CONFIDENCE REGIONS FOR THE SLOPE IN A LINEAR ERRORS-IN-VARIABLES REGRESSION MODEL

by

Leon Jay Gleser
Purdue University

Technical Report #82-23

Department of Statistics
Purdue University

July 1982

[1]Research supported by the National Science Foundation under Grant MCS 81-21948.

CONFIDENCE REGIONS FOR THE SLOPE IN A LINEAR ERRORS-IN-VARIABLES REGRESSION MODEL

by

Leon Jay Gleser
Purdue University

1. INTRODUCTION

The linear errors-in-variables model of interest in this paper is the following. Independent pairs (x_i, y_i) of random variables are observed, where

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} u_i \\ a+bu_i \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \end{pmatrix}, \quad i = 1,\ldots,N, \tag{1.1}$$

and

$$e_1,\ldots,e_N,\ f_1,\ldots,f_N \text{ are independent } N(0,\sigma^2) \text{ random variables.}$$

The means u_i of the x_i's are unknown scalars, but are not of central interest in the problem. The intercept a, slope b, and error variance σ^2 are also unknown, and are the parameters to be estimated.

Over 100 years ago, Adcock (1878) considered the model (1.1), and proposed estimating the parameters a and by by choosing a line $\{(x,a+bx): x \text{ real}\}$ such that the sum of squared distances from the observed points (x_i, y_i) to the line along perpendiculars to the line is minimized. This approach contrasts with classical least squares theory, where squared distances to the line are measured along perpendiculars to the x-axis. Adcock's methodology has been used in physical science and engineering applications, but it is only in the past 30 years that statisticians have given serious theoretical attention to inference for this model. Stimulus for this increased interest has come in part from awareness that the model (1.1) is closely related to important econometric and psychometric models, and in part from greater concern among scientists about measurement errors in variables previously assumed to be precisely measured.

Excellent reviews of the literature for the model (1.1), and various related models, can be found in Moran (1971), Anderson (1976), Kendall and Stuart (1979), and Bentler (1980). Additional theoretical material appears in Gleser (1981). Almost all previous research in this area has concerned point estimation for a, b, and σ^2, and the distributions of such estimators. As noted by the discussants of Anderson's (1976) paper, very little work has been done to study confidence regions for these parameters. This is in spite of the fact that users of the model (1.1) frequently express interest in confidence interval methodology for the slope b.

The purpose of this paper is twofold. First, to point out that no universally satisfactory confidence interval procedure for b can exist, at least if this procedure is based (as are most currently proposed confidence region procedures) on the sample cross product matrix

$$W = \begin{pmatrix} \sum_{i=1}^{N} (x_i-\bar{x})^2 & \sum_{i=1}^{N} (x_i-\bar{x})(y_i-\bar{y}) \\ \sum_{i=1}^{N} (x_i-\bar{x})(y_i-\bar{y}) & \sum_{i=1}^{N} (y_i-\bar{y})^2 \end{pmatrix} . \tag{1.2}$$

To that end, in Section 2, the following theorem is proved.

Theorem 1. <u>Any confidence interval</u> [L(W),U(W)] <u>for</u> b <u>which has coverage probability at least</u> $1-\alpha$, $0 < \alpha < \frac{1}{2}$, <u>for all parameter values must satisfy</u>

$$P\{[L(W),U(W)] = (-\infty,\infty)\} > 0$$

<u>for every parameter value</u> (and thus have infinite expected length).

Contrariwise, any confidence interval which almost surely has finite length for all parameter values must have confidence level (the infimum of the probabilities of coverage over the parameter space) equal to 0.

In Section 3, the advantages and disadvantages of three confidence region procedures for b are discussed. The procedures described are those proposed by Creasy (1957) and Williams (1969), by Brown (1957), and by Gleser (1981). It is concluded that none of these procedures is fully satisfactory. However, in Section 4 it is argued that if an investigator is willing to make a very reasonable assumption about the parameters of the model the confidence interval procedure proposed by Gleser can give satisfactory performance.

Two additional facts about inference for the model (1.1) are brought out in the paper. In Section 3.2, is shown that the likelihood ratio test of H_0: $b = b_0$ v.s. H_1: $b \neq b_0$ cannot be made to have fixed level of significance α, $0 \leq \alpha < 1$, at least when the sample size N goes to infinity. Finally, in an appendix, it is shown that the maximum likelihood estimator $\hat{b}$ of b is median biased, thus currecting a misleading remark in Anderson (1976, pp. 9, 12).

2. PROOF OF THEOREM 1

It is well known that the matrix W defined by (1.2) has a noncentral Wishart distribution with $n = N - 1$ degrees of freedom, covariance parameter $\sigma^2 I_2$, and noncentrality parameter $\sigma^{-2} n\sigma_u^2 (1,b)'(1,b)$, where

$$\sigma_u^2 = \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \bar{u})^2 .$$

That is

$$W \sim \mathscr{W}\left(n, \sigma^2 I_2; \frac{n\sigma_u^2}{\sigma^2} \binom{1}{b}\binom{1}{b}'\right). \tag{2.1}$$

The parameters of this distribution are b, σ^2 and σ_u^2. Hence, the parameter space referred to in Theorem 1 can be described by

$$\Omega = \{\omega = (b,\sigma^2,\sigma_u^2): -\infty < b < \infty,\ \sigma^2 > 0,\ \sigma_u^2 > 0\}. \qquad (2.2)$$

It must be assumed that $\sigma_u^2 > 0$, or else the model is not identifiable. Nevertheless, since the confidence level of any confidence region C(W) for b is defined by

$$\text{confidence level of } C = \inf_{\omega \in \Omega} P_\omega\{b \in C(W)\},$$

this confidence level may be attained as a limit of coverage probabilities corresponding to a sequence of points $\omega = (b,\sigma^2,\sigma_u^2)$ in the parameter space Ω for which $\sigma_u^2 \to 0$. This is the key to the results stated in Theorem 1.

Define

$$\Gamma_b = (1+b^2)^{-1/2} \begin{pmatrix} 1 & -b \\ b & 1 \end{pmatrix}, \qquad V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} = \Gamma_b' W \Gamma_b. \quad (2.3)$$

Then

$$V \sim \mathcal{W}(n,\sigma^2 I_2,\ \delta^2\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}), \qquad \delta^2 = \frac{n(1+b^2)\sigma_u^2}{\sigma^2}. \qquad (2.4)$$

Note that the distribution of V depends only on the error variance σ^2 and the noncentrality parameter δ^2.

To make the argument easier to follow, note that it follows from (2.3) that without loss of generality all of the probabilities under discussion can be assumed to be defined on a probability space in which the sample space consists of nonnegative definite 2x2 matrices V, the family of probability measures is defined by (2.4), and

$$W = \Gamma_b V \Gamma_b'. \qquad (2.5)$$

To prove the first part of Theorem 1, suppose that the confidence interval $[L(W),U(W)]$ has confidence level $1-\alpha$, so that

$$P_\omega \{L(W) \leq b \leq U(W)\} \geq 1-\alpha, \text{ all } \omega = (b,\sigma^2,\sigma_u^2) \in \Omega. \qquad (2.6)$$

It follows from (2.5) that

$$P_\omega \{L(W) \leq b \leq U(W)\} = P_{\sigma^2,\delta^2}\{L(\Gamma_b V\Gamma_b') \leq b \leq U(\Gamma_b V\Gamma_b')\}, \qquad (2.7)$$

where from (2.4) the distribution of V depends only on σ^2 and δ^2. Thus, (2.6) and (2.7) imply that

$$1-\alpha \leq \liminf_{\sigma_u^2 \to 0} P_{\sigma^2,\delta^2} \{L(\Gamma_b V\Gamma_b') \leq b \leq U(\Gamma_b V\Gamma_b')\}$$

$$= P_{\sigma^2,0} \{L(\Gamma_b V\Gamma_b') \leq b \leq U(\Gamma_b V\Gamma_b')\},$$

since for fixed b,σ^2, $\sigma_u^2 \to 0$ if and only if $\delta^2 \to 0$, and the probabilities of the noncentral Wishart distribution (2.4) are continuous in δ^2. However, when $\delta^2 = 0$, V and $\Gamma_b V\Gamma_b'$ have the same central Wishart distribution, $\mathcal{W}(n, \sigma^2 I_2)$. Thus,

$$1-\alpha \leq P_{\sigma^2,0} \{L(V) \leq b \leq U(V)\}, \qquad \text{all } \sigma^2, b. \qquad (2.8)$$

The distribution of V in (2.8) is independent of b. Thus, taking $b \to -\infty$, $b \to \infty$ in (2.8) demonstrates that

$$P_{\sigma^2,0} \{L(V) = -\infty\} \geq 1-\alpha, \qquad P_{\sigma^2,0} \{U(V) = \infty\} \geq 1-\alpha,$$

and hence that for all $\sigma^2 > 0$,

$$P_{\sigma^2,0} \{L(\Gamma_b V\Gamma_b') = -\infty,\ U(\Gamma_b V\Gamma_b') = \infty\} = P_{\sigma^2,0} \{L(V) = -\infty, U(V) = \infty\} \geq 1-2\alpha. \quad (2.9)$$

Note that $1-2\alpha > 0$ when $0 \leq \alpha < 1/2$. It now follows from (2.9), the mutual absolute continuity of all the noncentral Wishart distributions (2.4) with the central Wishart distribution $\mathscr{W}(n,\sigma^2 I_2)$ and (2.7) that

$$P_\omega\{L(W) = -\infty,\ U(W) = \infty\} > 0, \quad \text{all } \omega = (b,\sigma^2,\sigma_u^2) \in \Omega,$$

as asserted by the first part of Theorem 1.

To prove the "contrariwise" part of Theorem 1, assume that

$$P_\omega\{-\infty < L(W) \leq U(W) < \infty\} = 1, \quad \text{all } \omega = (b,\sigma^2,\sigma_u^2) \in \Omega.$$

Using (2.7) and arguments similar to those leading to (2.8), it follows that for all σ^2,

$$1 = \lim_{\sigma_u^2 \to 0} P_{\sigma^2,\delta^2}\{-\infty < L(\Gamma_b V \Gamma_b') \leq U(\Gamma_b V \Gamma_b') < \infty\} = P_{\sigma^2,0}\{-\infty < L(V) \leq U(V) < \infty\}, \tag{2.10}$$

and also that for all fixed b,σ^2,

$$\lim_{\sigma_u^2 \to 0} P_\omega\{L(W) \leq b \leq U(W)\} = P_{\sigma^2,0}\{L(V) \leq b \leq U(V)\}.$$

But since (2.10) states that $L(V) > -\infty$ and $U(V) < \infty$ with probability 1 when V has the $\mathscr{W}(n,\sigma^2 I_2)$ distribution, and since this distribution is independent of b, taking $b \to -\infty$ or $b \to \infty$ shows that

$$\inf_b \lim_{\sigma_u^2 \to 0} P_\omega\{L(W) \leq b \leq U(W)\} = \inf_b P_{\sigma^2,0}\{L(V) \leq b \leq U(V)\} = 0.$$

The proof is now completed by noting that

$$0 \leq \inf_\omega P_\omega\{L(W) \leq b \leq U(W)\} \leq \inf_b \lim_{\sigma_u^2 \to 0} P_\omega\{L(W) \leq b \leq U(W)\} = 0,$$

so that the confidence level of $[L(W),U(W)]$ for b is

$$\inf_{\omega \in \Omega} P_\omega\{L(W) \leq b \leq U(W)\} = 0$$

as asserted. □

The proof of Theorem 1 is, admittedly, rather technical. A little intuition from classical regression theory may thus be more convincing. Suppose, therefore, that after the data (x_i, y_i), $1 \le i \le N$, were obtained, you somehow learned the values of $u_1, \ldots, u_N$ and of σ^2. In this case, you certainly would ignore the x_i-values in favor of the u_i's, and estimate b by the usual least squares estimator

$$\hat{b}_{LS} = \frac{\sum_{i=1}^{N} y_i(u_i - \bar{u})}{n\sigma_u^2} .$$

The $100(1-\alpha)\%$ confidence interval

$$\hat{b}_{LS} \pm z(1-\tfrac{\alpha}{2}) \frac{\sigma}{\sqrt{n\sigma_u^2}} , \qquad (2.11)$$

where $z(1-\gamma)$ is the $100(1-\gamma)$ percentile of the $N(0,1)$ distribution, has many optimal properties in terms of balancing coverage probability and expected length. However, note that even in this ideal case, the length of (2.11) becomes infinite as $\sigma^{-2}\sigma_u^2 \to 0$. The proof of Theorem 1 reflects this insight, since for fixed b, n, taking σ_u^2 to 0 is equivalent to taking $\sigma^{-2}\sigma_u^2$ to 0.

It is thus apparent that for confidence <u>interval</u> estimation of b under the model (1.1) to give useful results in all cases, one must be prepared to assume that $\sigma^{-2}\sigma_u^2$ is bounded away from 0. More will be said about this assumption in Section 4.

3. AVAILABLE CONFIDENCE REGION PROCEDURES

3.1. The Creasy-Williams procedure

Perhaps the most frequently mentioned confidence region for b is one proposed independently by Creasy (1957) and by Williams (1969). Let r(b) be the sample correlation coefficient between $by_i + x_i$ and $y_i - bx_i$,

$1 \leq i \leq N$, and let $F_{1,n-1}(1-\alpha)$ be the $100(1-\alpha)$ percentile of the central F distribution with 1 and n - 1 degrees of freedom. The Creasy-Williams (CW) confidence region C_1 for b is then

$$C_1 = \{b: (n-1)r^2(b)[1-r^2(b)]^{-1} \leq F_{1,n-1}(1-\alpha))\} . \qquad (3.1)$$

The region C_1 can be thought of as the collection of all b for which the standard two-sided α-level test of

$$H_0^{(b)}: \ \rho_{by+x,y-bx} = 0 \quad \text{v.s.} \quad H_1^{(b)}: \ \rho_{by+x,y-bx} \neq 0$$

fails to reject $H_0^{(b)}$. This region does have confidence $1 - \alpha$ of covering the true value of b, but also has the following very undesirable property:

$$b \in C_1 \text{ if and only if } -\frac{1}{b} \in C_1. \qquad (3.2)$$

That (3.2) holds can be seen most easily by noting that

$$\rho_{by+x,y-bx} = -\rho_{x-(y/b),y+(x/b)},$$

and remembering that C_1 is based on a two-sided test of $H_0^{(b)}$. Because of property (3.2), the region C_1 cannot be used to test the sign of b. Further, the only way that C_1 can be an interval is for C_1 to equal $[-\infty,\infty]$. This last is an event which has positive probability under all parameter values.

Although the CW region has some undesirable properties as a region for b, it appears to yield a useful $100(1-\alpha)\%$ confidence interval for

$$\theta = \tan^{-1}(b),$$

the angle that the line a + bu makes with the u-axis. [Anderson 1976; Moran 1971.] Anderson (1976) gives a compelling argument for the superiority of θ to b as a parametrization of the model (1.1). In brief, since the model (1.1) could just as easily be defined by

$$x_i = \frac{1}{b} t_i - \frac{a}{b} + e_i, \quad y_i = t_i + f_i, \quad t_i = a + bu_i, \qquad (3.3)$$

the treatment of x_i as an observed value of an "independent" variable u, and y_i as an observed value of a "dependent" variable a + bu, is not inherent in the model (1.1) [which is symmetric in its treatment of the variables], but is instead arbitrarily imposed by the statistician upon the model. On the other hand, the angle θ is well defined whichever variant, (1.1) or (3.3), of the model is used. Nevertheless, investigators who use the model (1.1) as a more realistic alternative to the classical assumption of error-free "independent" variables x will typically want to estimate a slope b, and will not find θ an especially satisfying parameter for this purpose. (In particular, if the sign of b is important, this information cannot always be obtained from a confidence interval for θ.)

3.2. Brown's procedure and a likelihood ratio test

For σ^2 known, Brown (1957) suggests the confidence region:

$$C_2 = \{b: [(1+b^2)\sigma^2]^{-1} \sum_{i=1}^{N} [y_i-\bar{y}-b(x_i-\bar{x})]^2 \leq \chi_n^2(1-\alpha)\}, \qquad (3.4)$$

where $\chi_n^2(1-\alpha)$ is the 100(1-α) percentile of the chi-squared distribution with n degrees of freedom. Similar regions, in more general contexts, have been proposed by Villegas (1964) and Lord (1973). These last two authors allow σ^2 to be unknown, but require an independent estimator of σ^2 (perhaps based on independent replications of (x_i,y_i)).

In terms of the sample cross-product matrix W defined by (1.2), the region (3.4) becomes

$$C_2 = \left\{b: \frac{(-b,1)W\binom{-b}{1}}{\sigma^2(1+b^2)} \leq \chi_n^2(1-\alpha)\right\}.$$

Let

$$W = GDG', \; G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} : \; 2\times 2 \text{ orthogonal},$$
$$D = \text{diag}(d_1, d_2), \; d_1 \geq d_2 > 0, \tag{3.5}$$

be the spectral decomposition of W. Gleser (1981) shows that in terms of (3.5), the maximum likelihood estimator of b is

$$\hat{b} = \frac{g_{21}}{g_{11}} = -\frac{g_{12}}{g_{22}} . \tag{3.6}$$

Combining (3.5) and (3.6), it is easily seen that

$$W = \frac{1}{(1+\hat{b}^2)} \begin{pmatrix} 1 & -\hat{b} \\ \hat{b} & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 & \hat{b} \\ -\hat{b} & 1 \end{pmatrix} , \tag{3.7}$$

and hence that

$$C_2 = \left\{ b: \; \frac{d_2}{\sigma^2} + \frac{(\hat{b}-b)^2(d_1-d_2)}{(1+\hat{b}^2)(1+b^2)\sigma^2} \leq \chi_n^2(1-\alpha) \right\} . \tag{3.8}$$

Note that C_2 is the empty set if $d_2/\sigma^2 > \chi_n^2(1-\alpha)$. This is an event having positive probability for all parameter values. Indeed, an argument parallel to that in Healy (1979) shows that the rejection region of the likelihood ratio test of

H_0: The model (1.1) holds v.s. H_1: $(x_i, y_i)' \sim BVN((\psi_{1i}, \psi_{2i})', \sigma^2 I_2)$

consists of precisely those data points (x_i, y_i), $1 \leq i \leq N$, for which C_2 is empty.

Let

$$s = \frac{\sigma^2 \chi_n^2(1-\alpha) - d_2}{d_1 - d_2} .$$

It is easily shown that

$$C_2 \text{ is } \begin{cases} \text{the real line} & , \text{ if } s > 1, \\ (-\infty, L(\hat{b},d_1,d_2)] \cup [U(\hat{b},d_1,d_2), \infty), & \text{if } 1 \geq s > \frac{1}{1+\hat{b}^2}, \\ [L(\hat{b},d_1,d_2), U(\hat{b},d_1,d_2)] & , \text{ if } \frac{1}{1+\hat{b}^2} > s > 0 \\ \text{empty} & , \text{ if } s < 0. \end{cases}$$

Each of the above possibilities occurs with positive probability for all parameter values. Thus, Brown's procedure shares some of the unpleasant properties of the CW procedure. Like the CW procedure, Brown's procedure does have fixed coverage probability $1 - \alpha$ for all parameter values. This fact can most easily be seen by noting that in terms of V defined by (2.3),

$$\frac{(-b,1)W\binom{-b}{1}}{(1+b^2)\sigma^2} = \frac{v_{22}}{\sigma^2} \sim \chi_n^2 ,$$

the distributional assertion following directly from (2.4).

Brown's procedure is, of course, inapplicable when σ^2 is unknown. One can substitute a consistent estimator for σ^2 into (3.8), but then the fixed coverage probability of the procedure is not guaranteed, except perhaps in large samples. Brown originally defined C_2 to be the collection of all b values for which a standard α-level ANOVA test [see (3.4)] of

$$H_0^{(b)}: \quad E(y_i - bx_i) = \text{constant, all } i,$$

versus general alternatives failed to reject $H_0^{(b)}$. When σ^2 is unknown, no such test can be formed. However, one can try to generalize Brown's procedure to the unknown-σ^2 case by letting C_2 be the collection of all b_0 values for which the likelihood ratio test of

$$H_0: \quad b = b_0 \quad \text{v.s.} \quad H_1: \quad b \neq b_0, \tag{3.9}$$

in the context of the model (1.1), fails to reject H_0 at level of significance α. It is not hard to show that this region has the form (3.8) with

σ^2 replaced by its maximum likelihood estimator $\hat{\sigma}^2 = (2N)^{-1}d_2$, and with the cutoff point $\chi_n^2(1-\alpha)$ replaced by the $100(1-\alpha)$ percentile $K(1-\alpha; b_0, \sigma^2, \sigma_u^2)$ of the resulting statistic

$$2N\left(1+ \frac{(\hat{b}-b_0)^2(d_1-d_2)}{(1+\hat{b}^2)(1+b_0^2)d_2}\right) = T_N(b_0).$$

That is

$$\begin{aligned} C_2^* &= \{b: \frac{2N(-b,1)W\binom{-b}{1}}{d_2(1+b^2)} \le K(1-\alpha;\ b,\sigma^2,\sigma_u^2)\} \\ &= \{b:\ T_N(b) \le K(1-\alpha;\ b,\sigma^2,\sigma_u^2)\}. \end{aligned} \qquad (3.10)$$

Using the results of Gleser (1981), it can be shown that if

$$\lim_{N\to\infty} \sigma_u^2 = \Delta > 0, \qquad (3.11)$$

then

$$N\left(\frac{T_N(b_0)}{2N} - 1\right) \xrightarrow{L} \left(1 + \frac{\sigma^2}{(1+b_0^2)\Delta}\right)^2 \chi_1^2, \quad \text{as} \quad N \to \infty. \qquad (3.12)$$

Since $\sigma^{-2}\Delta$ can have any value in the interval $(0,\infty)$, the result (3.12) indicates that, at least asymptotically, one cannot find a cutoff point $K(1-\alpha;\ b_0)$ independent of σ^2 and σ_u^2 which will give the confidence region $\{b:\ T_N(b) \le K(1-\alpha;\ b_0)\}$ a fixed (or even bounded) probability of coverage for all parameter values. The same argument shows that, at least asymptotically, <u>one cannot construct an α-level</u>, $0 \le \alpha < 1$, <u>likelihood ratio test of (3.9)</u>.

Although it may yet be possible to adapt Brown's procedure to cover the case when σ^2 is unknown, the prospect does not seem promising. Further, any such adaptation that starts with the statistic

$$\frac{(-b,1)W\binom{-b}{1}}{(1+b^2)}$$

will, because of the term $(1+b^2)$ in the denominator, yield regions for b which are not intervals.

3.3. Gleser's Procedure

Under the assumption (3.11), Gleser (1981) proposed the following large-sample confidence interval for b:

$$C_3 = \{b: \frac{n(\hat{b}-b)^2(d_1-d_2)^2}{(1+\hat{b}^2)^2 d_1 d_2} \leq \chi_1^2(1-\alpha)\}. \tag{3.13}$$

When (3.11) holds, b, σ^2 are fixed, and $\lim_{N\to\infty} \sigma_u^2 = \Delta > 0$,

$$\lim_{N\to\infty} P_{b,\sigma^2,\sigma_u^2}\{b \in C_3\} = 1-\alpha, \text{ all } b, \sigma^2 \Delta, \tag{3.14}$$

as shown in Corollary 4.3 of Gleser (1981).

On the other hand, Theorem 2 asserts that for any fixed cutoff value k, and any finite sample size N = n+1,

$$\inf_{\omega \in \Omega} P_\omega [b \in \{b_0: \frac{n(\hat{b}-b_0)^2(d_1-d_2)^2}{(1+\hat{b}^2)^2 d_1 d_2} \leq k\}] = 0. \tag{3.15}$$

This apparent contradiction between (3.14) and (3.15) stems from the fact that the former result involves limits in N for fixed parameters, while the latter result involves limits in parameter values ω for fixed N. The conflict in these results serves as a warning that in errors-in-variables models one must be very careful with limiting arguments. In particular, one cannot generally interchange limits in N and limits in parameter values.

The region C_3 can be regarded as an adjustment to the Creasy-Williams procedure. To see this, note that the region C_1 can be expressed, using (3.9), as follows:

$$C_1 = \{b: \frac{(n-1)[(1,b)W\binom{-b}{1}]^2}{|W|(1+b^2)^2} \leq F_{1,n-1}(1-\alpha)\}$$

$$= \{b: [\frac{(n-1)(1+b\hat{b})^2}{n(1+b^2)^2}] \frac{n(\hat{b}-b)^2(d_1-d_2)^2}{(1+\hat{b}^2)^2 d_1 d_2} \leq F_{1,n-1}(1-\alpha)\}. \quad (3.16)$$

Comparing (3.13) and (3.16), it is seen that the region C_3 replaces the term $[n(1+b^2)^2]^{-1}(n-1)(1+b\hat{b})^2$ in C_1 by 1, which is the probability limit of this term as $N \to \infty$ (Gleser 1981). Such a substitution has no effect upon the asymptotic ($N \to \infty$) coverage probability of C_1, but does lose the good <u>global</u> finite-sample control of coverage probabilities provided by the CW procedure. (However, see Section 4.) In effect, the confidence procedure (3.13) gives up control of coverage probability in order to provide interval estimators for b.

The procedure (3.13) can also be viewed as an adjustment to the generalized Brown procedure (3.10). Here, the adjustment not only provides interval regions for b, but also controls asymptotic probability of coverage by dividing $n(\hat{b}-b)^2$ by a correct consistent estimator of the variance of the asymptotic distribution of $n^{\frac{1}{2}}(\hat{b}-b)$.

The major drawback of the confidence interval C_3 is that it fails to control global (over the parameter space) finite sample coverage probabilities. In the next section (Section 4), it is shown that under a reasonable restriction on the parameter space, this procedure can be adjusted to provide control of coverage probabilities in finite samples.

4. CONTROL OF COVERAGE PROBABILITIES FOR GLESER'S PROCEDURE

Since the cutoff $\chi_1^2(1-\alpha)$ in the large-sample region (3.13) may not be the correct value to use in finite samples, replace (3.13) by the more general region:

$$C_3^*(k) = \{b: \frac{(\hat{b}-b)^2(d_1-d_2)^2}{(1+\hat{b}^2)^2 d_1 d_2} \leq k\}. \tag{4.1}$$

Although k will depend upon N, this dependence is suppressed in our notation for the sake of convenience.

Let $\tau > 0$ be a positive constant. Define

$$\rho(k,\tau,N) = \inf\,[P_\omega\{b \in C_3^*(k)\}: -\infty < b < \infty,\ \tau \leq \sigma^{-2}\sigma_u^2 < \infty]. \tag{4.2}$$

Let χ_{N-1}^2 denote a chi-squared random variable with N-1 degrees of freedom and let r be a random variable, independent of χ_{N-1}^2, with the density

$$f_N(r) = \frac{(1-r^2)^{\frac{1}{2}(N-4)}}{B(\frac{1}{2},\frac{1}{2}(N-2))}, \quad -1 \leq r \leq 1. \tag{4.3}$$

That is, r has the density of the sample correlation coefficient based on a sample of N bivariate normal observations with covariance matrix equal to the identity matrix I_2. Define

$$L(k,\tau,N) = P\{\frac{r^2}{1-r^2} \leq k,\ \chi^2_{N-1} \leq (N-1)\tau\ (\frac{1}{r} + \frac{1}{r^2}[k(1-r^2)]^{\frac{1}{2}})^2\}. \tag{4.4}$$

The main result of this section is the following theorem.

Theorem 2. For all k, $0 < k < \infty$,

$$L(k,\tau,N) \leq \rho(k,\tau,N) \leq L(k,\tau,N) + \tfrac{1}{2}P\{\frac{r^2}{1-r^2} > k\}. \tag{4.5}$$

Hence, if it is known that $\sigma^{-2}\sigma_u^2 \geq \tau$, then for every α, $0 < \alpha < 1$, and every N, there exists a critical value $k^* = k(1-\alpha,\tau,N)$ satisfying

$$L(k^*,\tau,N) = 1 - \alpha$$

for which the confidence interval $C_3^*(k^*)$ has confidence level at least $1-\alpha$.

The discussion after the proof of Theorem 1 indicates why the requirement that $\sigma^{-2}\sigma_u^2 \geq \tau > 0$ is necessary for the confidence interval $C_3^*(k^*)$ to have a confidence level $1-\alpha$ strictly greater than 0. In many applications, an investigator may be able to provide a value of τ from knowledge of the nature of the experiment, and previous experience with the measuring instruments. If instrumental variables (see Moran, 1971) are available, the data (x_i, y_i), $1 \le i \le N$, itself can be used to find a lower confidence bound for $\sigma^{-2}\sigma_u^2$ which can serve as the value of τ. Since the method described for choosing k in Theorem 2 is conservative, if an investigator is willing to be imprecise about the value of α desired, the value of τ can also be chosen imprecisely.

The steps of the proof of Theorem 2 are outlined below in the form of lemmas. Technical details, if needed for the proofs of the lemmas, are given in the appendix.

4.1 Proof of Theorem 2.

Recall that $V = \Gamma_b' W \Gamma_b$ is defined in (2.3). Using (2.3) and (3.7), it is easily shown that

$$V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} = \frac{1}{1+\hat{b}_0^2} \begin{pmatrix} 1 & -\hat{b}_0 \\ \hat{b}_0 & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 & \hat{b}_0 \\ -\hat{b}_0 & 1 \end{pmatrix}, \tag{4.6}$$

where

$$\hat{b}_0 = \frac{\hat{b}-b}{1+b\hat{b}}, \quad \hat{b} = \frac{\hat{b}_0+b}{1-b\hat{b}_0}. \tag{4.7}$$

Let $n = N - 1$,

$$r = \frac{v_{12}}{(v_{11}v_{22})^{\frac{1}{2}}}, \qquad h = \frac{v_{11}}{v_{22}}. \tag{4.8}$$

Lemma 1. The random variables r and h are statistically independent, with r having the density $f_N(r)$ defined by (4.3), and h having the noncentral $F_{n,n}(\delta^2)$ distribution with noncentrality parameter

$$\delta^2 = \frac{n\,\sigma_u^2\,(1+b^2)}{\sigma^2}.$$

Proof Define $z = v_{11}^{-1/2}v_{12}$, $w = v_{22}-z^2$. Using (2.4) and Theorem 2.2 of Gleser (1976), v_{11}, z, and w are mutually statistically independent with

$$v_{11} \sim \sigma^2\chi_n^2(\delta^2), \quad z \sim N(0,\sigma)^2, \quad w \sim \sigma^2\chi_{n-1}^2,$$

where $\chi_d^2(\delta^2)$ is the noncentral chi squared distribution with d degrees of freedom and noncentrality parameter δ^2. However,

$$r = \frac{z}{\sqrt{w + z^2}}, \qquad h = \frac{v_{11}}{w + z^2}$$

and it is well known that r and $w + z^2$ are independent. The conclusion of the lemma now follows by standard distributional arguments. □

It follows from (4.6), (4.7) and (4.8) that

$$\frac{(\hat{b}-b)^2\,(d_1-d_2)^2}{(1+\hat{b}^2)^2\,d_1d_2} = \frac{(1-b\hat{b}_0)^2 v_{12}^2}{|V|} = \frac{r^2}{(1-r^2)}(1-b\hat{b}_0)^2.$$

Hence (see (4.1)),

$$P_{b,\sigma^2,\sigma_u^2}\{b \in C_3^*(k)\} = P_{\delta^2}\{(\frac{r^2}{1-r^2})\,(1-b\hat{b}_0)^2 \leq k\}, \tag{4.9}$$

where from (4.6) and Lemma 1, the probability on the right-hand side of (4.9) depends only on δ^2 (and b).

Using (4.6) and (4.7), it can be shown that

$$\hat{b}_0 = \frac{1-h + [(h-1)^2 + 4hr^2]^{1/2}}{2rh^{1/2}}. \tag{4.10}$$

For fixed $r \neq 0$,

$r\hat{b}_0$ is strictly decreasing and continuous in h,

$$\lim_{h \to 0} r\hat{b}_0 = \infty, \qquad \lim_{h \to \infty} r\hat{b}_0 = 0, \tag{4.11}$$

and consequently $r\hat{b}_0 \geq 0$.

Lemma 2. For $rb \neq 0$, define

$$q_i(r,b) = \frac{r}{b} + (-1)^i \left|\left(\frac{r}{b}\right)\right| \left(\frac{k(1-r^2)}{r^2}\right)^{1/2}, \qquad i=1,2.$$

Then for all b, δ^2,

$$P_{\delta^2}\{\frac{r^2}{1-r^2} \leq k,\ 0 \leq r\hat{b}_0 \leq q_2(r,b)\} \leq P_{\delta^2}\{(\frac{r^2}{1-r^2})\ (1-b\hat{b}_0)^2 \leq k\}$$

$$\leq P_{\delta^2}\{\frac{r^2}{1-r^2} \leq k,\ 0 \leq r\hat{b}_0 \leq q_2(r,b)\} + (1/2)P\{\frac{r^2}{1-r^2} > k\}. \tag{4.12}$$

Proof Note that when $b \neq 0$,

$$\frac{r^2}{1-r^2}\ (1-b\hat{b}_0)^2 = \frac{r^2}{1-r^2}\ (1-(r\hat{b}_0)(\frac{b}{r}))^2 \leq k$$

if and only if

$$q_1\ (r,b) \leq r\hat{b}_0 \leq q_2(r,b). \tag{4.13}$$

However, by (4.11), $r\hat{b}_0 \geq 0$, while $r^2(1-r^2)^{-1} \leq k$ implies that

$q_1(r,b) \leq 0$ and $q_2(r,b) \geq 0$. This establishes the left-hand inequality in

(4.12) for $b \neq 0$. Noting that $\lim_{b \to 0} q_2(r,b) = \infty$ when $(1-r^2)^{-1}r^2 \leq k$, it is easily established that this left-hand inequality is an equality when $b = 0$. (Therefore, the right-hand inequality in (4.12) also holds when $b = 0$, since $P\{(1-r^2)^{-1}r^2 \geq k\} \geq 0$.)

To prove the right-hand inequality of (4.12) when $b \neq 0$, let

$$A = \{\frac{r^2}{1-r^2}(1-b\hat{b}_0)^2 \leq k\} = \{q_1(r,b) \leq r\hat{b}_0 \leq q_2(r,b)\},$$

$$B = \{\frac{r^2}{1-r^2} \leq k,\ 0 \leq r\hat{b}_0 \leq q_2(r,b)\}.$$

Then since $r\hat{b}_0 \geq 0$,

$$A \cap B^c = \{q_1(r,b) \leq r\hat{b}_0 \leq q_2(r,b),\ \frac{r^2}{1-r^2} > k\}.$$

However, when $rb < 0$ and $(1-r^2)^{-1}r^2 > k$, $q_2(r,b) < 0$, and $r\hat{b}_0 > q_2(r,b)$.

Thus,

$$A \cap B^c = \{rb \geq 0,\ q_1(r,b) \leq r\hat{b}_0 \leq q_2(r,b),\ (1-r^2)^{-1}r^2 > k\}$$

$$\subset \{rb \geq 0,\ (1-r^2)^{-1}r^2 > k\}.$$

Since the density $f_N(r)$ of r is symmetric about 0 (see (4.3) and Lemma 1), and does not depend upon b or δ^2,

$$P_{\delta^2}\{A \cap B^c\} \leq P\{rb > 0,\ (1-r^2)^{-1}r^2 > k\} = (1/2)P\{(1-r^2)^{-1}r^2 > k\}.$$

This establishes the desired inequality. □

Note that $\hat{b}_0$ in (4.10) is a function only of r and h, while $q_2(r,b)$ is a function only of r (and b). From Lemma 1, r and h are independent, r has a continuous distribution independent of b, σ^2, and σ_u^2, and h has the $F_{n,n}(\delta^2)$ distribution which depends upon b, σ^2, and σ_u^2 only through $\delta^2 = n(1+b^2)(\sigma^{-2}\sigma_u^2)$.

When $\sigma^{-2}\sigma_u^2 \geq \tau$,

$$\delta^2 \geq n(1+b^2)\tau.$$

Finally, note from (4.11) that when $r \neq 0$ is fixed, $r\hat{b}_0$ is strictly decreasing in h. From these remarks, and the well-known fact that the $F_{n,n}(\delta^2)$ distribution has monotone likelihood ratio in δ^2, it can be straightforwardly shown that for all b, σ^2, σ_u^2 for which $\sigma^{-2}\ \sigma_u^2 \geq \tau$,

$$P_{\delta^2}\{\frac{r^2}{1-r^2} \leq k,\ 0 \leq r\hat{b}_0 \leq q_2(r,b)\}$$

$$\geq P_{\delta^2=n(1+b^2)\tau}\{\frac{r^2}{1-r^2} \leq k,\ 0 \leq r\hat{b}_0 \leq q_2(r,b)\}. \quad (4.14)$$

Lemma 3. *The quantity*

$$\psi(b) = P_{n\,\tau\,(1+b^2)}\{\frac{r^2}{1-r^2} \leq k,\ 0 \leq r\hat{b}_0 \leq q_2(r,b)\} \quad (4.15)$$

is nondecreasing in b *for* $b \leq 0$, *nonincreasing in* b *for* $b \geq 0$.

Proof. See the appendix. □

It follows from Lemma 3 that

$$\inf_{-\infty < b < \infty} \Psi(b) = \min \{\lim_{b \to -\infty} \Psi(b),\ \lim_{b \to \infty} \Psi(b)\}, \quad (4.16)$$

Lemma 4. *For all* k, τ, n,

$$\lim_{b \to \pm\infty} \Psi(b) = L(k,\ \tau,\ N).$$

Proof. See the appendix. □

The inequality (4.5) in Theorem 2 now follows as a direct consequence of Lemmas 2, 3 and 4, and the fact that P $\{(1-r^2)^{-1}r^2 > k\}$ does not depend for its value on b or δ^2.

The function $L(k,\tau,N)$ is clearly an increasing continuous function of k, with $L(0,\tau,N) = 0$, $\lim_{k\to\infty} L(k, \tau, N) = 1$. Hence for any α, $0 < \alpha < 1$, we can find $k^* = k(1-\alpha, \tau, N)$, $0 < k^* < \infty$, satisfying $L(k^*, \tau, N) = 1 - \alpha$. This completes the proof of Theorem 2.

4.2. Discussion

Note from (4.4) that

$$L(k^*, \tau, N) \leq P\{ \frac{r^2}{1-r^2} \leq k^*\}.$$

In cases where k^* is chosen so that $L(k^*, \tau, N) = 1-\alpha$ and α is small (α = .01, .05),

$$(1/2)P\{\frac{r^2}{1-r^2} > k^*\} \leq (1/2)\alpha,$$

and the inequality (4.5) shows that the confidence level $\rho(k^*, \tau, N)$ of the region $C_3^*(k^*)$ will be very close to $1 - \alpha$. To be precise,

$$1 - \alpha \leq \rho(k^*, \tau, N) \leq 1 - \frac{\alpha}{2}.$$

Thus, although the method of choosing k^* indicated in Theorem 2 is conservative, this conservativeness is not likely to be of great practical importance.

The function $L(k, \tau, N)$ has been tabulated for various values of k, n = N-1 and $n\tau$ in Shyr (1983), and from these calculations values of k^* have been obtained. Shyr's calculations have concentrated on cases where τ is small, since (as seen below) these are the situations where modification of the asymptotic confidence region (3.13) is most needed. To give an idea of the modification to the asymptotic region C_3 required when τ is small, Table 1 compares k* to the corresponding value $k = n^{-1} \chi_1^2 (1-\alpha)$ for selected values of $1 - \alpha$, τ and n. From this table, it is clear that when τ is small (small values of $\sigma^{-2} \sigma_u^2$ cannot be ruled out), a controlled probability of coverage is only possible at the expense of a considerable increase in the width of the confidence interval.

Table 1. Values of k* and $n^{-1} \chi_1^2(1-\alpha)$ for Various Values of $1 - \alpha$, n, and τ

τ	n	$1 - \alpha = .90$		$1 - \alpha = .95$		$1 - \alpha = .99$	
		k*	$\frac{\chi_1^2(1-\alpha)}{n}$	k*	$\frac{\chi_1^2(1-\alpha)}{n}$	k*	$\frac{\chi_1^2(1-\alpha)}{n}$
.01	10	9.9046	0.2706	20.8063	0.3841	68.4376	0.6635
	50	0.3578	0.0541	0.7425	0.0768	2.2686	0.1327
.05	10	2.2272	0.2706	4.7507	0.3841	15.7813	0.6635
	20	.5554	0.1353	1.1705	0.1921	3.6789	0.3317
.10	10	1.2862	0.2706	2.7490	0.3841	9.1235	0.6635
	20	.3740	0.1353	.7450	0.1921	2.2875	0.3317
.20	10	.8868	0.2706	1.7699	0.3841	5.6855	0.6635
	25	.2183	0.1082	.3832	0.1537	1.0476	0.2654
.50	10	.6689	0.2706	1.2010	0.3841	3.4360	0.6635

On the other hand, there are cases where one can be fairly certain, based on experience with the experimental procedure and the theory underlying the experiment, that the variability in the unknown means u_i accounts for half or more of the total variability of the observed x_i's. In this case the lower bound τ in Theorem 1 is at least 1, and it would be of interest to know how much is lost in confidence level performance if the asymptotic region C_3 is used. Equation (4.5) of Theorem 2 allows us to check this performance by providing an upper and lower bound to the minimum probability of coverage

$$\rho(n^{-1}\chi_1^2(1-\alpha),\ \tau,\ N) = \inf[P_{b,\sigma^2,\sigma_u^2}\{b \in C_3\}: -\infty < b < \infty\ \sigma^{-2}\sigma_u^2 \geq \tau]$$

of the region C_3. Table 2 gives values of the lower bound $L(n^{-1}\chi_1^2(1-\alpha),\ \tau, N)$ and upper bound

$$L(n^{-1}\chi_1^2(1-\alpha),\ \tau,\ N) + (1/2)P\{\frac{r^2}{1-r^2} > \frac{\chi_1^2(1-\alpha)}{n}\}$$

to $\rho(n^{-1}\chi_1^2(1-\alpha)\ \tau, N)$ when $1 - \alpha = .90, .95. .99$, $n = N-1 = 10, 12, 15, 25, 30, 50$, and $\tau = .25, .50, 1.00, 2.00$. The values of τ other than 1.0 are used so that insight can be gained into the performance of C_3.

Place Table 2 here

From Table 2, it can be seen that the confidence level of C_3 can be markedly less that $1 - \alpha$ in the worst case ($n = 10$, $\tau = 0.25$), but increases as either n or τ increase. Indeed, the strategy of calculating a 99% confidence interval in the hope of actually getting a confidence level of 95% would work reasonably well when $\tau \geq 1$.

One interesting fact noted in the calculations leading to Table 2, but not immediately apparent in the Table, is that as n or τ increase, $L(k, \tau, N)$ and $P\{(1-r^2)^{-1}r^2 \leq k\}$ approximate each other more and more closely in value. This is predictable from the expression (4.4) for $L(k, \tau, N)$, where it can also be seen that

$$L(k,\ \tau,\ N) \leq P\{(1-r^2)^{-1}r^2 \leq k\}.$$

It is also predictable since we know that C_3 and the CW region C_1 are asymptotically equivalent to one another ($N \to \infty$), and the random variable

Table 2. Lower and Upper Bounds for Confidence Level of C_3

		Desired Exact Value $1 - \alpha$ for Confidence Level					
τ	n	$1 - \alpha = .90$		$1 - \alpha = .95$		$1 - \alpha = .99$	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
0.25	10	.7471	.8235	.8050	.8530	.8780	.8966
	12	.7610	.8327	.8188	.8625	.8903	.9060
	15	.7759	.8429	.8335	.8731	.9032	.9163
	25	.8033	.8633	.8497	.8853	.9263	.9356
	30	.8114	.8696	.8682	.9002	.9328	.9414
	50	.8304	.8852	.8865	.9156	.9477	.9547
0.50	10	.7677	.8441	.8265	.8745	.8990	.9176
	12	.7812	.8528	.8397	.8834	.9103	.9260
	15	.7954	.8624	.8535	.8931	.9219	.9350
	25	.8209	.8808	.8780	.9114	.9418	.9512
	30	.8282	.8863	.8850	.9169	.9473	.9559
	50	.8449	.8997	.9008	.9299	.9593	.9663
1.00	10	.7855	.8619	.8447	.8927	.9160	.9346
	12	.7983	.8700	.8571	.9008	.9262	.9419
	15	.8118	.8788	.8699	.9095	.9365	.9495
	25	.8352	.8952	.8921	.9255	.9535	.9628
	30	.8418	.9000	.8982	.9302	.9580	.9665
	50	.8564	.9112	.9119	.9410	.9675	.9745
2.00	10	.8002	.8766	.8595	.9075	.9290	.9476
	12	.8125	.8841	.8712	.9147	.9383	.9540
	15	.8251	.8921	.8831	.9227	.9474	.9604
	25	.8467	.9066	.9031	.9365	.9619	.9713
	30	.8526	.9108	.9084	.9403	.9657	.9742
	50	.8654	.9202	.9203	.9494	.9733	.9803

r has the same distribution as the random variable r(b) in (3.1). Nevertheless, the calculations give some concrete evidence in favor of using

$$\tilde{k} = \frac{F_{1,n-1}(1-\alpha)}{n-1}$$

in (4.1) as a hedge between the extremes of using the asymptotic region C_3 and the region $C_3*(k*)$. Note that

$$\frac{F_{1,n-1}(1-\alpha)}{n-1} \geq \frac{\chi_1^2(1-\alpha)}{n},$$

so that using $\tilde{k}$ in (4.1) yields broader intervals than C_3 (but also a greater confidence level). Of course, when n is reasonably large, the difference between $\tilde{k}$ and $n^{-1}\chi_1^2(1-\alpha)$ becomes negligible since

$$\lim_{n\to\infty} n\left(\frac{F_{1,n-1}(1-\alpha)}{n-1} - \frac{\chi_1^2(1-\alpha)}{n}\right) = 0.$$

Use of $\tilde{k}$ does, however, seem to have merit for small values of n (say $10 \leq n \leq 50$) at least for values of τ no smaller that the smallest value of τ ($\tau = 0.25$) covered in the calculations for Table 2.

5. SUMMARY

Although Theorem 1 shows that no confidence interval for b based on W can control both coverage probability and expected length over all parameter values, Theorem 2 demonstrates that it is possible to control the coverage probability for the confidence interval procedure C_3^* when the standardized variability $\sigma^{-2}\sigma_u^2$ of the unknown means u_i has a known lower bound τ. Further, the calculations in Section 4.2 show that the asymptotic confidence interval C_3 gives a reasonable control of coverage probability when $\tau \geq 1$, or N is large. If an investigator sees no way to determine a lower bound τ, to $\sigma^{-2}\sigma_u^2$, he or she can either use the asymptotic confidence interval C_3, hoping that N is large enough for (3.14) to hold to an approximation, or can use the Creasy-Williams region C_1. Unless a better generalization of Brown's region C_2 than (3.10) can be found for unknown-σ^2 cases, use of Brown's methodology must be reserved for known-σ^2 cases (or cases where a good independent estimator of σ^2 is available). Users of the Creasy-Williams or Brown procedures should be willing to accept the non-interval forms of these confidence regions, or, alternatively, convince themselves that the confidence intervals for the angle θ available from the Creasy-Williams approach can satisfy their needs.

APPENDIX

<u>Proof of Lemma 3</u>. Using Lemma 1, it can be shown that

$$\Psi(b) = \int_{\frac{r^2}{1-r^2} \le k} [P_{n\tau(1+b^2)}\{h: \; r\hat{b}_0(h) \le q_2(b,r)\}] f_N(r)\,dr. \qquad (A.1)$$

Take $(d/db)\Psi(b)$. Since the probability in square brackets in (A.1) is bounded, it is easy to show that the derivative can be taken inside the integral sign. However, letting

$$M_{\delta^2}(q) = P_{\delta^2}\{h: \; r\hat{b}_0(h) \le q\},$$

it is seen that

$$\frac{d}{db} P_{n\tau(1+b^2)}\{h: \; r\hat{b}_0(h) \le q_2(b,r)\}$$

$$= \frac{d}{db} M_{n\tau(1+b^2)}(q_2(b,r))$$

$$= \frac{d}{db} q_2(b,r) \frac{d}{dq} M_{n\tau(1+b^2)}(q_2(b,r))$$

$$+ \frac{d(n\tau(1+b^2))}{db} \frac{d}{d\delta^2} M_{\delta^2}(q_2(b,r)).$$

Since $r\hat{b}_0$ is decreasing in h and h has monotone likelihood ratio in δ^2, $(d/d\delta^2)M_{\delta^2}(q) \leq 0$. Since $M_{\delta^2}(q)$ is a c.d.f.,$(d/dq)M_{\delta^2}(q) \geq 0$. Now

$$\frac{d}{db}[n\,\tau(1+b^2)] = 2n\tau\, b$$

while for rb≠0, $r^2(1-r^2)^{-1} \leq k$,

$$\frac{d}{db}\, q_2(b,r) = \frac{d}{db}\left(\frac{r}{b} + \left|\frac{r}{b}\right| \left[\frac{k(1-r^2)}{r^2}\right]^{1/2}\right)$$

$$= \begin{cases} -\left(1 + \left[\frac{k(1-r^2)}{r^2}\right]^{1/2}\right)\frac{r}{b^2}, & \frac{r}{b} > 0 \\ \left(\left[\frac{k(1-r^2)}{r^2}\right]^{1/2} - 1\right)\frac{r}{b^2}, & \frac{r}{b} < 0 \end{cases}$$

$$= \begin{cases} < 0 & b > 0 \\ > 0 & b < 0. \end{cases}$$

Combining these results, for b≠0

$$\frac{d}{db} P_{n\,\tau(1+b^2)}\{h: \; r\hat{b}_0(h) \leq q_2(b,r)\} = \begin{cases} \leq 0, \; b > 0, \\ \geq 0, \; b < 0, \end{cases} \quad \text{all } r \neq 0.$$

Thus, for b≠0, $(d/db)\Psi(b)$ is ≤ 0 for $b > 0$, ≥ 0 for $b < 0$. Since it is easy to show that $\Psi(b) \leq \Psi(0)$, all b, this completes the proof of Lemma 3. □

<u>Proof of Lemma 4</u>. From (A.1) and the dominated convergence theorem,

$$\lim_{b\to\pm\infty} \Psi(b) = \int_{\frac{r^2}{1-r^2} \leq k} \left[\lim_{b\to\pm\infty} P_{n\tau\,(1+b^2)}\{r\hat{b}_0(h) \leq q_2(b,r)\}\right] f_N(r)\,dr. \tag{A.2}$$

Note that for $r \neq 0$, $r^2(1-r^2)^{-1} \leq k$,

$$\lim_{|b|\to\infty} b\, q_2(b,r) = \lim_{|b|\to\infty} |b|\left(\frac{r}{b} + \frac{r}{b}\left[\frac{k(1-r^2)}{r^2}\right]^{1/2}\right)$$

$$= \begin{cases} r + [k(1-r^2)]^{1/2}, & b \to \infty. \\ -r + [k(1-r^2)]^{1/2}, & b \to -\infty. \end{cases} \tag{A.3}$$

Also, since $h \sim F_{n,n}(n\tau(1+b^2))$, it can be shown that

$$\operatorname*{plim}_{|b|\to\infty} h = \infty, \qquad \frac{h}{b^2} \xrightarrow{L} \frac{n\tau}{\chi_n^2}, \text{ as } b^2 \to \infty, \tag{A.4}$$

where χ_n^2 has the central chi-squared distribution with n degrees of freedom (independent of r, since h is). However, for $r \neq 0$

$$r\hat{b}_0 = \frac{1 - h \pm \sqrt{(h-1)^2 + 4r^2h}}{2h^{1/2}} = \frac{(h-1)}{2h^{1/2}}\left[-1 + \left(1 + \frac{4r^2h}{(h-1)^2}\right)^{1/2}\right]$$

$$= r^2h^{-1/2}(1 + o(1)), \quad h \to \infty.$$

It thus follows from (A.4) that for $r \neq 0$

$$|b|r\hat{b}_0 \xrightarrow{L} r^2\left(\frac{\chi_n^2}{n\tau}\right)^{1/2}. \tag{A.5}$$

Since the chi-squared distributions are continuous, it follows from (A.3) and (A.5) that for $r \neq 0$, $r^2(1-r^2)^{-1} \leq k$,

$$\lim_{b\to\pm\infty} P_{n\tau(1+b^2)}\{r\hat{b}_0(h) \leq q_2(b,r)\}$$

$$= P\{\chi_n^2 \leq \frac{n\tau}{r^2}(\pm r + [k(1-r^2)]^{1/2})^2\}. \tag{A.6}$$

However, $f_{r}(r)$ is symmetric about $r=0$. Hence, from (A.2) and (A.6), the conclusion of Lemma 4 follows. □

Proof that b is median biased. Note that from (4.7)

$$P\{\hat{b} > b\} = P\{\frac{\hat{b}-b}{1+b^2} > 0\} = P\{\frac{\hat{b}_0}{1-b\hat{b}_0} > 0\}$$

$$= P\{1 - b\hat{b}_0 > 0,\ \hat{b}_0 > 0\} + P\{1 - b\hat{b}_0 < 0,\ \hat{b}_0 < 0\}. \quad (A.7)$$

Note from (4.10) that $\hat{b}_0$ and r always have the same sign. Lemma 1 and (4.3) show that $P\{r<0\}=1/2$. Thus, $P\{\hat{b}_0<0\}=1/2$. Assume $b>0$. Then $\hat{b}_0<0$ implies $1-b\hat{b}_0>0$. Consequently (A.7) becomes

$$P\{\hat{b} > b\} = P\{\frac{1}{b} > \hat{b}_0 > 0\} + P\{\hat{b}_0 < 0\} \geq P\{\hat{b}_0 < 0\} = \frac{1}{2}$$

The inequality is strict unless $P\{\frac{1}{b} > \hat{b}_0 > 0\} = 0$, which is easily shown not to be true.

For $b < 0$, similar arguments show

$$P\{\hat{b} > b\} = P\{\hat{b}_0 > 0\} + P\{\frac{1}{b} < \hat{b}_0 < 0\} > P\{\hat{b}_0 > 0\} = \frac{1}{2}.$$

Of course, when $b=0$, $P\{\hat{b} > b\} = P\{\hat{b}_0 > 0\} = \frac{1}{2}$. Thus, $\hat{b}$ is median biased except when $b=0$.

REFERENCES

ADCOCK, R.J. (1878), "A problem in least squares", The Analyst, 5, 53-54.

ANDERSON, T.W. (1976), "Estimation of linear functional relationships: approximate distributions and connection with simultaneous equations in econometrics (with discussion)", Journal of the Royal Statistical Society, Series B, 38, 1-36.

BENTLER, P.M. (1980), "Multivariate analysis with latent variables: Causal modeling", Annual Review of Psychology, 31, 419-456.

BHARGAVA, ANIL K. and LAUD, P.W. (1979), "Some simulation results on asymptotic confidence intervals for parameters in a simple errors in variables model", Technocal Report, Northern Illinois University.

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

CREASY, M.A. (1957), "Confidence limits for the gradient in the linear structural relationship", Journal of the Royal Statistical Society, Series B, 18, 65-69.

GLESER, LEON JAY (1981), "Estimation in a multivariate 'errors in variables' regression model: Large sample results", Annals of Statistics, 9, 24-44.

HEALY, JOHN D. (1979), "On Kristof's test for a linear relation between true scores of two measures", Psychometrika, 44, 235-238.

KENDALL, MAURICE G. and STUART, A. (1979), The Advanced Theory of Statistics (Vol. II, Fourth Edition), New York: Macmillan.

LORD, FREDERIC M. (1973), "Testing if two measuring procedures measure the same dimension", Psychological Bulletin, 79, 71-72.

MORAN, P.A.P. (1971), "Estimating structural and functional relationships", Journal of Multivariate Analysis, 1, 232-255.

SHYR, J-Y (1983). Unpublished technical report.

VILLEGAS, C. (1964), "Confidence region for a linear relation", Annals of Mathematical Statistics 35, 780-788.

WILLIAMS, E.J. (1969), "A note on regression methods in calibration", Technometrics, 11, 189-192.