THE NONEXISTENCE OF $100(1-\alpha)\%$ CONFIDENCE SETS OF FINITE EXPECTED DIAMETER IN ERRORS-IN-VARIABLES AND RELATED MODELS[1]

by

Leon Jay Gleser and J. T. Hwang
Purdue University Cornell University

Technical Report #85-15

Department of Statistics
Purdue University

July 1985

[1]Research of Leon Jay Gleser was supported by NSF Grant D.M.S. 8501966. Research of J. T. Hwang was supported by NSF Grant MCS-8301139.

THE NONEXISTENCE OF 100(1-α)% CONFIDENCE SETS OF FINITE EXPECTED DIAMETER IN ERRORS-IN-VARIABLES AND RELATED MODELS[1]

by

Leon Jay Gleser
Purdue University

and

J. T. Hwang
Cornell University

ABSTRACT

Confidence intervals are widely used in statistical practice as indicators of precision for related point estimators, or as estimators in their own right. In the present paper it is shown that for some models, including most linear and non-linear errors-in-variables regression models, and for certain estimation problems arising in the context of classical linear models, such as the inverse regression problem, it is impossible to construct confidence intervals for key parameters which have both positive confidence and finite expected length. The results are generalized to cover general confidence sets for both scalar and vector parameters.

AMS 1980 subject classification. Primary 62F25; secondary 62F11, 62H12, 62H99.

Key words and phrases. Confidence intervals, confidence sets, confidence regions, coverage, expected length, expected diameter, errors-in-variables regression, inverse regression, calibration, principal components analysis, Fieller's method.

---

[1]Research of Leon Jay Gleser was supported by NSF Grant D.M.S. 8501966
Research of J. T. Hwang was supported by NSF Grant MCS-8301139.

1. Introduction. Confidence intervals are widely used in statistical practice as indicators of precision for related point estimators (which are often, but not always, midpoints of such intervals), or as estimators (interval estimators) in their own right. Suppose we observe data Y obeying a parametric model whose probabilities $P_\theta\{Y \text{ in } A\}$ are indexed by a parameter θ, where θ is an element of a parameter space Θ. A confidence interval $[L(Y),U(Y)]$ for a scalar function $\gamma(\theta)$ of θ is defined by (measurable) functions $L(Y),U(Y)$ of Y. The coverage probability $p(\theta)$ of $[L(Y),U(Y)]$ is defined for each θ in Θ by

$$p(\theta) = P_\theta\{L(Y) \leq \gamma(\theta) \leq U(Y)\},$$

and the confidence (confidence level) of the confidence interval by

$$1-\alpha = \inf_{\theta \in \Theta} p(\theta). \tag{1.1}$$

The expected length $\lambda(\theta)$ of $[L(Y),U(Y)]$ is defined by

$$\lambda(\theta) = E_\theta[U(Y) - L(Y)].$$

If the confidence of the interval is large (e.g., $1-\alpha = .95$) and the expected length $\lambda(\theta)$ is small for all θ, then the interval $[L(Y),U(Y)]$ is regarded as a good frequentist interval estimator of

$\gamma(\theta)$. Alternatively, the high confidence and small expected length of the interval $[L(Y),U(Y)]$ can be regarded as evidence of the accuracy or precision of the point estimator

$$\hat{\gamma} = \tfrac{1}{2}(L(Y) + U(Y))$$

for $\gamma(\theta)$, or indeed of the accuracy of any point estimator $\tilde{\gamma}(Y)$ for which $L(Y) \leq \tilde{\gamma}(Y) \leq U(Y)$.

Similarly, for any r-dimensional vector function $\underset{\sim}{\gamma}(\theta)$ of θ, $r \geq 1$, we might seek to simultaneously estimate the components $\gamma_1(\theta),\ldots,\gamma_r(\theta)$ of $\underset{\sim}{\gamma}(\theta)$ by a confidence set $C(Y)$. In this case, the coverage probability $p(\theta)$ of the set equals $P_\theta\{\gamma(\theta) \in C(Y)\}$ and the confidence $1-\alpha$ of $C(Y)$ is defined by (1.1). The diameter $d(Y)$ of $C(Y)$ is the maximum (supremum) distance between any two points in $C(Y)$.

In most parametric and nonparametric problems, it is possible to find $1-\alpha$ confidence intervals of finite expected length for parameters or parametric functions of interest. However, there are important exceptions to this rule.

For example, consider the classical simple linear errors-in-variables model (Anderson, 1984) in which we observe pairs (y_i,x_i) of random variables satisfying the model:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 u_i + e_{1i}, \\ x_i &= u_i + e_{2i}, \qquad i = 1,2,\ldots,n, \end{aligned} \tag{1.2}$$

where $(e_{1i},e_{2i})'$ are i.i.d. with common mean vector $(0,0)'$ and common covariance matrix $\Sigma_e = \sigma_e^2 I_2$. Here, β_0, β_1 and σ_e^2 are parameters of basic interest in the model. The quantities u_i are usually assumed to be either fixed constants (functional case) or i.i.d. random variables with mean μ and variance σ_u^2 (structural case). For the functional case of this model, Gleser (1982) has shown that:

(1) any $1-\alpha$ confidence interval for β_1, $0 \leq \alpha < 1$, must have infinite expected length,

(2) contrariwise, any confidence interval for β_1 of finite expected length must have confidence $1-\alpha = 0$.

Gleser confined his proof to intervals which are based on the data only through the first and second sample moments and cross moments of (y_i,x_i). This leaves open the question of whether there exist confidence intervals or confidence sets for β_1 having positive confidence $(1-\alpha > 0)$ and finite expected length or diameter which are based on more elaborate use of the data (such as through use of the Jackknife or Bootstrap). In Section 3, we show that no such confidence intervals or sets can exist. In Section 3, we also verify similar nonexistence assertions concerning confidence intervals or sets for arbitrary linear combinations of β_0 and β_1 in both the functional and structural cases of this model.

An analysis of Gleser's (1982) argument reveals that the key to his results is that the model (1.2), by suitable choice of the "nuisance parameters u_i(in the functional case) or σ_u^2 (in the structural case) can be made arbitrarily "close to the model

$$y_i = \beta_0 + \beta_1\mu + e_{1i}$$

$$x_i = \mu + e_{2i}, \qquad i = 1,2,\ldots,n,$$

for which β_0 and β_1 are not identifiable. This suggests that more general results are possible.

Thus, let Y be a random element of a probability space $(\mathcal{Y}, \mathfrak{F})$, with $\mathfrak{F}$ a sigma-field of measurable subsets of $\mathcal{Y}$. Let ζ be a sigma-finite measure on $(\mathcal{Y},\mathfrak{F})$, and let Y have probabilities determined by one of a parametric class of densities $f(Y|\theta)$ relative to ζ, with common support $\mathcal{Y}^* \subseteq \mathcal{Y}$. Thus

$$P_\theta\{Y \in A\} = \int_A f(Y|\theta)d\zeta(Y).$$

Assume that $\theta = (\theta_1, \theta_2)$ takes values in

$$\Theta = \Theta_1 \times \Theta_2,$$

where Θ_1 is a subset of p-dimensional Euclidean space E^p, and Θ_2 is a subset of q-dimensional Euclidean space E^q.

The following Theorem is the main result of our paper.

<u>THEOREM 1</u>. Let $\gamma(\theta_1)$ be a scalar function of $\theta_1 \in \Theta_1$. Suppose that there exists a subset Θ_1^* of Θ_1 and a point θ_2^* in the closure $\overline{\Theta}_2$ of Θ_2 such that

$$\gamma(\theta_1) \text{ has unbounded range over } \theta_1 \in \Theta_1^* , \tag{1.3}$$

and such that for each fixed $\theta_1 \in \Theta_1^*$, $Y \in \mathcal{Y}$,

$$\lim_{\theta_2 \to \theta_2^*} f(Y|(\theta_1,\theta_2)) = f(Y|\theta_2^*) \tag{1.4}$$

exists, is a density for Y relative to ζ, and is independent of θ_1. Then, every confidence set C(Y) for $\gamma(\theta_1)$ with confidence $1-\alpha > 0$ satisfies

$$P_{(\theta_1,\theta_2)}(d(Y) = \infty) > 0 \tag{1.5}$$

for all $(\theta_1,\theta_2) \in \Theta$, where d(Y) is the diameter of C(Y). [Consequently, $E_\theta[d(Y)] = \infty$ for all $\theta = (\theta_1,\theta_2) \in \Theta$.] Contrariwise, if C(Y) is a confidence set for $\gamma(\theta_1)$ whose diameter is finite with probability one for all $\theta \in \Theta$, then the confidence level $1-\alpha$ of C(Y) equals 0.

Theorem 1 deals with confidence set estimation of scalar parametric functions. However, this theorem is also applicable to vector-valued parametric functions $\underset{\sim}{\gamma}(\theta_1)$ because of the following theorem.

<u>THEOREM 2</u>. Let Y be a random vector whose distribution depends on an unknown vector parameter θ. Let $\underset{\sim}{\gamma}(\theta)$ be an m-dimensional vector-valued function of θ. If for some constant m-dimensional vector a it can be shown that no confidence set for $a'\underset{\sim}{\gamma}(\theta)$ with positive confidence and finite expected diameter exists, then the same conclusion holds for any confidence set C(Y) for $\underset{\sim}{\gamma}(\theta)$.

Theorems 1 and 2 are proven in Section 2. This section, which is technical in nature, can be skipped by anyone interested only in applications of the main results.

In Section 3, it is shown how Theorems 1 and 2 apply to the simple errors-in-variables models (1.2) in both functional and structural cases, and to various generalizations of these models, including nonlinear errors-in-variables models and estimation of principal component vectors. In Section 3, Theorems 1 and 2 are also applied to the classical inverse-regression (calibration) problem, and to the more general problem of estimating ratios of slopes in classical multiple linear regression. Finally, Section 4 briefly discusses some alternatives to frequentist confidence interval estimation.

2. Proofs.

Proof of Theorem 1. By (1.4), for every $\theta_1 \in \Theta_1^*$,

$$\lim_{\theta_2 \to \theta_2^*} \int_{\mathcal{Y}} f(Y|(\theta_1,\theta_2))d\zeta(Y) = 1 = \int_{\mathcal{Y}} f(Y|\theta_2^*)d\zeta(Y) = \int_{\mathcal{Y}} \lim_{\theta_2 \to \theta_2^*} f(Y|(\theta_1,\theta_2))d\zeta(Y). \quad (2.1)$$

Also, since $\Theta_1^* \subset \Theta_1$,

$$\Theta^* \equiv \Theta_1^* \times \Theta_2 \subset \Theta_1 \times \Theta_2 = \Theta .$$

It follows from (1.1) that

$$\begin{aligned} 1 - \alpha &= \inf_{(\theta_1,\theta_2) \in \Theta} P_{(\theta_1,\theta_2)}(\gamma(\theta_1) \in C(Y)) \\ &\le \inf_{(\theta_1,\theta_2) \in \Theta^*} P_{(\theta_1,\theta_2)}(\gamma(\theta_1) \in C(Y)). \end{aligned}$$

Fix $\theta_1 \in \Theta_1^*$. Since $C(Y)$ is asserted to have positive confidence,

$$0 < 1-\alpha \leq \inf_{(\theta_1,\theta_2)\in\Theta^*} P_{(\theta_1,\theta_2)}(\gamma(\theta_1)\in C(Y))$$

$$\leq \lim_{\theta_2\to\theta_2^*} P_{(\theta_1,\theta_2)}(\gamma(\theta_1)\in C(Y)) \qquad (2.2)$$

$$= \lim_{\theta_2\to\theta_2^*} \int_{\mathcal{Y}} I(\{\gamma(\theta_1)\in C(Y)\}) f(Y|(\theta_1,\theta_2))d\zeta(Y),$$

where I(A) is the indicator function of the set A. Note that for any set A in $\mathcal{Y}$,

$$0 \leq I(A)\ f(Y|(\theta_1,\theta_2) \leq f(Y|(\theta_1,\theta_2)).$$

Thus by (2.1), (2.2) and a well-known extension of the Lebesgue dominated convergence theorem, for each $\theta_1 \in \Theta_1^*$

$$0 < 1-\alpha \leq \lim_{\theta_2\to\theta_2^*} \int_{\mathcal{Y}} I(\{\gamma(\theta_1)\in C(Y)\}) f(Y|(\theta_1,\theta_2))d\zeta(Y)$$

$$= \int_{\mathcal{Y}} I(\{\gamma(\theta_1)\in C(Y)\}) f(Y|\theta_2^*)d\zeta(Y). \qquad (2.3)$$

Since the range of $\gamma(\theta_1)$ over $\theta_1 \in \Theta_1^*$ is infinite, we can find a sequence of values of θ_1 in Θ_1^* such that either $\gamma(\theta_1) \to \infty$ or $\gamma(\theta_1) \to -\infty$. Assume that we can take $\gamma(\theta_1) \to \infty$ (the proof when $\gamma(\theta_1) \to -\infty$ is similar). Let

$$U(Y) = \max\{g\colon g\in C(Y)\},$$
$$L(Y) = \min\{g\colon g\in C(Y)\}.$$

Then by (2.3) and the Lebesgue dominated convergence theorem,

$$0 < 1-\alpha \leq \lim_{\gamma(\theta_1) \to \infty} \int_{\mathcal{Y}} I(\{\gamma(\theta_1) \in C(Y)\})\ f(Y|\theta_2^*)d\zeta(Y)$$

$$\leq \lim_{\gamma(\theta_1) \to \infty} \int_{\mathcal{Y}} I(\{\gamma(\theta_1) \leq U(Y)\}) f(Y|\theta_2^*)d\zeta(Y)$$

$$= \int_{\mathcal{Y}} I(\{U(Y) = \infty\})\ f(Y|\theta_2^*)d\zeta(y)$$

$$\leq \int_{\mathcal{Y}} I(\{U(Y)-L(Y) = \infty\})\ f(Y|\theta_2^*)d\zeta(Y).$$

Let

$$S = \{Y:U(Y) - L(Y) = \infty\},$$
$$T = \{Y:f(Y|\theta_2^*) > 0\}. \qquad (2.4)$$

We have shown that

$$0 < 1-\alpha \leq \int_S f(Y|\theta_2^*)d\zeta(Y) = \int_{S \cap T} f(Y|\theta_2^*)d\zeta(Y), \qquad (2.5)$$

and it follows from (1.4) that the support T of $f(Y|\theta_2^*)$ is contained in the common support $\mathcal{Y}^*$ of the $f(Y|(\theta_1,\theta_2))$, $(\theta_1,\theta_2) \in \Theta$. Hence, it follows that for any $(\theta_1,\theta_2) \in \Theta$,

$$P_{(\theta_1,\theta_2)}(S) = \int_S f(Y|(\theta_1,\theta_2))d\zeta(Y)$$

$$\geq \int_{S \cap T} \left[\frac{f(Y|(\theta_1,\theta_2))}{f(Y|\theta_2^*)}\right] f(Y|\theta_2^*)d\zeta(Y)$$

$$> 0.$$

This completes the proof of the first part of Theorem 1. The "contrariwise" part of Theorem 1 follows directly as the contrapositive of the first part of Theorem 1. Hence, the proof of Theorem 1 is complete. □

Proof of Theorem 2. Let C(Y) be a confidence set for $\underset{\sim}{\gamma}(\theta)$ with positive confidence $1 - \alpha > 0$. Let

$$C_a(Y) = \{a'g: g \in C(Y)\}.$$

That is, $C_a(Y)$ is the Scheffé projection (Scheffé, 1959) of C(Y) for estimation of $a'\underset{\sim}{\gamma}(\theta)$. Clearly

$$0 < 1 - \alpha \leq \inf_{\theta \in \Theta} P_\theta(\underset{\sim}{\gamma}(\theta) \in C(Y)) \leq \inf_{\theta \in \Theta} P_\theta\{a'\underset{\sim}{\gamma}(\theta) \in C_a(Y)\}.$$

By assumption, every confidence set for $a'\underset{\sim}{\gamma}(\theta)$ having positive confidence $1 - \alpha$ must have infinite diameter with positive probability for all θ. The diameter of $C_a(Y)$ is obviously no larger than that of C(Y), and thus the proof of Theorem 2 is complete. □

3. Applications

3.1. Linear Errors-In-Variables Models. The simple linear errors-in-variables model (1.2) is a special case of the following multivariate linear errors-in-variables regression model. Let

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \underset{\sim}{0}_q \end{pmatrix} + \begin{pmatrix} B_1 \\ I_q \end{pmatrix} u_i + e_i, \quad i = 1,2,\ldots,n, \tag{3.1}$$

where I_q is the q-dimensional identity matrix and $\underset{\sim}{0}_q$ the q-dimensional zero column vector. In (3.1), β_0 and y_i, $1 \leq i \leq n$, are p-dimensional vectors; x_i and u_i, $1 \leq i \leq n$, are q-dimensional vectors; B_1 is a p x q matrix and the e_i, $1 \leq i \leq n$, are (p+q)-dimensional random vectors with mean vector $\underset{\sim}{0}_{p+q}$ and covariance matrix Σ_e.

Recall that any matrix can be represented as a vector using the "vec" notation. To avoid notational complexity, we do not do this. Instead, when we list the parameters making up the vectors θ_1, θ_2 in Theorem 1, it will always be understood that these are arrayed in vector form. Thus, for example, if we write $\theta_1 = (\beta_0, B_1)$, it is understood that θ_1 is a column vector composed of the elements of β_0, B_1.

To apply Theorem 1 to the model (3.1), let

$$Y = (y_1', x_1', y_2', x_2', \ldots, y_n', x_n')'.$$

For the functional case of the model (3.1), where the u_i's are unknown vector parameters, let

$$\theta_2 = (u_1, u_2, \ldots, u_n, \Sigma_e).$$

For the structural case of (3.1), where the u_i's are random vectors with unknown mean vector μ and unknown positive definite covariance matrix Σ_u, let

$$\theta_2 = (\mu, \Sigma_u, \Sigma_e).$$

In both cases, let

$\theta_1 = (\beta_0, B_1)$.

Suppose that we want to estimate an unbounded scalar function $\gamma(B_1)$ of B_1 -- for example, an element of B_1, a linear combination of elements of B_1, or perhaps the norm $[tr(B_1 B_1')]^{\frac{1}{2}}$ of B_1. To show that no nontrivial $(1 - \alpha > 0)$ confidence set for $\gamma(B_1)$ with finite expected diameter exists, we apply Theorem 1 with

$$\Theta_1^* = \{(\beta_0, B_1): \beta_0 = b\} \text{ , b a fixed p x 1 vector,}$$

and θ_2^* defined from θ_2 by setting $u_1 = u_2 = \ldots = u_n = \underset{\sim}{0}_q$ in the functional case and $\mu = \underset{\sim}{0}_q$, $\Sigma_u = \underset{\sim}{0}_{q \times q}$ in the structural case. ($\underset{\sim}{0}_{q \times q}$ is the q x q matrix of zeroes.) In both cases, functional and structural, it is easily seen that requiring $\theta_1 = (\beta_0, B_1)$ to be in Θ_1^* does not restrict B_1, and thus does not restrict the range of $\gamma(B_1)$. Further, when $\theta_2 = \theta_2^*$, $\theta_1 \in \Theta_1^*$, the distribution of Y depends only on the known vector b and on Σ_e, and is thus functionally independent of θ_1. (This assumes that the joint distribution of the e_i, $1 \leq i \leq n$, does not depend functionally on (β_0, B_1).) Consequently, Theorem 1 applies, and we can also conclude from Theorem 2 that no nontrivial confidence set with finite expected diameter exists for B_1.

Instead, suppose that we want to estimate an unbounded scalar function $\gamma(\beta_0)$ of β_0. In this case, define

$$\Theta_1^* = \{(\beta_0, B_1): \beta_0 + B_1 \underset{\sim}{1}_q = b\}\text{, b a fixed p x 1 vector,}$$

where $\underset{\sim}{1}_q$ is the q-dimensional vector $(1,1,1,\ldots,1)'$. Let θ_2^* be defined from θ_2 by setting $u_1 = u_2 = \ldots = u_n = \underset{\sim}{1}_q$ in the functional case and $\mu = \underset{\sim}{1}_q$, $\Sigma_u = \underset{\sim}{0}_{q \times q}$ in the structural case. Again, requiring θ_1 to be in Θ_1^* does not restrict the range of β_0, and thus of $\gamma(\beta_0)$. Further, when

$\theta_2 = \theta_2^*$, $\theta_1 \in \Theta_1^*$, the distribution of Y depends only on b and Σ_e. Consequently, Theorem 1 applies to $\gamma(\beta_0)$, and Theorem 2 shows the nonexistence of nontrivial confidence sets for β_0 with finite expected diameter.

In general, suppose that for a known (q+1) x r matrix A of rank $r \leq q$, we wish to estimate an unbounded scalar function of $(\beta_0,B_1)A$. For example, we might be interested in estimating a given linear combination of the elements of (β_0,B_1), say

$$c'(\beta_0,B_1)a \quad , \text{ c: p x 1, a: (q+1) x 1,}$$

in which case A = a and $\gamma(v) = c'v$. Since A has rank $r < q+1$, there exists a q-dimensional vector t such that the columns of A are linearly independent of $(1,t')'$. Define

$$\Theta_1^* = \{(\beta_0,B_1): \beta_0 + B_1 t = b\}, \text{ b a fixed p x 1 vector,}$$

and let θ_2^* be defined from θ_2 by letting $u_1 = \ldots = u_n = t$ in the functional case, and $\mu = t$ and $\Sigma_u = \underset{\sim}{0}_{q \times q}$ in the structural case. The applicability of Theorem 1 and Theorem 2 follows by the arguments stated previously. We thus can conclude from Theorem 2 that nontrivial $(1 - \alpha > 0)$ confidence sets for (β_0,B_1) with finite expected diameter do not exist.

The above assertions hold whether Σ_e is assumed known or unknown. (In the latter case, we may need conditions on Σ_e to make the model (3.1) identifiable.) However, Σ_e must be assumed to be positive definite in order that the model (3.1) does not degenerate to the standard linear model, where nontrivial confidence intervals for individual parameters in (β_0,B_1) having finite expected length are known to exist. Since our results hold for known

Σ_e, it follows that our nonexistence assertions about confidence sets for scalar and vector functions of (β_0, B_1) also hold in the contexts of generalizations of the model (3.1) which permit replications or use of instrumental variables in order to estimate Σ_e. (For examples of such models, see Anderson, 1984; Gleser, 1983.)

Note that use of Theorem 1 in this context does not require us to make any parametric assumption about the joint distribution of the errors e_i in (3.1). The e_i's do not have to be normally distributed, or independent, or even identically distributed. The e_i's do not even need to have common covariance matrix Σ_e. Of course, the more assumptions we make, the more striking are our nonexistence results! Still it is worth remarking that for Theorem 1 to hold it is sufficient that the joint density f(e) for $e = (e_1', e_2', \ldots, e_n')'$ satisfies the following conditions:

(i) f(e) is functionally independent of (β_0, B_1),

(ii) f(e) is continuous in e (permitting the limit as $\theta_2 \to \theta_2^*$ to hold and be a density),

(iii) The support of f(e) is n(p+q)-dimensional Euclidean space $E^{n(p+q)}$ (so that the densities for Y for all values of the parameters $\beta_0, B_1, u_1, \ldots, u_n$ have common support).

In the structural case, the u_i's are random vectors independent of the e_i's. The u_i's are usually assumed to be i.i.d. with a common q-variate normal distribution, but such an assumption is not needed in order to apply Theorem 1. The u_i's can be dependent and even have non-identical marginal distributions. In fact, the u_i's need not have common mean vector μ nor common covariance matrix Σ_u. If we assume that u_i has mean vector μ_i and positive definite matrix $\Sigma_u^{(i)}$, $i = 1,2,\ldots,n$, then we let

$$\theta_2 = (\mu_1,\mu_2,\ldots,\mu_n, \Sigma_u^{(1)},\ldots,\Sigma_u^{(n)}).$$

To apply Theorem 1, we define $\theta_2{}^*$ from θ_2 by setting

$$\mu_1 = \mu_2 = \ldots = \mu_n = \mu , \Sigma_u^{(1)} = \ldots = \Sigma_u^{(n)} = \underset{\sim}{0}_{q \times q},$$

where the vector μ depends upon which scalar function γ is of interest to us. (For example, if we are interested in $\gamma(B_1)$, we set $\mu = \underset{\sim}{0}_q$.) Consequently, the generality of allowing the u_i's to have possibly different mean vectors and/or covariance matrices is somewhat spurious. In any case, for Theorem 1 to apply in the structural case, it is sufficient that the class of distributions of the u_i's permit taking the limit $\theta_2 \to \theta_2{}^*$, and that the density f(e) of the vector of errors e has the properties (i), (ii), (iii) listed above for the functional case. [Of course, it is also implicit in our assumptions that the mean vector(s) and covariance matrix (matrices) of the u_i's do not depend functionally on (β_0,B_1).]

The key to our arguments in both the functional and structural cases of the linear multivariate errors-in-variables model (3.1) is that we can find a sequence of parameters tending to a limit for which the variability of the u_i's is equal to zero. Note that in the functional case, this limit lies in the interior of the parameter space; while in the structural case, the limit $\Sigma_u = \underset{\sim}{0}_{q \times q}$ is on the boundary of the parameter space.

3.2. Related Models.

Hwang (1984) considers a simple errors-in-variables model with a multiplicative error (rather than additive, as in (1.2) or (3.1)). Theorems 1 and

2 can be applied in the context of his model to show the nonexistence of nontrivial $(1 - \alpha > 0)$ confidence sets with finite expected diameter for linear combinations of the essential parameters.

Theorems 1 and 2 also apply to nonlinear functional and structural errors-in-variables models. Thus, suppose that

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} = \begin{pmatrix} h(\beta, u_i) \\ u_i \end{pmatrix} + e_i, \; i = 1,2,\ldots,n, \tag{3.2}$$

where y_i is a p-dimensional vector, β is an unknown m-dimensional vector, $h(\beta,u)$ is a known p-dimensional vector function of β and u which is continuous in u for all fixed β, x_i and u_i are q-dimensional vectors, and e_i is a (p+q)-dimensional random vector with mean vector $\underset{\sim}{0}_{p+q}$ and positive covariance matrix Σ_e. The u_i's can be unknown vector parameters (functional case) or random vectors with unknown mean vector μ and unknown positive definite covariance matrix Σ_e (structural case). Again, it will not matter whether Σ_e is known or unknown, and distributional assumptions for the u_i's and e_i's apart from those mentioned in Subsection 3.1 are not required.

We are interested in applying Theorem 1 to unbounded scalar functions $\gamma(\beta)$ of $\theta_1 = \beta$. Let $\theta_2 = (u_1, u_2, \ldots, u_n, \Sigma_e)$ in the functional case and $\theta_2 = (\mu, \Sigma_u, \Sigma_e)$ in the structural case. Once again we will let $\theta_2{}^*$ be defined from θ_2 by letting

$$u_1 = u_2 = \ldots = u_n = t \qquad \text{(functional case)},$$
$$\mu = t, \quad \Sigma_u = \underset{\sim}{0}_{q \times q} \qquad \text{(structural case)},$$

where the q-dimensional vector t is at our disposal. We let

$\Theta_1^* = \{\beta: h(\beta,t) = b\}$, b a fixed p x 1 vector.

Note that Θ_1^* defines a surface in the p-dimensional range Θ_1 of β. If $\gamma(\beta)$ for $\beta \in \Theta_1^*$ is unbounded in range for some t and b, Theorem 1 applies to show that no nontrivial confidence set for $\gamma(\beta)$ with finite expected diameter can exist. Giving general methods for finding t and b is beyond the scope of this paper. For particular cases, this is usually easy. As a simple example, if p = q = 1, $\beta = (\beta_0,\beta_1)'$, $\gamma(\beta) = \beta_1$ and

$$h(\beta,u) = \beta_0 \, e^{\beta_1 u} ,$$

then t = 0, b = 0, achieves the desired end, while if $\gamma(\beta) = \beta_0$, then t = 1, b = 1 will suffice.

Finally, it is well known that the structural form of the linear errors-in-variables model (3.1) is related to principal component analysis. Suppose that $y_1, y_2, \ldots, y_n$ are i.i.d. p-dimensional continuous random vectors with support E^p, mean vector μ, and unknown positive definite covariance matrix Σ. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ be the eigenvalues of Σ. Suppose that it is assumed that $\lambda_1 > \lambda_2$, so that the eigenvector x corresponding to λ_1 is uniquely defined up to a scalar multiple. Also suppose that x is not orthogonal to the vector (1,0,0,...,0)'. If we scale x so that

$$x = \begin{pmatrix} 1 \\ \beta \end{pmatrix} ,$$

where β is a (p-1)-dimensional vector, then the elements of β are the slopes of x relative to the last p-1 axes of p-dimensional Euclidean space and serve to define the first principal component of Σ. It is frequently desired to estimate the elements of β. In particular, confidence intervals for the elements $\beta_1,\ldots,\beta_{p-1}$ of β or a confidence set for β may be desired.

However, letting $\theta_1 = \beta$, $\theta_2 = (\lambda_1,\ldots,\lambda_p)$, $\Theta_1^* = \Theta_1$ and

$\theta_2{}^* = (\lambda,\lambda,\lambda,\ldots,\lambda)$ for some $\lambda > 0$, it is easily seen that Theorems 1 and 2 apply. Thus, no nontrivial confidence sets with finite expected diameter for elements of β can exist.

3.3. Ratios of Regression Parameters and Inverse Regression. Consider the classical multiple regression model

$$y_i = \beta_o + \sum_{j=1}^{p} \beta_j x_{ji} + e_i,\ i=1,2,\ldots,n, \qquad (3.3)$$

where the e_i have mean 0 and variance σ_e^2, and a joint distribution having the properties (i), (ii), (iii) mentioned in Subsection 3.1. For any given j_o, $0 \leq j_o \leq p$, define the ratios $\delta_j = \beta_{j_o}^{-1}\beta_j$, $j=0,\ldots,p$.

Applying Theorems 1 and 2 with

$$\theta_1 = (\delta_o,\ldots,\delta_p)',\ \theta_2 = (\beta_{j_o},\sigma_e^2)',\ \theta_2{}^* = (0,\sigma_e^2)\ ,$$

we see that no nontrivial confidence sets exist for δ_j, $j \neq j_o$, which have finite expected diameter. A special case of this problem (with p=1) is the inverse regression (discrimination, calibration) problem (Miller, 1981, p. 117; Hoadley, 1970; Seber, 1977, Chapter 7).

3.4. A Comment on Large Sample Approximations. In each of the above examples, our results are not due to lack of identifiability of the parameters. [In the functional errors-in-variables models, we can delete the line $u_1 = u_2 = \ldots = u_n$ from the parameter space and our conclusions still hold. The value $\theta_2{}^*$ simply becomes a boundary value of the parameter space.] Even after we impose identifiability restrictions on the parameters, the phenomenon persists. The reason for this phenomenon is stated in the discussion preceding the statement of Theorem 1: For fixed n, the confidence level of any confidence set is approached as a limit as $\theta_2 \to \theta_2^*$.

On the other hand, in each of our examples one can exhibit large sample approximate 100 $(1-\alpha)$%, $0 < \alpha < 1$, confidence intervals of finite length (almost surely) for any $\gamma(\theta_1)$. For example, a large sample confidence interval for β_1 can be constructed in the context of the model (1.2) [Anderson (1984)]. Although for each fixed (θ_1,θ_2) in $\Theta = \Theta_1 \times \Theta_2$ the coverage probability of the large sample confidence interval for $\gamma(\theta_1)$ converges to $1-\alpha$ as $n \to \infty$, Theorem 1 shows that for fixed n, no matter how large, the confidence level of this large sample confidence interval must equal 0. The technical reason for this apparent contradiction is that the limits as $n \to \infty$ and as $\theta_2 \to \theta_2^*$ cannot be interchanged. The practical conclusion from our arguments is <u>that large sample approximations</u> (asymptotic theory) <u>fail to uniformly approximate the finite sample distributions</u> over the parameter space Θ. To use large sample approximations for the models discussed in this section (and more generally in Theorem 1), one must have some information about the location of (θ_1,θ_2) in the parameter space (particularly how close θ_2 is to the points θ_2^*). This casts doubt upon the usefulness of large sample approximations in such models, at least when used for the purpose of forming confidence sets or assessing the accuracy of point estimators.

4. <u>Discussion</u>. The models and inference problems mentioned in Section 3 have wide applicability. Consequently, the nonexistence of nontrivial, finite-expected-length confidence intervals is of concern, at least to those statisticians who use confident intervals as frequentist indicators of precision, or as estimators in their own right.

Bayesian statisticians do not have similar problems. Bayesian credible intervals, having posterior probability $1-\alpha > 0$, always exist and have finite length with probability one (although some such intervals may have infinite expected length). However, such intervals solve a different problem. If reported by the experimenter, they give an interval of values for a parameter which has strong posterior support based on the experimenter's own prior distribution. The credible interval produced by a reader of an experimenter's results can be greatly different from that of the experimenter if that reader starts with a different prior distribution and the amount of information in the data is not great enough to overwhelm prior opinion. Hence, where this is possible, Bayesians should provide a family of credible intervals $I_\pi(Y)$ for a parameter, from which a reader with a particular prior π can select the appropriate interval to reflect the precision he or she finds in the data. The results of this paper (see Theorem 1) show that such intervals will be sensitive to the amount of prior probability mass or density for θ_2 in the neighborhood of θ_2^*.

The goal of the frequentist approach is to present intervals or measures of precision which have properties that can be stated independent of the prior opinions of investigators. Of necessity, this forces such measures to reflect "worst cases". In the case of the models described here, such worst cases occur when θ_2 is arbitrarily close to θ_2^*, and the data therefore provides vanishing information about θ_1.

However, some frequentists would point out that in real problems there is vague prior information about θ_2 (perhaps too crude to be quantitively modeled in the form of a prior distribution) which leads one to believe that θ_2 is bounded away from θ_2^*. In this case (specifically in the models presented in Section 3), it may be possible to exhibit intervals [L(Y),U(Y)] of finite expected length for which the coverage probability $p(\theta_1,\theta_2)$ is high as long as θ_2 stays away from θ_2^*. An example where this is true is given by Gleser (1982) for the model (1.2). Gleser shows that the usual large-sample interval for β_1 has confidence near $1-\alpha$ as long as $\sigma_u^2/\sigma_e^2 \geq 1$. Of course, asserting that $\sigma_u^2/\sigma_e^2 \geq 1$ changes the parameter space. One may be willing to make such an assumption in <u>evaluating</u> an estimator, but not in <u>constructing</u> it.

Alternatively, a frequentist can try to find arbitrary confidence sets which in some way reflect the information given by the data about possible values of $\gamma(\theta_1)$. This was apparently Fieller's approach (Fieller, 1954; see also Creasy, 1954, 1956). This approach starts by finding good α-level tests for the hypotheses $H_{0c}:\gamma(\theta_1) = c$, $-\infty < c < \infty$. One then lets a confidence set C(Y) for $\gamma(\theta_1)$ consist of all values c for which H_{0c} is not rejected by the data Y. In the context of the models discussed in Section 3, such regions C(Y), which are guaranteed to have confidence $1-\alpha$, are not always finite intervals, and indeed can have the form $(-\infty,L(Y)]\cup[U(Y),\infty)$, $L(Y) \leq U(Y)$. Note that Theorem 1 tells us that such confidence sets must have infinite expected diameter for all parameter values. Although the sets C(Y), provided they are based on good tests, do tell us what values of $\gamma(\theta_1)$ cannot be dismissed by the data, they only indirectly provide an idea of the accuracy with which $\gamma(\theta_1)$ is estimated. It is, for example, hard to imagine interpreting such regions in graphical summaries of results--e.g. comparison of $\gamma(\theta_1)$ across several populations. On the other hand, graphs showing point estimates with accompanying confidence intervals are easy to interpret. [Neyman (1954), in a discussion of the papers of Fieller (1954) and Creasy (1954), also worries about what confidence to assign to a region (interval) formed by an experimenter who repeats or continues experimentation until a set C(Y) which is a finite interval is obtained.]

Fieller's approach to estimating the δ_j's in the model (3.3) is given in most textbooks on linear models. A similar methodology is given by Williams (1959), Brown (1957), Gleser and Watson (1973), and Schneeweiss (1982) for the linear errors-in-variables regression models (3.1).

The present paper is intended only to point out a problem and summarize available solutions, not to propose a choice. The choice among these solutions (or others which may be proposed in the future) will depend upon one's statistical philosophy of inference, and on one's goals. Although the textbooks currently favor Fieller's approach, it is not at all clear that this approach is the most satisfactory for the goals of most practitioners. What the present paper does make clear is that the traditional confidence interval approach cannot work in the kinds of problems illustrated by the examples in Section 3.

REFERENCES

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

BROWN, R. L. (1957). The bivariate structural relation. Biometrika 44, 84-96.

CREASY, M. A. (1954). Limits for the ratio of means. J. Roy. Statist. Soc. B 16, 186-94.

CREASY, M. A. (1956). Confidence intervals for the gradient in the linear functional relationship. J. Roy. Statist. Soc. B 18, 65-69.

FIELLER, E. C. (1954). Some problems in interval estimation. J. Roy. Statist. Soc. B 16, 175-85.

GLESER, LEON JAY (1982). Confidence regions for the slope in a linear errors-in-variables regression model. Technical Report #82-23, Department of Statistics, Purdue University.

GLESER, LEON JAY (1983). Functional, structural and ultrastructural errors-in-variables models. 1983 Proc. Bus. Econ. Statist Sect., 57-66. American Statistical Association: Washington, D.C.

GLESER, L. J. and WATSON, G. S. (1973). Estimation of a linear transformation. Biometrika 60, 525-34.

HOADLEY, B. (1970). A Bayesian look at inverse linear regression. J. Amer. Statist. Assoc. 65, 356-69.

HWANG, JIUNN T. (1984). Multiplicative errors-in-variables models with applications to the recent data released by U.S. Department of Energy. Technical Report, Department of Mathematics, Cornell University.

MILLER, R. G., Jr. (1981). Simultaneous Statistical Inference (2nd ed.). McGraw-Hill: N.Y.

NEYMAN, J. (1954). Discussion on the symposium on interval estimation. J. Roy. Statist. Soc. B 16, 216-8.

SCHEFFE', HENRY (1959). The Analysis of Variance. John Wiley: N.Y.

SCHNEEWEISS, H. (1982). Note on Creasy's confidence limits for the gradient in the linear functional relationship. J. Multivariate Anal. 12, 155-7.

SEBER, G.A.F. (1977). Linear Regression Analysis. John Wiley: N.Y.

WILLIAMS, E. J. (1959). Regression Analysis. John Wiley: N.Y.