The Annals of Statistics
1987, Vol. 15, No. 1, 220–233

THE LIMITING DISTRIBUTION OF LEAST SQUARES IN AN ERRORS-IN-VARIABLES 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 $\hat{\beta}$ consistently estimate the corresponding linear combinations of β. In this paper, it is shown that under reasonable regularity conditions such linear combinations of $\hat{\beta}$ are (jointly) asymptotically normally distributed. Some methodological consequences of our results are given in a companion paper (Carroll, Gallo and Gleser (1985)).

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 overlook the measurement errors in the predictors, and choose the classical ordinary least squares (OLS) estimator of the parameters because of its familiarity and ease of use. 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), DeGracie and Fuller (1972) and Cronbach (1976). It is by now well-known that the OLS estimator $\hat{\beta}$ of the vector β of the slope and intercept parameters 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 (1972).

Received April 1985; revised April 1986.
[1] Research supported by NSF grant DMS-8121948.
[2] Research supported by Air Force Office of Scientific Research Contract AFOSR-F-49620-85-C-0144 and 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.

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.

The intent of the present paper is to go beyond consistency, and to determine when consistent linear combinations $c'\hat{\beta}$ of the OLS estimator are also asymptotically normally distributed. Numerous papers on this question dealing with various special cases have appeared in the literature (particularly the econometric literature), but no unified approach dealing with models allowing arbitrary collections of fixed and random predictors, measured both with and without error, appears to have been attempted.

The key to a unified approach is to establish asymptotic normality when all predictor variables are fixed. Let the rows of the matrix C be a basis for all linear combinations $c'\beta$ of β that are consistently estimated by $c'\hat{\beta}$. Under a reasonable extension of the regularity conditions given by Gallo (1982), it is shown in Theorem 2 of Section 2 that $n^{1/2}(C\hat{\beta} - C\beta)$ has an asymptotic multivariate normal distribution when all predictors are fixed. This result does not require that the random errors (errors of measurement, residual errors) be normally distributed, but only that these errors be sampled from a common population with finite second moments. In Section 3, Theorem 2 is utilized to extend the asymptotic multivariate normality results for $n^{1/2}(C\hat{\beta} - C\beta)$ to cases where some of the predictors (those measured with error, without error, or both) are random variables.

The nature of the limiting normal distribution of $n^{1/2}(C\hat{\beta} - C\beta)$ depends upon whether the predictors measured with error are random (*structural* errors-in-variables models) or fixed (*functional* 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. Asymptotic theory. Suppose that a dependent scalar variable y_i is related to a p-dimensional column vector f_{1i} of observable predictors and a q-dimensional column vector f_{2i} of latent (unobservable) predictors by the model

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

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

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

Here, β_1 is the p-dimensional column vector containing the unknown slopes of the observable predictors f_{1i}, and β_2 is the q-dimensional column vector of unknown slopes for the latent predictors f_{2i}. For fixed (f_{1i}', f_{2i}'), it is assumed that the random vectors $(e_i, u_i')'$ of errors of measurement are independently and identically distributed (i.i.d.) with mean vector $\mathbf{0}$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12}' & \Sigma_{22} \end{pmatrix},$$

where Σ_{22} is $q \times q$.

To state the model in more compact form, let $Y = (y_1, \dots, y_n)'$, $e = (e_1, \dots, e_n)'$, $F_1 = (f_{11}, \dots, f_{1n})'$, $F_2 = (f_{21}, \dots, f_{2n})'$, $X = (x_1, \dots, x_n)'$, $U = (u_1, \dots, u_n)'$ and

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Then

$$Y = F_1\beta_1 + F_2\beta_2 + e, \qquad X = F_2 + U, \tag{2.3}$$

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.

When Σ is totally unknown, the model (2.3) is not (necessarily) identifiable, and maximum likelihood estimators of β need not exist. However, the OLS estimator,

$$\hat{\beta} = \begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix}^{-1} \begin{pmatrix} F_1'Y \\ X'Y \end{pmatrix}, \tag{2.4}$$

is well defined in such contexts, and for this reason alone might be employed in practice. The results concerning consistency and asymptotic normality of linear combinations of $\hat{\beta}$ obtained in this paper apply whether or not Σ is known (or has known structure), and thus are applicable even in contexts where the maximum likelihood estimator of β does not exist.

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

$$\mathbf{f} = \{(f_{1i}', f_{2i}') : i = 1, 2, \dots\} \tag{2.5}$$

of *fixed* predictor values. These are the following.

ASSUMPTION 1.

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

ASSUMPTION 2.

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

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

We will make extensive use of the following results.

LEMMA 1. *Under* (2.3) *and Assumptions* 1 *and* 2,

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

in distribution as $n \to \infty$ *for all* $(q+1)$*-dimensional column vectors* t. *In particular,*

$$n^{-1/2}(F_1, F_2)'(e - U\beta_2) \to MVN\left(\mathbf{0}, \left[(1, -\beta_2')\Sigma\begin{pmatrix} 1 \\ -\beta_2 \end{pmatrix}\right]\Delta\right) \tag{2.6}$$

in distribution as $n \to \infty$.

PROOF. This is a direct consequence of Corollary 3.2 and the discussion following that corollary 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,

$$n^{-1}(e, U)'(e, U) = \Sigma + o_p(1), \tag{2.7}$$

while from Lemma 1, $n^{-1}(F_1, F_2)'(e, U) = O_p(n^{-1/2})$. From these facts, (2.3) and Assumption 1, the assertion of the lemma follows. □

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

THEOREM 1 (Gallo (1982)). *Under* (2.3) *and Assumptions* 1 *and* 2, *a necessary and sufficient condition for* $c'\hat{\beta}$ *to consistently estimate* $c'\beta$ (*that is,* $c'\hat{\beta} \to c'\beta$ *in probability, all* β, *all* Σ) *is* $c'M = \mathbf{0}$, *where*

$$M = \left(-\Delta_{12}'\Delta_{11}^{-1}, I_q\right)'$$

and I_q *is the* q*-dimensional identity matrix.*

PROOF. Note from (2.3) that

$$\frac{1}{n}\left[\begin{pmatrix} F_1'Y \\ X'Y \end{pmatrix} - \begin{pmatrix} F_1'F_1 & F_1'X \\ X'F_1 & X'X \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}\right] = \frac{1}{n}\begin{bmatrix} F_1' \\ F_2' \end{bmatrix}(e - U\beta_2) + \frac{1}{n}\begin{bmatrix} \mathbf{0} \\ U'(e - U\beta_2) \end{bmatrix}.$$

By Lemma 1, the first term on the right-hand side of the above equation is $O_p(n^{-1/2})$, while it follows from (2.7) that the second term on the right-hand side converges in probability to $(0', \gamma')'$, where

$$\gamma = \sigma_{12}' - \Sigma_{22}\beta_2.$$

From these facts, (2.4) and Lemma 2,

$$\hat{\beta} \to \beta + \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta'_{12} & \Delta_{22} + \Sigma_{22} \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ \gamma \end{pmatrix}, \quad \text{in probability.}$$

Let

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

and note that

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

Consequently, for any $(p+q)$-dimensional column vector c, $c'\hat{\beta} \to c'\beta + c'MQ\gamma$ in probability, and $c'\hat{\beta}$ consistently estimates $c'\beta$ if and only if $c'MQ\gamma = 0$ for all β, all Σ.

Clearly, $c'M = \mathbf{0}$ implies that $c'MQ\gamma = 0$ for all β, Σ, and hence that $c'\hat{\beta}$ consistently estimates $c'\beta$. On the other hand, if $c'\hat{\beta}$ consistently estimates $c'\beta$, then $c'MQ\gamma = 0$ for all β and Σ. In particular, for $\beta_2 = \Sigma_{22}^{-1}(\sigma'_{12} - M'c)$,

$$0 = c'MQ\gamma = c'MQM'c,$$

which, since $Q > 0$, implies that $c'M = \mathbf{0}$. This completes the proof. □

Note that

$$c'M = \mathbf{0} \Leftrightarrow c = d'\left[I_p, \Delta_{11}^{-1}\Delta_{12}\right], \quad \text{some } d.$$

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

$$C = \left[I_p, \Delta_{11}^{-1}\Delta_{12}\right]$$

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

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

Remark. Note that $C\beta$ is the limit of

$$\theta = \beta_1 + F_{11}^{-1}F_{12}\beta_2.$$

When both the error-free predictors f_1 and the error-prone predictors f_2 are normally distributed, θ is recognizable as the vector of slopes (unconditional slopes) of the linear regression of y on f_1 when f_2 is free to vary over its population. [In the special case where f_1 and f_2 are linearly independent ($\Delta_{12} = \mathbf{0}$), $\theta = \beta_1$.] Since f_1 is observable, it is clear from this remark that the OLS estimator for θ should be consistent. It is much less obvious, however, that θ is essentially the *only* linear transform of $\beta = (\beta'_1, \beta'_2)'$ which can be consistently estimated by ordinary least squares in the context of the model (2.3).

There are many situations where model (2.3) applies, and yet where inference concerning θ might be of interest. One such context, the analysis of covariance, has already been mentioned. One can also imagine experimental contexts where predictors known to be measured with error are used in an attempt to reduce variation (and thereby reduce the sample size needed for accurate inference), but where these same predictors would not be used in applications due to their high costs or to difficulties in obtaining measurements. In this case, the slope vector θ (rather than β_1 or $\beta = (\beta_1', \beta_2')'$) would be of primary interest to the investigators, and point and interval estimates of θ might be desired.

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

$$\begin{pmatrix} W_{1n} \\ W_{2n} \end{pmatrix} = \frac{1}{n}\left[\begin{pmatrix} F_1'Y \\ X'Y \end{pmatrix} - \begin{pmatrix} F_1'F_1 & F'X \\ X'F_1 & X'X \end{pmatrix}(\beta + MQ\gamma)\right].$$

Since $CM = \mathbf{0}$, it follows from (2.4) that

$$T_n = n^{1/2}(L_{1n}, L_{2n})\begin{pmatrix} W_{1n} \\ W_{2n} \end{pmatrix}. \tag{2.8}$$

LEMMA 3. *Under the assumptions of Lemma* 1,

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

and

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

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

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 (2.3) and the definition of W_{1n} that

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

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

LEMMA 4. *Under the assumptions of Lemma* 1,

$$W_{2n} = \left[\frac{1}{n}U'(e - U(\beta_2 + Q\gamma)) - \Delta_{22\cdot 1}Q\gamma\right] \tag{2.9}$$
$$- \left[\frac{1}{n}F_2'(F_2 - F_1\Delta_{11}^{-1}\Delta_{12}) - \Delta_{22\cdot 1}\right]Q\gamma + O_p(n^{-1/2})$$

and

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

PROOF. Use (2.3) and the definition of W_{2n} to write W_{2n} as the sum of the first two terms on the right-hand side of (2.9) plus

$$\frac{1}{n}F_2'(e - U(\beta_2 + Q\gamma)) - \frac{1}{n}U'(F_1, F_2)MQ\gamma.$$

It follows from Lemma 1 that this last term is $O_p(n^{-1/2})$, as asserted.

Facts about inverses of partitioned matrices, (2.3), and the definitions of C and L_{2n} yield

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

where

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

Lemma 2 can be applied to show that

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

while from Lemma 1, $n^{-1}F_1'U = O_p(n^{-1/2})$. 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. □

Assumption 1, Lemma 4 and (2.7) can be used to show that $W_{2n} = o_p(1)$. Let

$$Z_n = n^{-1/2}F_1'(F_2 - F_1\Delta_{11}^{-1}\Delta_{12}). \tag{2.10}$$

It follows from (2.8) and Lemmas 3 and 4 that

$$T_n = G_n - \left(\Delta_{11}^{-1} + o_p(1)\right)Z_nQ\gamma \tag{2.11}$$
$$-\left(\Delta_{11}^{-1} + o(1)\right)Z_n\left[Q^{-1} + o_p(1)\right]^{-1}\left(o_p(1)\right) + o_p(1).$$

A careful look at (2.11) 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 $\mathbf{f}$ defined by (2.5),

$$\lim_{n\to\infty} Z_n = \lim_{n\to\infty} n^{-1/2}F_1'\left(F_2 - F_1\Delta_{11}^{-1}\Delta_{12}\right) = Z(\mathbf{f}),$$

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

That Assumption 3, together with Assumptions 1 and 2, is *sufficient* for T_n to have a limiting multivariate normal distribution is clear from (2.11), Lemma 3 and Slutzky's theorem. This is our main result.

THEOREM 2. *Under Assumptions* 1, 2 *and* 3,

$$T_n = n^{1/2}(C\hat{\beta} - C\beta) \to MVN\left(-\Delta_{11}^{-1}Z(\mathbf{f})Q\gamma, (\eta'\Sigma\eta)\Delta_{11}^{-1}\right)$$

in distribution as $n \to \infty$, *where* $C = (I_p, \Delta_{11}^{-1}\Delta_{12})$, $Q = (\Delta_{22\cdot 1} + \Sigma_{22})^{-1}$, *and*

$$\gamma = (\sigma_{12}' - \Sigma_{22}\beta_2), \qquad \eta' = \left(1, -(\beta_2 + Q\gamma)'\right).$$

3. Discussion and extensions. Theorems 1 and 2 assume that the sequence $\mathbf{f}$ defined by (2.5) 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}'), $i = 1, 2, \dots,$ are random, a fairly easy argument can be used to extend Theorems 1 and 2 to apply unconditionally, provided that $\Delta_{11}^{-1}Z_nQ\gamma$, where $Z_n = Z_n(\mathbf{f})$ is defined by (2.10), has an asymptotic distribution.

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

$$A = \left\{\mathbf{s}\colon \lim_{n\to\infty} n^{-1}(F_1, F_2)'(F_1, F_2) = \Delta > 0,\ \lim_{n\to\infty} n^{-1/2}(F_1, F_2) = 0\right\}$$

satisfies

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

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

$$\lim_{n\to\infty} P\left\{\left[\operatorname{tr}(C\hat{\beta} - C\beta)'(C\hat{\beta} - C\beta)\right]^{1/2} > \varepsilon|\mathbf{s}\right\} = 0.$$

Thus, by the Lebesgue dominated convergence theorem, for all $\varepsilon > 0$,

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

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

In a similar fashion, it can be shown that the representation (2.11) 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 $\Delta_{11}^{-1}Z_nQ\gamma$ has a limiting distribution, the limiting distribution of T_n is the convolution of the limiting distributions of G_n and $-\Delta_{11}^{-1}Z_nQ\gamma$.

NOTE. The preceding 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 preceding analysis for some special cases of the model (2.3) that are commonly adopted in practice. Recall that if f_{2i}, $i = 1, 2, \ldots,$ are random vectors, the model (2.3) 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 accommodate 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:

(1) both h_i and f_{2i} are fixed (functional model);
(2) h_i random, f_{2i} fixed (functional model);
(3) h_i fixed, f_{2i} random (structural model);
(4) 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 $Z(\mathbf{f})$ 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 $Z(\mathbf{f})$ (or $\Delta_{11}^{-1}Z(\mathbf{f})Q\gamma$).

3.2. *h_i random and 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 (2.3), it is natural to assume that a similar linear model relates h_i to f_{2i}. Thus, we assume that

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

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

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

and that $\lim_{n\to\infty} n^{-1/2} \max[f_{21}, \ldots, f_{2n}] = 0$. For convenience, we here treat only the case $\mu = \mathbf{0}$. Results for the general case $\mu \neq \mathbf{0}$ can be obtained by replacing α by $\alpha + \psi\mu$, and Δ_{22} by $\Delta_{22} - \mu\mu'$ in the formulas that follow.

The strong law of large numbers shows that

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

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

$$\lim_{n\to\infty} n^{-1/2} \sum_{i=1}^{n} t_i f_{2i}' = \mathbf{0}$$

with probability one. Thus, (3.1) holds with

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

Note that

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

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

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

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

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

An application of Corollary 3.2, and the discussion following, in Gleser (1965) shows that the elements of $n^{-1/2} T' F_2 Q\gamma$ have an asymptotic multivariate normal distribution,

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

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

$$n^{1/2}(\mathbf{0}, I_{p-1})(C\hat{\beta} - C\beta) \to MVN(\mathbf{0}, \Theta), \tag{3.4}$$

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

$$\Theta = \begin{pmatrix} 1 \\ -(\beta_2 + Q\gamma) \end{pmatrix}' \Sigma \begin{pmatrix} 1 \\ -(\beta_2 + Q\gamma) \end{pmatrix} (0, I_{p-1}) \Delta_{11}^{-1} \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix} + (\gamma' Q \Delta_{22} Q \gamma) \Lambda^{-1}$$

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

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 (3.4) 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 (2.3).

3.3. *h_i fixed and f_{2i} random.* In analogy with the discussion in Section 3.2, we assume that

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

where the t_i are i.i.d. with common mean vector $\mathbf{0}$ 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 $\lim_{n\to\infty} n^{-1/2}\max[f_{11}, \ldots, f_{1n}] = \mathbf{0}$. An argument similar to that used in Section 3.2 shows 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

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

where $T' = (t_1, \ldots, t_n)$. Corollary 3.2 in Gleser (1965), and the discussion following that corollary, can be used to show that

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

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

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

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

$$C = (I_p, \psi), \qquad \Lambda = \Delta_{22\cdot 1}, \qquad \eta = \begin{pmatrix} 1 \\ -(\beta_2 + Q\gamma) \end{pmatrix}.$$

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 h_i and 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 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

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

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

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

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

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

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

$$\psi = \left(\mu_2 - \Phi_{12}'\Phi_{11}^{-1}\mu_1, \Phi_{12}'\Phi_{11}^{-1}\right), \qquad \Lambda = \Phi_{22 \cdot 1}.$$

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,

$$n^{1/2}(C\hat{\beta} - C\beta) \to MVN\left(0, \Delta_{11}^{-1}[\eta'\Sigma\eta + \gamma'Q\Phi_{22 \cdot 1}Q\gamma]\right) \tag{3.9}$$

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

3.5. *Conclusion.* The results (3.4), (3.7), and (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.* **80** 929–932.

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* (D. N. M. DeGruijter and L. J. Th. van der Kamp, eds.). Wiley, New York.

DEGRACIE, J. S. and FULLER, W. A. (1972). 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, Univ. of North Carolina, 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 Proc. Bus. and Econ. Statist. Sect.*, 57–66. Amer. Statist. Assoc., Washington.

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
321 PHILLIPS HALL 039A
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514

PAUL P. GALLO
GROUP LEADER, PRECLINICAL STATISTICS
LEDERLE LABORATORIES
PEARL RIVER, NEW YORK 10954