The Annals of Statistics
1987, Vol. 15, No. 4, 1351–1362

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

BY LEON JAY GLESER[1] AND JIUNN T. HWANG[2]

Purdue University and Cornell University

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 nonlinear 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.

1. Introduction. Confidence intervals are widely used in statistical practice as indicators of precision for related point estimators or as 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. For each θ in Θ, the coverage probability $p(\theta)$ of $[L(Y), U(Y)]$ is defined by $P_\theta\{L(Y) \le \gamma(\theta) \le U(Y)\}$ and the confidence (confidence level) of the confidence interval by

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

If the confidence of the interval is large (e.g., $1 - \alpha = 0.95$) and the expected length $E_\theta[U(Y) - L(Y)]$ 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 used as evidence of the accuracy of any point estimator $\tilde{\gamma}(Y)$ for which $L(Y) \le \tilde{\gamma}(Y) \le U(Y)$.

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

Received July 1985; revised June 1986.

[1]Research supported by National Science Foundation grant MCS-85-01966.
[2]Research supported by National Science Foundation grant MCS-83-01139.

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

Key words and phrases. Confidence intervals, confidence regions, coverage, estimation of mixing proportions, expected length, errors-in-variables regression, inverse regression, calibration, principal components analysis, von Mises distribution on the circle.

In many 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.

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

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

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, I_q represents the q-dimensional identity matrix and 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 both the functional and the structural cases of the model, it can be shown (Section 3) that

(1.3)
(a) any $1-\alpha$ confidence interval for β_1 $(0 \le \alpha < 1)$ must have infinite expected length;
(b) contrariwise, any confidence interval for β_1 of finite expected length must have confidence $1-\alpha = 0$.

The key to a proof of (1.3) is that (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

$$\begin{aligned} y_i &= \beta_0 + \beta_1 \mu + e_{1i}, \\ x_i &= \mu + e_{2i}, \qquad i = 1, 2, \dots, n, \end{aligned}$$

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 $(\mathscr{Y}, \mathscr{F})$, with $\mathscr{F}$ a sigma-field of measurable subsets of $\mathscr{Y}$. Let ζ be a sigma-finite measure on $(\mathscr{Y}, \mathscr{F})$ and let Y have probabilities determined by one of a parametric class of densities $f(Y|\theta)$ relative to ζ, with common support $\mathscr{Y}^* \subseteq \mathscr{Y}$. Thus $\mathscr{Y}^* = \{Y: f(Y|\theta) > 0\}$, all θ, and

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

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.

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

$$\gamma(\theta_1) \text{ has unbounded range over } \theta_1 \in \Theta_1^* \tag{1.4}$$

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

$$\lim_{\theta_2 \to \theta_2^*} f(Y|(\theta_1, \theta_2)) = f(Y|\theta_2^*) \tag{1.5}$$

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*

$$P_{(\theta_1, \theta_2)}(d(Y) = \infty) > 0 \tag{1.6}$$

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* 1 *for some* $\theta \in \Theta$, *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 $\gamma(\theta_1)$ because of the following theorem.

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

Theorems 1 and 2 are proven in Section 2. That 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 linear and nonlinear errors-in-variables models, estimation of principal component vectors and to the problem of estimating ratios of slopes in classical linear regression. Section 4 shows how Theorem 1 can be extended to cover estimation of parameters $\gamma(\theta_1)$ which have a finite range $[a, b]$. There, under the assumptions of Theorem 1, it is shown (Theorem 3) that any confidence interval for $\gamma(\theta_1)$ with confidence $1 - \alpha > 0.5$ must contain the interval $[a, b]$ with positive probability for all θ. That is, for all θ there is positive probability of obtaining a noninformative interval. Two examples of the applicability of Theorem 3 are given in Section 4: estimation of mixing proportions in mixtures of distributions and estimation of the location parameter in the von Mises distribution on the circle.

Finally, Section 5 comments briefly on the implications of the results of this paper for the use of large-sample theory in determining properties of point and interval estimators and for the sensitivity of Bayesian inferential methods to choice of the prior distribution.

The problem of constructing confidence intervals or confidence regions for parameters of interest in errors-in-variables regression models and in inverse regression (calibration) problems has a long and controversial history. See, for example, Fieller (1954), Neyman (1954), Creasy (1956), Miller (1981), Brown (1982) and Schneeweiss (1982) for discussions of such problems and proposals of possible methodology.

2. Proofs.

PROOF OF THEOREM 1. By (1.5), for every $\theta_1 \in \Theta_1^*$,

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

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

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

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

$$\begin{aligned} 0 < 1 - \alpha &\le \inf_{(\theta_1, \theta_2) \in \Theta^*} P_{(\theta_1, \theta_2)}(\gamma(\theta_1) \in C(Y)) \\ &\le \lim_{\theta_2 \to \theta_2^*} P_{(\theta_1, \theta_2)}(\gamma(\theta_1) \in C(Y)) \\ &= \lim_{\theta_2 \to \theta_2^*} \int_{\mathscr{Y}} I(\{\gamma(\theta_1) \in C(Y)\}) f(Y|(\theta_1, \theta_2))\, d\zeta(Y), \end{aligned} \tag{2.2}$$

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

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

Thus by (2.1), (2.2) and a generalization of the Lebesgue dominated convergence theorem [Billingsley (1986), Exercise 16.6 (a)], for each $\theta_1 \in \Theta_1^*$,

$$\begin{aligned} 0 < 1 - \alpha &\le \lim_{\theta_2 \to \theta_2^*} \int_{\mathscr{Y}} I(\{\gamma(\theta_1) \in C(Y)\}) f(Y|(\theta_1, \theta_2))\, d\zeta(Y) \\ &= \int_{\mathscr{Y}} I(\{\gamma(\theta_1) \in C(Y)\}) f(Y|\theta_2^*)\, d\zeta(Y). \end{aligned} \tag{2.3}$$

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

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

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

$$\begin{aligned} 0<1-\alpha &\le \lim_{\gamma(\theta_1)\to\infty} \int_{\mathscr{Y}} I(\{\gamma(\theta_1) \in C(Y)\}) f(Y|\theta_2^*)\, d\zeta(Y) \\ &\le \lim_{\gamma(\theta_1)\to\infty} \int_{\mathscr{Y}} I(\{\gamma(\theta_1) \le U(Y)\}) f(Y|\theta_2^*)\, d\zeta(Y) \\ &= \int_{\mathscr{Y}} I(\{U(Y) = \infty\}) f(Y|\theta_2^*)\, d\zeta(Y) \\ &\le \int_{\mathscr{Y}} I(\{U(Y) - L(Y) = \infty\}) f(Y|\theta_2^*)\, d\zeta(Y). \end{aligned}$$

Let

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

We have shown that

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

and it follows from (1.5) that the support T of $f(Y|\theta_2^*)$ is contained in the common support $\mathscr{Y}^*$ of the $f(Y|(\theta_1, \theta_2))$, $(\theta_1, \theta_2) \in \Theta$. Hence, it follows that for any $(\theta_1, \theta_2) \in \Theta$,

$$\begin{aligned} P_{(\theta_1,\theta_2)}(S) &= \int_S f(Y|(\theta_1, \theta_2))\, d\zeta(Y) \\ &\ge \int_{S\cap T} \left[\frac{f(Y|(\theta_1, \theta_2))}{f(Y|\theta_2^*)}\right] f(Y|\theta_2^*)\, d\zeta(Y) \\ &> 0. \end{aligned}$$

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. □

Remark. If we do not require that $f(Y|\theta)$ has common support for all $\theta \in \Theta$, then the conclusion (1.6) of Theorem 1 can be shown to hold for any (θ_1, θ_2), such that $\theta_1 \in \Theta_1^*$ and θ_2 is in the intersection of a certain neighborhood of θ_2^* (possibly depending on θ_1) and Θ_2.

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

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

That is, $C_{\mathbf{a}}(Y)$ is the Scheffé projection [Scheffé (1959)] of $C(Y)$ for estimation of $\mathbf{a}'\gamma(\theta)$. Clearly,

$$0 < 1 - \alpha = \inf_{\theta \in \Theta} P_\theta(\gamma(\theta) \in C(Y)) \leq \inf_{\theta \in \Theta} P_\theta\{\mathbf{a}'\gamma(\theta) \in C_{\mathbf{a}}(Y)\}.$$

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

3. Applications.

3.1. *Errors-in-variables models.* The simple linear errors-in-variables model (1.2) is a special case of the following nonlinear errors-in-variables model. Let

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} = \begin{pmatrix} h(\beta, u_i) \\ u_i \end{pmatrix} + e_i, \qquad i = 1, 2, \dots, n, \tag{3.1}$$

where y_i is a p-dimensional vector, x_i and u_i are q-dimensional vectors and further

(i) $h(\beta, u)$ is a known p-dimensional vector function of β and u which is continuous in u for all fixed β;
(ii) the e_i's are $(p + q)$-dimensional random vectors with zero mean vector $\mathbf{0}_{p+q}$ and positive definite 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 covariance matrix Σ_u (structural case).

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

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

$\theta_1 = \beta$ and

$$\theta_2 = (u_1, u_2, \dots, u_n, \Sigma_e) \qquad \text{(functional case)},$$
$$\theta_2 = (\mu, \Sigma_u, \Sigma_e) \qquad \text{(structural case)}.$$

Let θ_2^* be defined from θ_2 by letting

$$\begin{aligned} &u_1 = u_2 = \cdots = u_n = t \qquad \text{(functional case)}, \\ &\mu = t, \qquad \Sigma_u = \mathbf{0}_{q \times q} \qquad \text{(structural case)}, \end{aligned} \tag{3.2}$$

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

$$\Theta_1^* = \{\beta: h(\beta, t) = b\} \tag{3.3}$$

for b a fixed $p \times 1$ vector. Note that Θ_1^* defines a surface in the p-dimensional range Θ_1 of β.

If the scalar function $\gamma(\beta)$ of β is unbounded in range over $\beta \in \Theta_1^*$ for some t and b, Theorem 1 applies to show that no $1 - \alpha$ confidence set $(1 - \alpha > 0)$ 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. Two examples, the multivariate linear errors-in-variables model [which includes (1.2) as the special case where $p = q = 1$] and a simple nonlinear model, are given in the following text.

The multivariate linear errors-in-variables model is given by

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \mathbf{0}_q \end{pmatrix} + \begin{pmatrix} B_1 \\ I_q \end{pmatrix} u_i + e_i, \qquad i = 1, 2, \ldots, n, \tag{3.4}$$

where β_0 is an unknown $p \times 1$ vector and B_1 is an unknown $p \times q$ matrix. Note that (3.4) implicitly defines $h(\beta, u_i) = \beta_0 + B_1 u_i$ as a function of β, where

$$\beta = \text{vec}(\beta_0, B_1)$$

is the vector composed by stacking the columns of (β_0, B_1). Suppose that we wish to estimate the unbounded scalar function

$$\gamma(\beta) = \mathbf{c}'(\beta_0, B_1)\mathbf{a} \qquad \mathbf{c}\colon p \times 1, \quad \mathbf{a}\colon (q+1) \times 1, \tag{3.5}$$

of β. Choose t in (3.2) so that $(1, t')'$ and $\mathbf{a}$ are linearly independent. It is then easily seen that the function (3.5) is unbounded for $\beta \in \Theta_1^*$, where Θ_1^* is defined by (3.3), while the distribution of Y when $\theta_2 = \theta_2^*$ and $\beta \in \Theta_1^*$ depends only on b and Σ_e. Consequently, Theorem 1 applies to show that no $1 - \alpha$ $(1 - \alpha > 0)$ confidence interval with finite expected length exists for $\gamma(\beta)$ defined by (3.5). It then follows from Theorem 2 that nontrivial $(1 - \alpha > 0)$ confidence sets for (β, B_1) with finite expected diameter do not exist.

As an example of a nonlinear errors-in-variables model to which Theorems 1 and 2 apply, let $p = q = 1$, $\beta = (\beta_0, \beta_1)'$ and $h(\beta, u) = \beta_0 \exp\{\beta_1 u\}$ in (3.1). If $\gamma(\beta) = \beta_1$, then $t = 0$, $b = 0$ in (3.2) and (3.3) can be used to show the nonexistence of a nontrivial $(1 - \alpha > 0)$ confidence interval for β_1 with finite expected length, while if $\gamma(\beta) = \beta_0$ then $t = 1$, $b = 1$ will suffice.

The preceding 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 a standard nonlinear regression model, where nontrivial confidence intervals for individual elements of β having finite expected length can 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 also hold in the context 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 in the linear errors-in-variables case (3.4), see Anderson (1984) and 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 apply in the functional case of the model (3.1), it is sufficient that the joint density $f(e)$ for $e =$

$(e_1', e_2', \dots, e_n')'$ satisfies the following conditions:

(i) $f(e)$ is functionally independent of β;
(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 $\beta, u_1, \dots, u_n$ 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 nonidentical marginal distributions. In fact, the u_i's need not have common mean vector μ nor common covariance matrix Σ_u. 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 listed properties (i), (ii) and (iii) 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.]

The key to our arguments in both the functional and structural cases of the 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 $\Sigma_u = \mathbf{O}_{q\times q}$ is on the boundary of the parameter space.

NOTE. Hwang (1986) considers a simple errors-in-variables model with a multiplicative error [rather than additive, as in (1.2), (3.1) or (3.4)]. 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.

3.2. *Estimation of principal component vectors.* It is well known that the structural form of the linear errors-in-variables model (3.4) is related to principal component analysis. Suppose that $y_1, y_2, \dots, y_n$ are i.i.d. p-dimensional continuous random vectors with support E^p, mean vector μ and unknown positive definite covariance matrix Σ. Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p > 0$ 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 the first component of x is not zero. 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 $\beta_1, \dots, \beta_{p-1}$ of β or a confidence set for β may be desired.

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

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i, \qquad i = 1, 2, \dots, n, \tag{3.6}$$

where the e_i have mean 0, variance σ_e^2 and a joint distribution having the properties (i), (ii) and (iii) mentioned in Section 3.1. For any given j_0, $0 \le j_0 \le p$, define the ratios $\delta_j = \beta_{j_0}^{-1}\beta_j$, $j = 0, \dots, p$.

Applying Theorems 1 and 2 with

$$\theta_1 = (\delta_0, \dots, \delta_p)' \qquad \theta_2 = (\beta_{j_0}, \sigma_e^2)' \qquad \theta_2^* = (0, \sigma_e^2)',$$

we see that no nontrivial confidence sets exist for δ_j, $j \ne j_0$, which have finite expected diameter. A special case of this problem (with $p = 1$) is the *inverse regression* (discrimination, calibration) problem [Miller (1981), page 117; Brown (1982); Seber (1977), Chapter 7].

4. An extension of Theorem 1. The results so far in this paper have concerned scalar functions $\gamma(\theta_1)$ of θ_1 having infinite range. Suppose, instead, that we are interested in a scalar function $\gamma(\theta_1)$ which has a *finite* range $[a, b]$ over the domain Θ_1, $-\infty < a < b < \infty$. Since $[a, b]$ itself is a confidence interval for $\gamma(\theta_1)$ with confidence 1 and finite length $b - a$, we cannot obtain the conclusion reached in Theorem 1. However, since the range of $\gamma(\theta_1)$ is a finite interval, it makes intuitive sense to restrict attention only to confidence sets for $\gamma(\theta_1)$ which are intervals $[L(Y), U(Y)]$. In the context of Theorem 1, it is then possible to show that any confidence interval for $\gamma(\theta_1)$ with confidence $1 - \alpha$ exceeding 0.5 must have positive probability of containing the entire range $[a, b]$ of $\gamma(\theta_1)$.

THEOREM 3. *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*

$$\textit{the closure of } \{g\colon \gamma(\theta_1) = g, \theta_1 \in \Theta_1^*\} \textit{ equals } [a, b]$$

and such that for each fixed $\theta_1 \in \Theta_1^*$, $Y \in \mathscr{Y}$, the limit (1.5) *exists, is a density for Y relative to ζ and is independent of θ_1. Then, every confidence interval $[L(Y), U(Y)]$ for $\gamma(\theta_1)$ with confidence $1 - \alpha > \frac{1}{2}$ satisfies*

$$P_{(\theta_1, \theta_2)}\{L(Y) \le a < b \le U(Y)\} > 0 \tag{4.1}$$

for all $(\theta_1, \theta_2) \in \Theta$.

PROOF OF THEOREM 3. The proof of (4.1) follows the steps of the proof of Theorem 1 in Section 2 until we reach (2.3). Now we find a sequence of values θ_1

in Θ_1^* such that $\gamma(\theta_1) \to a$ and a sequence of values θ_1 in Θ_1^* such that $\gamma(\theta_1) \to b$. Using (2.3) and the Lebesgue dominated convergence theorem,

$$\tfrac{1}{2} < 1 - \alpha \le \int_{\mathscr{Y}} I(\{L(Y) \le a\}) f(Y|\theta_2^*)\, d\zeta(Y),$$

$$\tfrac{1}{2} < 1 - \alpha \le \int_{\mathscr{Y}} I(\{U(Y) \ge b\}) f(Y|\theta_2^*)\, d\zeta(Y).$$

Hence,

$$\int_{\mathscr{Y}} I(\{L(Y) \le a < b \le U(Y)\}) f(Y|\theta_2^*)\, d\zeta(Y) > 1 - \tfrac{1}{2} - \tfrac{1}{2} = 0. \tag{4.2}$$

Let T be defined by (2.4) and let

$$S^* = \{Y\colon [L(Y), U(Y)] \text{ contains } [a, b]\}.$$

Then (4.2) yields

$$0 < \int_{S^*} f(Y|\theta_2^*)\, d\zeta(Y) = \int_{S^* \cap T} f(Y|\theta_2^*)\, d\zeta(Y).$$

The remainder of the proof is similar to that of the first part of Theorem 1. □

4.1. *Estimation of mixture proportions.* Let $g(y|\eta)$, $\eta \in H$, be an identifiable parametric family of density functions indexed by the q-dimensional parameter η. Let

$$f(y|\theta_1, \eta_1, \eta_2) = \theta_1 g(y|\eta_1) + (1 - \theta_1) g(y|\eta_2),$$

where $\theta_1 \in [0, 1]$, $\eta_1 \ne \eta_2$, $\eta_1, \eta_2 \in H$. Let $Y = (y_1', y_2', \dots, y_n')'$, where $y_1, y_2, \dots, y_n$ is a random sample (i.i.d.) from $f(y|\theta_1, \eta_1, \eta_2)$. Here, the mixing parameter θ_1, $0 \le \theta_1 \le 1$, is unknown, as is the "nuisance" parameter $\theta_2 = (\eta_1', \eta_2')'$. If for each y, $g(y|\eta)$ is continuous at η_0, then Theorem 3 applies [with $\theta_2^* = (\eta_0', \eta_0')'$] to show that every $1 - \alpha$ confidence interval for θ_1 (with $1 - \alpha > \frac{1}{2}$) has positive probability of containing $[0, 1]$, all $\theta = (\theta_1, \eta_1', \eta_2')'$.
Similar results clearly hold for the estimation of any mixing proportion θ_{1j} in the model

$$f(y|\theta_1, \eta_1, \eta_2, \dots, \eta_m) = \sum_{j=1}^{m} \theta_{1j} g(y|\eta_j),$$

where $\theta_1 = (\theta_{11}, \dots, \theta_{1m})$, $m \ge 2$, $\theta_{1j} \ge 0$, all j and $\sum_{j=1}^{m} \theta_{1j} = 1$.

4.2. *Estimation of the location parameter of the von Mises distribution on the circle.* Let y have density

$$f(y|\theta_1, \theta_2) = \kappa(\theta_2)\exp\{\theta_2 \cos(y - \theta_1)\}, \qquad y \in [0, 2\pi), \tag{4.3}$$

where $\theta_1 \in [0, 2\pi)$, $-\infty < \theta_2 < \infty$, $\theta_2 \ne 0$. Then for $Y = (y_1', y_2', \dots, y_n')'$, a random sample (i.i.d.) from the density $f(y|\theta_1, \theta_2)$, we conclude from Theorem 3 (with $\theta_2^* = 0$) that every $1 - \alpha$ confidence interval for θ_1, with $1 - \alpha \geq \frac{1}{2}$, must contain $[0, 2\pi)$ with positive probability, all $\theta_1 \in [0, 2\pi)$, $-\infty < \theta_2 < \infty$, $\theta_2 \neq 0$.

5. Comments. In each of the examples in Sections 3 and 4, our results are not due to lack of identifiability of the parameters. [In the functional errors-in-variables models, we can delete the line $u_1 = u_2 = \cdots = u_n$ from the parameter space and our conclusions still hold. The value θ_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 bounded above by the limit of coverage probabilities 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 that *large sample approximations* (asymptotic theory) *fail to uniformly approximate the finite sample distributions* over the parameter space Θ. To use large sample approximations for the models discussed in Sections 3 and 4 (and more generally in Theorems 1, 2 and 3), 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.

The models and inference problems mentioned in Sections 3 and 4 have wide applicability. Consequently, the nonexistence of nontrivial finite-expected-length confidence intervals is of concern, at least to those statisticians who use confidence intervals as frequentist indicators of precision or as estimators in their own right. Since confidence sets (particularly confidence intervals) follow from a frequentist approach to inference, our discussion in this paper has been confined to frequentist measures of accuracy (coverage probability, confidence, expected length or diameter). No attempt has been made to take a Bayesian perspective on this problem. However, the results of this paper do serve as a warning to Bayesians that Bayesian methods of inference will be sensitive to the amount of prior probability mass or density for θ_2 in neighborhoods of the values θ_2^*.

Acknowledgments. We wish to thank the referees for their helpful comments. In particular, Theorem 3 and the examples in Section 4 were added as a result of their remarks.

REFERENCES

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

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

BROWN, P. J. (1982). Multivariate calibration (with discussion). *J. Roy. Statist. Soc. Ser. B* **44** 287–321.

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

FIELLER, E. C. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B* **16** 175–185.

GLESER, L. J. (1983). Functional, structural and ultrastructural errors-in-variables models. *1983 Proc. Bus. Econ. Statist. Sect.* 57–66. Amer. Statist. Assoc., Washington.

HWANG, J. T. (1986). Multiplicative errors-in-variables models with applications to recent data released by the U.S. Department of Energy. *J. Amer. Statist. Assoc.* **81** 680–688.

MILLER, R. G., JR. (1981). *Simultaneous Statistical Inference*, 2nd ed. McGraw-Hill, New York.

NEYMAN, J. (1954). Discussion on the symposium on interval estimation. *J. Roy. Statist. Soc. Ser. B* **16** 216–218.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

SCHNEEWEISS, H. (1982). Note on Creasy's confidence limits for the gradient in the linear functional relationship. *J. Multivariate Anal.* **12** 155–157.

SEBER, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF MATHEMATICS
WHITE HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7901