The Annals of Statistics
1993, Vol. 21, No. 4, 1780–1791

BOOTSTRAP CONFIDENCE REGIONS FOR FUNCTIONAL RELATIONSHIPS IN ERRORS-IN-VARIABLES MODELS

BY JAMES G. BOOTH AND PETER HALL

University of Florida, and Australian National University and CSIRO and CSIRO

We suggest bootstrap methods for constructing confidence bands (and intervals) for an unknown linear functional relationship in an errors-in-variables model. It is assumed that the ratio of error variances is known to lie within an interval $\Lambda = [\lambda_1, \lambda_2]$. A confidence band is constructed for the range of possible linear relationships when $\lambda \in \Lambda$. Meaningful results are obtained even in the extreme case $\Lambda = [0, \infty]$, which corresponds to no assumption being made about Λ. The bootstrap bands have several interesting features, which include the following: (i) the bands do not shrink to a line as $n \to \infty$, unless Λ is a singleton (i.e., $\lambda_1 = \lambda_2$); (ii) percentile-t versions of the bands enjoy only first-order coverage accuracy, not the second-order accuracy normally found in simpler statistical problems.

1. Introduction. Linear regression models describe linear functional relationships when there is observation error in only one variable, the regressor Y. The regressand X is assumed to be measured precisely. On the other hand, errors-in-variables models apply to situations where both regressor and regressand are subject to error. For example, the observed value of X may represent the outcome of a sample survey or some other sampling experiment and therefore may be a somewhat imprecise measure of the true variable, say, U, in the linear relationship.

More concisely, we assume that the true linear relationship is given by $V = a + bU$, where a and b are fixed and U and V are only observed subject to error. The actual observed values are $X = U + \xi$ and $Y = V + \eta$, where U, ξ and η are independent random variables with means μ, 0 and 0 and variances σ_U^2, σ_ξ^2 and σ_η^2. From a set of random data $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ generated by this model, we wish to estimate the functional relationship $v = a + bu$.

It is well known [see, e.g., Kendall and Stuart (1979), Chapter 29] that a and b are often not identifiable. For example, they are not identifiable in the case where U, ξ and η have a joint normal distribution. However, if the variance ratio $\lambda = \sigma_\eta^2/\sigma_\xi^2$ is known, then both a and b may be estimated consistently, assuming only that the variances σ_U^2, σ_ξ^2 and σ_η^2 are finite. In practice it is often possible to determine bounds for λ, for example, $\frac{1}{2} \le \lambda \le 2$ or even $0 \le \lambda \le \infty$. The first of these bounds expresses a belief that the unknown variables σ_ξ^2 and σ_η^2 differ by no more than a factor of 2; the second bound amounts to no constraint at all on either σ_ξ^2 or σ_η^2.

Received May 1991; revised June 1992.

AMS 1991 *subject classifications*. Primary 62J05; secondary 62G15.

Key words and phrases. Bootstrap, confidence band, confidence interval, Edgeworth expansion, errors-in-variables model, functional relationship, percentile method, percentile-t method, structural relationship.

If one assumes that $\lambda \in \Lambda = [\lambda_1, \lambda_2]$, then it is possible to estimate a range of functional relationships, given by $\{\hat{a}_\lambda + \hat{b}_\lambda u,\ \lambda \in \Lambda\}$, say, within which the true relationship $a + bu$ might lie. Here, $\hat{a}_\lambda$ and $\hat{b}_\lambda$ represent estimates of a and b, respectively, under the hypothesis that the true variance ratio equals λ. Our purpose in this paper is to suggest more concise ways in which sample information about the formula $a + bu$ might be represented over a range of values u, say, $u \in \mathscr{U}$, under the assumption that $\lambda \in \Lambda$.

Quite apart from the considerable practical importance of errors-in-variables models, for example, in statistics [Fuller (1987)], economics [Zellner (1970), Goldberger (1972), Griliches (1974) and Maddala (1988), Chapter 11] and the sciences generally [e.g., Jones (1979)], this problem has a number of intrinsically interesting features which set it aside from related work on confidence regions and motivate its study. First of all, confidence bands in this problem do not shrink to a line as $n \to \infty$, but instead converge to another band $\mathscr{B}^\circ$, say, of nonzero width. This property reflects uncertainty as to the true value of the variance ratio. Second, the confidence bands, and their limit $\mathscr{B}^\circ$, are well defined and readily interpretable even in the extreme case where $\Lambda = [0, \infty]$ and $\mathscr{U} = (-\infty, \infty)$. Third, there is no preeminent classical competitor to the bootstrap in this context, since it is not practically feasible to construct a band based on the normal approximation. Fourth, the percentile-t bootstrap method produces a confidence band that is only first-order correct, not second-order correct. This is a consequence of the fact that although Studentizing standardizes the variances of statistics used to construct the confidence band, it does not adequately standardize the covariances. A second level of bootstrap simulation is suggested for calibrating the basic bootstrap confidence band and improving coverage accuracy, much as suggested by Loh (1987) in simpler problems.

To describe our proposal, write a_λ and b_λ for values taken by a and b when the variance ratio equals λ. (That is, a_λ and b_λ are the limits of $\hat{a}_\lambda$ and $\hat{b}_\lambda$ as $n \to \infty$. If λ_0 denotes the true value of λ, then $a_{\lambda_0} = a$ and $b_{\lambda_0} = b$.) We shall suggest ways of constructing a confidence band $\mathscr{B}$, with the property that, for a given probability α (such as $\alpha = 0.95$), the pair $(u, a_\lambda + b_\lambda u)$ lies inside $\mathscr{B}$ for all $\lambda \in \Lambda$ and all $u \in \mathscr{U}$. In particular, if our claim that $\lambda \in \Lambda$ is correct, then the band $\mathscr{B}$ will contain the linear relationship, over the range $u \in \mathscr{U}$, with probability at least α.

In practice, construction of $\mathscr{B}$ requires a substantial amount of information about a particular stochastic process, and that information would usually not be readily available, even under the assumption that (U, ξ, η) are normally distributed. Therefore, we suggest bootstrap methods for computing an approximation $\hat{\mathscr{B}}$ to $\mathscr{B}$. The boundaries of $\mathscr{B}$ and $\hat{\mathscr{B}}$ are $O_p(n^{-1})$ apart, and the coverage level of $\hat{\mathscr{B}}$ is asymptotically correct. If $\Lambda = [\lambda_1, \lambda_2]$, then both may be regarded as confidence bands for the region.

$$\begin{aligned}\mathscr{B}^\circ = \big\{(u,v):\, u \in \mathscr{U},\, &\min\big(a_{\lambda_1} + b_{\lambda_1}u,\, a_{\lambda_2} + b_{\lambda_2}u\big) \\ &\le v \le \max\big(a_{\lambda_1} + b_{\lambda_1}u,\, a_{\lambda_2} + b_{\lambda_2}u\big)\big\}\end{aligned} \tag{1.1}$$

to which they converge.

Section 2 introduces the basic estimators and describes methods for constructing confidence bands. In Section 3 we summarize the results of a simulation study and consider an application to porosity and permeability measurements on sandstone core samples. Section 4 presents a theoretical comparison of the "ideal" band $\mathscr{B}$ and its bootstrap version $\hat{\mathscr{B}}$.

2. Methodology.

2.1. *Estimators of a and b.* If the true value λ_0 of the variance ratio $\lambda = \sigma_\eta^2/\sigma_\xi^2$ were known, then the value of b could be expressed as b_{λ_0}, where

$$b_\lambda = H_\lambda + \left(H_\lambda^2 + \lambda\right)^{1/2} \operatorname{sgn}(\sigma_{XY}),$$
$$H_\lambda = \left(\sigma_Y^2 - \lambda\sigma_X^2\right)/(2\sigma_{XY}),$$
$$\sigma_X^2 = \operatorname{var}(X), \qquad \sigma_Y^2 = \operatorname{var}(Y), \qquad \sigma_{XY} = \operatorname{cov}(X, Y).$$

The maximum likelihood estimator of b_λ in a normal model, and also the least squares estimator when both σ_ξ^2 and σ_η^2 are known, is

$$\hat{b}_\lambda = \hat{H}_\lambda + \left(\hat{H}_\lambda^2 + \lambda\right)^{1/2} \operatorname{sgn}(S_{XY}),$$

where $\hat{H}_\lambda = (S_Y^2 - \lambda S_X^2)/(2S_{XY})$, $S_X^2 = n^{-1}\Sigma(X_i - \bar{X})^2$, $S_Y^2 = n^{-1}\Sigma(Y_i - \bar{Y})^2$ and $S_{XY} = n^{-1}\Sigma(X_i - \bar{X})(Y_i - \bar{Y})$. The analogous estimator of a is $\hat{a}_\lambda = \bar{Y} - \hat{b}_\lambda\bar{X}$, which converges to $a_\lambda = a + (b - b_\lambda)\mu$ as $n \to \infty$. Of course, $a_\lambda = a$ if λ is the true value of the variance ratio. See Kendall and Stuart [(1979), Chapter 29, especially page 405] for a detailed development of these results.

The differences $\hat{a}_\lambda - a_\lambda$ and $\hat{b}_\lambda - b_\lambda$ are asymptotically normally distributed with zero mean and variance of order n^{-1}, in a very wide range of circumstances. For example, it is sufficient that $\sigma_{XY} \neq 0$ and (X, Y) have finite fourth moments.

Throughout Section 2 we shall assume that $\sigma_{XY} > 0$. In that case, assuming all moments of X and Y are finite, we have $P(S_{XY} > 0) = 1 - O(n^{-C})$ for all $C > 0$. Therefore, the sign of σ_{XY} may be determined with considerable accuracy, and so we may take

$$b_\lambda = H_\lambda + \left(H_\lambda^2 + \lambda\right)^{1/2}, \qquad \hat{b}_\lambda = \hat{H}_\lambda + \left(\hat{H}_\lambda^2 + \lambda\right)^{1/2}.$$

The case where $\sigma_{XY} < 0$ may obviously be treated by changing the sign of X or Y, and b_λ is not well defined if $\sigma_{XY} = 0$.

Our estimator of the functional relationship $v = a + bu$ is $v = \hat{a}_\lambda + \hat{b}_\lambda u$, assuming that the variance ratio equals λ. Note that $\hat{b}_\lambda$ is a decreasing function of λ.

2.2. *Variance of the functional relationship estimator.* For any random variable W, with finite expectation, define the operator δ by $\delta W = W - EW$. Thus, for example, $\delta(\delta W)^j = (\delta W)^j - E(\delta W)^j$. Similarly, if $W_1, \ldots, W_n$ are random variables, define $dW_i = W_i - \overline{W}$, $i = 1, \ldots, n$. For $j, k \geq 1$, let $\overline{X}_j = n^{-1}\Sigma\delta(\delta X_i)^j$, $\overline{Y}_j = n^{-1}\Sigma\delta(\delta Y_i)^j$ and $\overline{Z}_{jk} = n^{-1}\Sigma\delta\{(\delta X_i)^j(\delta Y_i)^k\}$. Then a little

asymptotic analysis shows that

$$\hat{b}_\lambda - b_\lambda = K_\lambda\left(\tfrac{1}{2}\overline{Y}_2 - \tfrac{1}{2}\lambda\overline{X}_2 - H_\lambda\overline{Z}_{11}\right) + O_p(n^{-1}),$$

where $K_\lambda = \sigma_{XY}^{-1} b_\lambda (H_\lambda^2 + \lambda)^{-1/2}$. Hence,

$$\begin{aligned}\Delta_\lambda(u) &= \hat{a}_\lambda + \hat{b}_\lambda u - (a_\lambda + b_\lambda u)\\ &= \overline{Y}_1 - b_\lambda\overline{X}_1 + K_\lambda\left(\tfrac{1}{2}\overline{Y}_2 - \tfrac{1}{2}\lambda\overline{X}_2 - H_\lambda\overline{Z}_{11}\right)(u-\mu) + O_p(n^{-1}).\end{aligned} \tag{2.1}$$

It follows that the asymptotic variance of $n^{1/2}\,\Delta_\lambda(u)$ is given by

$$\begin{aligned}w_\lambda(u)^2 &= \operatorname{var}\Big[\delta Y_i - b_\lambda\,\delta X_i\\ &\qquad + K_\lambda\,\delta\left\{\tfrac{1}{2}(\delta Y_i)^2 - \tfrac{1}{2}\lambda(\delta X_i)^2 - H_\lambda\,\delta X_i\,\delta Y_i\right\}(u-\mu)\Big]\\ &= \operatorname{var}\{Q_{1i} + Q_{2i}(u-\mu)\},\end{aligned}$$

say. A consistent estimator of $w_\lambda(u)^2$ is therefore

$$\hat{w}_\lambda(u)^2 = n^{-1}\sum_{i=1}^{n}\left\{\hat{Q}_{1i} + \hat{Q}_{2i}(u - \overline{X})\right\}^2,$$

where $\hat{Q}_{1i} = dY_i - \hat{b}_\lambda\,dX_i$,

$$\hat{Q}_{2i} = \hat{K}_\lambda\,d\left\{\tfrac{1}{2}(dY_i)^2 - \tfrac{1}{2}\lambda(dX_i)^2 - \hat{H}_\lambda\,dX_i\,dY_i\right\}$$

and $\hat{K}_\lambda = S_{XY}^{-1}\hat{b}_\lambda(\hat{H}_\lambda^2 + \lambda)^{-1/2}$. It is straightforward to prove that the estimator

$$T_\lambda(u) = n^{1/2}\,\Delta_\lambda(u)/\hat{w}_\lambda(u) \tag{2.2}$$

is asymptotically normal $N(0,1)$.

2.3. *The cases* $\lambda = 0, \infty$. The cases $\lambda = 0, \infty$ correspond to the models

$$X_i = -ab^{-1} + b^{-1}Y_i + \xi_i, \qquad Y_i = a + bX_i + \eta_i,$$

respectively; that is, they correspond to regression of X on Y and of Y on X, respectively. The corresponding estimators of b are $\hat{b}_0 = S_Y^2/S_{XY}$ and $\hat{b}_\infty = S_{XY}/S_X^2$. Clearly, $\hat{w}_0^2(u)$ is well defined and consistently estimates $n^{1/2}\,\Delta_0(u)$. Furthermore, $\hat{w}_\lambda(u) \to \hat{w}_\infty(u)$ as $\lambda \to \infty$, where

$$\hat{w}_\infty(u)^2 = n^{-1}\sum_{i=1}^{n}\left\{dY_i - \hat{b}_\infty\,dX_i - \hat{b}_\infty\,d\left(S_X^{-2}(dX_i)^2 - S_{XY}^{-1}\,dX_i\,dY_i\right)(u - \overline{X})\right\}^2$$

and consistently estimates the asymptotic variance of $n^{1/2}\,\Delta_\infty(u)$.

Thus, the estimators $\hat{a}_\lambda$, $\hat{b}_\lambda$ and $\hat{w}_\lambda$ are as valid when $\lambda = 0$ or ∞ as they are for $0 < \lambda < \infty$. Note that the classical regression estimator $\hat{b}_\infty$ is a lower bound to the set $\{\hat{b}_\lambda,\ \lambda \geq 0\}$.

2.4. *The concept of confidence bands.* Let $\Lambda = [\lambda_1, \lambda_2]$ and $\mathscr{U} = [u_1, u_2]$ be closed subintervals of $[0, \infty]$ and $(-\infty, \infty)$, respectively. We shall construct a confidence band for the functional relationship $v = a + bu$, over values u in

the interval $\mathscr{U}$, under the hypothesis that $\lambda \in \Lambda$. Either of Λ or $\mathscr{U}$ may be degenerate or unbounded; in the case where $\mathscr{U} = \{u\}$, a singleton, we obtain a confidence interval rather than a confidence band. There is no particular reason to insist that $\mathscr{U}$ be an interval. Indeed, our methodology applies equally to any finite union of intervals, although practical considerations usually demand that $\mathscr{U}$ be contiguous.

Let $T_\lambda(u)$ be as at (2.2), and let α be the coverage level desired of a confidence band. We wish to determine random functions $g_1 \le g_2$ such that

$$P\{g_1(u) \le a_\lambda + b_\lambda u \le g_2(u), \text{ for all } u \in \mathscr{U} \text{ and all } \lambda \in \Lambda\} = \alpha. \tag{2.3}$$

The confidence band would then be

$$\mathscr{B} = \{(u, v): g_1(u) \le v \le g_2(u), u \in \mathscr{U}\}. \tag{2.4}$$

Let t be the solution of the equation

$$P\Big\{\sup_{u \in \mathscr{U},\, \lambda \in \Lambda} |T_\lambda(u)| \le t\Big\} = \alpha, \tag{2.5}$$

and define

$$g_1(u) = \inf_{\lambda \in \Lambda} \big\{\hat{a}_\lambda + \hat{b}_\lambda u - n^{-1/2} t \hat{w}_\lambda(u)\big\}, \tag{2.6a}$$

$$g_2(u) = \sup_{\lambda \in \Lambda} \big\{\hat{a}_\lambda + \hat{b}_\lambda u + n^{-1/2} t \hat{w}_\lambda(u)\big\}. \tag{2.6b}$$

Then (2.3) holds.

2.5. *Bootstrap confidence bands.* In practice, the value of t defined by (2.5) would usually be unobtainable. However, t may be estimated by bootstrap methods, as follows. Let $\mathscr{X}^* = \{(X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*)\}$ denote a resample drawn randomly, with replacement, from $\mathscr{X} = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. Define $\hat{a}_\lambda^*, \hat{b}_\lambda^*, \hat{w}_\lambda^*, \ldots$ as for $\hat{a}_\lambda, \hat{b}_\lambda, \hat{w}_\lambda, \ldots$, except that $\mathscr{X}^*$ should replace $\mathscr{X}$ throughout. Put either

$$T_\lambda^*(u) = n^{1/2}\big\{\hat{a}_\lambda^* + \hat{b}_\lambda^* u - \big(\hat{a}_\lambda + \hat{b}_\lambda u\big)\big\}/\hat{w}_\lambda(u) \tag{2.7}$$

or

$$T_\lambda^*(u) = n^{1/2}\big\{\hat{a}_\lambda^* + \hat{b}_\lambda^* u - \big(\hat{a}_\lambda + \hat{b}_\lambda u\big)\big\}/\hat{w}_\lambda^*(u). \tag{2.8}$$

For either of these definitions, let the bootstrap estimate $\hat{t}$ of t be the solution of the equation

$$P\Big\{\sup_{u \in \mathscr{U},\, \lambda \in \Lambda} |T_\lambda^*(u)| \le \hat{t} \,\Big|\, \mathscr{X}\Big\} = \alpha, \tag{2.9}$$

and let $\hat{g}_1$ and $\hat{g}_2$ denote the versions of g_1 and g_2 in which t is replaced by $\hat{t}$. Then the bootstrap version of $\mathscr{B}$ is

$$\hat{\mathscr{B}} = \{(u, v): \hat{g}_1(u) \le v \le \hat{g}_2(u), u \in \mathscr{U}\}. \tag{2.10}$$

TABLE 1
Parameter configurations used in simulations

Population	$(p_\xi, \sigma_{\xi 1}, \sigma_{\xi 2})$	$(p_\eta, \sigma_{\eta 1}, \sigma_{\eta 2})$	λ_0	θ_0 / π	ρ_0
1	(0.50, 0.20, 0.50)	(0.50, 0.50, 0.75)	2.80	0.391	0.788
2	(0.50, 0.10, 0.20)	(0.25, 0.20, 0.50)	7.90	0.460	0.903
3	(1.0, 0.10, —)	(1.0, 0.50, —)	25.0	0.487	0.890
4	(0.25, 0.50, 0.75)	(0.50, 0.75, 0.95)	1.51	0.314	0.624

If definition (2.7) is employed for T_λ^*, then we are using the percentile method; the method is percentile-t if we invoke (2.8). In simpler applications of the bootstrap, the percentile-t method usually produces confidence regions $\hat{\mathscr{B}}$ whose boundaries are second-order correct for those of the "ideal" region $\mathscr{B}$; that is, the boundaries differ only in terms of order $n^{-3/2}$. However, as we shall show in Section 4, the multivariate nature of the present problem means that the percentile-t method produces only first-order correct regions, with the boundaries differing by order n^{-1}.

3. Numerical work. Tables 1–3 summarize the results of a simulation study in which the random variable U is assumed to have a standard normal distribution and ξ and η are each mixtures of two zero-mean normal variables. More precisely, the distribution of (U, ξ, η) has the form

$$U \sim Z(0,1), \tag{3.1a}$$

$$\xi \sim \Delta_\xi Z\left(0, \sigma_{\xi 1}^2\right) + (1 - \Delta_\xi) Z\left(0, \sigma_{\xi 2}^2\right), \tag{3.1b}$$

$$\eta \sim \Delta_\eta Z\left(0, \sigma_{\eta 1}^2\right) + (1 - \Delta_\eta) Z\left(0, \sigma_{\eta 2}^2\right), \tag{3.1c}$$

where $\sim$ denotes distributional equivalence, and where Δ_ξ, Δ_η, $Z(0,1)$, $Z(0, \sigma_{\xi i}^2)$ and $Z(0, \sigma_{\eta i}^2)$, $i = 1, 2$, are independent random variables; Δ_ξ and Δ_η are Bernoulli variables with associated success probabilities p_ξ and p_η, respec-

TABLE 2
Empirical coverages of bootstrap confidence regions with nominal coverage level $\alpha = 0.9$ *(1000 simulations)*

Population	Θ / π	Percentile $n = 25, 50$	Percentile-t $n = 25, 50$
1	[0.00, 0.50]	0.867, 0.881	0.933, 0.920
	[0.25, 0.45]	0.865, 0.873	0.912, 0.907
2	[0.00, 0.50]	0.857, 0.875	0.922, 0.919
	[0.40, 0.50]	0.858, 0.853	0.927, 0.892
3	[0.00, 0.50]	0.872, 0.881	0.926, 0.914
	[0.45, 0.50]	0.861, 0.857	0.908, 0.887
4	[0.00, 0.50]	0.885, 0.900	0.935, 0.920
	[0.25, 0.40]	0.872, 0.882	0.904, 0.897

TABLE 3
Summary statistics for simulations from population 1

			Percentile			Percentile-t		
n	Θ/π	u	Mean $\hat{g}_1(u)$	Mean $\hat{g}_2(u)$	Std. dev. $\hat{g}_2 - \hat{g}_1$	Mean $\hat{g}_1(u)$	Mean $\hat{g}_2(u)$	Std. dev. $\hat{g}_2 - \hat{g}_1$
25	[0.00, 0.50]	0.0	−0.563	0.549	1.228	−0.663	0.649	1.419
		1.0	0.285	2.327	2.649	0.160	2.484	2.867
		2.0	0.829	4.391	4.705	0.624	4.658	5.076
	[0.25, 0.45]	0.0	−0.415	0.405	0.859	−0.480	0.470	0.990
		1.0	0.417	1.771	1.552	0.326	1.872	1.712
		2.0	1.051	3.340	2.671	0.901	3.510	2.933
50	[0.00, 0.50]	0.0	−0.352	0.340	0.708	−0.384	0.371	0.773
		1.0	0.522	1.949	1.490	0.487	2.001	1.582
		2.0	1.206	3.733	2.645	1.150	3.820	2.799
	[0.25, 0.45]	0.0	−0.273	0.263	0.543	−0.294	0.284	0.586
		1.0	0.598	1.542	0.976	0.569	1.575	1.040
		2.0	1.341	2.953	1.673	1.294	3.009	1.782

tively, and $Z(\mu, \sigma^2)$ represents a normal random variable with mean μ and variance σ^2.

Bivariate (X, Y)-samples $(X_1, Y_1), \ldots, (X_n, Y_n)$ were generated by first generating n independent random vectors (U_i, ξ_i, η_i), $i = 1, \ldots, n$, with distribution (3.1) and setting

$$X_i = U_i + \xi_i, \qquad Y_i = a + bU_i + \eta_i, \qquad i = 1, \ldots, n.$$

The parameters a and b of the linear functional relationship $v = a + bu$ are fixed throughout the study at 0 and 1, respectively, and $\mathscr{U}$ is fixed as the interval $[-3, 3]$. Four different configurations of the population parameter vectors $(p_\xi, \sigma_{\xi 1}, \sigma_{\xi 2})$ and $(p_\eta, \sigma_{\eta 1}, \sigma_{\eta 2})$ are considered. These are given in Table 1 together with the corresponding values of λ_0, $\theta_0 = \tan^{-1}(\lambda_0)$ and $\rho_0 = \text{cor}(X, Y)$. The empirical coverages in Table 2 are based on 1000 simulated samples of sizes $n = 25$ and 50 from each population. Bootstrap confidence bands were constructed using $B = 200$ resamples from each simulated (X, Y)-sample.

Computation of the bootstrap confidence region $\hat{\mathscr{B}}$ involves finding the supremum of $T_\lambda^*(u)$ over all $\lambda \in \Lambda$ and $u \in \mathscr{U}$. For fixed λ, $T_\lambda^*(u)$ is the ratio of a linear function in u to the square root of a positive quadratic function in u. Using this fact it is easy to show that, for each fixed λ, $T_\lambda^*(u)$ either decreases in u on each side of a unique maximum or increases on each side of a unique minimum. Hence, for each fixed $\lambda \in \Lambda$, it is only necessary to compute $|T_\lambda^*(u)|$ at three points when searching for its supremum over the interval $\mathscr{U}$, the maximum or minimum value and the endpoints of the interval. Maximization over Λ for each fixed value of u is achieved in practice by conducting a grid search. Since Λ can be an infinite interval, it is more convenient to work with the parameter $\theta = \tan^{-1}\lambda$, which takes values in the

closed interval $[0, \frac{1}{2}\pi]$. Let $\Theta = \tan^{-1}\Lambda$. All of the results presented in this section are based on grid searches over 11 equally spaced points over the range of Θ. This last comment also applies to the computation of $\hat{g}_1(u)$ and $\hat{g}_2(u)$, the bootstrap versions of (2.6a) and (2.6b). In practice, the effect of grid size can be assessed by comparing results for a fine and a course grid of θ values.

Some summary statistics for the simulations from the first population are presented in Table 3. The corresponding statistics for negative values of u are omitted because of an approximate symmetry in the results about $u = 0$. For example, with $n = 25$ and $\Theta = [0.25\pi, 0.45\pi]$, the mean values of $\hat{g}_1(-1)$ and $\hat{g}_2(-1)$ were -1.787 and -0.417, respectively. Notice that the percentile-t confidence bands are wider than the percentile bands. This fact is reflected in the empirical coverages tabulated in Table 2; the empirical coverages of the percentile-t bands always exceed those of their percentile counterparts. The summary statistics for simulations from the other three populations were qualitatively similar to those for population 1, the main difference being that the bands were narrower for populations 2 and 3 and wider for population 4, reflecting the size of the error variances σ_ξ^2 and σ_η^2 in these cases.

As in more standard bootstrap problems [see, e.g., Loh (1987)], the coverage accuracy of the bootstrap confidence region (2.10) can be improved by iteration. For $0 < \alpha < 1$, let $\pi(\alpha) = P(\mathscr{B}^\circ \subset \hat{\mathscr{B}}_\alpha)$ denote the true coverage of the bootstrap confidence region for $\mathscr{B}^\circ$ with target coverage level α, as defined in (2.10), and let $\beta(\alpha)$ be the target coverage level for which $\pi\{\beta(\alpha)\} = \alpha$. Then $\hat{\mathscr{B}}_{\beta(\alpha)}$ is a confidence region for $\mathscr{B}^\circ$ with exact coverage level α. Unfortunately, since the precise form of the coverage function π is unknown, the exact value of $\beta(\alpha)$ is unobtainable. However, the function π can be estimated by its bootstrap version

$$\hat{\pi}(\alpha) = P\left(\hat{\mathscr{B}}^\circ \subset \hat{\mathscr{B}}_\alpha^* | \mathscr{X}\right), \tag{3.2}$$

where $\hat{\mathscr{B}}^\circ$ is the region between the estimated linear relationships corresponding to the two endpoints of Λ, and $\hat{\mathscr{B}}_\alpha^*$ is the version of $\hat{\mathscr{B}}_\alpha$ computed using a resample $\mathscr{X}^*$ in place of the original sample $\mathscr{X}$. Then the value $\hat{\beta}(\alpha)$ which satisfies $\hat{\pi}\{\hat{\beta}(\alpha)\} = \alpha$ is an estimate of $\beta(\alpha)$, and we call the region $\hat{\mathscr{B}}_{\hat{\beta}(\alpha)}$ an iterated bootstrap confidence region for $\mathscr{B}^\circ$. It is shown in Section 4 that the coverage level of $\hat{\mathscr{B}}_\alpha$ is in error by a term of order n^{-1}. On the other hand, the iterated bootstrap confidence region has coverage error of only $o(n^{-1})$; that is,

$$P\left\{\mathscr{B}^\circ \subset \hat{\mathscr{B}}_{\hat{\beta}(\alpha)}\right\} = \alpha + o(n^{-1}).$$

This result holds for iterated versions of both percentile and percentile-t bootstrap confidence regions.

Jones (1979) estimates an assumed linear functional relationship between the porosity (%) and log permeability (md.) of a certain kind of sandstone. His estimates are based on measurements of porosity and permeability taken from 101 core plugs. Both measurements were subject to error and so an error-in-variables approach is required. Graphs of percentile, percentile-t and iterated percentile confidence bands for $\mathscr{B}^\circ$ corresponding to two different choices of Λ

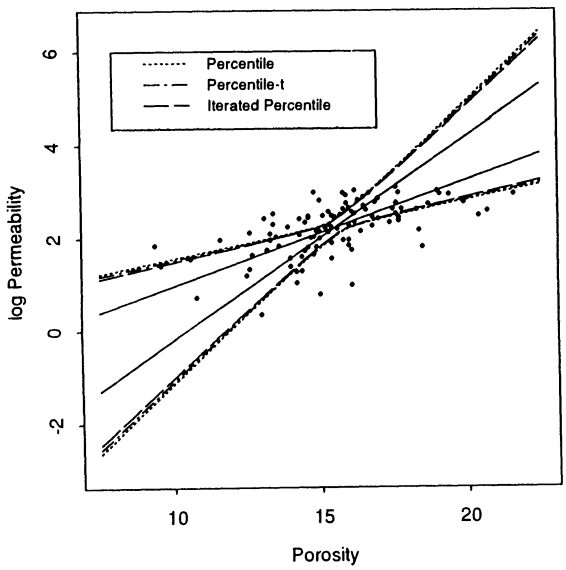

FIG. 1. *Confidence bands for linear functional relationships between porosity and the logarithm of permeability of sandstone core samples*: $\alpha = 0.9$, $\lambda = [0.0, 0.1]$.

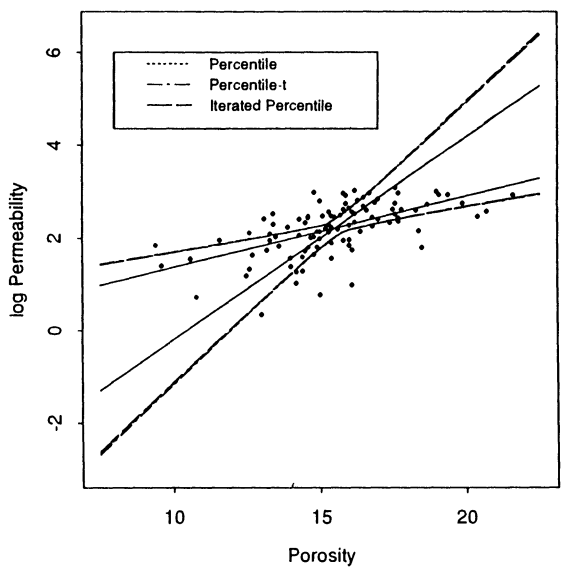

FIG. 2. *Confidence bands for linear functional relationships between porosity and the logarithm of permeability of sandstone core samples*: $\alpha = 0.9$, $\lambda = [0.0, \infty]$.

for this problem are depicted in Figures 1 and 2. The values of $\hat{\beta}(0.90)$ corresponding to $\Lambda = [0.0, 0.1]$ and $\Lambda = [0.0, \infty]$ are 0.84 and 0.87, respectively. The choice of $\Lambda = [0.0, 0.1]$ in Figure 1 is motivated by two different estimates, $\hat{\lambda} = 0.049$ and $\hat{\lambda} = 0.065$ given by Jones. In each figure the region $\hat{\mathscr{B}}^\circ$ between the two intersecting lines in the interior of the confidence bands is the sample estimate of the target region $\mathscr{B}^\circ$.

4. Accuracy with which $\hat{\mathscr{B}}$ approximates $\mathscr{B}$. Recall that $\hat{\mathscr{B}}$ is a bootstrap approximation to the "ideal" confidence band $\mathscr{B}$, the latter being defined by (2.4), with g_1 and g_2 given by (2.6). Both g_1 and g_2 depend on the value of $t = t(n)$, which converges to a finite number t^0 as $n \to \infty$, given by

$$P\Big\{ \sup_{u \in \mathscr{U},\, \lambda \in \Lambda} \big| T_\lambda^0(u) \big| \le t^0 \Big\} = \alpha,$$

where $\{T_0(u),\ \lambda \in \Lambda,\ u \in \mathscr{U}\}$ denotes a Gaussian process with zero mean and the same covariance structure as the limit as $n \to \infty$ of the covariance of $\{T_\lambda(u),\ \lambda \in \Lambda,\ u \in \mathscr{U}\}$.

Since $\lim t(n)$ is finite, the boundary of the confidence band $\mathscr{B}$ is $O_p(n^{-1/2})$ away from that of the "limiting" band $\mathscr{B}^0$; that is, the set of values (u, v) satisfying

$$\min\big(a_{\lambda_1} + b_{\lambda_1} u, a_{\lambda_2} + b_{\lambda_2} u\big) \le v \le \max\big(a_{\lambda_1} + b_{\lambda_1} u, a_{\lambda_2} + b_{\lambda_2} u\big)$$

for all $u \in \mathscr{U}$. We shall show that, under appropriate regularity conditions, the bootstrap estimate $\hat{t}$ of t differs from t only in terms of order $n^{-1/2}$. It then follows that the boundary of $\hat{\mathscr{B}}$ is $O_p(n^{-1})$ away from that of $\mathscr{B}$ and that it converges to that of $\mathscr{B}^0$ as $n \to \infty$.

Thus, the bootstrap band $\hat{\mathscr{B}}$ is first-order correct for $\mathscr{B}$: Its boundary agrees with that of $\mathscr{B}$ in terms of first order in $n^{-1/2}$. When one appreciates that even the asymptotic value t^0 could usually not be computed without a procedure of numerical approximation, such as simulation, it becomes clear that first-order correctness is a valuable property.

It might be thought that, if $\hat{t}$ is defined according to the percentile-t prescription at (2.8) and (2.9), then the bootstrap band $\hat{\mathscr{B}}$ should be second-order correct for $\mathscr{B}$; that is, the boundaries of $\mathscr{B}$ and $\hat{\mathscr{B}}$ should differ only in terms of order $n^{-3/2}$. This is not the case. The operation of Studentizing ensures only that asymptotic variances are constant, and asymptotic covariances will usually depend on unknowns. Thus, Studentizing does not have the effect of pivoting, unlike its role in simpler statistical problems [see, e.g., Hall (1988)].

We now state our main result.

THEOREM. *Assume that the distribution of (U, ξ, η) is constrained by the property that (X, Y) is nonsingular, and that, for some $C > 0$ sufficiently large, $E(\|X\|^C + \|Y\|^C) < \infty$. Then for either of the two prescriptions of the confidence band $\mathscr{B}$ and its bootstrap counterpart $\hat{\mathscr{B}}$, the boundary of $\hat{\mathscr{B}}$ is distant an amount of size n^{-1}, in probability, from that of $\mathscr{B}$.*

A proof of the theorem consists of showing that for either prescription of $\hat{t}$ from Section 2, $\hat{t} - t$ is of size $n^{-1/2}$ in probability. It follows that $\hat{t} \to t^0$ in probability. That result, together with the observation that t^0 $(= \lim t)$ is a strictly increasing function of α, is sufficient to imply that

$$P\{\hat{g}_1(u) \le a_\lambda + b_\lambda u \le \hat{g}_2(u), \text{ for all } u \in \mathscr{U} \text{ and all } \lambda \in \Lambda\} \to \alpha$$

as $n \to \infty$. Therefore, the bootstrap band $\hat{\mathscr{B}}$ has asymptotically correct coverage. A longer argument, based on considerations of symmetry and of the parity of terms in Edgeworth expansions, may be used to prove that

$$P\{\hat{g}_1(u) \le a_\lambda + b_\lambda u \le \hat{g}_2(u), \text{ for all } u \in \mathscr{U} \text{ and all } \lambda \in \Lambda\} = \alpha + O(n^{-1}).$$

The coverage accuracy may be improved to $o(n^{-1})$ by using the iterated bootstrap, as in Section 3.

We claim that, except for a small number of values of λ, which do not usually include the true value of λ, the probability that either $\mathscr{B}$ or $\hat{\mathscr{B}}$ covers a specific line segment,

$$\mathscr{L}_\lambda = \{(u, v): v = a_{\lambda_0} + b_{\lambda_0} u\},$$

converges to unity as $n \to \infty$. Indeed, if all moments of (X, Y) are finite, then the convergence is at rate $O(n^{-C})$ for all $C > 0$. The exceptional set of λ's is the union of $\{\lambda_1, \lambda_2\}$ with the set of solutions of the equation

$$2\left\{H_\lambda + \left(H_\lambda^2 + \lambda\right)^{1/2}\right\} \frac{\partial H_\lambda}{\partial \lambda} + 1 = 0, \tag{4.1}$$

this set having at most two elements. To appreciate the argument, observe that if $\operatorname{sgn}(\sigma_{XY}) > 0$, then

$$a_\lambda + b_\lambda u = a + b\mu + \left\{H_\lambda + \left(H_\lambda^2 + \lambda\right)^{1/2}\right\}(u - \mu).$$

Therefore, $\mathscr{B}^\circ$ represents the locus of a sequence of line segments over the domain $\mathscr{U}$, each segment being on a line that passes through $(\mu, a + b\mu)$ and has gradient $s(\lambda) = H_\lambda + (H_\lambda^2 + \lambda)^{1/2}$. If $\lambda \ne \lambda_1$ or λ_2 and λ is an element of $[\lambda_1, \lambda_2]$, but is not a turning point of $s(\lambda)$, then $\mathscr{L}_\lambda$ is an interior point of $\mathscr{B}^\circ$. The set of turning points of $s(\lambda)$ is just the set of solutions to (4.1).

Acknowledgments. We are grateful to Dr. T. S. Breusch for helpful and stimulating conversations, to Dr. T. Jones for providing the measurements on sandstone core samples and to two referees and an Associate Editor for their helpful suggestions.

REFERENCES

EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82** 171–200.

FULLER, W. A. (1987). *Measurement Error Models*. Wiley, New York.

GOLDBERGER, A. S. (1972). Maximum likelihood estimation of regression models containing unobserved variables. *Internat. Econom. Rev.* **13** 1–15.

GRILICHES, Z. (1974). Errors-in-variables and other unobservables. *Econometrica* **42** 971–998.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **50** 35–45.

JONES, T. A. (1979). Fitting straight lines when both variables are subject to error. *Math. Geol.* **11** 1–25.

KENDALL, M. G. and STUART, A. (1979). *The Advanced Theory of Statistics* **1**, 4th ed. Griffin, London.

LOH, W. Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82** 155–162.

MADDALA, G. S. (1988). *Introduction to Econometrics*. Macmillan, New York.

ZELLNER, A. (1970). Estimation of regression relationships containing unobservable independent variables. *Internat. Econom. Rev.* **11** 441–454.

DEPARTMENT OF STATISTICS
FLOYD HALL
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4
CANBERRA A.C.T. 2601
AUSTRALIA