The Annals of Statistics
1988, Vol. 16, No. 1, 174–187

CONFIDENCE REGIONS IN MULTIVARIATE CALIBRATION

BY SAMUEL D. OMAN

Hebrew University

The multivariate calibration problem is considered, in which a sample of n observations on vectors $\xi_{(i)}$ (of "true values") and $Y_{(i)}$ (of less accurate but more easily obtained values) are to be used to estimate the unknown ξ corresponding to a future Y. It is assumed that $Y = BX + \varepsilon$, where ε is multivariate normal and $X = h(\xi)$ for known h. Current methods for obtaining a confidence region C for ξ, which consist of computing a region R for X and then taking $C = h^{-1}(R)$, have the disadvantage that although the region R might be nicely behaved, the region C need not be. An alternative method is proposed which gives a well-behaved region (corresponding to the uniformly most accurate translation-invariant region when h is linear, B is known and the covariance matrix of ε is a known multiple of the identity). An application is given to the estimation of gestational age using ultrasound fetal bone measurements.

1. Introduction. An example of the calibration problem in its simplest setting is as follows: Observations $(\xi_1, y_1), \ldots, (\xi_n, y_n)$ are given, in which the ξ_i are highly accurate (yet expensive) measurements of a quantity, while the y_i are less accurate (yet also less expensive) measurements. The objective is to compute a calibration curve, to be used with a future y measurement (say y_0) in order to estimate the corresponding ξ_0.

If the (ξ_i, y_i) are a random sample from a bivariate distribution and if (ξ_0, y_0) is also to be drawn from the same distribution, then we may simply compute the regression of the ξ_i on the y_i. If, however, these assumptions are not met (for example, the ξ_i might be laboratory values chosen to evenly cover the range of ξ, without regard to the frequency distribution of future ξ's), then other methods are called for. In particular, if the data are modeled as

$$y_i = \beta_0 + \beta_1 \xi_i + \varepsilon_i, \qquad 1 \le i \le n, \tag{1.1}$$

for independent $N(0, \sigma^2)$ error terms ε_i, then a confidence region for ξ_0 may be obtained by inverting the distribution of $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 \xi_0$, where the $\hat{\beta}_i$ are the maximum likelihood estimates of the β_i [Eisenhart (1939) and Fieller (1954)]. The region may be peculiarly shaped (e.g., two semiinfinite intervals), but is a finite interval when the regression is significant at the corresponding level.

A useful generalization of (1.1) is the model

$$Y_{(i)} = g(\xi_{(i)}) + \varepsilon_{(i)}, \qquad 0 \le i \le n, \tag{1.2a}$$

where now the $Y_{(i)}$ are q-vectors, the $\xi_{(i)}$ are r-vectors $(r \le q)$ and the $\varepsilon_{(i)}$ are

Received April 1985; revised May 1987.
AMS 1980 *subject classifications.* Primary 62F25; secondary 62H99.
Key words and phrases. Calibration, confidence regions.

independent $N(0, \Gamma)$ error terms. g is of the form

$$g(\xi) = B_{q\times p}X, \tag{1.2b}$$

where

$$X = X(\xi) = \left[h_1(\xi), \ldots, h_p(\xi)\right]^t \tag{1.2c}$$

for known functions h_j (e.g., squares or logarithms of components of ξ) which enable (1.2) to be expressed as a standard multivariate linear regression model in the unknown parameters B and Γ.

Brown (1982) discusses this model and presents the following procedure for obtaining a $1 - \alpha$ level confidence region C for $\xi_{(0)}$, the value of ξ corresponding to a new observation $Y_{(0)}$. Use $Y_{(0)}$ to determine a $1 - \alpha$ level confidence region R for $X_{(0)} = X(\xi_{(0)})$ (the specific form of R is given in the next section) and take

$$C = \{\xi\colon X(\xi) \in R\}. \tag{1.3}$$

Brown (1982) provides conditions (analogous to those in the univariate case) guaranteeing that R is a bounded ellipsoid, but points out that even when these conditions are satisfied, the region C for $\xi_{(0)}$ may be peculiarly behaved.

In this paper, an alternative confidence region D for $\xi_{(0)}$ is developed. D is conservative [although it is exact if $n = \infty$, i.e., when B and Γ in (1.2) are known, as sometimes occurs in chemical applications, for example, Spjøtvoll, Martens and Volden (1982)] while C given by (1.3) is exact. On the other hand, our procedure results in regions which tend to be better behaved and more easily interpretable than those obtained from (1.3). Also, if $n = \infty$, $\Gamma = I$ and the functions h_j in (1.2c) are linear, then D is the uniformly most accurate translation-invariant region for ξ, while C is not (unless $q = p = r$).

In the next section, (1.3) is explicitly defined and discussed, and the proposed method is motivated. The procedure is explicitly defined in Section 3 and illustrated in Section 4. Section 5 compares the proposed method with asymptotic results obtained by Fujikoshi and Nishii (1984) and by Brown and Sundberg (1987).

In the problem of Section 4 (which motivated the present paper), the objective was to determine whether combining ultrasound measurements of two fetal bone lengths (as opposed to using one measurement only) would extend the period during which ultrasound techniques provide sufficiently precise estimates of gestational age. This question could not be satisfactorily answered using (1.3), because of the peculiar behavior of the confidence regions, whereas it can be answered using the proposed procedure.

Although Brown (1982) and Hunter and Lamboy (1981) contain numerous references to related work in calibration, it appears that Brown's method (1982) is the only one proposed to date for obtaining exact confidence regions using the model (1.2) [in particular, when the functions h_j in (1.2c) are nonlinear]. Brown also discusses Bayesian approaches to the problem, as do Hunter and Lamboy in the univariate case.

2. Discussion of C and motivation for D. Let $\hat{B}$ denote the least-squares estimate of B from the multivariate regression of the $Y_{(i)}$ on the $X_{(i)} = X(\xi_{(i)})$, $1 \le i \le n$, denote $\hat{y}(\xi) = \hat{B}X(\xi)$ and define the SSE matrix $S = \sum_{i=1}^{n}[Y_{(i)} - \hat{B}X_{(i)}][Y_{(i)} - BX_{(i)}]^t$ and the "X^tX" matrix $M = \sum_{i=1}^{n}X_{(i)}X_{(i)}^t$. To simplify notation we shall occasionally omit the index (0) [e.g., from $\xi_{(0)}$] and suppress the dependence of quantities on ξ.

Since the covariance between the ith and jth columns of $\hat{B}$ is $(M^{-1})_{ij}\Gamma$, one has

$$\Gamma^{-t/2}\hat{B}v \sim N(\Gamma^{-t/2}Bv, v^tM^{-1}vI) \tag{2.1}$$

for any constant vector $v \in \mathbb{R}^p$ and square root $\Gamma^{1/2}$ of Γ ($\Gamma^{t/2}\Gamma^{1/2} = \Gamma$). Also, S has the Wishart distribution

$$S \sim W(n - p, \Gamma),$$

independently of $\hat{B}$. Using this, Brown (1982) proves that for $Y = Y_{(0)}$ independent of $\{Y_{(i)}: 1 \le i \le n\}$,

$$\frac{1}{q}V \equiv \left(\frac{m}{q}\right)\frac{1}{1 + a(X)}(Y - \hat{B}X)^tS^{-1}(Y - \hat{B}X) \sim F(q, m), \tag{2.2}$$

where

$$m = n - p - q + 1$$

and $a(X) = X^tM^{-1}X$. A $1 - \alpha$ level confidence region R for X is then the set of all X such that

$$V \le qF_\alpha(q, m), \tag{2.3}$$

where F_α denotes the upper α fractile. Consequently, (1.3) defines a $1 - \alpha$ level confidence region C for ξ. Also, a point estimate ξ^* may be obtained by minimizing V as a function of ξ.

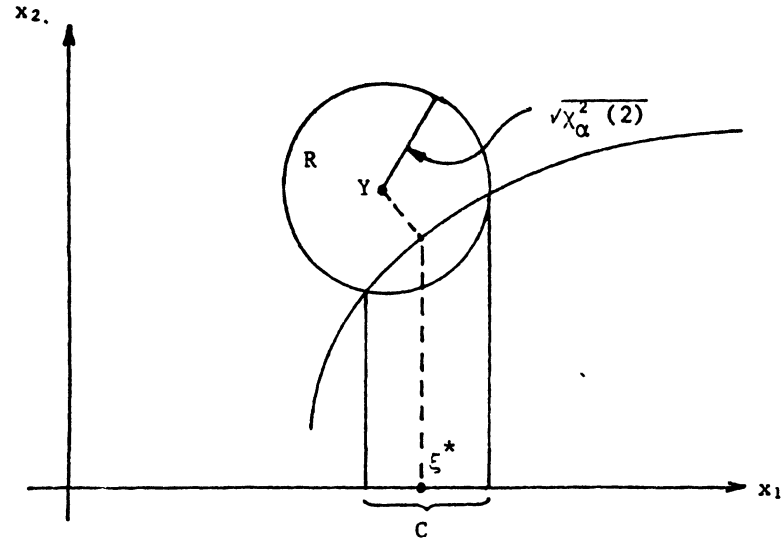
Brown (1982) provides conditions sufficient to guarantee that R is an ellipse, but also observes that the corresponding region for ξ need not be well behaved. To see this, consider the case $q = p = 2$, $r = 1$, $n = \infty$, $B = \Gamma = I$ (known) and $X(\xi) = (\xi, \phi(\xi))^t$ for a strictly increasing, differentiable function ϕ. Let $\chi_\alpha^2(m)$ denote the upper α fractile of the $\chi^2(m)$ distribution. Then R, C and ξ^* are as shown in Figure 1(a). This procedure has several undesirable properties:

1. C may be empty, even though ξ^* is defined. Although empty confidence regions are theoretically permissible, they are unpleasant to explain.

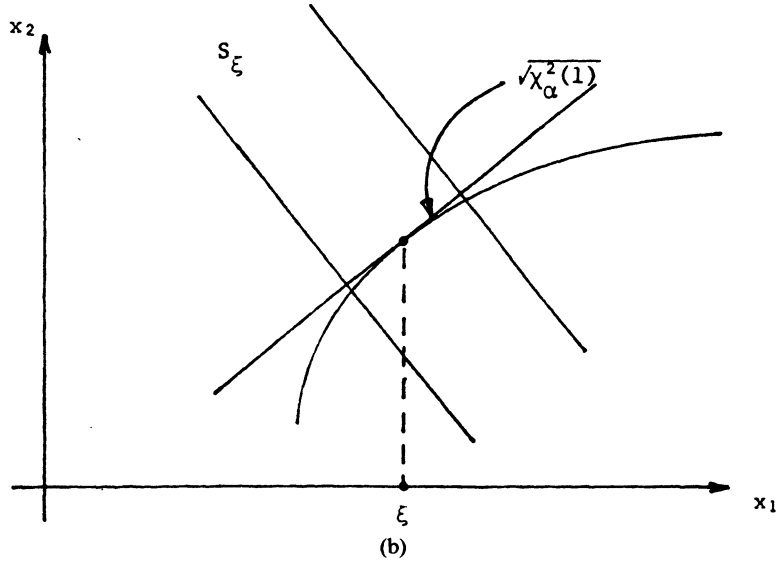
2. Suppose for the moment that ϕ is linear. After a transformation, the problem is equivalent to obtaining a confidence region for ξ in the assumed model

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \xi \\ 0 \end{pmatrix}, I\right).$$

In this case, Y_2, the distance from Y to the horizontal axis, should be used to test the validity of the assumed model, large distances leading to rejection. However, once the model is accepted, Y_2 has no further role to play, and a confidence interval for ξ should be of constant width centered at Y_1.



(a)



(b)

FIG. 1.

For general ϕ, this suggests that atypical Y should lead to questioning the assumed model, although not necessarily to larger confidence regions once the validity of the model is accepted. With C, however, as Y moves away from the curve, the confidence interval becomes increasingly *shorter*, thus giving the false impression that accurate inference about ξ is being made when, in fact, the entire model is suspect.

3. Since C is short with low probability and long with high probability, it should be possible to derive an alternative procedure with lower expected length.

4. Suppose $\phi(\xi) = b\xi$. The optimal (uniformly most accurate translation-invariant) procedure then gives the confidence interval

$$\left\{\xi: \frac{\left[Y_1 + bY_2 - (1 + b^2)\xi\right]^2}{1 + b^2} \leq \chi^2_{1-\alpha}(1)\right\} \tag{2.4}$$

of constant length

$$2\left[\chi^2_{1-\alpha}(1)/(1 + b^2)\right]^{1/2}. \tag{2.5}$$

On the other hand, C gives an interval with varying length whose maximum value is $2[\chi^2_{1-\alpha}(2)/(1 + b^2)]^{1/2}$ and whose expectation is necessarily greater than (2.5).

The idea of our proposed procedure is indicated in Figure 1(b). For given ξ, construct the tangent to the curve at $(\xi, \phi(\xi))$ and the corresponding strip S_ξ such that $P_\xi(Y \in S_\xi) = 1 - \alpha$. Then for given Y, $D = \{\xi: Y \in S_\xi\}$ is a $1 - \alpha$ level confidence region for ξ. Explicitly,

$$D = \left\{\xi: \frac{\left[Y_1 - \xi + \phi'(\xi)(Y_2 - \phi(\xi))\right]^2}{1 + \phi'(\xi)^2} \leq \chi^2_\alpha(1)\right\}. \tag{2.6}$$

Comparing D with C, note the following:

1. D is never empty, as it always contains the (possibly multivalued) estimate $\hat{\xi}$ which minimizes the distance from Y to the curve.
2. If the curvature of $(\cdot, \phi(\cdot))$ is not too great relative to the width of the strips S_ξ, D should be an interval whose length either increases or decreases slowly as Y moves away from the curve. It is difficult to give precise results in general, but an illustration is provided in Section 4.
3. D's maximum length is roughly proportional to $(\chi^2_\alpha(1))^{1/2}$, as opposed to $(\chi^2_\alpha(2))^{1/2}$ for C. For arbitrary $q \geq r$, if D and C are approximately spherical, then the ratio of their maximum volumes should be of the order $[\chi^2_\alpha(r)/\chi^2_\alpha(q)]^{r/2}$, which is quite small if $q \gg r$.
4. If $\phi(\xi) = b\xi$, then (2.6) reduces to the optimal (2.4).

3. The region *D*. To define D in the general case, let Ω denote the set of all ξ for which (1.2) is assumed, and assume the h_i to be differentiable for $\xi \in \Omega$. In the $r = 1$ dimensional case, it is reasonable to assume that there is no $\xi \in \Omega$ which is simultaneously a critical point for all the functions g_i. Analogously, for general r we require that there be no $\xi \in \Omega$ and nonzero vector $c \in \mathbb{R}^r$ such that

the directional derivative of each $\hat{g}_i$ along c at ξ is zero. Explicitly, if

$$\hat{G}_{q\times r} = \hat{G}(\xi) = \left(\frac{\partial \hat{g}_i}{\partial \xi_j}(\xi)\right), \tag{3.1}$$

then for each $\xi \in \Omega$, $\hat{G}$ is assumed to be almost surely of full rank r.

As in (2.6), define the statistic

$$U = (n - p - r + 1)[J(Y - \hat{g})]^t (JSJ)^- [J(Y - \hat{g})], \tag{3.2}$$

where J is the orthogonal projection matrix onto the column space of $\hat{G}$ and where $(\ )^-$ denotes a generalized inverse; note that both $\hat{g}$ and J depend on ξ and $\hat{B}$.

D is to be of the form $D = \{\xi\colon U \le \text{const.}\}$. Observe that D so defined is the set of all ξ such that the projection of Y onto the tangent plane to $\hat{g}(\cdot)$ [at $\hat{g}(\xi)$] is within a specified distance (measured in an appropriate norm) of $\hat{g}(\xi)$. In other words, analogous to Figure 1(b), $\xi \in D$ exactly when Y is in an ellipsoidally cross-sectioned cylinder which is perpendicular to the tangent plane to $\hat{g}$ at $\hat{g}(\xi)$ and which is "centered" at $\hat{g}(\xi)$.

The constant used to define D cannot be a percentile of U's distribution, since as a result of the dependence of J on $\hat{B}$ this distribution depends on B and Γ. However, we may proceed as follows. Let χ^2_m denote a central $\chi^2(m)$ variable and $\chi^2_m(\lambda)$ denote a noncentral $\chi^2_m(m, \lambda)$ variable (with mean $m + \lambda$). Define

$$G_{m,n} = \frac{\chi^2_m}{\chi^2_n/n}$$

and

$$G_{m,n}(\lambda) = \frac{\chi^2_m(\lambda)}{\chi^2_n/n}, \tag{3.3}$$

where the variables in the ratios are independent. Then for $a \ge 0$, define

$$K_a(x) = K(m, n, l, a, x) = \int_0^\infty P[G_{m,n}(\lambda a) \le x]\, P(\chi^2_l \in d\lambda). \tag{3.4}$$

THEOREM. *For any x,*

$$P[G_{r,(n-p-r+1)} \le x] \ge P(U \le x) \ge K(r, n - p - r + 1, q, a(\xi), x),$$

where

$$a(\xi) = X(\xi)^t M^{-1} X(\xi). \tag{3.5}$$

Thus, if x_α is such that

$$K_{a(\xi)}(x_\alpha) = 1 - \alpha, \tag{3.6}$$

then

$$D = \{\xi\colon U \le x_\alpha) \tag{3.7}$$

is a confidence region for ξ of level at least $1 - \alpha$.

The proof uses the following result, which is also useful in computations.

LEMMA. *U may be written as*

$$U = (n - p - r + 1)(Y - \hat{B}X)^t Q_{(1)} A^{-1} Q_{(1)}^t (Y - \hat{B}X), \tag{3.8}$$

where the columns of the $q \times r$ matrix $Q_{(1)}$ are an orthogonal basis for the column space of $\hat{G}$ and where $A = Q_{(1)}^t S Q_{(1)}$.

PROOF. Use $Q_{(1)}$ to form an orthogonal matrix $Q = (Q_{(1)}|Q_{(2)})$, so that

$$J = QLQ^t$$

for

$$L = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

Writing $JSJ = Q \cdot LQ^tSQL \cdot Q^t$ shows that

$$(JSJ)^- = Q \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q^t$$

for A as defined following (3.8). Substituting in (3.2) then gives

$$U = (n - p - r + 1) Z^t A^{-1} Z, \tag{3.9}$$

where

$$Z = Q_{(1)}^t (Y - \hat{g}),$$

proving the lemma. □

PROOF OF THE THEOREM. Condition on $\hat{B}$ in (3.9) (recall that $Q_{(1)}$ does not depend on S or Y). Z and A are independent given $\hat{B}$, with

$$Z|\hat{B} \sim N\big(Q_{(1)}^t(g - \hat{g}), Q_{(1)}^t \Gamma Q_{(1)}\big)$$

and

$$A|\hat{B} \sim W\big(n - p, Q_{(1)}^t \Gamma Q_{(1)}\big).$$

It follows that $U|\hat{B}$ is a multiple of Hotelling's T^2. Explicitly [see Muirhead (1982), page 211],

$$U|\hat{B} =_D G_{r,(n-p-r+1)}(\Lambda),$$

where $G_{m,n}(\lambda)$ is defined in (3.3) and

$$\Lambda = \big[Q_{(1)}^t(g - \hat{g})\big]^t \big(Q_{(1)}^t \Gamma Q_{(1)}\big)^{-1} \big[Q_{(1)}^t(g - \hat{g})\big].$$

Thus,

$$P(U \le x) = \int_0^\infty P\big[G_{r,(n-p-r+1)}(\lambda) \le x\big] P(\Lambda \in d\lambda). \tag{3.10}$$

Now write

$$\Lambda = \big[\Gamma^{-t/2}(\hat{g} - g)\big]^t N (N^tN)^{-1} N^t \big[\Gamma^{-t/2}(\hat{g} - g)\big],$$

where $\Gamma^{t/2}\Gamma^{1/2} = \Gamma$ and $N = \Gamma^{1/2} Q_{(1)}$. Because of the dependence of $\hat{g}$ and N on

$\hat{B}$, the distribution of Λ is intractable. However, we may clearly write

$$0 \le \Lambda \le \|\Gamma^{-t/2}(\hat{g} - g)\|^2 \equiv W.$$

Since the integrand in (3.10) is decreasing in λ, it follows that

$$\begin{aligned} P\big[G_{r,(n-p-r+1)} \le x\big] &\ge P(U \le x) \\ &\ge \int_0^\infty P\big[G_{r,(n-p-r+1)}(\lambda) \le x\big] P(W \in d\lambda). \end{aligned} \tag{3.11}$$

From (2.1) one has

$$W = \|\Gamma^{-t/2}(\hat{B} - B)X(\xi)\|^2 \sim a\chi^2(q)$$

for

$$a = [X(\xi)]^t M^{-1} X(\xi);$$

substituting into (3.11) then proves the theorem. □

A convenient computational formula is

$$\begin{aligned} K(m, n, q, a, x) = &\frac{1}{(a+1)^{q/2}\Gamma(q/2)} \\ &\times \sum_{j=0}^{\infty} \frac{P(G_{m+2j,n} \le x)\Gamma(q/2 + j)}{j!} \left(\frac{a}{a+1}\right)^j, \end{aligned} \tag{3.12}$$

which is obtained using the mixture representation

$$P\big[G_{m,n}(a\lambda) \le x\big] = e^{-a\lambda/2} \sum_{j=0}^{\infty} (a\lambda/2)^j P\big[G_{m+2j,n} \le x\big]/j!$$

and integrating out λ in (3.4).

In practice, the function $K_a(x)$ must be evaluated by numerical methods. Since $a(\xi)$ in (3.5) is of order n^{-1}, the series in (3.12) converges rapidly for n sufficiently large. Also, it is easy to obtain analytic bounds on the error incurred by truncating the summation. Note from (3.4) that $K_a(x)$ is monotone decreasing in a, and thus $x_\alpha = x_\alpha(a)$ defined by (3.6) is monotone increasing in a.

To implement the procedure one may first compute the $x_\alpha(a)$ for a grid of values of a in $[0, a_0]$, where $a_0 = \sup_{\xi \in \Omega} a(\xi)$. For given Y, one may then determine D sequentially by computing $U(\xi)$ [using (3.8)] for various ξ's and checking whether $U(\xi) \le x_\alpha(a(\xi))$. This is essentially the way (2.3) is used to determine the region C in (1.3), except that in (2.3) the critical point does not depend on ξ. An alternative method, not requiring inversion of K_a to determine x_α, is to write $U \le x_\alpha = K_a^{-1}(1 - \alpha) \Leftrightarrow K_a(U) \le 1 - \alpha$. For each ξ, one may thus simply evaluate $k_0 = K_{a(\xi)}(U(\xi))$ and declare $\xi \in D$ if $k_0 \le 1 - \alpha$.

Two further remarks are in order.

1. For $n = \infty$ [i.e., when B and Γ are known; cf. Spjotvoll, Martens and Volden (1982)], D is exact. Indeed, if $(n - p - r + 1)^{-1}S$ is replaced by Γ and $\hat{B}$ by B in (3.8), then clearly $U =_D G_{q,(n-p-r+1)}$, and since $a = 0$, (3.4) shows that $K_0(x)$ is U's distribution function. For finite n, the theorem's upper bound on $P(U \le x)$ may be used to measure the degree of conservatism of D.

2. If in addition, the functions h_j in (1.2c) are linear and $q = r$, then $\hat{G}$ in (3.1) is invertible and $J = I$ in (3.2). $n^{-1}U$ then reduces to V in (2.2) and the theorem gives the same region as (2.3).

4. Examples. We first illustrate our method on a problem, described in Oman and Wax (1984, 1985), of estimating gestational age by ultrasound measurements of two fetal bone lengths: the femur length F and the biparietal diameter (BPD) (essentially the diameter of the skull along the axis between the ears). (F, BPD) measurements were obtained for $n = 1114$ women whose week of pregnancy W was accurately determined, with the object of forming a calibration curve to estimate the W_0 corresponding to a future pair (F_0, BPD_0) of measurements.

Since increasing (with W) biological variability limits the usefulness of estimates based on F_0 alone to about the 31st week of pregnancy, a central question asked by the researchers was whether the confidence interval for W_0 obtained by combining F_0 and BPD_0 would be short enough (less than 5.5 weeks long) to justify using ultrasound based estimates of gestational age past the 31st week. As F and BPD are highly correlated, it was not clear whether using both measurements would result in significantly shorter confidence intervals.

In an attempt to answer this question, Brown's (1982) method was used to compute confidence intervals for W_0 for various combinations of F_0 and BPD_0. The results were summarized in Oman and Wax [(1985), Figure 8], which presents the rounded lengths of these intervals. The difficulties in interpreting this figure were discussed in Section 4.7 of that article. In particular, the fact that longer intervals were obtained for typical (F, BPD) combinations necessitated the paradoxical "conclusion" that intervals based on F and BPD could be used only through the 26th week, as opposed to the 31st week for intervals based on F alone.

Figure 2 shows the widths of the confidence intervals obtained using the present method, for all (F, BPD) combinations in a 99% confidence ellipse. The intervals are better behaved than those of Oman and Wax, their length tending to increase as weeks of pregnancy increase. Figure 2 thus permits the conclusion that F and BPD can be used through approximately the 31st week, indicating that the use of F and BPD together does not extend the period of usefulness of ultrasound techniques over that obtained from F alone. Although disappointing, this conclusion is at least nonparadoxical and can be explained by the high correlation between F and BPD.

We now describe the implementation of the theorem in obtaining Figure 2. In terms of the model (1.2), $Y = (F, BPD)^t$, $\xi = W$ and $X = (1, \xi, \xi^2)^t$. The matrices at the beginning of Section 2 are

$$\hat{B} = \begin{bmatrix} -42.917 & 4.514 & -0.0402 \\ -39.187 & 5.292 & -0.0492 \end{bmatrix}$$

and

$$S = \begin{bmatrix} 8281.00 & 4{,}900.48 \\ & 14{,}484.23 \end{bmatrix}.$$

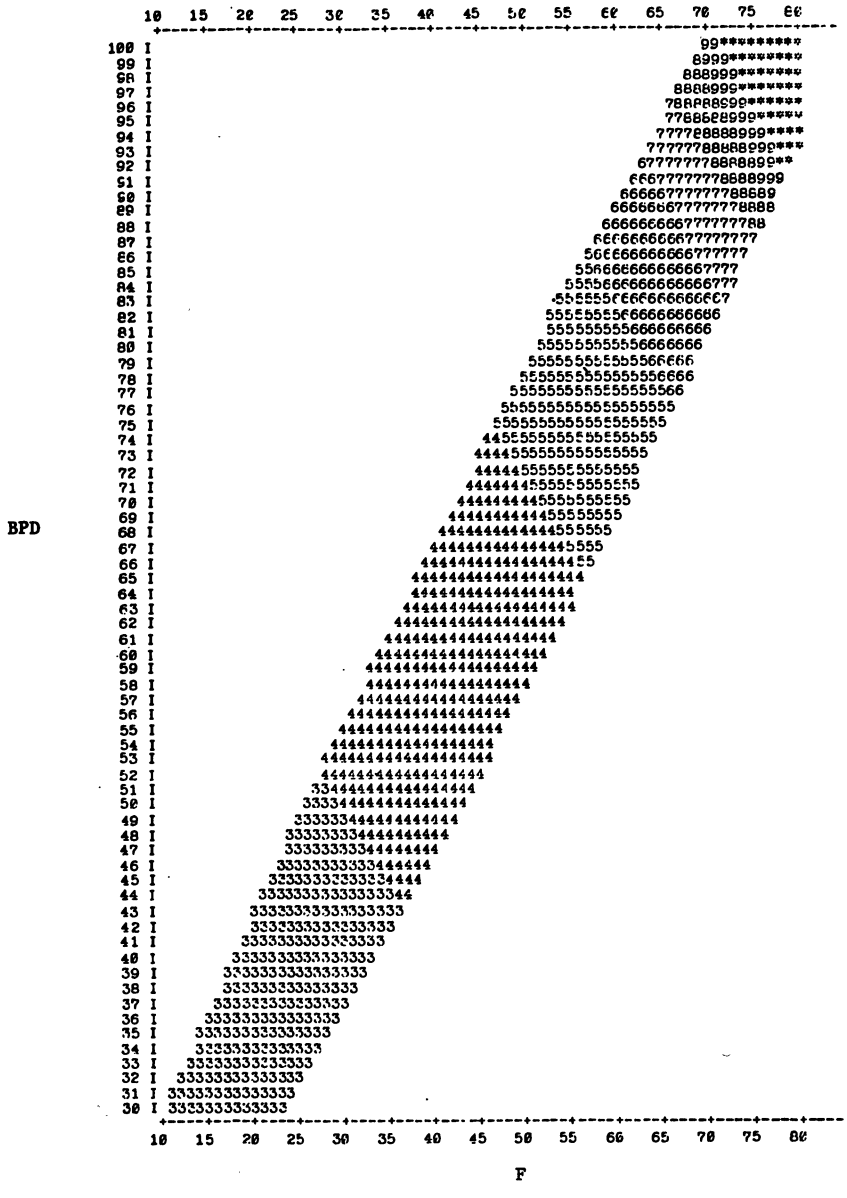


FIG. 2. *Confidence interval lengths (weeks). F is the femur length (mm) and BPD is the biparietal diameter (mm). Entry is the rounded length of the 95% confidence interval for a given (F, BPD) combination using the quadratic model.*

We first compute a_0 as described following (3.12). In (3.5),

$$a(\xi) = a(W) = \frac{1}{1114} + \begin{bmatrix} W - m_1 \\ W^2 - m_2 \end{bmatrix}^t \tilde{M}^{-1} \begin{bmatrix} W - m_1 \\ W^2 - m_2 \end{bmatrix},$$

where $\tilde{M}$ is the "X^tX" matrix [see Oman and Wax (1984)] for the centered observations and m_1, m_2 denote the sample means. We obtain $\max_{14 \le W \le 41} a(W) = a_0 = 0.0117$.

If we now determine x^* so that $K_{a_0}(x^*) = 0.95$ and use x^* instead of $x_{0.05}(a(W))$ in (3.7), then the monotonicity of $x_{0.05}(a)$ guarantees that D is (slightly, since $a_0 \approx 0$) conservative. To compute x^*, in (3.12) write

$$K(1, 1114, 2, a_0, x) = \frac{1}{a_0 + 1} \sum_{j=0}^{\infty} P(G_{1+2j,1114} \le x) \left(\frac{a_0}{a_0 + 1} \right)^j.$$

Truncating the sum after two terms gives an error of at most 1.337×10^{-4}, and approximating the G's by χ^2 variables gives $x^* \approx 3.93$.

Next, note that in (3.1)

$$\hat{G} = \hat{B} \begin{bmatrix} 0 \\ 1 \\ 2W \end{bmatrix} \in \mathbb{R}^2;$$

thus, in (3.8) A is a scalar and we may take $Q_{(1)} = \hat{G}$. The result is

$$U(W) = 1111 \frac{\left[\Sigma_{i=1}^2 (Y_i - \hat{\beta}_{i1} - \hat{\beta}_{i2} W - \hat{\beta}_{i3} W^2)(\hat{\beta}_{i2} + 2\hat{\beta}_{i3} W) \right]^2}{\Sigma_{i=1}^2 \Sigma_{j=1}^2 (\hat{\beta}_{i2} + 2\hat{\beta}_{i3} W) s_{ij} (\hat{\beta}_{j2} + 2\hat{\beta}_{j3} W)},$$

which is easily computed. Comparing $U(W)$ with x^* for a grid of W values then gives the desired interval.

We next examine the conservatism of our regions using Brown's (1982) paint finish data. In Brown's Section 5.3, pigmentation is estimated via a quadratic model with orthogonal polynomials, with 27 observations on P (pigmentation) and the response variables Y_1, Y_4. Thus $n = 27$, $p = 3$, $q = 2$ and $r = 1$.

Letting $\xi = P - 1$, we find in (3.5) that

$$a(\xi) = \left[1/27 + \xi^2/18 + (2 - 3\xi^2)^2/54 \right].$$

For a grid of values of ξ covering the range of the experimental data, Table 1 contains ξ, the critical point $x_\alpha(\xi)$ in (3.6) [obtained by truncating (3.12) at $j = 3$, which results in an error of at most 0.001 in computing K] and $P[G_{1,24} \le x_\alpha(\xi)]$, the upper bound on the coverage probability of D which results from the theorem. We remark that we also used our procedure to compute 95% confidence regions for ξ using the nine values of Y considered by Brown, and obtained intervals too wide to be of any use. This may be due to near zero (statistically speaking) components of $\hat{\beta}$. Specifically, the quadratic coefficient in the regression of Y_4 on ξ and ξ^2 is highly nonsignificant (p-value = 0.94), and its small value causes a "flatness" in the second component of the estimate $\hat{g}$ of (1.2b). As a result, for a wide range of ξ values U_ξ in (3.7) remains less than the critical point $x_\alpha(\xi)$ and our method results in overly wide intervals.

TABLE 1
Confidence region characteristics for the paint data

ξ	$x_{0.05}(\xi)$[a]	$P[G_{1,24} \le x_{0.05}(\xi)]$[b]
0.0	5.19	0.968
0.1	5.18	0.968
0.2	5.14	0.967
0.3	5.08	0.966
0.4	5.01	0.965
0.5	4.93	0.964
0.6	4.87	0.963
0.7	4.85	0.963
0.8	4.87	0.963
0.9	4.98	0.965
1.0	5.19	0.968

[a] $x_{0.05}(\xi)$ is an upper bound on the 95th fractile of U's distribution.
[b] $P[G_{1,24} \le x_{0.05}(\xi)]$ is an upper bound on the coverage probability of the nominal 95% interval.

5. Comparison with other methods. Suppose for the moment that $p = r = 1$, $X(\xi) = \xi$, and that $B = \beta$ and Γ are known. Defining $\xi^* = \beta^t\Gamma^{-1}Y/\beta^t\Gamma^{-1}\beta$, so that $Y^* = \xi^*\beta$ is the projection (in the norm determined by Γ^{-1}) of the new observation Y onto the span of β, we may write the log-likelihood (except for terms constant in ξ) as

$$(Y - \xi\beta)^t\Gamma^{-1}(Y - \xi\beta) = (Y - \xi^*\beta)^t\Gamma^{-1}(Y - \xi^*\beta) + (\xi - \xi^*)^2\beta^t\Gamma^{-1}\beta. \tag{5.1}$$

Thus ξ^*, the sufficient statistic for ξ, is independent of the ancillary statistic $(Y - \xi^*\beta)^t\Gamma^{-1}(Y - \xi^*\beta)$, which is the squared Γ^{-1} distance from Y to the calibration curve. In particular, the size of a confidence interval for ξ should not increase with this distance.

Now suppose that β and Γ are unknown, denote their estimates from the calibration experiment by $\hat{\beta}$ and S and define $\hat{\xi} = \hat{\beta}^tS^{-1}Y/\hat{\beta}^tS^{-1}\hat{\beta}$. Equation (5.1) suggests that

$$(Y - \xi\beta)^t\Gamma^{-1}(Y - \xi\beta) \approx (Y - \hat{\xi}\hat{\beta})S^{-1}(Y - \hat{\xi}\hat{\beta}) + (\xi - \hat{\xi})^2\hat{\beta}^tS^{-1}\hat{\beta}$$
$$= H + Q.$$

Thus, approximately, inference for ξ should be based on Q and be independent of H. To emphasize this point, consider the simpler problem in which n observations $Y_i \sim N(\mu_i, \sigma^2)$ are available, with σ^2 and μ_i unknown (but with replications). Let s^2 denote the usual estimate of σ^2. Suppose now that it is desired to estimate μ_0, based on $Y_0 \sim N(\mu_0, \sigma^2)$. The width of the confidence interval for μ_0 should depend on s^2 and not on the distance of Y_0 from the estimates $\hat{\mu}_i$; the proper role of this distance is in testing the consistency of Y_0 with the n observations Y_i.

A diagnostic statistic based on H has been proposed by Brown and Sundberg (1987), for the case of general p, q and r [but with the h_i in (1.2c) assumed linear]. Williams (1959) also considers testing consistency of $Y_{(0)}$ with the calibration data. Also, a confidence region whose size increases with H has been obtained by Brown and Sundberg (1987) using a profile likelihood approach. Their results are asymptotic and again assume linear h_i in (1.2c).

Confidence regions based on Q (again, assuming arbitrary dimensions) have been obtained by Fujikoshi and Nishii (1984); however, they require h_i in (1.2c) to be linear, and their results are asymptotic due to the intractable distribution of Q.

Now Q may be written as $(\hat{Y} - \xi\hat{\beta})'S^{-1}(\hat{Y} - \xi\hat{\beta})$, where $\hat{Y}$ is the projection onto $\hat{\beta}$ in the norm determined by S^{-1}; it is the use of this norm which complicates the distribution of Q. In our approach, we compute the Euclidean projection $\tilde{Y}$ and then normalize $\tilde{Y} - \xi\hat{\beta}$ by its approximate covariance matrix. Although the resulting score statistic is not invariant under arbitrary linear transformations, it has a distribution sufficiently tractable to make possible nonasymptotic results which can be generalized to arbitrary dimensions and, more importantly, to general functions h_i in (1.2c).

Finally, suppose in (1.2) that B is known and $\Gamma = \sigma^2 I$. Then estimating $\xi_{(0)}$, based on the observation $Y_{(0)}$, may be viewed as a problem in nonlinear regression, using the q observations in $Y_{(0)}$. This suggests trying to adapt results on exact confidence regions in nonlinear regression [e.g., see Hartley (1964)] to the present case. However, once B and Γ are assumed unknown, the distributions involved become intractable.

Acknowledgments. I would like to thank Larry Brown for numerous helpful suggestions during the course of this research, Alan Fenech, Ingram Olkin and Robert Shumway for several useful conversations and Israel Einot for programming assistance. I would also like to thank an Associate Editor and two referees for several very helpful comments.

REFERENCES

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

BROWN, P. J. and SUNDBERG, R. (1987). Confidence and conflict in multivariate calibration. *J. Roy. Statist. Soc. Ser. B* **49** 46–57.

EISENHART, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Statist.* **10** 162–186.

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

FUJIKOSHI, Y. and NISHII, R. (1984). On the distribution of a statistic in multivariate inverse regression analysis. *Hiroshima Math. J.* **14** 215–225.

HARTLEY, H. O. (1964). Exact confidence regions for the parameters in nonlinear regression laws. *Biometrika* **51** 347–353.

HUNTER, W. G. and LAMBOY, W. F. (1981). A Bayesian analysis of the linear calibration problem (with discussion). *Technometrics* **23** 323–350.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

OMAN, S. D. and WAX, Y. (1984). Estimating fetal age by ultrasound measurements: An example of multivariate calibration. *Biometrics* **40** 947–960.

OMAN, S. D. and WAX, Y. (1985). Estimating fetal age by ultrasound measurements: An example of multivariate calibration (correction). *Biometrics* **41** 821–822.

SPJØTVOLL, E., MARTENS, H. and VOLDEN, R. (1982). Restricted least squares estimation of the spectra and concentration of two unknown constituents available in mixtures. *Technometrics* **24** 173–180.

WILLIAMS, E. J. (1959). *Regression Analysis*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA AT SAN DIEGO
LA JOLLA, CALIFORNIA 92093