THE LIMITING DISTRIBUTION OF LEAST SQUARES IN AN ERRORS-IN-VARIABLES LINEAR REGRESSION MODEL

by

Leon Jay Gleser[1], Purdue University

Raymond J. Carroll[2], University of North Carolina

Paul P. Gallo
Lederle Laboratories

Technical Report #85-4

Department of Statistics
Purdue University

Revised: April, 1985

THE LIMITING DISTRIBUTION OF LEAST SQUARES
IN AN ERRORS-IN-VARIABLES LINEAR REGRESSION MODEL

BY LEON JAY GLESER[1], RAYMOND J. CARROLL[2], AND PAUL P. GALLO
Purdue University, University of North Carolina, and Lederle Laboratories

It is well-known that the ordinary least squares (OLS) estimator $\hat{\beta}$ of the slope and intercept parameters β in a linear regression model with errors of measurement for some of the independent variables (predictors) is inconsistent. However, Gallo (1982) has shown that certain linear combinations of β are consistently estimated by the corresponding linear combinations of $\hat{\beta}$. In this paper, it is shown that under reasonable regularity conditions such linear combinations are (jointly) asymptotically normally distributed. Some methodological consequences of our results are given in a companion paper (Carroll, Gallo and Gleser, 1985).

[1]Research of Leon Jay Gleser supported by NSF Grant DMS-8121948.

[2]Research of Raymond J. Carroll supported by Air Force Office of Scientific Research Contract AFOSR-F-49620-82-C-009 and sponsored by the United States Army under Contract No. DAAG 29-80-C-0041.

AMS (1980) subject classifications. Primary 60F05, 62J05; secondary 62F10, 62H99.

Key words and phrases. Regression, functional models, structural models, instrumental variables, ordinary least squares estimators, consistency, asymptotic normality.

1. Introduction. There is a substantial literature concerning linear regression when some of the predictors (independent variables) are measured with error. Such models are of importance in econometrics (instrumental variables models), psychometrics (correction for attenuation, models of change), and in instrumental calibration studies in medicine and industry. Recent theoretical work concerning maximum likelihood estimation in such models appears in Healy (1980), Fuller (1980), and Anderson (1984), while Reilly and Patino-Leal (1981) take a Bayesian approach.

In an applied context, an investigator may either overlook the measurement errors in the predictors, or choose the classical ordinary least squares (OLS) estimator of the parameters because of its familiarity and ease of use. Certainly, the methodology of classical least squares theory (confidence intervals, multiple comparisons, tests of hypotheses, residual analysis) is considerably more developed than the corresponding errors-in-variables methodology, particularly in samples of moderate size. If the OLS estimator is used, what are the consequences?

Cochran (1968) has given a general discussion of the consequences of using the OLS estimator in errors-in-variables contexts. For the special case of the analysis of covariance (ANCOVA), where the covariates are measured with error, detailed investigations have been done by Lord (1960), De Gracie and Fuller (1972) and Cronbach (1976). It is by now well-known that the OLS estimator $\hat{\beta}$ of the slope and intercept parameters β in such errors-in-variables models is inconsistent; that is, $\hat{\beta}$ does not tend in probability to β as the sample size n becomes infinitely large. However, in ANCOVA with covariates measured with error but balanced (in terms of means) across the design, the OLS estimator of the design effects is known to be consistent. This is shown in the two-treatment case by Cochran (1968) and DeGracie and Fuller (1980).

More generally, Gallo (1982) has shown that for general linear errors-in-variables regression models, certain linear combinations $c'\hat{\beta}$ of the OLS estimator are consistent estimators of the corresponding linear combinations of β. Gallo's result is reproduced in Section 2 as Theorem 1.

Let the rows of C be a basis for all linear combinations $c'\beta$ of β that are consistently estimated by $c'\hat{\beta}$. In the present paper, it is shown that under a reasonable extension of the regularity conditions given by Gallo (1982), $n^{\frac{1}{2}}(C\hat{\beta}-C\beta)$ has a limiting asymptotic multivariate normal distribution (Theorem 2 of Section 2). This result does not require that the random errors (errors of measurement, residual errors) are normally distributed, but only that these errors are sampled from a common population with finite second moments. However, Theorem 2 does assume that all predictors are fixed. In Section 3, Theorem 2 is extended to cases where some of the predictors are random variables.

The nature of the limiting normal distribution of $n^{\frac{1}{2}}(C\hat{\beta}-C\beta)$ depends upon whether the predictors measured with error are random (<u>structural</u> errors-in-variables models) or fixed (<u>functional</u> errors-in-variables models). In the former case, the limiting normal distribution has a zero mean vector, while in the latter case the mean vector need not be zero (and is a function of unknown parameters). A companion paper (Carroll, Gallo and Gleser, 1985) uses these results to compare the asymptotic efficiencies of the OLS and maximum likelihood estimators of $C\beta$ when the errors-in-variables model is of the structural kind.

2. <u>Asymptotic Theory</u>. Suppose that a dependent scalar variable y_i is related to a vector f_{1i}: p×1 of observable predictors and a vector f_{2i}: q×1 of latent (unobservable) predictors by the model

$$(2.1)\qquad y_i = f_{1i}'\beta_1 + f_{2i}'\beta_2 + e_i, \quad i = 1,2,\ldots,n,$$

and that f_{2i} is observed with error by x_i, where

$$(2.2)\qquad x_i = f_{2i} + u_i, \quad i = 1,2,\ldots,n.$$

For <u>fixed</u> (f_{1i}', f_{2i}') it is assumed that

$$(2.3)\qquad \begin{pmatrix} e_i \\ u_i \end{pmatrix}, \quad 1 \leq i \leq n, \text{ are i.i.d.}$$

with mean vector 0 and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12}' & \Sigma_{22} \end{pmatrix} \quad ; \quad \Sigma_{22}: \; q\times q.$$

To state the model in vector-matrix form, let

$$Y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \quad F_1 = \begin{pmatrix} f_{11}' \\ \cdot \\ \cdot \\ \cdot \\ f_{1n}' \end{pmatrix}, \quad F_2 = \begin{pmatrix} f_{21}' \\ \cdot \\ \cdot \\ \cdot \\ f_{2n}' \end{pmatrix}, \quad X = \begin{pmatrix} x_1' \\ \cdot \\ \cdot \\ \cdot \\ x_n' \end{pmatrix},$$

$$e = \begin{pmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix}, \quad U = \begin{pmatrix} u_1' \\ \cdot \\ \cdot \\ \cdot \\ u_n' \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Then

$$Y = F_1\beta_1 + F_2\beta_2 + e, \quad X = F_2 + U, \tag{2.4}$$

where the rows of $E = (e,U)$ are i.i.d. random vectors with mean vector 0 and covariance matrix Σ.

Note. It is assumed that all design (dummy) variables are included in F_1. This eliminates the need for separately including an intercept term in the model.

The OLS estimator of β for the model (2.4) is

$$\hat{\beta} = \begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix}^{-1} \begin{pmatrix} F_1'Y \\ X'Y \end{pmatrix} . \tag{2.5}$$

2.1 Asymptotic Consistency. To give asymptotic results about $\hat{\beta}$, we need to make some assumptions about the sequence

$$\underset{\sim}{f} = \{(f_{1i}', f_{2i}'): \; i = 1,2,\ldots\} \tag{2.6}$$

of fixed predictor values. These are the following.

Assumption 1.

$$\lim_{n\to\infty} n^{-1} \begin{bmatrix} F_1'F_1 & F_1'F_2 \\ F_2'F_1 & F_2'F_2 \end{bmatrix} = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{12}' & \Delta_{22} \end{bmatrix} \equiv \Delta, \quad \Delta > 0.$$

Assumption 2.

$$\lim_{n\to\infty} n^{-\frac{1}{2}} \max\, [F_1, F_2] = 0,$$

where for any matrix $A = ((a_{ij}))$, $\max\,(A) = \max |a_{ij}|$.

We will make extensive use of the following results.

Lemma 1. Under (2.4) and Assumptions 1 and 2, for all (q+1)-dimensional column vectors t,

$$n^{-\frac{1}{2}}\ (F_1, F_2)'\ (e,U)t \ \to\ MVN(0,(t'\Sigma t)\Delta)$$

in distribution as $n\to\infty$. In particular,

$$n^{-\frac{1}{2}}\,(f_1, F_2)'\ (e-U\beta_2) \to MVN\left(0,\left[(1,-\beta_2')\Sigma\begin{pmatrix}1\\ -\beta_2\end{pmatrix}\right]\Delta\right) \tag{2.7}$$

in distribution as $n\to\infty$.

Proof. This is a direct consequence of Corollary 3.2 and the discussion following in Gleser (1965). □

Lemma 2. Under the assumptions of Lemma 1,

$$n^{-1}\begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix} = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{12}' & \Delta_{22} + \Sigma_{22} \end{pmatrix} + o_p(1).$$

Proof. From the weak law of large numbers,

$$n^{-1}(e,U)'(e,U) = \Sigma + o_p(1)\ , \tag{2.8}$$

while from Lemma 1,

$$n^{-1} F_2'U = 0_p(n^{-\frac{1}{2}}).$$

From these facts, (2.4) and Assumption 1, the assertion of the lemma directly follows. □

The following theorem is a restatement of the result of Gallo (1982) mentioned in Section 1.

Theorem 1 (Gallo, 1982). Under (2.4) and Assumptions 1 and 2,

$$c'\hat{\beta} \xrightarrow{p} c'\beta \quad \Leftrightarrow \quad c'\begin{pmatrix} -\Delta_{11}^{-1}\Delta_{12} \\ I_q \end{pmatrix} = 0,$$

where I_q is the q-dimensional identity matrix.

Proof. Note from (2.4) that

$$\frac{1}{n}\left[\begin{pmatrix} F_1'Y \\ F_2'Y \end{pmatrix} - \begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}\right]$$

$$= \frac{1}{n}\begin{bmatrix} F_1'(e-U\beta_2) \\ F_2'(e-U\beta_2) + U'(e-U\beta_2) \end{bmatrix}.$$

However, Lemma 1 implies that

$$\frac{1}{n}\begin{pmatrix} F_1' \\ F_2' \end{pmatrix}(e-U\beta_2) = 0_p(n^{-\frac{1}{2}}),$$

while it follows from (2.8) that

$$\frac{1}{n}U'(e-U\beta_2) = \sigma_{12}' - \Sigma_{22}\beta_2 + o_p(1).$$

From these facts, (2.5) and Lemma 2 it follows that

$$(2.9) \qquad \hat{\beta} = \beta + \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{12}' & \Delta_{22} + \Sigma_{22} \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ \sigma_{12}' - \Sigma_{22}\beta_2 \end{pmatrix} + o_p(1).$$

Let

$$Q = (\Sigma_{22} + \Delta_{22.1})^{-1}, \qquad \Delta_{22.1} = \Delta_{22} - \Delta_{12}'\Delta_{11}^{-1}\Delta_{12}.$$

Then

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{12}' & \Delta_{22} + \Sigma_{22} \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ I_q \end{pmatrix} = \begin{bmatrix} -\Delta_{11}^{-1}\Delta_{12} \\ I_q \end{bmatrix} Q$$

and it follows from (2.9) that

$$(2.10) \qquad c'\hat{\beta} \xrightarrow{p} c'\beta + c' \begin{bmatrix} -\Delta_{11}^{-1} \Delta_{12} \\ I_q \end{bmatrix} Q\,(\sigma_{12}' - \Sigma_{22}\beta_2).$$

Thus,

$$c'\hat{\beta} \xrightarrow{p} c'\beta \Leftrightarrow c' \begin{bmatrix} \Delta_{11}^{-1} \Delta_{12} \\ I_q \end{bmatrix} Q(\sigma_{12}' - \Sigma_{22}\beta) = 0, \quad \text{all } \beta,\ \Sigma.$$

Clearly

$$c' \begin{bmatrix} -\Delta_{11}^{-1} \Delta_{12} \\ I_q \end{bmatrix} = 0 \Rightarrow c' \begin{bmatrix} -\Delta_{11}^{-1} \Delta_{12} \\ I_q \end{bmatrix} Q(\sigma_{12}' - \Sigma_{22}\beta) = 0$$

for all β, Σ. On the other hand, if

$$\beta_2 = \Sigma_{22}^{-1}\sigma_{22}' - \Sigma_{22}^{-1}(-\Delta_{12}'\Delta_{11}^{-1}, I_q)c,$$

then

$$0 = c' \begin{bmatrix} -\Delta_{11}^{-1} & \Delta_{12} \\ I_q & \end{bmatrix} Q(\sigma_{12}' - \Sigma_{22}\beta) \Rightarrow c' \begin{pmatrix} -\Delta_{11}^{-1} & \Delta_{12} \\ I_q & \end{pmatrix} Q \begin{pmatrix} -\Delta_{11}^{-1} & \Delta_{12} \\ I_q & \end{pmatrix}' c = 0$$

$$\Rightarrow c' \begin{pmatrix} -\Delta_{11}^{-1} \Delta_{12} \\ I_q \end{pmatrix} = 0,$$

since $Q > 0$. This completes the proof. □

Note that

$$c' \begin{bmatrix} -\Delta_{11}^{-1} & \Delta_{12} \\ I_q & \end{bmatrix} = 0 \Leftrightarrow c = d' [I_p, \Delta_{11}^{-1} \Delta_{12}], \quad \text{some } d.$$

From this fact, it is easily seen that the rows of

$$C = (I_p, \Delta_{11}^{-1} \Delta_{12})$$

serve as a basis for the linear manifold of all c such that $c'\beta$ is consistent for $c'\hat{\beta}$. This motivates consideration of the limiting distribution of

$$T_n = n^{\frac{1}{2}} C(\hat{\beta}-\beta).$$

2.2 Asymptotic Normality of T_n. Rather than state our main result (Theorem 2) at once, we first derive a representation for T_n that leads us to the extra assumption needed to obtain asymptotic normality of T_n.

Let

$$(L_{1n}, L_{2n}) = C \left[\frac{1}{n} \begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix} \right]^{-1}$$

and

$$\begin{pmatrix} W_{1n} \\ W_{2n} \end{pmatrix} = \frac{1}{n} \left[\begin{pmatrix} F_1'Y \\ X'Y \end{pmatrix} - \begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix} \left(\beta + \begin{pmatrix} -\Delta_{11}^{-1} \Delta_{12} \\ I_q \end{pmatrix} \gamma \right) \right]$$

where

$$\gamma = Q(\sigma_{12}' - \Sigma_{22}\,\beta_2).$$

Since

$$C\begin{pmatrix} -\Delta_{11}^{-1}\,\Delta_{12} \\ I_q \end{pmatrix} = 0,$$

it follows from (2.5) that

$$T_n = n^{\frac{1}{2}}\,(L_{1n}, L_{2n})\begin{pmatrix} W_{1n} \\ W_{2n} \end{pmatrix}. \tag{2.11}$$

Lemma 3. Under the assumptions of Lemma 1,

$$L_{1n} = \Delta_{11}^{-1} + o_p(1),$$

and

$$G_n = n^{\frac{1}{2}}\,L_{1n}(W_{1n} + \frac{1}{n}\,F_1'(F_2 - F_1\,\Delta_{11}^{-1}\,\Delta_{12})\gamma)$$

$$\longrightarrow \ \mathrm{MVN}(0, \left\{\begin{bmatrix} 1 \\ -(\beta_2+\gamma) \end{bmatrix}' \ \Sigma \ \begin{bmatrix} 1 \\ -(\beta_2+\gamma) \end{bmatrix}\right\} \Delta_{11}^{-1})$$

in distribution as $n\to\infty$.

Proof. The first assertion is a direct consequence of Lemma 2 and the fact that

$$C\begin{pmatrix}\Delta_{11} & \Delta_{12} \\ \Delta_{12}' & \Delta_{22}+\Sigma_{22}\end{pmatrix}^{-1} = (\Delta_{11}^{-1}, 0)$$

Note from (2.4) and the definition of W_{1n} that

$$W_{1n} + \frac{1}{n} F_1'(F_2 - F_1\Delta_{11}^{-1}\Delta_{12})\gamma = \frac{1}{n} F_1'(e,U)\begin{pmatrix}1 \\ -(\beta_2+\gamma)\end{pmatrix}.$$

The second assertion of the lemma now follows from this representation, Lemma 1, the first assertion of the lemma and Slutzky's Theorem. □

<u>Lemma 4</u>. Under the assumptions of Lemma 1,

$$\begin{aligned} W_{2n} &= [\tfrac{1}{n} U'(e-U(\beta_2+\gamma)) - \Delta_{22.1}\gamma] \\ &\quad - [\tfrac{1}{n} F_2'(F_2-F_1\Delta_{11}^{-1}\Delta_{12}) - \Delta_{22.1}]\gamma + O_p(n^{-\frac{1}{2}}) \end{aligned} \tag{2.12}$$

and

$$L_{2n} = -(\tfrac{1}{n} F_1'F_1)^{-1}(\tfrac{1}{n} F_1'(F_2-F_1\Delta_{11}^{-1}\Delta_{12}))[Q^{-1}+o_p(1)]^{-1} + O_p(n^{-\frac{1}{2}}).$$

<u>Proof</u>. Using (2.4) and the definition of W_{2n}, we can write W_{2n} as the sum of the first two terms on the right-hand side of (2.12) plus

$$\frac{1}{n} F_2'(e-U(\beta_2+\gamma)) - \frac{1}{n} U'(F_1,F_2)\begin{pmatrix}-\Delta_{11}^{-1}\Delta_{12} \\ I_p\end{pmatrix}\gamma .$$

Using Lemma 1, this last term can be shown to be $O_p(n^{-\frac{1}{2}})$, as asserted.

From facts about inverses of partitioned matrices, the definitions of C and L_{2n} and (2.4),

$$L_{2n} = -(\tfrac{1}{n} F_1'F_1)^{-1} [\tfrac{1}{n} F_1'(F_2-F_1 \Delta_{11}^{-1} \Delta_{12}) + \tfrac{1}{n} F_1'U] A_n$$

where

$$A_n^{-1} = \tfrac{1}{n}(X'X - X'F_1 (F_1'F_1)^{-1} F_1'X).$$

Using Lemma 2, it is easily shown that

$$A_n^{-1} = \Delta_{22.1} + \Sigma_{22} + o_p(1) = Q^{-1} + o_p(1).$$

Using Lemma 1,

$$n^{-1} F_1'U = O_p(n^{-\frac{1}{2}}).$$

Since $n^{-1} F_1'F_1 = \Delta_{11} + o(1)$ by Assumption 1, the representation for L_{2n} given by the lemma follows from Slutzky's Theorem. □

Using (2.8), Assumption 1 and Lemma 4, it is straightforward to show that $W_{2n} = o_p(1)$. Let

$$Z_n = n^{-\frac{1}{2}} F_1'(F_2-F_1 \Delta_{11}^{-1} \Delta_{12}). \tag{2.13}$$

It follows from (2.11) and Lemmas 3 and 4 that

$$T_n = G_n - (\Delta_{11}^{-1}+o_p(1)) Z_n\gamma - (\Delta_{11}^{-1}+o(1))Z_n[Q^{-1} + o_p(1)]^{-1}(o_p(1)) + o_p(1). \tag{2.14}$$

A careful look at (2.14) shows that for T_n to converge in distribution for all β, Σ it is necessary that Z_n be $O(1)$. Thus, we are led to make the following assumption

Assumption 3. For every sequence $\underset{\sim}{f}$ defined by (2.6),

$$\lim_{n\to\infty} Z_n = \lim_{n\to\infty} n^{-\frac{1}{2}} F_1'(F_2 - F_1\Delta_{11}^{-1}\Delta_{12}) = Z(\underset{\sim}{f})$$

where the limit $Z(\underset{\sim}{f})$ may depend on $\underset{\sim}{f}$.

That Assumption 3, together with Assumptions 1 and 2, is sufficient for T_n to have a limiting multivariate normal distribution is clear from (2.13), Lemma 3 and Slutzky's Theorem. This is our main result.

Theorem 2. Under Assumptions 1, 2 and 3,

$$T_n = n^{\frac{1}{2}}(C\hat{\beta} - C\beta) \to MVN(-\Delta_{11}^{-1} Z(\underset{\sim}{f})\gamma, (\eta'\Sigma\eta)\Delta_{11}^{-1})$$

in distribution as $n\to\infty$, where $C = (I_p, \Delta_{11}^{-1}\Delta_{12})$,

$$\gamma = (\Sigma_{22} + \Delta_{22.1})^{-1}(\sigma_{12}' - \Sigma_{22}\beta_2), \quad \eta' = (1, -(\beta_2+\gamma)').$$

3. Discussion and Extensions. Theorems 1 and 2 assume that the sequence $\underset{\sim}{f}$ defined by (2.6) is a sequence of fixed vectors. If elements of the vectors (f_{1i}', f_{2i}') in this sequence are random variables, one can think of these results as being conditional limit theorems.

When components of each (f_{1i}', f_{2i}'), $i = 1,2,\ldots$, are random, a fairly easy argument can be used to extend Theorems 1 and 2 to apply unconditionally, provided that $\Delta_{11}^{-1} Z_n\gamma$, where $Z_n = Z_n(\underset{\sim}{f})$ is defined by (2.13), has an asymptotic distribution.

Thus, let s_i represent the random part of (f_{1i}', f_{2i}') and let $\underset{\sim}{s} = \{s_i, i=1,2,...\}$. Distributional assumptions about the s_i yield a probability measure $\mu(\underset{\sim}{s})$ over the sequences $\underset{\sim}{s}$. Suppose that

$$A = \{\underset{\sim}{s}: \lim_{n\to\infty} n^{-1}(F_1,F_2)'(F_1,F_2) = \Delta > 0, \qquad \lim_{n\to\infty} n^{-\frac{1}{2}}(F_1,F_2) = 0\}$$

satisfies

$$\int_A d\mu(\underset{\sim}{s}) = 1. \tag{3.1}$$

In other words, Assumptions 1 and 2 are satisfied with probability one. Then Theorem 1 shows that for all $\underset{\sim}{s}$ in A, all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{[\mathrm{tr}(C\hat{\beta}-C\beta)'(C\hat{\beta}-C\beta)]^{\frac{1}{2}} > \varepsilon | \underset{\sim}{s}\} = 0.$$

Thus, by the Lebesgue Dominated Convergence Theorem, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{[\mathrm{tr}(C\hat{\beta}-C\beta)'(C\hat{\beta}-C\beta)]^{\frac{1}{2}} > \varepsilon\} = 0,$$

and hence $C\hat{\beta}$ converges unconditionally in probability to $C\beta$. This shows that Theorem 1 holds unconditionally (over $\underset{\sim}{s}$).

In a similar fashion, it can be shown that the representation (2.14) for T_n holds unconditionally, that G_n in that representation has the limiting multivariate normal distribution described in Lemma 3, and that G_n and Z_n are asymptotically statistically independent. Consequently, if $\Delta_{11}^{-1} Z_n \gamma$ has a limiting distribution, the limiting distribution of T_n is the convolution of the limiting distributions of G_n and $\Delta_{11}^{-1} Z_n\gamma$.

Note: The above discussion is only a sketch of the arguments needed, and skips over such details as measurability. A more extensive discussion in a similar context can be found in Gleser (1983).

We will now follow the steps of the above analysis for some special cases of the model (2.4) which are commonly adopted in practice. Recall that if f_{2i}, $i = 1,2,\ldots$, are random vectors, the model (2.4) is called a structural linear errors-in-variables regression model, while if the f_{2i}, $i = 1,2,\ldots$, are vectors of constants, the model is that of a functional linear errors-in-variables regression model. Mixes of these cases, where some elements of f_{2i} are fixed and some elements are random, are also possible. Further, the elements of f_{1i} (except for the first component, which is always equal to 1 to accomodate an intercept term) can also be fixed or random. Let

$$f_{1i} = \begin{pmatrix} 1 \\ h_i \end{pmatrix} .$$

We will consider the following cases:

(a) both h_i and f_{2i} fixed (functional model),

(b) h_i random, f_{2i} fixed (functional model),

(c) h_i fixed, f_{2i} random (structural model),

(d) both h_i and f_{2i} random (structural model).

3.1 Both h_i and f_{2i} fixed. Theorems 1 and 2 already summarize what we can say about this case. Although Theorem 2 has some technical interest, it is unfortunately rather useless for statistical applications. Unless we are in the unlikely case where we either know the limit $Z(\underset{\sim}{f})$ or can consistently estimate this quantity, we cannot use Theorem 2 to construct large-sample confidence regions for $C\beta$. Recall that $\{f_{2i}, i = 1,2,\ldots\}$ is a sequence of unknown parameters, and that the individual vectors f_{2i} in this sequence cannot be consistently estimated. Thus, very strong assumptions are needed to permit us to consistently estimate $Z(\underset{\sim}{f})$ (or $\Delta_{11}^{-1} Z(\underset{\sim}{f})\gamma$).

3.2. $\underline{h_i}$ random and $\underline{f_{2i}}$ fixed. Here, we can assume that the vectors h_i are mutually statistically independent, but must consider the possibility that the distribution of h_i depends upon f_{2i}, $i = 1,2,\ldots$. (That is, the h_i's are not identically distributed.) Given the linear form of (2.4), it is natural to assume that a similar linear model relates h_i to f_{2i}. Thus, we assume that

$$h_i = \alpha + \psi f_{2i} + t_i, \quad i = 1,2,\ldots \tag{3.2}$$

where the t_i's are i.i.d. with mean vector 0 and covariance matrix Λ. We also assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_{2i} = \mu, \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_{2i}f_{2i}' = \Delta_{22} > 0 \tag{3.3}$$

and that $\lim_{n\to\infty} n^{-\frac{1}{2}} f_{2i} = 0$, all i. By letting $f_{2i} \to f_{2i} - \mu$, $\alpha \to \alpha + \psi\mu$, $\Delta_{22} \to \Delta_{22} - \mu\mu'$, we can let $\mu = 0$ without loss of generality.

The strong law of large numbers shows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} t_i = 0, \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} t_i t_i' = \Lambda$$

with probability one. Using (3.2), (3.3) and Theorem 3 of Chow (1966),

$$\lim_{n\to\infty} n^{-\frac{1}{2}} \sum_{i=1}^{n} t_i f'_{2i} \left(\frac{1}{n} \sum_{j=1}^{n} f_{2j} f'_{2j}\right)^{-\frac{1}{2}} = 0$$

with probability one. Thus (3.1) holds with

$$\Delta = \begin{pmatrix} 1 & \alpha' & 0 \\ \alpha & \alpha\alpha' + \psi\Delta_{22}\psi' + \Lambda & \psi\Delta_{22} \\ 0 & \Delta_{22}\psi' & \Delta_{22} \end{pmatrix}.$$

Note that

$$\Delta_{11}^{-1} \Delta_{12} = \begin{bmatrix} -\alpha' \\ I_{p-1} \end{bmatrix} \quad [\psi\Delta_{22}\psi' + \Lambda]^{-1} \psi\Delta_{22}.$$

Let $1'_n = (1,1,\ldots 1)$ and $T' = (t_1,\ldots,t_n)$. Then

$$\begin{aligned} Z_n &= n^{-\frac{1}{2}} F'_1 (F_2 - F_1 \Delta_{11}^{-1} \Delta_{12}) \\ &= n^{-\frac{1}{2}} \begin{pmatrix} 1'_n \\ \alpha 1'_n + \psi F'_2 + T' \end{pmatrix} (F_2 \Gamma - T\Omega) \end{aligned}$$

where

$$\Gamma = I_q - \psi'\Omega, \qquad \Omega = [\psi\Delta_{22}\psi' + \Lambda]^{-1} \psi\Delta_{22}.$$

It is apparent that, in general, extra conditions on both F_2 and the higher order moments of the common distribution of the t_i's are needed to permit Z_n to have a limiting distribution.

However, consider the special case $\psi = 0$. In this case the random parts h_i of f_{1i} are i.i.d. random vectors independent of the f_{2i}'s, and

$$\Delta_{11}^{-1} Z_n\gamma = n^{-\frac{1}{2}} \Delta_{11}^{-1} \begin{pmatrix} 1_n'F_2 \\ \alpha 1_n'F_2 + T'F_2 \end{pmatrix}\gamma = n^{-\frac{1}{2}} \begin{pmatrix} 1_n'F_2\gamma \\ \Lambda^{-1}T'F_2\gamma \end{pmatrix}$$

Using Corollary 3.2 and the discussion following in Gleser (1965), it can be shown that the elements of $n^{-\frac{1}{2}} T'F_2\gamma$ have an asymptotic multivariate normal distribution:

$$n^{-\frac{1}{2}} T'F_2\gamma \to MVN(0,(\gamma'\Delta_{22}\gamma)\Lambda).$$

Although we could impose the condition that $n^{-\frac{1}{2}} 1_n'F_2\gamma = O(1)$, this is a rather restrictive condition, and still leaves us the problem of estimating the limit of $n^{-\frac{1}{2}} 1_n'F_2\gamma$ in statistical applications. Instead, we settle for a more restricted result:

$$(3.4) \qquad n^{\frac{1}{2}} (0,I_{p-1})(C\hat{\beta}-C\beta) \to MVN(0,\dot{\Theta})$$

in distribution as $n\to\infty$, where

$$\Theta = \begin{pmatrix} 1 \\ -(\beta_2+\gamma) \end{pmatrix}' \Sigma \begin{pmatrix} 1 \\ -(\beta_2+\gamma) \end{pmatrix} (0,I_{p-1})\Delta_{11}^{-1} \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix}$$

$$+ (\gamma'\Delta_{22}\gamma)\Lambda^{-1}$$

$$= \Lambda^{-1} \left[\begin{pmatrix} 1 \\ -(\beta_2+\gamma) \end{pmatrix}' \Sigma \begin{pmatrix} 1 \\ -(\beta_2+\gamma) \end{pmatrix} + \gamma'\Delta_{22}\gamma \right],$$

since $\Lambda^{-1} = (0,I_{p-1})\Delta_{11}^{-1}(0,I_{p-1})'$. In this context ($\psi$=0), it is worth noting that

$$(0,I_{p-1})\ C = (0,I_{p-1})(I_p,-\Delta_{11}^{-1}\Delta_{12})$$

$$= (0,I_{p-1},0),$$

so that the result concerns the estimates of the slopes $(0,I_{p-1})\beta_1$ of the y_i on the h_i (the random part of f_{1i}) in (2.4).

3.3 $\underline{h_i}$ <u>fixed and</u> $\underline{f_{2i}}$ <u>random</u>. In analogy with the discussion in Section 3.2, we assume that

$$f_{2i} = \psi f_{1i} + t_i, \qquad i = 1,2,..., \tag{3.5}$$

where the t_i are i.i.d. with common mean vector 0 and covariance matrix Λ. (Here, since the first element of f_{1i} is always 1, there is no need for a separate intercept term.) Assumption (3.5) is commonly adopted in instrumental variables approaches to errors in variables models in econometrics, and in ANCOVA with measurement errors in the covariates.

We also assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_{1i}f_{1i}' = \Delta_{11} > 0 \tag{3.6}$$

and that $\lim_{n\to\infty} n^{-\frac{1}{2}} f_{1i} = 0$, all i. Following steps similar to those used in Section 3.2, we can show that (3.1) holds with

$$\Delta = \begin{pmatrix} \Delta_{11} & \Delta_{11}\psi' \\ \psi\Delta_{11} & \psi\Delta_{11}\psi'+\Lambda \end{pmatrix} .$$

Hence,

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

Note that

$$Z_n = n^{-\frac{1}{2}} F_1'(F_2-F_1\Delta_{11}^{-1}\Delta_{12}) = n^{-\frac{1}{2}} F_1'T,$$

where $T' = (t_1,\ldots,t_n)$. Applying Corollary 3.2 and the following discussion in Gleser (1965),

$$\Delta_{11}^{-1}Z_n\gamma \rightarrow MVN(0,\Delta_{11}^{-1}(\gamma'\Lambda\gamma))$$

in distribution as $n\to\infty$. Consequently,

$$n^{\frac{1}{2}} (C\hat{\beta}-C\beta) \rightarrow MVN(0,\Delta_{11}^{-1} [\eta'\Sigma\eta + \gamma'\Lambda\gamma]) \tag{3.7}$$

in distribution as $n\to\infty$. It is worth noting that here

$$C = (I_p,\psi), \qquad \Lambda = \Delta_{22.1}, \qquad \eta = \begin{pmatrix} 1 \\ -(\beta_2+\gamma) \end{pmatrix}.$$

When $\psi = 0$, there is a close parallel between (3.4) and (3.7). Note also that in this case $C\beta = \beta_1$.

Even when $\psi \neq 0$ (the distribution of f_{2i} depends on f_{1i}), the result (3.7) was obtained without the need to make extra assumptions on the higher moments of the common distribution of the t_i, in contrast to our conclusions in the case of Section 3.2.

3.4 Both $\underline{h}_i$ and $\underline{f}_{2i}$ random. In this case it is more natural to make assumptions concerning (h_i', f_{2i}'), $i = 1,2,\ldots$. We assume that these vectors are i.i.d. with a common mean vector μ and a common covariance matrix Φ. The strong law of large numbers now shows that (3.1) holds with

$$\Delta = \begin{pmatrix} 1 & \mu' \\ \mu & \Phi+\mu\mu' \end{pmatrix}.$$

Let $\mu' = (\mu_1', \mu_2')$ and

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}' & \Phi_{22} \end{pmatrix}$$

where μ_1, Φ_{11} are the common mean vector and covariance matrix of the h_i's. Thus,

$$\Delta_{11}^{-1}\Delta_{12} = \begin{pmatrix} 1 & \mu_1' \\ \mu_1 & \Phi_{11}+\mu_1\mu_1' \end{pmatrix}^{-1} \begin{pmatrix} \mu_2' \\ \Phi_{12}+\mu_1\mu_2' \end{pmatrix}$$

$$= \begin{pmatrix} \mu_2' - \mu_1' \Phi_{11}^{-1} \Phi_{12} \\ \Phi_{11}^{-1} \Phi_{12} \end{pmatrix}.$$

Let $H' = (h_1, h_2, \ldots, h_n)$. Then

$$Z_n = n^{-\frac{1}{2}} \begin{pmatrix} 1_n' \\ H' \end{pmatrix} (F_2 - 1_n(\mu_2' - \mu_1' \Phi_{11}^{-1} \Phi_{12}) - H \Phi_{11}^{-1} \Phi_{12}) \, .$$

The Central Limit Theorem shows that the first row of Z_n has an asymptotic multivariate normal distribution. For the remaining rows of Z_n to be asymptotically multivariate normally distributed, additional assumptions on the higher moments of the joint distribution of (h_i', f_{2i}') are needed. To avoid such assumptions, we can assume that

$$\text{(3.8)} \qquad f_{2i} = \mu_2 - \Phi_{12}' \Phi_{11}^{-1} \mu_1 + \Phi_{12}' \Phi_{11}^{-1} h_i + t_i, \; i = 1, 2, \ldots,$$

where the t_i's are i.i.d. with mean vector 0 and covariance matrix

$$\Phi_{22.1} = \Phi_{22} - \Phi_{12}' \, \Phi_{11}^{-1} \, \Phi_{12}$$

and statistically independent of the h_i's. If we condition on the h_i's, (3.8) is the model (3.5) with

$$\psi = (\mu_2 - \Phi_{12}' \Phi_{11}^{-1} \mu_1, \;\; \Phi_{12}' \Phi_{11}^{-1}), \;\; \Lambda = \Phi_{22.1} .$$

We can now use the results of Section 3.2, noting that with probability one (over sequences $h_1, h_2, \ldots$)

$$\lim_{n\to\infty} \frac{1}{n} F_1'F_1 = \lim_{n\to\infty} \frac{1}{n}(1_n,H)'\,(1_n,H)$$

$$= \begin{pmatrix} 1 & \mu_1' \\ \mu_1 & \Phi_{11}+\mu_1\mu_1' \end{pmatrix} = \Delta_{11}.$$

Thus, conditional on the h_i's,

(3.9) $$n^{\frac{1}{2}}\,(C\hat{\beta}-C\beta) \to MVN(0,\Delta_{11}^{-1}[\eta'\Sigma\eta+\gamma'\Phi_{22.1}\gamma])$$

in distribution as $n\to\infty$. Using the arguments given at the beginning of this section about converting conditional limiting results to unconditional limiting results, we conclude that (3.9) also holds unconditionally.

3.5 Conclusion. The results (3.4), (3.7), (3.9) can be used to construct large sample confidence ellipsoids for $C\beta$ based on the OLS estimator $C\hat{\beta}$ provided that consistent estimators can be found for the covariance matrices of the asymptotic normal distributions. It should be noted that in general $C\beta$ is a function not only of β, but also of $\Delta_{11}^{-1}\Delta_{12}$, which need not be a known matrix.

REFERENCES

ANDERSON, T. W. (1984). Estimating linear statistical relationships. Ann. Statist. 12, 1-45.

CARROLL, R. J., GALLO, P. and GLESER, L. J. (1985). Comparisons of least squares and errors-in-variables regression with special reference to randomized analysis of covariance. J. Amer. Statist. Assoc. To appear.

CHOW, Y. S. (1966). Some convergence theorems for independent random variables. Ann. Math. Statist. 37, 1482-1493.

CRONBACH, L. J. (1976). On the design of educational measures. In Advancement in Psychological and Educational Measurement, ed. D.N.M. DeGruijter and L. J. Th. van der Kamp, John Wiley and Sons, New York.

DeGRACIE, J. S. and FULLER, W. A. (1980). 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. (1980). Properties of some estimators for the errors-in-variables model. Ann. Statist. 8, 407-422.

GALLO, P. P. (1982). Properties of estimators in errors-in-variables regression models. Ph.D. dissertation, University of North Carolina at Chapel Hill.

GLESER, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. Ann. Math. Statist. 36, 463-467.

GLESER, L. J. (1983). Functional, structural and ultrastructural errors-in-variables models. In 1983 Proceedings Bus. and Econ. Statist. Sect., American Statistical Association, Washington, D.C., 57-66.

HEALY, J. D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship. J. Multivariate Anal. 10, 243-251.

LORD, F. M. (1960). Large-sample covariance analysis when the control variable is fallible. J. Amer. Statist. Assoc. 55, 307-321.

REILLY, P. M. and PATINO-LEAL, H. (1981). A Bayesian study of the errors-in-variables model. Technometrics 23, 221-231.

LEON JAY GLESER
Department of Statistics
Purdue University
West Lafayette, Indiana 47907

RAYMOND J. CARROLL
Department of Statistics
The University of North Carolina
321 Phillips Hall 039A
Chapel Hill, North Carolina 27514

PAUL P. GALLO
Group Leader, Preclinical Statistics
Lederle Laboratories
Pearl River, New York 10954