random matrices  $\delta$ ,  $[\varepsilon_1 U_1], \dots, [\varepsilon_r U_r]$  are mutually independent and that the other specifications of Case No. 1 hold. (The normality-MLE has not been calculated for a case such as this in which  $\delta \neq 0$ ).

Case No. 3 (Equal replication, no equation error) This is the same situation as in Case No. 2 except that  $\delta = 0$ , i.e., apart from measurement error the underlying relation is exactly linear.

Let  $\theta_R$  be the smallest eigenvalue of  $T_1^{-1} T_2$ , where

$$T_1 = \sum_{i=1}^r \sum_{j=1}^r [C_i Y_i]' (\delta_{ij} - r^{-1}) I_N [C_j Y_j]$$

$$T_2 = T_1 + r^{-1} \left( \sum_{i=1}^r [C_i Y_i]' \right) R \left( \sum_{i=1}^r [C_i Y_i] \right)$$

( $\delta_{ij}$  is the Kronecker delta, the indicator of  $i = j$ ). Then with

$$m_{ij} = (r(\theta_R - 1) \delta_{ij} - \theta_R)$$

we define

$$\hat{\beta}_{MR} = \left( \sum_{j=1}^r \sum_{k=1}^r m_{jk} C_{j*}' C_{k*} \right)^{-1} \sum_{j=1}^r \sum_{k=1}^r m_{jk} C_{j*}' Y_k \quad (2.5)$$

This is the normality-MLE in the replication case, and has been derived by Anderson (1951) and Healy (1980). Note that assumption (2.2) is unnecessary here.

The estimates in all cases above have been shown to be consistent for  $\hat{\beta}$  as  $N \rightarrow \infty$ ; conditions on  $X$  weaker than (2.1) were obtained by Gallo (1982b). In Case No. 2,  $\hat{\beta}$  has been shown by Fuller (1980) to have a limiting normal distribution when  $U$  and  $\varepsilon$  are normally distributed; under non-normality, Fuller (1975) has some related results, although our proofs are different. The MLE in Case No. 1 was demonstrated by Gleser (1981) to be asymptotically

normal, but the proof contains a slight error. (In particular, Gleser's Lemma 4.1 is contradicted by the following example: let  $\{y_k\}$  be a sequence of independent random variables assuming values  $-2^{k/2}$ ,  $0$ ,  $2^{k/2}$  with probabilities  $2^{-(k+1)}$ ,  $1-2^{-k}$ ,  $2^{-(k+1)}$ , respectively. According to the lemma,  $N^{-1/2} \sum_{k=1}^N y_k$  is asymptotically standard normal, yet it can be shown that this quantity is  $o_p(1)$ .) A simple remedy would require that the errors have finite moments of order greater than four, an assumption we would like to avoid if possible.

Finally, there are practical problems where it is known in advance that  $\Sigma_{\epsilon U} = 0$ . In this instance, the estimators (2.3) and (2.5) can be altered to a form in which they are more efficient. Our main qualitative comparisons and conclusions (Sections 4-5) are unaffected.

### 3. Asymptotic Normality

In this section we state the form of the asymptotic distributions of the estimators. The proofs are technical and are delayed until Section 6.

#### Theorem 1

(No. 1) In Case No. 1,  $N^{1/2} (\hat{\beta}_M - \beta)$  is asymptotically normally distributed with mean zero. If the third and fourth moments of the joint distribution of the rows of  $U$  and  $\epsilon$  are the same as those of the normal distribution then the asymptotic covariance matrix of  $N^{1/2} (\hat{\beta}_M - \beta)$  is

$$\text{Cov} (\hat{\beta}_M) = d \{ \Delta^{-1} + \Delta^{-1} \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \Delta^{-1} \} \quad (3.1)$$

where

$$d = [\beta_2' \ -1] \ \Sigma \ [\beta_2' \ -1]'$$

$$Q = ([I_{p_2} \ \beta_2] \ \Sigma^{-1} [I_{p_2} \ \beta_2]')^{-1}.$$



(No. 2) In Case No. 2,  $N^{1/2} (\hat{\beta}_R - \beta)$  is asymptotically normally distributed with mean zero and covariance matrix

$$\begin{aligned} \text{Cov} (\hat{\beta}_R) = & \Delta^{-1} (r^{-1} (d-\sigma_\delta^2) \Delta + \sigma_\delta^2 (\Delta + r^{-1} \Sigma_{u^*})) \\ & + r^{-1} (r-1)^{-1} ((d-\sigma_\delta^2) \Sigma_{u^*} + D_R)) \Delta^{-1}, \end{aligned} \quad (3.2)$$

where

$$D_R = (\Sigma_{u^*} \beta - \Sigma_{\epsilon u^*}) (\Sigma_{u^*} \beta - \Sigma_{\epsilon u^*})'.$$

(No. 3) In Case No. 3,  $N^{1/2} (\hat{\beta}_R - \beta)$  is asymptotically normally distributed with mean zero and covariance matrix

$$\text{Cov} (\hat{\beta}_{MR}) = (dr^{-1}) \Delta^{-1} + (r-1)^{-1} \Delta^{-1} \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix} \Delta^{-1}. \quad (3.3)$$

Again, note that although two of our three estimates are normality-MLE's, we do not assume in any part of Theorem 1 that the errors are normally distributed. The assumption made in part (No. 1) of the theorem that the error distribution moments are those of the normal distribution is not necessary for the asymptotic normality of  $\hat{\beta}_M$ ; nevertheless, the limit variance depends on the third and fourth error moments and is in general quite unwieldy. In stating the theorem we thus assume that the moments are those of the normal distribution (as did Gleser (1981)) since this yields a concise expression more easily compared with those of other estimates. We have made no further assumptions on the errors beyond those of Section 2; in particular, in Cases No. 2 and 3 we require only two finite moments.

#### 4. Comparisons for the Linear Functional Relationship

We consider in this section Case No. 3, the linear functional relationship with no equation error. In this case the asymptotic covariances (3.2) and (3.3) are comparable.

Theorem 2 For Case No. 3, (even under non-normal distributions), the normality-MLE (2.5) is asymptotically no worse than the moment estimator (2.4), i.e.,

$\text{Cov}(\hat{\beta}_R) - \text{Cov}(\hat{\beta}_{MR})$

is positive semi-definite.

The proof is given in Section 7. However, there is an important special case where the result is obvious, when  $\Sigma = \sigma^2 I_{p_2+1}$ . Then

$$\text{Cov}(\hat{\beta}_R) - \text{Cov}(\hat{\beta}_{MR}) = 2r^{-1} (r-1)^{-1} \sigma^4 \Delta^{-1} \begin{bmatrix} 0 & 0 \\ 0 & \beta_2 \beta_2' \end{bmatrix} \Delta^{-1};$$

of course, what is most interesting about Theorem 2 is that the normality-MLE is the (asymptotic) winner over method of moments even at non-normal distributions. To get some idea of whether this result holds in small samples, we performed the following Monte-Carlo study.

All calculations were done at the NIH computing center. Random numbers were generated using the IMSL routines GGNPM and GGUBS. There were 500 Monte-Carlo replications. The true model was simple linear regression following the format of Section 2 with  $r=s=2$  replications. The intercept was 10 and the slope was -4. In the notation of Section 2,  $X_1$  is a column vector of  $N=40$  ones,  $\beta_1 = 10$ ,  $\beta_2 = -4$  and  $X_2$  is a column vector obtained as the values of  $X_2$  from Table 1 of Jobson and Fuller (1980).

Although the estimates were calculated in the forms which do not assume  $\Sigma_{\epsilon U} = 0$ , we fixed  $\Sigma_{\epsilon U} = 0$  and performed the Monte-Carlo study. The rows of the error terms  $S$ ,  $V_i$  and  $U_i$  were thus generated independently; all three were either normally distributed or had a contaminated normal distribution. In general, any random variable was either  $N(0, \sigma^2)$  (Normal) or  $N(0, \sigma^2)$  with probability 0.9 and  $N(0, c^2 \sigma^2)$  with probability 0.1 (Contaminated Normal), with

Equation Error ( $\delta$ ):  $\sigma_\delta^2 = 0$  or 4,  $c=5$

Y Measurement Error ( $V$ ):  $\sigma_V^2 = 4$ ,  $c=5$

X Measurement Error ( $U$ ):  $\sigma_U^2 = 1, 2$  or 4,  $c=3$

The normality-MLE is calculated assuming  $\sigma_\delta^2 = 0$ , so the results for  $\sigma_\delta^2 = 4$  give some idea of its robustness against this assumption. The normality-MLE was computed by (2.5). The moments estimator was (2.4), with exception that the term in (2.4)

$$\sum_{j=1}^r \sum_{\substack{k=1 \\ j \neq k}}^r C_j' C_k^*$$

was replaced by Fuller's modification (Fuller, (1980), page 414-415) using Fuller's  $\alpha=1$ . This modification is crucial to get the best performance of the moments estimator. Although this modification occurs with negligible probability asymptotically, we found in our study that it occurred often.

Finally, we studied a generalization of the moments estimator (2.4) introduced by Carroll and Gallo (1982) and designed to be robust against departures from normality in  $\delta$  and  $V$ . If Fuller's modification was necessary, we used his estimator. Otherwise, we solved

$$0 = \sum_{i=1}^N Z_{i2} \Psi \left( \frac{Y_{i1} - \beta_1 - \beta_2 Z_{i1}}{\hat{\sigma}} \right) + Z_{i1} \Psi \left( \frac{Y_{i2} - \beta_1 - \beta_2 Z_{i2}}{\hat{\sigma}} \right) \quad (4.1)$$

where  $\{Z_{ij}\}$  are the individual elements of  $C_j$  and  $\hat{\sigma}$  is the median absolute residual from the moments fit divided by .6745. It is easy to solve (4.1) by

iteratively reweighted least squares, although some care must be taken. To this date there is no estimator with distributional robustness properties which includes as a special case the normality-MLE, although such an estimator will surely be soon discovered. For Case No. 1, Brown (1982) has suggested such a class, but his proof of consistency is in error and his estimator is not consistent and asymptotically normal in general.

Mean square error (MSE) efficiencies relative to the normality-MLE are given in Table 1, along with the percentage of times Fuller's modification was used. Efficiencies are also given in Table 2 for the 95th percentile of the absolute errors for the different estimators. The first twelve lines of each table are for the situation of no equation error ( $\delta = \sigma_\delta^2 = 0$ ), assumed in calculating the normality MLE. Note as suggested by Theorem 2 that the normality-MLE generally dominates the moments estimator (but not vastly so), even at non-normal distributions. The Carroll-Gallo estimator is generally the best, even when compared to the normality-MLE and even though it is meant to improve the moments estimator, not the normality-MLE. The Carroll-Gallo estimator does lose some efficiency when the measurement error in  $X_2$  becomes very large; this is a reflection of the fact that the "asymptotically negligible" Fuller modification is needed 30%-50% of the time. Clearly, it would be helpful to have a distribution-robust generalization of the normality-MLE. Further work should also focus on bounded influence (Carroll and Gallo (1982) make one simple suggestion along these lines).

The last twelve lines of Tables 1 and 2 reflect the situation  $\sigma_\delta^2 = 4$ , i.e., there is substantial equation error. Here the normality-MLE calculated assuming  $\sigma_\delta^2 = 0$  does particularly poorly. Clearly, the normality-MLE is not robust against violations of the linear functional relationship (no equation error). Certainly, the Monte-Carlo suggests the need for calculation and study of the normality-MLE in the general Case No. 2. In the absence of such a general estimator, we would in practice favor the moments estimator (2.4) with Fuller's modification over the estimator (2.5), especially for small

samples or if substantial equation error is a possibility. Wolter and Fuller (1982 a,b) also present Monte-Carlo which emphasizes that the moments estimator can be superior in practice.

We consider the results for the Carroll-Gallo estimator to be very encouraging, but further development is clearly needed. On an interim basis, our estimators should be considered a supplement to and not a replacement for Fuller's modified MME.

### 5. Comparisons with the Least Squares Estimator

It is well-known that the ordinary least squares estimate computed as if the observed values  $C_*$  were the exact values of interest,

$$\hat{\beta}_L = (C_*' C_*)^{-1} C_*' Y,$$

is inconsistent for  $\beta$  in EIV models. There are, however, situations in which least squares can consistently estimate particular parameters of interest. One such situation is two-class analysis of covariance. (DeGracie and Fuller (1972) considered this situation in an EIV context). The most important parameter is often the treatment difference; it turns out that this is consistently estimated by least squares as long as subjects are assigned to treatments in such a way that the difference between the treatment means of the covariate approaches zero, as would occur in a randomized or matched study. More generally, Gallo (1982a) has shown

Theorem 3 Let  $\gamma' = [\gamma_1' \ \gamma_2']$ , where  $\gamma_1$  is a  $(p_1 \times 1)$  vector and  $\gamma_2$  is a  $(p_2 \times 1)$  vector. Then

$$\gamma' \hat{\beta}_L \stackrel{P}{\rightarrow} \gamma' \beta \text{ for all } \beta, \Sigma \text{ iff} \tag{5.1}$$

$$\gamma_2' = \gamma_1' \Delta_{11}^{-1} \Delta_{12}.$$

The advantage of such consistent contrasts is that we can consistently estimate important parameters without needing replication or the often artificial assumption (2.2). The LSE can only be considered if it is as good an estimator as the EIV methods such as (2.3) - (2.5).

For example, first consider Case No. 1. General comparison of  $\hat{\beta}_L$  and  $\hat{\beta}_M$  turns out to be difficult in our model, because the limit distribution of  $\hat{\beta}_L$  is not easily calculated. The following heuristic calculations are informative. Suppose

$$N^{1/2} (\gamma_1' (X_1' X_1)^{-1} X_1' X_2 - \gamma_2') \rightarrow 0 ; \quad (5.2)$$

(this implies (5.1)). Then

Theorem 4 Under (5.2),  $N^{1/2} \gamma' (\hat{\beta}_L - \beta)$  is asymptotically normal with mean zero and covariance  $\text{Cov} (\hat{\beta}_L, \gamma)$ . Further,

$$\text{Cov} (\hat{\beta}_L, \gamma) \leq \gamma' \text{Cov} (\hat{\beta}_M) \gamma, \quad (5.3)$$

i.e., the LSE has no larger asymptotic variance than the normality-MLE. Equality holds if and only if  $\Sigma_{\epsilon U} = \Sigma_U \beta_2$ .

The proof of Theorem 4 is in Section 7. Since (5.2) cannot be guaranteed, the relevance of Theorem 4 is heuristic. For example, in the ANOCOVA example mentioned previously, (5.2) implies that the true mean covariable difference is  $o(N^{-1/2})$ , which is not assured even if the observed means are set equal for all  $N$ . The Monte-Carlo reported later supports Theorem 4, but other theoretical calculations are also possible.

Consider balanced two-class analysis of covariance such as might occur in a randomized study. In the notation of Section 2.

$$\begin{aligned}
 X_1' &= \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & -1 & 1 & -1 & \dots & -1 \end{bmatrix} \\
 (5.3) \quad X_2' &= \begin{bmatrix} Z_1 & Z_2 & \dots & Z_N \end{bmatrix} \\
 \beta_1' &= (\mu \quad \alpha).
 \end{aligned}$$

The parameter of interest is  $\alpha = (0 \ 1 \ 0) \beta$ . If the design is actually randomized, the  $\{Z_i\}$  would be i.i.d. with mean  $\mu_Z$  and variance  $\sigma^2$ . This is not our functional model because  $X_2$  is not a vector of fixed constants, nevertheless Theorem 1 is true in this instance. Denoting the estimates of  $\alpha$  by  $\hat{\alpha}_L$ ,  $\hat{\alpha}_M$ ,  $\hat{\alpha}_R$  and  $\hat{\alpha}_{MR}$  for the LSE and (2.3) - (2.5) respectively, one can use Theorem 1 to prove that  $\hat{\alpha}_L$  is always as least as good asymptotically as the others. If by an appropriately standardized version of  $\hat{\alpha}$  we mean  $N^{1/2} (\hat{\alpha} - \alpha)$ , then

Theorem 5 Appropriately standardized, in either Case No. 1 or Case No. 3, the LSE  $\hat{\alpha}_L$  always has smaller asymptotic variance than  $\hat{\alpha}_M$ ,  $\hat{\alpha}_R$  or  $\hat{\alpha}_{MR}$ .

Theorem 5 seems to imply that in such randomized studies, one is better off not using EIV techniques. Also, in terms of inference, detailed calculations enable one to prove

Theorem 6 Consider Case No. 1 (this result holds for Case No. 3 as well).  $N^{1/2} (\hat{\alpha}_L - \alpha)$  is asymptotically normal with mean zero and variance  $\sigma_L^2$ . Let  $\hat{\sigma}_L^2$  be the usual estimate of the variance of  $N^{1/2} (\hat{\alpha}_L - \alpha)$ . Then

$$\hat{\sigma}_L^2 \xrightarrow{P} \sigma_L^2.$$

Thus, while errors-in-variables make the LSE less efficient, the inferences one normally makes using the LSE are asymptotically correct for the treatment effect in randomized two class ANOCOVA.

We also performed a small Monte-Carlo study of two class ANOCOVA when  $N=40$  and  $r=s=2$ . The random variables for measurement and equation error were generated as in Section 4, although we only studied the case of no equation error. The covariables of (5.3) were normally distributed with mean zero and variance 9. We chose  $\beta' = (4 \ 4 \ 4)$  so that  $\alpha = 4$ .

In Table 3, for estimating the treatment effect  $\alpha$  we report the MSE efficiencies of the LSE relative to the normality-MLE. The results are strikingly in accord with Theorem 5.

As seen in Table 3, even when there is no equation error, the LSE is much better than the normality-MLE for estimating the treatment effect, while the LSE is much worse for estimating the often less important covariable effect.

The preceding results apply only to balanced analysis of covariance. If the covariates are not balanced across treatments, the LSE will inconsistently estimate the treatment effect, with possibly disastrous consequences.

## 6. Proof of Main Result

In proving Theorem 1, we will make use of the following result.

Lemma 1 Let  $\{X_i\}$  and  $\{Y_i\}$  be two sequences of random variables, each i.i.d. with zero mean, positive finite variances  $\sigma_x^2$  and  $\sigma_y^2$ , respectively, and  $\text{Cov}(X_i, Y_j) = \delta_{ij} \sigma_{xy}$ . Let  $\{a_i\}$  be a sequence of constants satisfying

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n a_i^2 = a^2, \quad 0 < a^2 < \infty. \quad (6.1)$$

Then with

$$S_n^2 = \sigma_x^2 \sum_{i=1}^n a_i^2 + n \sigma_y^2 + 2 \sigma_{xy} \sum_{i=1}^n a_i, \quad S_n^2 > 0,$$

$$S_n^{-1} \sum_{i=1}^n (a_i X_i + Y_i)$$



converges in distribution to a standard normal random variable.

The proof is straightforward, and is contained in Gallo (1982a).

We now outline the proof of Theorem 1. Again, complete details are provided by Gallo (1982a).

Proof of Theorem 1 Part (i): let

$$\begin{aligned} W_* &= [C_* Y]' [C_* Y] \\ S_1 &= [I_{p_2} \beta_2]' \Sigma_0^{-1} \\ S_2 &= (S_1 [I_{p_2} \beta_2]')^{-1} S_1 \\ S_3 &= - S_2 [0 \beta_1]'. \end{aligned}$$

The following representation is essentially an extension of one obtained by Gleser (1981):

$$N^{1/2} (\hat{\beta}_M - \beta) = - \Delta^{-1} \begin{bmatrix} I_{p_1} & 0 \\ S_3 & S_2 \end{bmatrix} N^{-1/2} (W_* - E(W_*)) [\beta' \ -1] + o_p(1) \quad (6.2)$$

as long as

$$(W_* - E(W_*)) [\beta' \ -1]' = o_p(N^{1/2}). \quad (6.3)$$

Thus, finding a limit distribution for  $\hat{\beta}_M$  reduces to finding one for the term in (6.3). Letting

$$H = [X \ X\beta], \quad G = [U_* \ \epsilon]$$

with  $H_i'$ ,  $G_i'$  the  $i^{\text{th}}$  rows of these matrices, and noting that  $H [\beta' \ -1]' = 0$ , for all  $\gamma \in \mathbb{R}^{p+1}$  we have

$$\gamma' (W_* - E(W_*)) [\beta' \ -1]' = \sum_{i=1}^N \gamma' (H_i' G_i' + G_i' G_i' - \Sigma_*) [\beta' \ -1]'$$

If  $\gamma' = [\beta' \ -1]$ , then each  $\gamma' H_i = 0$  and limiting normality follows trivially since we have a sum of i.i.d. random variables with finite variance. For all other  $\gamma$ ,

$$N^{-1} \sum_{i=1}^N (\gamma' H_i)^2 = N^{-1} \gamma' H' H \gamma + \gamma' [I_p \ \beta]' \Delta [I_p \ \beta] \gamma > 0,$$

and the sequences  $\{G_i' [\beta' \ -1]'\}$  and  $\{\gamma' (G_i \ G_i' - \Sigma_*) [\beta' \ -1]'\}$  are each i.i.d. with zero mean and finite variances. Thus, Lemma 1 applies, and after some algebra we obtain

$$N^{-1/2} (W_* - E(W_*)) \rightarrow N(0, d([I_p \ \beta]' \Delta [I_p \ \beta] + \Sigma_*) + \Sigma_* [\beta' \ -1]' [\beta' \ -1] \Sigma_*).$$

The result (3.1) follows using (6.2), after some more algebra.

Part (ii):

$$\begin{aligned} \hat{\beta} &= \left( \sum_{i \neq j} \sum C_{i*}' C_{j*} \right)^{-1} \left( \sum_{i \neq j} \sum C_{i*}' Y_j \right) \\ N^{1/2} (\hat{\beta} - \beta) &= N \left( \sum_{i \neq j} \sum C_{i*}' C_{j*} \right)^{-1} N^{-1/2} \sum_{i \neq j} \sum C_{i*}' (Y_j - C_{j*} \beta) \\ &= r^{-1} (r-1)^{-1} \Delta^{-1} N^{-1/2} \sum_{i \neq j} \sum C_{i*}' (Y_j - C_{j*} \beta) + o_p(1) \end{aligned}$$

as long as

$$\sum_{i \neq j} \sum C_{i*}' (Y_j - C_{j*} \beta) = o_p(N^{1/2}).$$

With  $G_j = [U_{j*} \ \epsilon_j]$  and  $U_{jk*}'$ ,  $X_k'$ ,  $G_{jk}'$  the  $k^{\text{th}}$  rows of  $U_{j*}$ ,  $X$  and  $G_j$ , we have,

for all  $\gamma \in \mathbb{R}^p$ ,

$$\gamma' \sum_{i \neq j} \sum C_{i*}' (Y_j - C_{j*} \beta) = - \gamma' \sum_{k=1}^N ((r-1) X_k \sum_{j=1}^r G_{jk}' + \sum_{i \neq j} \sum U_{ik*} G_{jk}') [\beta' \ -1]'$$

The sequences  $\{\sum_{j=1}^r G_{jk}' [\beta' \ -1]'\}$  and  $\{\gamma' \sum_{i \neq j} \sum U_{ik*} G_{jk}' [\beta' \ -1]'\}$  are each i.i.d.

with zero means and finite variances; also,

$$N^{-1} \sum_{k=1}^N ((r-1) \gamma' X_k)^2 = N^{-1} (r-1)^2 \gamma' X' X \gamma + (r-1)^2 \gamma' \Delta \gamma;$$

thus Lemma 1 applies as before, and the result (3.2) follows.

Part (iii): This is similar to part (i); with  $H$  and  $G_j$  as in part (ii), let

$$T_{1*} = r^{-1} (r-1) \sum_{i=1}^r G_i' G_i - r^{-1} \sum_{i \neq j} \sum G_i' G_j$$

$$T_{2*} = \sum_{i=1}^r (H + G_i)' (H + G_i).$$

Analogously to (6.2), we can obtain the following: if

$$(T_{2*} - r(r-1)^{-1} T_{1*}) [\beta' \ -1]' = o_p(N^{1/2}),$$

then

$$N^{1/2} (\hat{\beta}_{MR} - \beta) = r^{-1} \Delta^{-1} \begin{bmatrix} I_{p1} & 0 \\ S_3 & S_2 \end{bmatrix} N^{-1/2} (T_{2*} - r(r-1)^{-1} T_{1*}) [\beta' \ -1]' + o_p(1).$$

For all  $\gamma \in \mathbb{R}^p$ ,

$$\gamma' (T_{2*} - r(r-1)^{-1} T_{1*}) [\beta' \ -1]' =$$

$$\gamma' \left( \sum_{i=1}^r H' G_i + (r-1)^{-1} \sum_{i \neq j} \sum G_i' G_j \right) [\beta' \ -1]'$$

As before, the sequences

$$\left\{ \sum_{i=1}^r G_{ik}' [\beta' \ -1]' \right\} \text{ and } \left\{ \gamma' \sum_{i \neq j} \sum G_{ik}' G_{jk}' [\beta' \ -1]' \right\}$$

are i.i.d., each with zero mean, and the sequence  $\{\gamma' H_k\}$  is just as in part (i); so the result again follows from Lemma 1.

## 7. Other Details

In this section we supply the details of the proofs of Theorems 2 and 4.

Proof of Theorem 2 With  $\sigma_\delta^2 = 0$ , the limiting covariance matrix of  $\hat{\beta}_R$  in (3.2) becomes

$$\Delta^{-1} (r^{-1} d \Delta + r^{-1}(r-1)^{-1} (d\Sigma_{u^*} + D_R)) \Delta^{-1}.$$

Comparing this with (3.3), since  $D_R$  is positive semidefinite, it will suffice to show that for all  $\gamma \in \mathbb{R}^{p_2}$ ,

$$\gamma' ([I_{p_2} \ \beta] \Sigma^{-1} [I_{p_2} \ \beta_2]')^{-1} \gamma' \leq \gamma' \Sigma_u \gamma. \quad (7.1)$$

Now since  $\Sigma^{-1}$  can be expressed as

$$\begin{bmatrix} \Sigma_u^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \Sigma_u^{-1} & \Sigma_{\epsilon u} \\ -1 & \end{bmatrix} (\sigma^2 - \Sigma_{\epsilon u}' \Sigma_u^{-1} \Sigma_{\epsilon u})^{-1} [\Sigma_{\epsilon u}' \Sigma_u^{-1} \quad -1]$$

we have

$$[I_{p_2} \ \beta_2] \Sigma^{-1} [I_{p_2} \ \beta_2]' = \Sigma_u^{-1} + (\sigma^2 - \Sigma_{\epsilon u}' \Sigma_u^{-1} \Sigma_{\epsilon u})^{-1} (\Sigma_u^{-1} \Sigma_{\epsilon u} - \beta_2) (\Sigma_u^{-1} \Sigma_{\epsilon u} - \beta_2)'$$

$$\gamma' [I_{p_2} \ \beta_2] \Sigma^{-1} [I_{p_2} \ \beta_2]' \gamma \geq \gamma' \Sigma_u^{-1} \gamma \quad (7.2)$$

Equation (7.1) now follows immediately from (7.2) and Graybill (1969), Theorem 12.2.14 (5).

Proof of Theorem 4 Recall that

$$\hat{\beta} = (C_*' C_*)^{-1} (C_*' Y).$$

It is easily shown that

$$N^{-1} C_*' C_* \xrightarrow{P} \Delta + \Sigma_{u^*}, \quad N^{-1} C_*' Y \xrightarrow{P} \Delta\beta + \Sigma_{\epsilon u^*}$$

so

$$\hat{\beta}_L \xrightarrow{P} (\Delta + \Sigma_{u^*})^{-1} (\Delta\beta + \Sigma_{\epsilon u^*}) = \beta_L, \text{ say.}$$

Let

$$\hat{\beta}_L = [\hat{\beta}_{1L}' \hat{\beta}_{2L}']', \quad \beta_L = [\beta_{1L}' \beta_{2L}']';$$

we have

$$\hat{\beta}_{iL} \xrightarrow{P} \beta_{iL}, \quad i = 1, 2 \quad (7.3)$$

Thus for all  $\gamma \in \mathbb{R}^p$ ,

$$\begin{aligned} \gamma' \hat{\beta}_L &= \gamma_1' \hat{\beta}_{1L} + \gamma' \hat{\beta}_{2L} \\ &= \gamma_1' \beta_1 + \gamma_2' \beta_2 + (\gamma_1' (X_1' X_1)^{-1} X_1' X_2 - \gamma_2') (\beta_2 - \hat{\beta}_{2L}) + \gamma_1' (X_1' X_1)^{-1} X_1' (\epsilon - U \hat{\beta}_{2L}) \end{aligned}$$

so if  $\gamma$  satisfies (5.2), using (7.3) we obtain

$$N^{1/2} \gamma' (\hat{\beta}_L - \beta) = N^{1/2} \gamma_1' (X_1' X_1)^{-1} X_1' (\epsilon - U \beta_{2L}) + o_p(1).$$

It follows that  $N^{1/2} \gamma' (\hat{\beta}_L - \beta)$  is asymptotically normal, since the elements of  $\epsilon - U \beta_{2L}$  are i.i.d. with zero mean and finite variance, and the elements of  $\gamma_1' (X_1' X_1)^{-1} X_1'$  satisfy the Noether condition. Also,

$$\begin{aligned} \text{Var} (N^{1/2} \gamma_1' (X_1' X_1)^{-1} X_1' (\epsilon - U \beta_{2L})) \\ = N \gamma_1' (X_1' X_1)^{-1} \gamma + d_L \gamma_1' \Delta_{11}^{-1} \gamma_1 \end{aligned}$$

with  $d_L = [\beta_{2L}' \ -1] \Sigma [\beta_{2L}' \ -1]'$ .

Noting that for  $\gamma$  satisfying (5.2)

$$\gamma' \Delta^{-1} \gamma = \gamma_1' \Delta_{11}^{-1} \gamma_1,$$

we obtain

$$N^{1/2} \gamma' (\hat{\beta}_L - \beta) \stackrel{D}{\rightarrow} N(0, d_L \gamma' \Delta^{-1} \gamma).$$

Now for  $\gamma$  satisfying (5.2) the limit variance of  $\hat{\beta}_M$  in (3.1) becomes  $d \gamma' \Delta^{-1} \gamma$ , so we need only show that  $d_L \leq d$ . With

$$M = \Delta_{22} - \Delta_{21} \Delta_{11}^{-1} \Delta_{12}$$

$$q = (M + \Sigma_U)^{-1} [\Sigma_U \Sigma_{\epsilon U}] [\beta_2' \quad -1]'$$

we can, after some algebra show that

$$d - d_L = q' (2M + \Sigma_U) q$$

which is non-negative for all  $q$  since  $M$  and  $\Sigma_U$  are positive definite;  $d = d_L$  only if  $q = 0$ , that is, only if  $\Sigma_{\epsilon U} = \Sigma_U \beta_2$ , completing the proof.

TABLE 1

$\delta$ AND V ERROR DISTRIBUTION	$\sigma_{\delta} = 4$ ( $\sigma_{\delta} = 4$ THROUGHOUT)	U ERROR DISTRIBUTION	$\sigma_u$	MSE EFFICIENCIES FOR INTERCEPT				MSE EFFICIENCIES FOR SLOPE		
				LSE	MOMENTS	CARROLL/ GALLO	% FULLER MODIFICATION	LSE	MOMENTS	CARROLL/ GALLO
N	0	N	1	.90	.98	.95	2.8	.89	.98	.95
N	0	N	2	.52	.98	.97	31.2	.52	.98	.98
N	0	N	4	.23	1.02	1.02	50.0	.23	1.03	1.04
N	0	CN	1	.68	.95	1.07	13.4	.69	.95	1.07
N	0	CN	2	.31	.92	1.10	45.0	.31	.91	1.10
N	0	CN	4	.16	.88	1.08	57.8	.16	.89	1.09
CN	0	N	1	.96	.98	.147	00.0	.97	.98	1.44
CN	0	N	2	.63	.94	1.03	06.4	.62	.94	1.02
CN	0	N	4	.27	.96	.95	36.2	.26	.96	.96
CN	0	CN	1	.81	.96	1.54	01.0	.82	.96	1.53
CN	0	CN	2	.37	.84	1.04	19.4	.37	.84	1.05
CN	0	CN	4	.16	.80	.93	47.6	.17	.80	.94
N	4	N	1	1.15	1.19	1.18	00.0	1.17	1.21	1.20
N	4	N	2	.79	1.24	1.22	04.6	.79	1.25	1.24
N	4	N	4	.30	1.20	1.19	34.8	.30	1.21	1.20
N	4	CN	1	1.02	1.17	1.22	00.4	1.01	1.18	1.25
N	4	CN	2	.46	1.07	1.23	22.2	.45	1.06	1.23
N	4	CN	4	.19	1.00	1.22	49.4	.19	.99	1.21
CN	4	N	1	1.28	1.27	2.00	00.0	1.34	1.32	2.01
CN	4	N	2	1.57	1.92	2.33	00.2	.69	2.01	2.40
CN	4	N	4	.71	1.79	1.81	12.2	.74	1.85	1.87
CN	4	CN	1	1.48	1.51	2.20	09.0	1.52	1.55	2.27
CN	4	CN	2	1.14	1.76	2.15	03.2	1.16	1.82	2.23
CN	4	CN	4	.40	1.27	1.46	29.0	.40	1.29	1.49

TABLE 2

S AND V ERROR DISTRIBUTION	$\sigma_s = 4$ ( $\sigma_v$ THROUGHOUT)	U ERROR DISTRIBUTION	$\sigma_u$	95th PERCENTILE OF ABSOLUTE ERRORS - EFFICIENCIES - INTERCEPT			95th PERCENTILE OF ABSOLUTE ERRORS - EFFICIENCIES - SLOPE		
				LSE	MOMENTS	CARROLL/ GALLO	LSE	MOMENTS	CARROLL/ GALLO
N	0	N	1	1.00	.96	1.00	.99	.99	.98
N	0	N	2	.81	1.01	.98	.82	1.00	.99
N	0	N	4	.68	1.08	1.08	.67	1.09	1.09
N	0	CN	1	.83	1.01	1.07	.79	.97	1.03
N	0	CN	2	.64	.98	1.08	.60	.94	1.03
N	0	CN	4	.53	.96	1.02	.52	.90	.98
CN	0	N	1	.97	1.00	1.20	1.02	.96	1.26
CN	0	N	2	.87	.96	1.04	.86	.97	1.01
CN	0	N	4	.69	.97	.96	.68	.97	.96
CN	0	CN	1	.92	1.00	1.30	.92	1.02	1.28
CN	0	CN	2	.65	.94	1.00	.68	.93	1.02
CN	0	CN	4	.58	.92	1.00	.55	.89	1.00
N	4	N	1	1.00	1.15	1.14	1.09	1.11	1.09
N	4	N	2	.95	1.11	1.13	.97	1.14	1.13
N	4	N	4	.78	1.14	1.14	.79	1.17	1.14
N	4	CN	1	1.04	1.12	1.11	1.06	1.14	1.14
N	4	CN	2	.73	1.07	1.11	.76	1.06	1.15
N	4	CN	4	.59	1.00	1.11	.62	1.06	1.16
CN	4	N	1	1.09	1.09	1.45	1.17	1.14	1.42
CN	4	N	2	1.34	1.37	1.53	1.34	1.39	1.51
CN	4	N	4	1.06	1.32	1.31	1.09	1.34	1.35
CN	4	CN	1	1.20	1.19	1.48	1.25	1.20	1.53
CN	4	CN	2	1.13	1.34	1.50	1.13	1.33	1.46
CN	4	CN	4	.82	1.19	1.22	.83	1.12	1.25



TABLE 3

Efficiencies of the LSE Relative to the Normality-MLE  
for the ANOCOVA Experiment

X,u Measurement Error Distribution	$\sigma_u$	MSE Efficiency, Treatment Effect	MSE Efficiency, Covariable Effect
N	1	1.02	0.80
N	2	1.07	0.36
N	4	1.34	0.15
CN	1	1.03	0.65
CN	2	1.16	0.29
CN	4	1.78	0.21

## References

- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficient for multivariate normal distributions. Ann. Math. Statist. 22 327-351.
- Blomquist, N. (1977). On the relation between change and initial value. J. Amer. Statist. Assoc. 72 746-749.
- Brown, M. (1982). Robust line estimation with errors in both variables. J. Amer. Statist. Assoc. 77 71-79.
- Carroll, R. J. and Gallo, P. P. (1982). Some aspects of robustness in the functional errors-in-variables regression model. To appear in Comm. in Statist., Series A.
- DeGracie, J. S. and Fuller, W. A. Estimation of the Slope and Analysis of Covariance when the Concomitant Variable is Measured with Error. J. Amer. Statist. Assoc. 67 930-937.
- Fuller, W. A. (1975). Regression analysis for sample survey. Sankhya, Series C. 37 117-132.
- Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. Ann. Statist. 8 407-422.
- Gallo, P. P. (1982a). Properties of estimators in errors-in-variables regression models. Ph.D. dissertation, University of North Carolina at Chapel Hill.
- Gallo, P. P. (1982b). Consistency of regression estimates when some variables are subject to error. Comm. in Statist., Series A. 11 973-983.
- Gleser, L. J. (1981). Estimation in a multivariate "errors in variables" regression models: large sample results. Ann. Statist. 9 24-44.
- Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Co., Belmont, Cal.
- Healy, J. D. (1975). Estimation and tests for unknown linear restrictions in multivariate linear models. Ph.D. dissertation, Purdue University.

- Healy, J. D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship. J. Multivariate Anal. 10 243-251
- Jobson, J. D. and Fuller, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. J. Amer. Statist. Assoc. 75 176-181.
- Nussbaum, M. (1978). Asymptotic efficiency of estimators of a multivariate linear functional relation. Akademie der Wissenschaften der DDR, Berlin.
- Reilly, P. M. and Patino-Leal, H. (1981). A Bayesian study of the errors-in-variables model. Technometrics 23 221-231.
- Wolter, K. M. and Fuller, W. A. (1982 a). Estimation of the quadratic errors-in-variables model. Biometrika 69 175-182.
- Wolter, K. M. and Fuller, W. A. (1982 b). Estimation of non-linear errors-in-variables models. Ann. Statist. 10 539-548.
- Wu, M., Ware, J. H. and Feinlieb, M. (1980). On the relation between blood pressure change and initial value. Jour. of Chronic Diseases 33 637-644.