The Annals of Statistics
1994, Vol. 22, No. 1, 49–66

CONFIDENCE REGIONS IN LINEAR FUNCTIONAL RELATIONSHIPS

BY HEPING ZHANG

Yale University

A unified approach to deriving confidence regions in linear functional relationship models is presented, based on the conditional likelihood ratio method of Knowles, Siegmund and Zhang. In the case of a single latent predictor, the confidence region for the slope produced by this approach is the familiar one of Fieller and Creasy. However, here it is shown how to derive a confidence region for the slope, when it is known that the slope is positive, that improves on merely intersecting the region for an unrestricted slope with $(0, \infty)$. A geometric interpretation is given for Fieller–Creasy confidence region for the ratio of population means (Fieller–Creasy problem). Regions are also derived for simultaneous estimation of the slope and intercept in the model with a single latent predictor, and for the slopes in a model with two latent predictors.

1. Introduction. Starting with Adcock (1878), there has been much interest in the following model,

$$\begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \begin{pmatrix} \mu_i \\ \alpha + \beta\mu_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}, \qquad i = 1, \dots, N, \tag{1}$$

where $\mu_1, \dots, \mu_N$ are unknown parameters and $(\varepsilon'_{1i}, \varepsilon'_{2i})'$, $i = 1, \dots, N$, are i.i.d. $N(0, \Sigma)$. It is called the linear functional relationship model [see, e.g., Anderson (1976, 1984), Fuller (1987), Gleser and Hwang (1987), Kendall and Stuart (1979), Johansen (1984) and Seber and Wild (1989)]. Although it is called linear, this model is a nonlinear regression model. In the literature, the identifiability of model (1) is a concern. A widely used assumption is that $\Sigma = \sigma^2\Sigma_0$, with Σ_0 known. See Creasy (1956), Fuller (1987) and Gleser (1987) for more information. In what follows, without loss of generality we assume that $\Sigma = \sigma^2 I_2$.

The maximum likelihood estimates of the parameters are known; their asymptotic behavior has been studied by many authors. A property of the present model, which causes difficulties for asymptotic theory, is that the number of nuisance parameters tends to infinity as the number of observations tends to infinity. For more details, we refer to Anderson (1976), Fuller (1987), Kendall and Stuart (1979), Johansen (1984) and Seber and Wild (1989).

Finding confidence regions for the slope β is of particular interest in the literature. There are two general approaches, one based on asymptotic theory [cf. Gleser (1987)] and the other using the idea of constructing a suitable

Received December 1990; revised August 1993.
[1]Research supported in part by NSF Grant DMS-89-96526-03.
AMS 1991 *subject classifications*. Primary 62F25; secondary 62E15, 62J02.
Key words and phrases. Linear functional relationship, confidence region, upcrossings.

pivotal statistic. In the 1950's, Fieller (1954), Williams (1955) and Creasy (1956) developed a constructive approach that makes use of a pivotal statistic. Schneeweiss (1982) offers a very neat summary for this approach and also extends the idea to multivariate models. This approach is straightforward if one is able to decide what kind of statistic to use as a pivot. However, except that one wants to choose a statistic whose distribution is easy to find and free of nuisance parameters, no standard is provided in the literature for how to generate the pivotal statistic. Another problem of this approach is lack of flexibility because confidence regions cannot be naturally adjusted when models are changed. This will be clearer when we deal with a variety of concrete problems later on.

This paper presents a unified approach adapted from Knowles, Siegmund and Zhang (1991) to constructing confidence regions for the parameters in a linear functional relationship model. This approach is based on a conditional likelihood ratio test of specified values for the parameters in question. Although the general method of Knowles, Siegmund and Zhang (1991) yields conservative confidence regions, the regions obtained here are exact. Our approach can be used as a guide to generating pivotal statistics; for instance, the Fieller–Creasy–Williams confidence region for the slope β is obtained by our method in Section 2. In addition to its conceptual unity, the conditional likelihood ratio approach has the advantage of being able to incorporate restrictions of range for parameters and yields confidence regions with uniform conditional coverage probability.

The paper is organized as follows. Section 2 contains the main results. It begins with the derivation of the conditional likelihood ratio (CLR) confidence region for the slope β and shows how to obtain a region for β when $\beta > 0$ that improves on merely intersecting the region for unconstrained β with $(0, \infty)$. A joint confidence region for the slope β and intercept α is also constructed. In Section 3, a geometric interpretation is presented for somewhat related Fieller–Creasy problem of estimating ratios of means. Finally, Section 4 extends the method to a multivariate linear functional relationship model.

2. Confidence regions.

2.1. *CLR confidence region for* β. To simplify notation, we rewrite model (1) as

$$\mathbf{y} = \begin{pmatrix} \mathbf{0} & I_N \\ \mathbf{1}_N & \beta I_N \end{pmatrix} \begin{pmatrix} \alpha \\ \mu \end{pmatrix} + \varepsilon, \tag{2}$$

where $\mathbf{0} = (0, \dots, 0)'$, $\mathbf{1}_N = (1, \dots, 1)'$, I_N is the $N \times N$ identity matrix, $\mathbf{y} = (y_{11}, \dots, y_{1N}, y_{21}, \dots, y_{2N})'$, $\varepsilon = (\varepsilon_{11}, \dots, \varepsilon_{1N}, \varepsilon_{21}, \dots, \varepsilon_{2N})'$ and $\mu = (\mu_1, \dots, \mu_N)'$. Let

$$X(\beta) = (1 + \beta^2)^{-1/2} \begin{pmatrix} I_N & -\beta N^{-1/2}\mathbf{1}_N \\ \beta I_N & N^{-1/2}\mathbf{1}_N \end{pmatrix} \tag{3}$$

and $P(\beta) = X(\beta)X'(\beta)$. For fixed $\beta, P(\beta)$ is the projection matrix onto the space spanned by the columns of the design matrix in model (2), and hence the residual sum of squares (RSS) after minimization with respect to α and μ is

$$\text{(4)} \qquad \text{RSS}(\beta) = \mathbf{y}'[I_{2N} - P(\beta)]\mathbf{y}.$$

For generic vectors **a** and **b**, let

$$s_{\mathbf{ab}} = \mathbf{a}'(I_N - \mathbf{1}_N\mathbf{1}_N'/N)\mathbf{b},$$

and let $s_{\mathbf{a}}^2 = s_{\mathbf{aa}}$. Let $\mathbf{y}_1 = (y_{11}, \ldots, y_{1N})'$ and $\mathbf{y}_2 = (y_{21}, \ldots, y_{2N})'$. It follows from (3) and (4) that

$$\text{(5)} \qquad \text{RSS}(\beta) = \left(\beta^2 s_{\mathbf{y}_1}^2 - 2\beta s_{\mathbf{y}_1\mathbf{y}_2} + s_{\mathbf{y}_2}^2\right) \Big/ \left(1 + \beta^2\right).$$

Setting the derivative of (5) with respect to β to zero, we have

$$\text{(6)} \qquad \min_\beta \text{RSS}(\beta) = \left[s_{\mathbf{y}_1}^2 + s_{\mathbf{y}_2}^2 - \sqrt{\left(s_{\mathbf{y}_2}^2 - s_{\mathbf{y}_1}^2\right)^2 + 4s_{\mathbf{y}_1\mathbf{y}_2}^2}\right] \Big/ 2.$$

The log-likelihood ratio statistic for testing $H_0: \beta = \beta_0$ is

$$\text{(7)} \qquad L(\beta_0) = -N \log\left[\min_\beta \text{RSS}(\beta) \Big/ \text{RSS}(\beta_0)\right].$$

Note that when $\beta = \beta_0$, $\|\mathbf{y}\|$ and $\mathbf{y}'X(\beta_0)$ are sufficient statistics for the nuisance parameters α, μ and σ^2, and then our conditional likelihood ratio $100(1 - \nu)\%$ confidence region for β is the set of all β_0 such that

$$\text{(8)} \qquad Pr_{\beta_0}\left[L(\beta_0) > \widehat{L} \mid \|\mathbf{y}\|, \mathbf{y}'X(\beta_0)\right] > \nu,$$

where $\widehat{L}$ is the observed value of $L(\beta_0)$. By conditioning on the sufficient statistics, the conditional probability in (8) does not depend on the nuisance parameters.

Define r as the sample correlation coefficient of

$$\mathbf{w} = (\mathbf{y}_2 - \beta_0\mathbf{y}_1) \Big/ \sqrt{1 + \beta_0^2} \quad \text{and} \quad \mathbf{v} = (\beta_0\mathbf{y}_2 + \mathbf{y}_1) \Big/ \sqrt{1 + \beta_0^2},$$

namely, $r = s_{\mathbf{wv}}/(s_{\mathbf{w}}s_{\mathbf{v}})$; r is the pivotal statistic used by Fieller, Creasy and Williams. Also see Schneeweiss (1982). Note that

$$\mathbf{y}_1 = (\mathbf{v} - \beta_0\mathbf{w}) \Big/ \sqrt{1 + \beta_0^2},$$
$$\mathbf{y}_2 = (\mathbf{w} + \beta_0\mathbf{v}) \Big/ \sqrt{1 + \beta_0^2}.$$

Then

$$\text{(9)} \qquad (1 + \beta_0^2)s_{\mathbf{y}_1}^2 = s_{\mathbf{v}}^2 - 2\beta_0^2 s_{\mathbf{wv}} + \beta_0^2 s_{\mathbf{w}}^2,$$

$$\text{(10)} \qquad (1 + \beta_0^2)s_{\mathbf{y}_1\mathbf{y}_2} = \beta_0(s_{\mathbf{v}}^2 - s_{\mathbf{w}}^2) + (1 - \beta_0^2)s_{\mathbf{wv}},$$

$$\text{(11)} \qquad (1 + \beta_0^2)s_{\mathbf{y}_2}^2 = \beta_0^2 s_{\mathbf{v}}^2 + 2\beta_0 s_{\mathbf{wv}} + s_{\mathbf{w}}^2.$$

It follows from (6) and (9)–(11) that

$$\min_{\beta} \ \mathrm{RSS}(\beta) = \left[s_{\mathbf{w}}^2 + s_{\mathbf{v}}^2 - \sqrt{\left(s_{\mathbf{w}}^2 - s_{\mathbf{v}}^2\right)^2 + 4s_{\mathbf{wv}}^2} \ \right] \Big/ 2. \tag{12}$$

By conditioning on the sufficient statistics in (8), $s_{\mathbf{w}}$ and $s_{\mathbf{v}}$ are fixed, and therefore, $\min_\beta$ RSS(β) [and $L(\beta_0)$ in (7)] is a monotonic function of r^2. The definition of r implies that, under the null hypothesis, the following are true: (i) its distribution is independent of nuisance parameters α, μ, and σ^2, and hence is the same as its conditional distribution given the sufficient statistics in (8) via Basu's theorem; (ii) it is distributed the same as the sample correlation coefficient of two independent, standard normal variables. Therefore, the (conditional and unconditional) distribution of r can be written as

$$F_{N-2}(x) = \int_{-1}^{x} \frac{\Gamma((N-1)/2)}{\pi^{1/2}\Gamma[(N-2)/2]} \left(1-u^2\right)^{(N-4)/2} du, \qquad |x| \le 1. \tag{13}$$

So we can summarize our discussion with the following theorem.

THEOREM 1. *Using previous notation, we have*

$$Pr_{\beta_0}\left[L(\beta_0) > \widehat{L} \mid \|\mathbf{y}\|, \mathbf{y}'X(\beta_0)\right] = 2F_{N-2}\left(-|r_{\mathrm{obs}}|\right),$$

where r_{obs} is the observed value of r.

Now, we see that the Fieller–Creasy–Williams confidence set is also the CLR one. Unfortunately, it is known that the use of r is problematic. The distribution of r is the same under $\beta = \beta_0$ and $-1/\beta_0$. This implies that the confidence set for β obtained from r contains $-1/\beta_0$ if it does for β_0 and is disjoint unless it is the whole line. To avoid this problem, a naive way taken by Creasy (1956) and Kendall and Stuart (1979) is to ignore those portions of the confidence set for β which are not around the MLE $\widehat{\beta}$. Since $\widehat{\beta}$ is a consistent estimate of β [cf. Fuller (1987)], there is no loss essentially by doing that when one has enough data. There are other proposals in the literature on modifying r to resolve the problem of having disconnected intervals [see, e.g., Fuller (1987)].

In many applications, we may have an a priori idea about the slope β. For example, let us consider $\beta > 0$; the situation for $\beta \in (B_1, B_2)$ is similar. With the restriction $\beta > 0$, $\min_\beta$ RSS(β) in (7) is replaced with $\min_{\beta>0}$ RSS(β). Some calculations lead to

$$\min_{\beta>0} \ \mathrm{RSS}(\beta) = \begin{cases} \min_\beta \ \mathrm{RSS}(\beta), & \text{if } \widehat{\beta} > 0, \text{ that is, } s_{\mathbf{y}_1\mathbf{y}_2} > 0, \\ \min\left(s_{\mathbf{y}_1}^2, s_{\mathbf{y}_2}^2\right), & \text{otherwise.} \end{cases}$$

Let $d = s_{\mathbf{w}}/s_{\mathbf{v}} - s_{\mathbf{v}}/s_{\mathbf{w}}$. It follows from (9)–(12) that

$$(14)\quad Pr_{\beta_0}\left[L(\beta_0) > \widehat{L} \mid \|\mathbf{y}\|, \mathbf{y}'X(\beta_0)\right]$$

$$(15)\quad = P\left[r^2 > r_{\text{obs}}^2 \text{ and } (1-\beta_0^2)r > \beta_0 d\right]$$

$$(16)\quad +P\left[r > \frac{(1+\beta_0^2)\sqrt{d^2+4r_{\text{obs}}^2} + (\beta_0^2-1)d}{4\beta_0} \text{ and } (1-\beta_0^2)r \le \beta_0 d\right]$$

$$(17)\quad +P\left[r < \frac{-(1+\beta_0^2)\sqrt{d^2+4r_{\text{obs}}^2} + (\beta_0^2-1)d}{4\beta_0} \text{ and } (1-\beta_0^2)r \le \beta_0 d\right].$$

Therefore, (14) can be computed numerically through the distribution of r, although it is analytically more complicated than what is given in Theorem 1. For convenience, let $\mathcal{I}_p(\nu)$ and $\mathcal{I}(\nu)$ denote the $1-\nu$ confidence sets for β when the prior information that $\beta > 0$ is and is not available, respectively. Since $\min_{\beta>0}$ RSS$(\beta) \ge \min_\beta$ RSS(β), $\mathcal{I}_p(\nu) \subset \mathcal{I}(\nu) \cap (0,\infty)$. Let

$$p = F_{N-2}\left[-|s_{\mathbf{y}_1\mathbf{y}_2}|/(s_{\mathbf{y}_1}s_{\mathbf{y}_2})\right].$$

When β_0 tends to ∞, the p-value at β_0 without the prior information (monotonically) tends to $2p$; however, the p-value at β_0 with the prior information tends to p. Note that $2p$ is also the significance level for testing hypothesis $\beta = 0$ against $\beta \neq 0$, and p is the level for testing $\beta = 0$ against $\beta > 0$. When $2p$ is much smaller than ν (i.e., $\beta = 0$ is rejected at a very significant level), then $\mathcal{I}_p(\nu)$ and $\mathcal{I}(\nu) \cap (0,\infty)$ are intervals and hardly different. When $2p$ is much larger than ν (i.e., $\beta = 0$ cannot be rejected at all), then both $\mathcal{I}_p(\nu)$ and $\mathcal{I}(\nu) \cap (0,\infty)$ can be disconnected. In this case, we need to question our prior information $\beta > 0$ or be careful with our data. The discrepancy arises when $2p$ is at a marginal level. More precisely, when $\nu/2 < p \le \nu, \mathcal{I}_p(\nu)$ can be an interval, but $\mathcal{I}(\nu) \cap (0,\infty)$ is not. When $p \approx \nu/2$, the left end point of $\mathcal{I}_p(\nu)$ is further away from zero than that of $\mathcal{I}(\nu) \cap (0,\infty)$. To illustrate these, let us examine two simulated data sets. Take $\mu_i = i/2$, for $i = 1, \ldots, 20$; y_{1i} is the sum of μ_i and a normal random number, and y_{2i} is the sum of $0.2\mu_i$ and a normal random number. Table 1 lists the generated $\mathbf{y}_1$ and $\mathbf{y}_2$. Numerical computations give $p = 0.02$, and this implies that $\beta = 0$ can be rejected at the significance level 0.05 from a two-sided test. Moreover, $\mathcal{I}_p(0.05) = (0.039, 0.363)$ and $\mathcal{I}(0.05) \cap (0,\infty) = (0.009, 0.363)$. So, $\mathcal{I}_p(0.05)$ is about 8.5% shorter than $\mathcal{I}(0.05) \cap (0,\infty)$. Next, another set of $\mathbf{y}_1$ and $\mathbf{y}_2$ is generated. This time, we have $p = 0.038$. The two-sided test rejects the hypothesis: $\beta = 0$ at level 0.05 but not at 0.1. $\mathcal{I}_p(0.05) = (0.018, 0.495)$ is again an interval, but $\mathcal{I}(0.05) \cap (0,\infty) = (0, 0.495) \cup (37.2, \infty)$ is disjoint. A simple S code is available upon request from the author.

TABLE 1
Simulated data

Set 1		Set 2	
$\mathbf{y}_1$	$\mathbf{y}_2$	$\mathbf{y}_1$	$\mathbf{y}_2$
1.229	0.59252	−0.0276	−0.6906
−0.518	−1.44312	2.2531	2.8548
1.195	0.98665	2.2623	−0.0731
2.957	−0.34819	1.9414	−0.5859
5.481	2.47289	2.4705	−0.1612
3.552	−0.00811	4.7233	−1.4533
3.469	0.93196	4.1708	3.0119
4.956	−0.27197	4.3310	2.3677
4.164	1.13357	3.0909	−0.5976
5.118	1.95072	5.3934	−0.9106
5.703	1.66155	5.0477	1.2049
6.226	2.00264	6.8991	1.3481
6.871	0.67623	7.5014	1.1699
6.896	1.21482	6.0355	−0.4378
8.848	2.45088	8.6153	1.5291
7.025	0.99575	8.2811	1.7803
8.793	−0.36699	10.7966	1.1784
7.400	2.27000	11.1708	0.3557
10.265	2.01203	9.3575	3.7896
11.171	0.69594	11.0389	2.5982

2.2. *Joint confidence region for* (α, β). The log-likelihood ratio statistic for testing $H_0: \beta = \beta_0, \alpha = \alpha_0$ is

$$L(\beta_0, \alpha_0) = -N \log\left[\min_{\beta} \mathrm{RSS}(\beta)/\mathrm{RSS}(\beta_0, \alpha_0)\right], \tag{18}$$

where $\mathrm{RSS}(\beta_0, \alpha_0)$ is the residual sum of squares due to model (2) for $(\beta, \alpha) = (\beta_0, \alpha_0)$. Our CLR $1 - \nu$ confidence region for (β, α) is the set of all (β_0, α_0) such that

$$Pr_{H_0}\left[L(\beta_0, \alpha_0) > \widehat{L} \mid \mathcal{S}\right] > \nu, \tag{19}$$

where $\mathcal{S} = \{\|\mathbf{y}_1\|^2 + \|\mathbf{y}_2 - \alpha_0 \mathbf{1}_N\|^2, \mathbf{y}_1 + \beta_0 \mathbf{y}_2\}$ is a set of sufficient statistics with respect to μ and σ^2 when $\alpha = \alpha_0$ and $\beta = \beta_0$, and $\widehat{L}$ is the observed value of $L(\beta_0, \alpha_0)$. Observe that

$$L(\beta_0, \alpha_0) = L(\beta_0) + \tilde{L}(\beta_0, \alpha_0),$$

where $\tilde{L}(\beta_0, \alpha_0) = -N \log[\mathrm{RSS}(\beta_0)/\mathrm{RSS}(\beta_0, \alpha_0)]$.

To evaluate the left-hand side of (19), the following two facts are important. First,

$$\tilde{L}(\beta_0, \alpha_0) = N \log\left[1 + \frac{t_{N-1}^2}{N-1}\right],$$

where $t_{N-1}^2 = N(N-1)\left(\bar{\mathbf{w}} - \alpha_0/\sqrt{1+\beta_0^2}\right)^2/s_{\mathbf{w}}^2$ has Student's t-distribution with $N-1$ degrees of freedom and is independent of $\mathcal{S}$. Second, given $\mathcal{S}$ and $\tilde{L}(\beta_0, \alpha_0)$, $L(\beta_0)$ is a monotonic function of r which is independent of $\tilde{L}(\beta_0, \alpha_0)$ and $\mathcal{S}$.

Let $c^2 = 1 - \exp(-\widehat{L}/N)$ and $w = 1 -$ RSS$(\beta_0, \alpha_0)(1-c^2)/s_{\mathbf{v}}^2$. To summarize our discussion, we have [see Zhang (1991) for more details] the following theorem.

THEOREM 2. *Using previous notation, the left-hand side of* (19) *equals*

$$Pr_{H_0}\left[\tilde{L}(\beta_0, \alpha_0) > \widehat{L} \mid \mathcal{S}\right] + Pr_{H_0}\left[L(\beta_0) > \widehat{L} - \tilde{L}(\beta_0, \alpha_0); \tilde{L}(\beta_0, \alpha_0) \le \widehat{L} \mid \mathcal{S}\right]$$
$$= 2F_{N-1}(-c) + \int_{-c}^{c} F_{N-2}\left[\sqrt{(c^2 - x^2)w/(1-x^2)}\right] f_{N-1}(x)\, dx,$$

where F_{N-2} is defined by (13), *F_{N-1} can be obtained by replacing $N-2$ with $N-1$ in* (13) *and f_{N-1} is the density function corresponding to F_{N-1}.*

REMARK 1. Based on the discussion in subsection 2.1, we expect that the joint CLR confidence region for (β, α) may have two (or more) disjoint pieces. The one, say R_1, around $(\widehat{\beta}, \widehat{\alpha})$ can be easily obtained by a contouring algorithm [see, e.g., Zhang (1991)]. When R_1 is found, let β_1 and β_2 be the minimum and maximum of β in R_1, respectively. Now, increase β_0 from β_2 and also decrease it from β_1 further and further. For any β_0, if the left-hand side of (19) is greater than ν for some α_0, use the contouring algorithm to find one piece of the confidence region nearby. This search usually is not difficult because the conditional probability in (19) may have only two or so maxima.

REMARK 2. In applications, some components of μ, defined in (2), may be assumed to be the same. Then, corresponding to (2), we have the replication model

$$\mathbf{y} = \alpha\begin{pmatrix} 0 \\ \mathbf{1}_N \end{pmatrix} + \begin{pmatrix} I_N \\ \beta I_N \end{pmatrix}\begin{pmatrix} \mathbf{1}_{n_1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{1}_{n_p} \end{pmatrix}\mu + \varepsilon, \tag{20}$$

where $\sum_1^p n_i = N$. When the n_i's are equal, modification of the approach suggested here for model (20) is straightforward. The case of unequal n_i will be discussed in a separate paper.

3. Geometric interpretation of the Fieller–Creasy problem. Fieller (1954) considered the following model

$$y_i = \begin{cases} \mu + \varepsilon_i, & \text{for } i \le N, \\ \mu\beta + \varepsilon_i, & \text{for } N < i \le 2N, \end{cases} \tag{21}$$

where we assume that the ε_i's are i.i.d. $N(0, \sigma^2)$. This model comes from a situation where one is concerned with the ratio of two population means, and it is of particular interest in the literature [see, e.g., Cook and Witmer (1985)] because its simple structure allows many authors to illustrate constructively and geometrically their methodology. In contrast to model (1), model (21) does not include the intercept term α, and μ is not a vector but a scalar parameter. These two models have a common feature that both independent and dependent variables (corresponding to regression model) involve measurement errors. Although the geometry of model (21) is not completely the same as that for model (1), understanding model (21) is still helpful for getting more insight into model (1).

3.1. *The unconstrained case.* First, consider $-\infty < \mu, \beta < \infty$. Normalizing the observed vector $\mathbf{y} = (y_1, \dots, y_{2N})'$, we get a unit length vector, say $\mathbf{u}$. Partition $\mathbf{u}$ into two N-vectors $\mathbf{u}_{(1)}$ and $\mathbf{u}_{(2)}$. Then $\widehat{\beta} \doteq \bar{\mathbf{u}}_{(2)}/\bar{\mathbf{u}}_{(1)}$. The r defined in subsection 2.1 corresponds to

$$\frac{\left(\bar{\mathbf{u}}_{(2)} - \beta_0 \bar{\mathbf{u}}_{(1)}\right)^2}{1 - N\left(\bar{\mathbf{u}}_{(1)} + \bar{\mathbf{u}}_{(2)}\right)^2}. \tag{22}$$

Now, introduce a curve γ on the unit sphere as

$$\gamma(\beta) = (1, \dots, 1, \beta, \dots, \beta)' \Big/ \sqrt{N(1+\beta^2)}.$$

The sufficient statistics in the present model are $\|y\|$ and $\gamma'(\beta_0)\mathbf{u}$ when $\beta = \beta_0$. Let $z = \gamma'(\beta_0)\mathbf{u}$ and $w = \gamma'(\widehat{\beta})\mathbf{u}$. The $100(1-\nu)\%$ CLR confidence interval for β is the set of β_0 for which

$$Pr\left[\max_{\beta} |\gamma'(\beta)\mathbf{u}| > w \,\Big|\, \gamma(\beta_0)'\mathbf{u} = z\right] > \nu, \tag{23}$$

where $\mathbf{u}$ is uniformly distributed on a unit sphere in R^{2N}. Let $p(\beta_0)$ denote the conditional tail probability in (23). Note that $\gamma(\beta)$ is a geodesic curve (zero geodesic curvature) ending at

$$(0, \dots, 0, 1, \dots, 1)'/N^{1/2} \quad \text{and} \quad -(0, \dots, 0, 1, \dots, 1)'/N^{1/2}.$$

By rotation, we can assume that $\gamma(\beta)$ is a semequator and $\gamma(\beta_0)$ is the first coordinate axis $\mathbf{e}_1$. The set, denoted by $\mathcal{A}$, of $\mathbf{u}$ for which

$$\gamma(\beta_0)'\mathbf{u} = z$$

is a circle perpendicular to $\mathbf{e}_1$ with radius $\sqrt{1-z^2}$, that is, geodesic radius $\cos^{-1} z$.

Since the union of γ and $-\gamma$ constitutes a full equator, $p(\beta_0)$ is the sum of volumes of two congruent caps on $\mathcal{A}$; see Figure 1(a). The condition $\gamma(\beta)'\mathbf{u} > w$ implies that the caps are in the tube around γ with geodesic radius $\cos^{-1} w$.

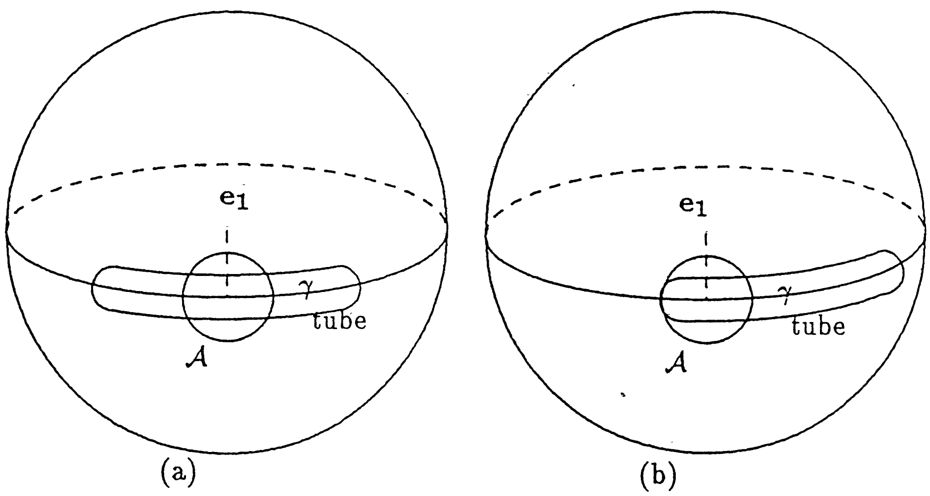

FIG. 1. *Geometric interpretation for Fieller–Creasy problem.*

The boundary of the caps satisfies $|u_2| = \sqrt{w^2 - z^2}$. Since the radius of $\mathcal{A}$ is $\sqrt{1 - z^2}$, we have

$$p(\beta_0) = Pr\left[|u_1^{(2N-1)}| > \sqrt{\frac{w^2 - z^2}{1 - z^2}}\right], \tag{24}$$

where $u_1^{(2N-1)}$ is the first coordinate of a point uniformly distributed on a unit sphere in R^{2N-1}.

It is obvious that (24) gives the same answer as if we start from r in (22). For model (1), the confidence interval for β is always disjoint except when it is a whole line; whereas the disjoint intervals occur for model (21) only when the estimate of μ is close to zero. In fact, if μ is near zero, there is an identifiability problem for β because β can be $+\infty$ or $-\infty$. In this case, one should regard $+\infty$ and $-\infty$ as the same point for which the confidence interval for β is no longer disjoint.

3.2. *The constrained case.* Now, suppose $\hat{\beta} > 0$. For simplicity, let us assume that $\bar{\mathbf{u}}_{(2)}/\bar{\mathbf{u}}_{(1)} > 0$. Otherwise, we take $\hat{\beta} = 0$. Then r in (22) is the same as before; apparently it has a defect, for $\arg\max_\beta, \gamma(\beta)'\mathbf{u}$ may not be in $(0, \infty)$ for an arbitrary point $\mathbf{u}$.

In the current situation, the curve γ traverses

$$\mathbf{p}_1 = (0, \ldots, 0, 1, \ldots, 1)'/N^{1/2} \quad \text{and} \quad \mathbf{p}_2 = (1, \ldots, 1, 0, \ldots, 0)'/N^{1/2}.$$

By rotation, I can still assume that γ is one quarter of the equator and $\gamma(\beta_0) =$

$\mathbf{e}_1$. Hence γ and $-\gamma$ are two disjoint pieces of the equator. Let

$$\delta_1 = \gamma(\beta_0)'\mathbf{p}_1 = \beta_0 \Big/ \sqrt{1+\beta_0^2},$$
$$\delta_2 = \gamma(\beta_0)'\mathbf{p}_2 = 1 \Big/ \sqrt{1+\beta_0^2}.$$

Without loss of generality I consider only $z > 0$.

CASE 1 $(\delta_1 \le z/w$ and $\delta_2 \le z/w)$. The conditional sphere $\mathcal{A}$ intersects the tube $\max_\beta \gamma(\beta)'\mathbf{u} > w$, resulting in two maximum caps of the same volume; see Figure 1(a). Thus,

$$p(\beta_0) = Pr\left[\left|u_1^{(2N-1)}\right| > \sqrt{\frac{w^2 - z^2}{1-z^2}}\right].$$

This is the same as with the unconstrained case.

CASE 2 $(\delta_1 \le z/w$ and $\delta_2 > z/w)$. The intersection gives one large cap near either $\mathbf{p}_1$ or $-\mathbf{p}_1$ and another smaller cap near the other end; see Figure 1(b). We have

$$p(\beta_0) = Pr\left[u_1^{(2N-1)} > \sqrt{\frac{w^2 - z^2}{1-z^2}}\right] + Pr\left[u_1^{(2N-1)} > \frac{w - z\delta_2}{\sqrt{(1-\delta_2^2)(1-z^2)}}\right].$$

CASE 3 $(\delta_1 > z/w$ and $\delta_2 \le z/w)$. Similar to Case 2, we get

$$p(\beta_0) = Pr\left[u_1^{(2N-1)} > \sqrt{\frac{w^2 - z^2}{1-z^2}}\right] + Pr\left[u_1^{(2N-1)} > \frac{w - z\delta_1}{\sqrt{(1-\delta_1^2)(1-z^2)}}\right].$$

CASE 4 $(\delta_1 > z/w$ and $z/w)$.

$$p(\beta_0) = Pr\left[u_1^{(2N-1)} > \min\left\{\frac{w - z\delta_1}{\sqrt{(1-\delta_1^2)(1-z^2)}}, \frac{w + z\delta_2}{\sqrt{(1-\delta_2^2)(1-z^2)}}\right\}\right]$$
$$+ Pr\left[u_1^{(2N-1)} > \min\left\{\frac{w + z\delta_1}{\sqrt{(1-\delta_1^2)(1-z^2)}}, \frac{w - z\delta_2}{\sqrt{(1-\delta_2^2)(1-z^2)}}\right\}\right].$$

Thus the classical confidence interval for β is always conservative since it assigns the maximum probability $Pr[|u_1^{(2N-1)}| > \sqrt{(w^2 - z^2)/(1-z^2)}]$ to $p(\beta_0)$.

These exercises show how to adjust $p(\beta_0)$ when the restriction $\beta > 0$ is introduced. The identifiability of β is not an issue, but boundary effects become very important.

4. A multivariate model. In this section, we discuss the linear functional model with two latent predictors. The extension to the general multivariate model is straightforward but the computation would be more complicated.

Analogous to model (1) is

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{pmatrix} = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \alpha + \beta\mu_{1i} + \xi\mu_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix}, \qquad i = 1, \dots, N. \tag{25}$$

Like (2), model (25) can be written as

$$\mathbf{y} = \begin{pmatrix} \mathbf{0}_N & I_N & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{0}_N & I_M \\ \mathbf{1}_N & \beta I_N & \xi I_N \end{pmatrix} \begin{pmatrix} \alpha \\ \mu_1 \\ \mu_2 \end{pmatrix} + \varepsilon. \tag{26}$$

We assume that $\varepsilon \sim N(0, \sigma^2 I_{3N})$ for the sake of simplicity. The joint confidence region for (β, ξ) is of interest.

For fixed (β, ξ), the residual sum of squares due to (26) is $\mathbf{y}'(I - H(\beta, \xi))\mathbf{y}$; here $H(\beta, \xi) = X(\beta, \xi)X'(\beta, \xi)$ with

$$X(\beta, \xi) = \begin{pmatrix} \delta^{-1}I_N & -\beta\xi(\delta\Delta)^{-1}I_N & -\beta(N\Delta)^{-1}\mathbf{1}_N \\ \mathbf{0}_N & \delta\Delta^{-1}I_N & -\xi(N\Delta)^{-1}\mathbf{1}_N \\ \beta\delta^{-1}I_N & -\xi(\delta\Delta)^{-1}I_N & (N\Delta)^{-1}\mathbf{1}_N \end{pmatrix}, \tag{27}$$

where $\delta = \sqrt{1 + \beta^2}$ and $\Delta = \sqrt{1 + \beta^2 + \xi^2}$. Note that $H(\beta, \xi)$ is the projection matrix onto the space spanned by the columns of the design matrix in (26) or, equivalently, by the orthonormal columns of $X(\beta, \xi)$. In fact, the column vectors of $X(\beta, \xi)$ are obtained by the Gram–Schmidt scheme, starting from the second column of the design matrix in (26) to the last one and ending up with the first one.

Given (β, ξ), the sufficient statistics with respect to the nuisance parameters σ^2, α, μ_1, and μ_2 are $\|\mathbf{y}\|$ and $X'(\beta, \xi)\mathbf{y}$. The conditional likelihood $1 - \nu$ confidence region for (β, ξ) is the set of all (β_0, ξ_0) such that

$$Pr_{(\beta_0, \xi_0)}\left[L(\beta_0, \xi_0) > \widehat{L} \,\middle|\, \|\mathbf{y}\|, X'(\beta_0, \xi_0)\mathbf{y}\right] > \nu, \tag{28}$$

where

$$L(\beta_0, \xi_0) = -1.5N \log\left(\frac{\|\mathbf{y}\|^2 - \max_{\beta, \xi} \mathbf{y}'H(\beta, \xi)\mathbf{y}}{\|\mathbf{y}\|^2 - \mathbf{y}'H(\beta_0, \xi_0)\mathbf{y}}\right)$$

and $\widehat{L}$ is the observed value of $L(\beta_0, \xi_0)$; $L(\beta_0, \xi_0)$ is an increasing function of $\max_{\beta,\xi} \mathbf{u}'H(\beta,\xi)\mathbf{u}$ (here $\mathbf{u} = \mathbf{y}/\|\mathbf{y}\|$) when the sufficient statistics are conditioned on their observed values, say, $X'(\beta_0,\xi_0)\mathbf{u} = \mathbf{v}$. By the conditioning on the sufficient statistics, we can assume the $\mathbf{y} \sim N(0, I_{3N})$, namely, $\mathbf{u}$ is uniformly distributed on a unit sphere in R^{3N}. The left-hand side of (28) equals

$$Pr_{(\beta_0,\xi_0)}\left[\max_{\beta,\xi} \mathbf{u}'H(\beta,\xi)\mathbf{u} > w \middle| X'(\beta_0,\xi_0)\mathbf{u} = \mathbf{v}\right] > \nu, \tag{29}$$

where w is the observed value of $\max_{\beta,\xi} \mathbf{u}'H(\beta,\xi)\mathbf{u}$.

To evaluate (29), choose an orthogonal matrix Q_0 such that $X'(\beta_0,\xi_0)Q_0$ is diagonal; that is,

$$Q_0 = \begin{pmatrix} \delta_0^{-1} & -\beta_0\xi_0(\delta_0\Delta_0)^{-1}I_N & -\beta_0(N\Delta_0)^{-1} \\ 0 & \delta_0\Delta_0^{-1} & -\xi_0(N\Delta_0)^{-1} \\ \beta_0\delta_0^{-1} & -\xi_0(\delta_0\Delta_0)^{-1} & (N\Delta_0)^{-1} \end{pmatrix} \otimes I_N, \tag{30}$$

where the operator $\otimes$ is the Kronecker product,

$$\delta_0 = \sqrt{1+\beta_0^2} \text{ and } \Delta_0 = \sqrt{1+\beta_0^2+\xi_0^2}.$$

It follows that

$$X'(\beta,\xi)Q_0 = \begin{pmatrix} \frac{1+\beta\beta_0}{\delta\delta_0}I_N & \frac{\xi_0(\beta-\beta_0)}{\delta\delta_0\Delta_0}I_N & \frac{\beta-\beta_0}{\delta\Delta_0}I_N \\ \frac{\xi(\beta_0-\beta)}{\delta_0\delta\Delta}I_N & \frac{(1+\beta\beta_0)\xi\xi_0+\delta^2\delta_0^2}{\delta\delta_0\Delta\Delta_0}I_N & \frac{(1+\beta\beta_0)\xi-\delta^2\xi_0}{\delta\Delta\Delta_0}I_N \\ \frac{\beta_0-\beta}{N^{1/2}\delta_0\Delta}\mathbf{1}_N' & \frac{(1+\beta\beta_0)\xi_0-\delta_0^2\xi}{N^{1/2}\delta_0\Delta_0\Delta}\mathbf{1}_N' & \frac{1+\beta\beta_0+\xi\xi_0}{N^{1/2}\Delta_0\Delta}\mathbf{1}_N' \end{pmatrix}; \tag{31}$$

and then $X'(\beta_0,\xi_0)Q_0\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2', N^{1/2}\bar{\mathbf{u}}_3)'$, where $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2', \mathbf{u}_3')'$.

Let

$$\theta_1 = \frac{(\beta-\beta_0)\Delta_0}{(1+\beta\beta_0+\xi\xi_0)\delta_0} \quad \text{and} \quad \theta_2 = \frac{\delta_0^2(\xi-\xi_0)-\beta_0\xi_0(\beta-\beta_0)}{(1+\beta\beta_0+\xi\xi_0)\delta_0}.$$

Using (31), after some tedious calculation (see Appendix 1) we have

$$\begin{aligned} \|X'(\beta,\xi)Q_0\mathbf{u}\|^2 = \frac{1}{1+\theta_1^2+\theta_2^2}\Big\{ & 2\theta_1(\mathbf{u}_1'\mathbf{u}_3 - N\bar{\mathbf{u}}_1\bar{\mathbf{u}}_3) + 2\theta_2(\mathbf{u}_2'\mathbf{u}_3 - N\bar{\mathbf{u}}_2\bar{\mathbf{u}}_3) \\ & + \|\mathbf{u}_1\|^2 + \|\mathbf{u}_2\|^2 + N\bar{\mathbf{u}}_3^2 - 2\theta_1\theta_2 s_{\mathbf{u}_1\mathbf{u}_2} \\ & + \theta_1^2(1-s_{\mathbf{u}_1}^2) + \theta_2^2(1-s_{\mathbf{u}_2}^2)\Big\}. \end{aligned} \tag{32}$$

Since the distribution of $\mathbf{u}$ is invariant under orthogonal transformation, $\mathbf{u}$ in (29) is now replaced with $Q_0\mathbf{u}$. The condition in (29) becomes $(\mathbf{u}_1', \mathbf{u}_2', N^{1/2}\bar{\mathbf{u}}_3')' = \mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2', v_{2N+1})'$. We see now that (32) depends on the distribution of $\mathbf{1}_N'\mathbf{u}_3$, $\mathbf{v}_1'\mathbf{u}_3$, and $\mathbf{v}_2'\mathbf{u}_3$, of which $\mathbf{1}_N'\mathbf{u}_3$ appears in the condition. We want to rotate $\mathbf{u}_3$

so that (32) depends on only the first two components of $\mathbf{u}_3$. Based on a Gram–Schmidt scheme, this is done by choosing the three basis vectors

$$\mathbf{e}_{2N+1} = \left(\mathbf{0}'_N, \mathbf{0}'_N, N^{-1/2}\mathbf{1}'_N\right)',$$

$$\text{(33)} \qquad \mathbf{e}_{2N+1} = \left(\mathbf{0}'_N, \mathbf{0}'_N, \frac{\mathbf{v}'_1 - \bar{\mathbf{v}}_1\mathbf{1}'_N}{s_{\mathbf{v}_1}}\right)',$$

$$\text{(34)} \qquad \mathbf{e}_{2N+3} = \left(\mathbf{0}'_N, \mathbf{0}'_N, \frac{\left(\mathbf{v}'_2 - \bar{\mathbf{v}}_2\mathbf{1}'_N - \left(s_{\mathbf{v}_1\mathbf{v}_2}/s_{\mathbf{v}_1}\right)\mathbf{e}'_{2N+2}\right)}{\sqrt{s^2_{\mathbf{v}_2} - s^2_{\mathbf{v}_1\mathbf{v}_2}/s^2_{\mathbf{v}_1}}}\right)'.$$

Equations (33) and (34) imply that $\left(\mathbf{v}'_1\mathbf{u}_3 - N\bar{\mathbf{v}}_1\bar{\mathbf{u}}_3\right)/s_{\mathbf{v}_1} = \mathbf{e}'_{2N+2}\mathbf{u}_3 \equiv u_1$ and

$$\frac{\mathbf{v}'_2\mathbf{u}_3 - N\bar{\mathbf{v}}_2\bar{\mathbf{u}}_3 - \left(s_{\mathbf{v}_1\mathbf{v}_2}/s_{\mathbf{v}_1}\right)u_1}{\sqrt{s^2_{\mathbf{v}_2} - s^2_{\mathbf{v}_1\mathbf{v}_2}/s^2_{\mathbf{v}_1}}} = \mathbf{e}'_{2N+3}\mathbf{u}_3 \equiv u_2,$$

where (u_1, u_2) are the first two coordinates of a point uniformly distributed on the unit sphere in R^{N-1}. After the rotation and via (32), the conditional distribution of $\mathbf{u}'H(\beta,\xi)\mathbf{u}$ in (29) is the same as the unconditional distribution of the following, denoted by $h(\theta_1, \theta_2)$,

$$\frac{1}{1+\theta_1^2+\theta_2^2}\left\{2\theta_1\sqrt{1-\|\mathbf{v}\|^2}s_{\mathbf{v}_1}u_1 + 2\theta_2\sqrt{1-\|\mathbf{v}\|^2}\left(\sqrt{s^2_{\mathbf{v}_2} - s^2_{\mathbf{v}_1\mathbf{v}_2}/s^2_{\mathbf{v}_1}}u_2 + \frac{s_{\mathbf{v}_1\mathbf{v}_2}}{s_{\mathbf{v}_1}}u_1\right) + \|\mathbf{v}\|^2 - 2\theta_1\theta_2 s_{\mathbf{v}_1\mathbf{v}_2} + \theta_1^2\left(1-s^2_{\mathbf{v}_1}\right) + \theta_2^2\left(1-s^2_{\mathbf{v}_2}\right)\right\}.$$

We leave it to Appendix 2 for evaluating

$$\text{(35)} \qquad Pr\left[\max_{\theta_1,\theta_2} h(\theta_1,\theta_2) > w\right].$$

REMARK 3. Extending an idea of Williams (1955), Schneeweiss (1982) suggested a pivotal statistic based on the empirical coefficient of multiple correlation to derive the confidence region for (β, ξ) for the model (26). Although his procedure is simpler, it is not obvious that the confidence region contains $(\widehat{\beta,\xi})$, and one degree of freedom is sacrificed in his procedure since he uses only $N-3$ degrees of freedom. If there are k latent predictors, then $k-1$ degrees of freedom will be wasted. This may imply a loss of information. In contrast, the approach presented here is based on sufficient statistics and uses all $N-2$ degrees of freedom after paying $2N+2$ degrees of freedom to estimate α, μ_1, μ_2 and σ.

APPENDIX 1

A.1. Derivation of (32). First, note that

$$\text{(36)} \qquad 1+\theta_1^2+\theta_2^2 = \frac{\Delta^2\Delta_0^2}{(1+\beta\beta_0+\xi\xi_0)^2} \quad \text{and} \quad \delta^2\delta_0^2 = (1+\beta\beta_0)^2 + (\beta-\beta_0)^2.$$

To find $\|X'(\beta,\xi)Q_0\mathbf{u}\|^2$ we expand it in terms of $\|\mathbf{u}_1\|^2$, $\|\mathbf{u}_2\|^2$, $\mathbf{u}_1'\mathbf{u}_2, \mathbf{u}_1'\mathbf{u}_3, \mathbf{u}_2'\mathbf{u}_3$, $\bar{\mathbf{u}}_1$, $\bar{\mathbf{u}}_2$, and $\bar{\mathbf{u}}_3$, and also note that $1-\|\mathbf{u}_3\|^2 = \|\mathbf{u}_1\|^2 + \|\mathbf{u}_2\|^2$.

The coefficient for $\|\mathbf{u}_1\|^2$ is

$$\begin{aligned}
&\frac{(1+\beta\beta_0)^2}{\delta^2\delta_0^2} + \frac{\xi^2(\beta-\beta_0)^2}{\delta^2\delta_0^2\Delta^2} - \frac{(\beta-\beta_0)^2}{\delta^2\Delta_0^2} - \frac{\left[(1+\beta\beta_0)\xi - \delta^2\xi_0\right]^2}{\delta^2\Delta^2\Delta_0^2} \\
&\quad = 1 - \frac{(\beta-\beta_0)^2}{\delta^2\delta_0^2} + \frac{\xi^2(\beta-\beta_0)^2}{\delta^2\delta_0^2\Delta^2} \\
&\qquad - \frac{(\beta-\beta_0)^2}{\delta^2\Delta_0^2} - \frac{\left[(1+\beta\beta_0)\xi - \delta^2\xi_0\right]^2}{\delta^2\Delta^2\Delta_0^2} \quad \text{[via (36)]} \\
&\quad = -\frac{(\beta-\beta_0)^2}{\Delta^2\delta_0^2} + 1 - \frac{(\beta-\beta_0)^2}{\delta^2\Delta_0^2} - \frac{\left[(1+\beta\beta_0)\xi - \delta^2\xi_0\right]^2}{\delta^2\Delta^2\Delta_0^2} \\
&\quad = -\frac{(\beta-\beta_0)^2}{\Delta^2\delta_0^2} + \frac{(1+\beta\beta_0+\xi\xi_0)^2}{\Delta^2\Delta_0^2} \quad \text{(note the coefficient for } \bar{\mathbf{u}}_1^2\text{)} \\
&\quad = \frac{-\theta_1^2+1}{1+\theta_1^2+\theta_2^2};
\end{aligned}$$

the coefficient for $\|\mathbf{u}_2\|^2$ is

$$\begin{aligned}
&\frac{\xi_0^2(\beta_0-\beta)^2}{\delta_0^2\delta^2\Delta_0^2} + \frac{\left[(1+\beta\beta_0)\xi\xi_0 + \delta^2\delta_0^2\right]^2}{\delta^2\delta_0^2\Delta^2\Delta_0^2} - \frac{(\beta-\beta_0)^2}{\delta^2\Delta_0^2} - \frac{\left[(1+\beta\beta_0)\xi - \delta^2\xi_0\right]^2}{\delta^2\Delta^2\Delta_0^2} \\
&\quad = -\frac{\left[(1+\beta\beta_0)\xi_0 - \delta_0^2\xi\right]^2}{\delta_0^2\Delta_0^2\Delta^2} + 1 - \frac{(\beta-\beta_0)^2}{\delta^2\Delta_0^2} - \frac{\left[(1+\beta\beta_0)\xi - \delta^2\xi_0\right]^2}{\delta^2\Delta^2\Delta_0^2} \\
&\quad = -\frac{\left[(1+\beta\beta_0)\xi_0 - \delta_0^2\xi\right]^2}{\delta_0^2\Delta_0^2\Delta^2} + \frac{(1+\beta\beta_0+\xi\xi_0)^2}{\Delta^2\Delta_0^2} \quad \text{(note the coefficient for } \bar{\mathbf{u}}_2^2\text{)} \\
&\quad = \frac{-\theta_2^2+1}{1+\theta_1^2+\theta_2^2};
\end{aligned}$$

the coefficient for $\mathbf{u}_1'\mathbf{u}_3$ is

$$\begin{aligned}
&\frac{(1+\beta\beta_0)(\beta-\beta_0)}{\delta^2\delta_0\Delta_0} + \frac{\xi(\beta_0-\beta)\left[(1+\beta\beta_0)\xi - \delta^2\xi_0\right]}{\delta_0\delta^2\Delta^2\Delta_0} \\
&\quad = \frac{(\beta_0-\beta)(1+\beta\beta_0+\xi\xi_0)}{\delta_0\Delta^2\Delta_0} \quad \text{(note the coefficient for } \bar{\mathbf{u}}_1\bar{\mathbf{u}}_3\text{)} \\
&\quad = \frac{\theta_1}{1+\theta_1^2+\theta_2^2};
\end{aligned}$$

the coefficient for $\mathbf{u}_2'\mathbf{u}_3$ is

$$\frac{\xi_0(\beta-\beta_0)^2}{\delta^2\delta_0\Delta_0^2}+\frac{[(1+\beta\beta_0)\xi\xi_0+\delta^2\delta_0^2][(1+\beta\beta_0)\xi-\delta^2\xi_0]}{\delta^2\delta_0\Delta^2\Delta_0^2}$$

$$=\frac{[(1+\beta\beta_0)\xi_0-\delta_0^2\xi][1+\beta\beta_0+\xi\xi_0]}{\delta_0\Delta_0^2\Delta^2}\quad \text{(note the coefficient for } \bar{\mathbf{u}}_2\bar{\mathbf{u}}_3)$$

$$=\frac{\theta_2}{1+\theta_1^2+\theta_2^2};$$

the coefficient for $\mathbf{u}_1'\mathbf{u}_2$ is

$$\frac{(1+\beta\beta_0)\xi_0(\beta_0-\beta)}{\delta^2\delta_0^2\Delta_0}+\frac{\xi(\beta-\beta_0)[(1+\beta\beta_0)\xi\xi_0+\delta^2\delta_0^2]}{\delta_0^2\delta^2\Delta\Delta_0}$$

$$=\frac{(\beta-\beta_0)[(1+\beta\beta_0)\xi_0-\delta_0^2\xi]}{\delta_0^2\Delta^2\Delta_0}\quad \text{(note the coefficient for } \bar{\mathbf{u}}_1\bar{\mathbf{u}}_2)$$

$$=\frac{\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}.$$

Equation (32) is an immediate consequence of these calculations.

APPENDIX 2

A.2. Evaluation of (35). Note the fact that, for some constants a, b and c,

$$\max_\theta\,(a\theta^2+b\theta+c)\big/(1+\theta^2)=\left[a+c+\sqrt{(a-c)^2+b^2}\right]\big/2. \tag{37}$$

and

$$\min_\theta\,(a\theta^2+b\theta+c)\big/(1+\theta^2)=\left[a+c-\sqrt{(a-c)^2+b^2}\right]\big/2. \tag{38}$$

Letting $\theta=\theta_2/\sqrt{1+\theta_1^2}$, we have

$$h(\theta_1,\theta_2)=(a\theta^2+b\theta+c)\Big/\left[(1+\theta^2)(1+\theta_1^2)\right], \tag{39}$$

where a, b and c are (functions of θ_1) defined by

$$a=(1+\theta_1^2)(1-s_{\mathbf{v}_2}^2),$$

$$b=2\sqrt{1+\theta_1^2}\left[\sqrt{1-\|\mathbf{v}\|^2}\left(\sqrt{s_{\mathbf{v}_2}^2-s_{\mathbf{v}_1\mathbf{v}_2}^2/s_{\mathbf{v}_1}^2}\,u_2+\frac{s_{\mathbf{v}_1\mathbf{v}_2}}{s_{\mathbf{v}_1}}u_1\right)-\theta_1 s_{\mathbf{v}_1\mathbf{v}_2}\right],$$

$$c=2\theta_1\sqrt{1-\|\mathbf{v}\|^2}s_{\mathbf{v}_1}u_1+\|\mathbf{v}\|^2+\theta_1^2(1-s_{\mathbf{v}_1}^2).$$

Maximizing h over θ first, we obtain from (37) and (39) that

$$\max_{\theta_1,\theta_2} h(\theta_1,\theta_2) = \max_{\theta_1} \frac{a+c+\sqrt{(a-c)^2+b^2}}{2(1+\theta_1^2)}. \tag{40}$$

For any θ_1, it follows from (38) that

$$1-s_{\mathbf{v}_2}^2 = \lim_{\theta_2\to\infty} h(\theta_1,\theta_2) \geq \min_{\theta_2} h(\theta_1,\theta_2) = \frac{a+c-\sqrt{(a-c)^2+b^2}}{2(1+\theta_1^2)}.$$

By the definition of $w, w > 1-s_{\mathbf{v}_2}^2$. Therefore,

$$w > \frac{a+c-\sqrt{(a-c)^2+b^2}}{2(1+\theta_1^2)}. \tag{41}$$

Relations (40) and (41) imply that

$$\max_{\theta_1,\theta_2} h(\theta_1,\theta_2) > w \tag{42}$$

if and only if

$$\exists\theta_1, \quad \frac{c+a+\sqrt{(a-c)^2+b^2}}{2} > (1+\theta_1^2)w > \frac{c+a-\sqrt{(a-c)^2+b^2}}{2}. \tag{43}$$

let $w_0 = w-1+s_{\mathbf{v}_2}^2$. Inequality (43) is the same as (note the definition of a)

$$\exists\theta_1, \quad \frac{c-a+\sqrt{(a-c)^2+b^2}}{2} > (1+\theta_1^2)w_0 > \frac{c-a-\sqrt{(a-c)^2+b^2}}{2}. \tag{44}$$

Note that $\left[c-a\pm\sqrt{(a-c)^2+b^2}\right]/2$ are roots of the quadratic equation $q(x) = -x^2+(c-a)x+b^2/4 = 0$, and $q[(1+\theta_1^2)w_0] > 0$ if and only if $(1+\theta_1^2)w_0$ is in between the two roots of $q(x)=0$. Therefore, (44) is equivalent to

$$\exists\theta_1, \quad -(1+\theta_1^2)^2w_0^2+(c-a)(1+\theta_1^2)w_0+b^2/4 > 0, \tag{45}$$

that is,

$$\exists\theta_1, \quad -(1+\theta_1^2)w_0^2+(c-a)w_0+b^2/[4(1+\theta_1^2)] > 0. \tag{46}$$

It is clear that (46) amounts to

$$\max_{\theta_1}\{-(1+\theta_1^2)w_0^2+(c-a)w_0+b^2/[4(1+\theta_1^2)]\} > 0. \tag{47}$$

Note that $c-a$ and $b^2/(1+\theta_1^2)$ are quadratic functions of θ_1. Then (47) involves maximization of the quadratic function

$$-\left[w_0\left(w+s_{\mathbf{v}_1}^2-1\right)-s_{\mathbf{v}_1\mathbf{v}_2}^2\right]\theta_1^2$$
$$+\,2\sqrt{1-\|\mathbf{v}\|^2}\left[\left(w_0 s_{\mathbf{v}_1}-\frac{s_{\mathbf{v}_1\mathbf{v}_2}^2}{s_{\mathbf{v}_1}}\right)u_1+s_{\mathbf{v}_1\mathbf{v}_2}\sqrt{s_{\mathbf{v}_2}^2-\frac{s_{\mathbf{v}_1\mathbf{v}_2}^2}{s_{\mathbf{v}_1}^2}}\,u_2\right]\theta_1$$
$$-\,w_0\left(w-\|\mathbf{v}\|^2\right)+\left(1-\|\mathbf{v}\|^2\right)\left[\sqrt{s_{\mathbf{v}_2}^2-\frac{s_{\mathbf{v}_1\mathbf{v}_2}^2}{s_{\mathbf{v}_1}^2}}\,u_2+\frac{s_{\mathbf{v}_1\mathbf{v}_2}}{s_{\mathbf{v}_1}}u_1\right]^2.$$

It is easy to see that the maximum is of the form

$$l_1u_1^2+l_2u_2^2+l_{12}u_1u_2-w_0\left(w-\|\mathbf{v}\|^2\right),$$

for some constants l_1, l_2 and l_{12} (not depending on u_1 and u_2). Let

$$A=\left\{(u_1u_2)\colon l_1u_1^2+l_2u_2^2+l_{12}u_1u_2>\lambda\right\},$$

where $\lambda=w_0(w-\|\mathbf{v}\|^2)>0$. Then the evaluation of (35) amounts to finding

$$Pr[A]=\frac{\Gamma\left((N-1)/2\right)}{\pi\Gamma\left((N-3)/2\right)}\int_A\left(1-u_1^2-u_2^2\right)^{(N-5)/2}du_1\,du_2.$$

Let $u_1=\rho\cos\zeta$ and $u_2=\rho\sin\zeta$, $\rho>0, \text{υ}<\zeta<2\pi$. Then

$$Pr[A]=\frac{\Gamma\left((N-1)/2\right)}{\pi\Gamma\left((N-3)/2\right)}\int\left(1-\rho^2\right)^{(N-5)/2}\rho\,d\rho\,d\zeta,$$

and the integration is taken over the domain

$$\left\{(\rho,\zeta)\colon \rho^2\left(l_1\cos^2\zeta+l_2\sin^2\zeta+l_{12}\sin\zeta\cos\zeta\right)>\lambda\right\}.$$

Hence

$$Pr[A]=\int_{C(\zeta)}\frac{1}{2\pi}\left\{\left(1-g(\zeta)\right)^+\right\}^{(N-3)/2}d\zeta,$$

where

$$g(\zeta)=\frac{\lambda}{l_1\cos^2\zeta+l_2\sin^2\zeta+l_{12}\sin\zeta\cos\zeta},$$

and

$$C(\zeta)=\{\zeta\colon l_1\cos^2\zeta+l_2\sin^2\zeta+l_{12}\sin\zeta\cos\zeta>0\}.$$

Note that $\sin^2\zeta=1-\cos^2\zeta$, and hence the equation $l_1\cos^2\zeta+l_2\sin^2\zeta+l_{12}\sin\zeta\cos\zeta=0$ has four roots (they are not necessarily distinct or real, although those real, distinct roots are relevant here), which can be found by solving a quartic equation; hence $C(\zeta)$ consists of at most three intervals in $(0,2\pi)$. It follows that $P[A]$ can be easily computed by one-dimensional numerical integration.

Acknowledgments. The author wishes to thank Professor David O. Siegmund for his supervision of this research and to thank Professors T. W. Anderson, Richard Olshen and Jianqing Fan for their valuable suggestions in connection with preparing the paper. The referee's illuminating comments were extremely helpful. The author is also grateful to the Editors for their help and support.

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 connections with simultaneous equations in econometrics. *J. Roy. Statist. Soc. Ser. B* **38** 1–36.

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

ANDERSON, T. W. and SAWA, T. (1982). Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient. *J. Roy. Statist. Soc. Ser. B* **44** 52–62.

BROWN, R. L. (1957). Bivariate structural relation. *Biometrika* **44** 84–96.

COOK, R. D. and WITMER, J. A. (1985). A note on parameter-effects curvature. *J. Amer. Statist. Assoc.* **80** 872–878.

CREASY, M. A. (1956). Confidence limits 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.

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

GLESER, L. J. (1981). Estimation in a multivariate "errors in variables" regression model: Large sample results. *Ann. Statist.* **9** 24–44.

GLESER, L. J. (1987). Confidence intervals for the slope in a linear errors-in-variables regression model. In *Advances in Multivariate Statistical Analysis* (A. K. Gupta, ed.) 85–109. Reidel, Dordrecht.

GLESER, L. J. and HWANG, J. T. (1987). The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models. *Ann. Statist.* **15** 1351–1362.

JOHANSEN, S. (1984). *Functional Relations, Random Coefficients, and Nonlinear Regression with Application to Kinetic Data*. Springer, New York.

JOHNSTONE, I. and SIEGMUND, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality. *Ann. Statist.* **17** 184–194.

JOLICOEUR, P. (1973). Imaginary confidence limits of the slope of the major axis of a bivariate normal distribution: A sampling experiment. *J. Amer. Statist. Assoc.* **68** 866–871.

KENDALL, M. G. and STUART, A. (1979). *The Advanced Theory of Statistics* **2**, 4th ed. Hafner, New York.

KNOWLES, M., SIEGMUND, D. and ZHANG, H. P. (1991). Confidence regions in semilinear regression. *Biometrika* **78** 15–31.

SCHNEEWEISS, H. (1982). Note on Creasy's confidence limits for the gradient in the linear functional relationship. *J. Multivariate Anal.* **12** 155–158.

SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

WILLIAMS, E. J. (1955). Significance tests for discriminant functions and linear functional relationships. *Biometrika* **42** 360–381.

ZHANG, H. (1991). Confidence regions in nonlinear regression and geometry. Ph. D. dissertation, Stanford Univ.

DEPARTMENT OF EPIDEMIOLOGY
AND PUBLIC HEALTH
AND DEPARTMENT OF STATISTICS
YALE UNIVERSITY
P.O. BOX 3333
NEW HAVEN, CONNECTICUT 06510