Confidence and Conflict in Multivariate Calibration

Philip J. Brown; Rolf Sundberg

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 49, No. 1. (1987), pp. 46-57.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281987%2949%3A1%3C46%3ACACIMC%3E2.0.CO%3B2-B

Journal of the Royal Statistical Society. Series B (Methodological) is currently published by Royal Statistical Society.

J. R. Statist. Soc. **B** (1987)
49, *No.* 1, *pp.* 46–57

Confidence and Conflict in Multivariate Calibration

By PHILIP J. BROWN† and ROLF SUNDBERG

The University of Liverpool, England — *University of Stockholm, Sweden*

[Received September 1985. Revised July 1986]

SUMMARY

Multivariate calibration involves using an estimated relationship between a multivariate response Y and an explanatory vector X to predict unknown X in future from further observed responses. In controlled calibration with multivariate normal errors, the profile likelihood function for the unknown X (denoted ξ) displays a term which measures the mutual inconsistency of the given response vector (denoted Z) in predicting ξ. This inconsistency diagnostic fundamentally differentiates the behaviour of likelihood based and Bayes "confidence" regions from those of the unconditional sampling approach. In addition the diagnostic serves to pinpoint an inadequate response vector Z.

Keywords: MULTIVARIATE CONTROLLED CALIBRATION; PROFILE LIKELIHOOD; LIKELIHOOD RATIO CONFIDENCE REGIONS; DIAGNOSTICS; CONFLICT

1. INTRODUCTION

Multivariate controlled calibration, as depicted in Brown (1982), takes training or calibrating data $Y_i(q \times 1)$ at fixed controlled values $X_i(p \times 1)$, $i = 1, \ldots, n$ to construct a relationship between Y and X. This is to be used in future to estimate ("predict") a new fixed but unknown X (denoted by the p-vector ξ) corresponding to the observed q-vector Z, this letter being used rather than Y to distinguish the prediction response from the Y of calibration.

The q response variables represent cheap or quick measurements on the item. There are p true characteristics of interest, X, which are expensive and laborious to measure. It is natural to insist that $q \geqslant p$, and often one ($p = 1$) characteristic only is of interest. Even when more than one ($p > 1$) characteristics are to be predicted, Brown (1982) suggests that it may yet be beneficial to treat the characteristics one at a time, forgetting the existence of the other $p - 1$ characteristics. Sundberg (1982, 1985) justifies this in some circumstances and investigates conditions for asymptotic mean square error improvement. We retain for the most part general p since even if the restriction to $p = 1$ characteristic were universally valid there would remain possible polynomial dependence on that characteristic, implying $p > 1$ components of X. In the polynomial case it is only necessary that the number of underlying regressor variables be at most q.

The assumed multivariate linear regression model for the n-sample calibrating data is

$$Y = 1\alpha' + XB + E, \tag{1.1}$$

where $Y(n \times q)$, $E(n \times q)$, are random matrices, $X(n \times p)$ is a fixed matrix of constants, 1 is an $n \times 1$ vector of ones and $\alpha(q \times 1)$, $B(p \times q)$ are unknown parameters. The error matrix E is such that with $E = (\varepsilon_1, \ldots, \varepsilon_n)'$,

$$E(\varepsilon_i) = 0,\ E(\varepsilon_i \varepsilon_i') = \Gamma,\ E(\varepsilon_i \varepsilon_j') = 0,\ i \neq j,\ i, j = 1, \ldots, n. \tag{1.2}$$

† *Address for correspondence*: Department of Statistics and Computational Mathematics, Victoria Building, The University, PO Box 147, Liverpool L69 3BX, UK.

 0035–9246/87/49046 $2.00

These are additionally multivariate normal. The model for $Z(q \times 1)$ is

$$Z = \alpha + B'\xi + \varepsilon \tag{1.3}$$

where $\varepsilon(q \times 1)$ satisfies the same assumptions as ε_i of (1.2), is normal and is independent of E.

We note that Brown (1982) considers the more general prediction problem in which $l > 1$ observations Z are made at the same ξ. This is a straightforward generalisation and for notational simplicity is avoided here. In this connection using $l > 1$ observations at a particular ξ is equivalent to a prediction error covariance matrix of Γ/l. Any common scaling of the covariance can be seen from general arguments or Brown (1982) to leave unaltered the standard controlled calibration estimator of ξ see later $\hat{\xi}$ as given by (3.1)). Thus concern that prediction errors may be on a different scale than those of calibration, as expressed by Rosenblatt and Spiegelman (1981) in the standard univariate case, whilst legitimate, is relevant only to interval estimates. In our multivariate formulation, however, there is no reason to suppose that prediction errors, if different from calibration errors, should be confined to common scale changes. From a less standard perspective emerging in this paper from the influence of an inconsistency diagnostic on the profile likelihood, prediction replicates may be specially useful.

In Section 2 we obtain the profile likelihood for ξ from data of model (1.1), (1.2) and (1.3). The form of the maximum likelihood estimator is derived in Section 3 and likelihood-based confidence regions are discussed in Section 4. Section 5 gives an illustrative example from Brown (1982).

2. THE PROFILE LIKELIHOOD

The unknown parameters of model (1.1)–(1.3) are α, B, Γ and ξ, where ξ is the parameter of interest and the others are nuisance parameters, whose removal is required for inference about ξ. One method of elimination is by means of a prior distribution on the parameters followed by integration out of the nuisance parameters in the posterior distribution. This approach was used in Brown (1982, Section 3). As an alternative here, avoiding the need for specification of a prior distribution, the profile or maximum relative likelihood (Kalbfleisch and Sprott, 1970), is determined. The method entails forming the maximised likelihood *as if ξ were known* and normalising by the likelihood maximised over all unknown parameters. The profile likelihood is thus a function of ξ which has a maximum value of one at a maximum likelihood estimate of ξ.

We will suppose for notational simplicity that X and Y of (1.1) have been centred, post hoc in the case of the observations Y, so that

$$\Sigma x_{ij} = 0, \quad j = 1, \ldots, p; \quad \Sigma Y_{ij} = 0, j = 1, \ldots, q. \tag{2.1}$$

It is assumed that the observations Z in (1.3) have *post hoc* been adjusted conformably.

When ξ is assumed known (1.1) and (1.3) have the same unknown parameters and they may be put together to form

$$Y_0 = 1\alpha_0' + X_0 B + E_0, \tag{2.2}$$

where Y_0, E_0 are $(n+1) \times q$ matrices, formed by augmenting Y, E with Z, ε, respectively. Also $X_0\{(n+1) \times p\}$ is the X matrix augmented by ξ and then centred by $-1\xi'/(n+1)$ so that

$$\alpha_0 = \alpha + B'\xi/(n+1). \tag{2.3}$$

Note that X_0 is a function of ξ, the parameter of interest. Now the maximised log-likelihood for model (2.2) under multivariate normality, in accordance with (1.2), is

$$\tfrac{1}{2}(n+1)[q\{1 + \ln(2\pi)\} - \ln\{|\hat{\Gamma}(\xi)|\}], \tag{2.4}$$

with

$$(n+1)\hat{\Gamma}(\xi) = (Y_0 - 1\hat{\alpha}_0' - X_0\hat{B}_0)'(Y_0 - 1\alpha_0' - X_0\hat{B}_0), \tag{2.5}$$

the $q \times q$ matrix of sum of products of error from a least squares fit, with the zero subscript on B serving as a reminder that this estimate depends on the temporarily assumed known ξ. Linear algebra enables (2.5) to be written as an explicit function of ξ. By introducing the idempotent $n \times n$ matrix given as

$$I - X_0(X_0'X_0)^{-1}X_0'$$

we may write the right hand side of (2.5) as

$$(Y_0 - 1Y_0')'[I - X_0(X_0'X_0)^{-1}X_0'](Y_0 - 1Y_0'). \tag{2.6}$$

Now from the facts,

$$(Y_0 - 1Y_0')'(Y_0 - 1Y_0') = Y'Y + nZZ'/(n+1),$$
$$(Y_0 - 1Y_0')'X_0 = Y'X + nZ\xi'/(n+1),$$
$$X_0'X_0 = X'X + n\xi\xi'/(n+1),$$

and the standard binomial inverse theorem, see for example Press (1972, p. 23), (2.6) becomes

$$\begin{aligned} &Y'Y + nZZ'/(n+1) \\ &\quad + \{Y'X + nZ\xi'/(n+1)\}G[n\xi\xi'/\{(n+)r(\xi)\} - G^{-1}]G\{Y'X + nZ\xi'/(n+1)\}', \end{aligned} \tag{2.7}$$

where

$$G = (X'X)^{-1}, \tag{2.8}$$

and

$$r(\xi) = 1 + n\xi'G\xi/(n+1). \tag{2.9}$$

We define the residual sum of products $S(q \times q)$ where

$$S = Y'Y - \hat{B}'X'Y \tag{2.10}$$

and

$$\hat{\Gamma} = S/n_0 \text{ with } n_0 = n - p - q, \tag{2.11}$$

and

$$\hat{B} = GX'Y, \tag{2.12}$$

the maximum likelihood estimator of B from the calibration training data solely. After some algebra (2.7) becomes

$$S + [n/\{(n+1)r(\xi)\}](Z - \hat{B}'\xi)(Z - \hat{B}'\xi)'\}.$$

This has been shown to be the right hand side of (2.5) and hence

$$|(n+1)\hat{\Gamma}(\xi)| = |S|\{1 + [n/\{(n+1)r(\xi)\}](Z - \hat{B}'\xi)'S^{-1}(Z - \hat{B}'\xi)\} \tag{2.13}$$

where we have used the determinantal identity,

$$|I + AB| = |I + BA|;$$

see for example Press (1972, p. 20).

Equation (2.13) together with (2.4) determine the profile log-likelihood up to an additive constant.

Remark 1. When there are $l \geqslant 1$ replicates in the prediction it may be similarly seen that the profile likelihood is proportional to

$$\left\{\frac{\sigma^2(\xi)}{\sigma^2(\xi)+(\bar{Z}-\hat{B}'\xi)'S_+^{-1}(\bar{Z}-\hat{B}'\xi)}\right\}^{(n+1)/2} \tag{2.14}$$

with $\sigma^2(\xi)=1/l+1/n+\xi'G\xi$ and the $q\times q$ matrix S_+ being formed by pooling the residual sum of products from both calibration and prediction experiments. This results in $n_0 = n-p-q+l-1$ degrees of freedom. The Bayes (integrated) likelihood corresponding to a Jeffrey's invariant vague prior distribution, (3.5) of Brown (1982), is almost of the form (2.14): the difference being that the powers of numerator and denominator of the Bayes posterior are $\frac{1}{2}n_0$ and $\frac{1}{2}(n_0+q)$ and not both $\frac{1}{2}(n+1)$ as in the profile likelihood. The additional $q/2$ power of the denominator in the integrated likelihood is crucial in determining whether the Bayes posterior is integrable with only a uniform 'vague' prior distribution on ξ: for $p=1$, $q\geqslant 2$ suffices.

Remark 2. (2.13) determines the profile likelihood also in the case of polynomial dependence of $E(Y)$ on X, that is when components of ξ are related.

Remark 3. If Γ were known the profile likelihood would be proportional to the exponential of $-\frac{1}{2}(n+1)tr\{\Gamma^{-1}\hat{\Gamma}(\xi)\}$, thus resulting also in a dependence on $\hat{\Gamma}(\xi)$. On the other hand B known and Γ unknown results in a standard regression likelihood from Z, augmented by information about Γ from the calibration. □

Formula (2.14) leads directly to the general form of the profile likelihood of ξ. The value of ξ for which it achieves its maximum is the maximum likelihood estimator. The generalised likelihood ratio of ξ or profile likelihood is (2.14) divided by the value of (2.14) at its maximum.

Values of ξ for which this ratio is greater than some constant are likelihood—based confidence regions of a size determined by the constant chosen.

3. MAXIMUM LIKELIHOOD ESTIMATION

3.1. *Maximum Likelihood Estimates: General Linear Case*

The classical estimator $\hat{\xi}$ of ξ is, given the centring of (2.1),

$$\hat{\xi}=H^{-1}\hat{B}\hat{\Gamma}^{-1}Z, \tag{3.1}$$

where

$$H=\hat{B}\hat{\Gamma}^{-1}\hat{B}'. \tag{3.2}$$

This form is suggested by the maximum likelihood estimator with B, Γ known and these replaced by the estimators B, Γ given by (2.12), (2.11) or equivalently (2.10). We shall see however that (3.1) is not generally the maximum likelihood estimator of ξ when $q>p$. The maximum likelihood estimator minimises (2.13) and this is proportional to

$$m+(R+|\xi-\hat{\xi}|^2_H)/r(\xi) \tag{3.3}$$

with

$$m=n_0(n+1)/n,$$

H given by (3.2), and with R, the prediction residual sum of products, given by

$$R=|Z-\hat{B}'\hat{\xi}|^2_{\hat{\Gamma}^{-1}}.$$

Here $|x|^2_A=x'Ax$, and in the sequel if A is omitted it is assumed to be the identity matrix. The statistic R may be interpreted as an inconsistency diagnostic.

Remark. In the case of polynomial dependence of EY on X, as in Remark 2 of the previous sections, the natural generalisation of R is

$$R = \operatorname*{Min}_{\xi} (Z - \hat{B}'\xi)'\hat{\Gamma}^{-1}(Z - \hat{B}'\xi),$$

the minimum squared distance from the prediction given by the calibration experiment, where minimisation is over the restricted possible variation of ξ. □

There exists a non-singular $p \times p$ matrix U which simultaneously diagonalises the symmetric matrices H and $nG/(n+1)$, that is

$$U'HU = I$$

$$nU'GU/(n+1) = D = \operatorname{diag}(g), \quad g_1 \geqslant \ldots \geqslant g_p \geqslant 0.$$

Now R, H are $O_p(1)$ whereas G in $r(\xi)$ given by (2.11) is $O_p(1/n)$ so that the g_j are $O_p(1/n)$. If

$$\mu = U^{-1}\xi, \quad \hat{\mu} = U^{-1}\hat{\xi}, \tag{3.4}$$

(3.3) becomes

$$m + f(\mu, R) \tag{3.5}$$

with

$$f(\mu, R) = (R + |\mu - \hat{\mu}|^2)/(1 + |\mu|_D^2). \tag{3.6}$$

Special case, $R = 0$.

In particular this is so for $q = p$, when the inconsistency diagnostic R is identically zero. When $R = 0$, (3.6) is a non-negative function of μ and is minimised by $\mu = \hat{\mu}$. Thus the maximum likelihood estimator and the classical estimator, (3.1), coincide.

Typical case, $R > 0$.

When $R > 0$ we first suppose without loss of generality that

$$g = g_p \leqslant g_j, \quad j = 1, \ldots, p-1.$$

It is easy to see that $\liminf_{|\mu| \to \infty} f(\mu, R) = 1/g$.

Equi-likelihood contours are given by $f(\mu, R) = c > 0$. It follows that, if the lower bound of attainable values of c is $< 1/g$, a finite maximum likelihood estimate exists. The equi-likelihood contours may be written,

$$R + |\mu - \hat{\mu}|^2 = c(1 + |\mu|_D^2)$$

or

$$|\mu - \hat{\hat{\mu}}|_A^2 = c + |\hat{\mu}|_{A^{-1}}^2 - R - |\hat{\mu}|^2, \tag{3.7}$$

with,

$$A = \operatorname{diag}(a),$$

a $p \times p$ diagonal matrix with jth diagonal elements

$$a_j = 1 - cg_j, \tag{3.8}$$

and the centre of the ellipsoid being

$$\hat{\hat{\mu}} = \hat{\hat{\mu}}(c) = A^{-1}\hat{\mu}. \tag{3.9}$$

The right hand side of (3.7) is a strictly increasing function of c. For $c = 1/g$, we know that

(3.7) has an unbounded solution. If c is reduced from $1/g$, the right hand side of (3.7) decreases monotonically with decreasing c. The smallest value, zero, of the left hand side at $\mu = \hat{\hat{\mu}}$ is independent of c and will therefore correspond to the smallest value of c as determined by the right hand side. It is achievable if $\hat{\mu}_p \neq 0$ since the right hand side ranges from ∞ down to $-R < 0$ as c ranges from $1/g$ to 0. Thus (3.7) has a minimising value c_L of c in the range $(0, 1/g)$ with $\mu = \hat{\hat{\mu}}$ finite.

The above argument required $\hat{\mu}_p \neq 0$, which is true with probability one. If the whole vector $\hat{\mu} = 0$ and $R < 1/g$, then $\hat{\hat{\mu}}$ is zero and $c_L = R$. If $\hat{\mu}_p = 0$ but $\hat{\mu} \neq 0$ a more complicated upper bound on R ensues but $R < 1/g$ is always sufficient.

In conclusion the maximum likelihood estimator of μ exists in all but some very special cases indicated in the above paragraph and is given as $\hat{\hat{\mu}}$ where

$$\hat{\hat{\mu}}_j = \hat{\mu}_j/(1 - c_L g_j).$$

Here, in general, c_L and hence $\hat{\hat{\mu}}$ cannot be expressed explicitly.

This form of maximum likelihood estimator does not revert to form (3.1) if Γ is known. As may be seen from Remark 3 at the end of section 2 the same form as above results with Γ replacing $\hat{\Gamma}$.

It may be noted that, since $c_L < 1/g$,

$$0 < 1 - c_L g_j < 1,$$

so that $\hat{\hat{\mu}}$ is an expansion of $\hat{\mu}$. When the calibration experiment is informative and accurate the g_j are small; for well behaved prediction R also will be small and $\hat{\hat{\mu}}$ and $\hat{\mu}$ will not differ by much. In this case $f(\mu, R)$ may be well approximated by its numerator as defined in (3.6). However, even with accurate calibration, a large value of R can shift the maximum likelihood estimator considerably from $\hat{\mu}$ and hence $\hat{\xi}$ of (3.1).

It should be observed that one is unlikely to want to use $\hat{\hat{\mu}}$ when R is significantly larger than would be expected given that Z comes from the same model as the calibration data. For tests of consistency, see Williams (1959, Ch. 9) and Naes (1983). For such a large R one might question the validity of the observation. Various strategies are then possible, including investigation of the individual error components and seeking further data.

One has to question whether in practical terms one would prefer the maximum likelihood estimator to that of the simpler $\hat{\xi}$ of (3.1). The answer must be in some doubt since the former offers an expansion of $\hat{\xi}$ rather than a shrinkage to zero favoured by any prior distribution such that future ξ are like past X, see Brown (1982, Section 3).

A further aspect of the maximum likelihood estimator for $q > p$ is that it needs strictly to be updated each time that a new ξ is to be estimated from a corresponding Z. While that might be technically sensible it does suggest considerable complications. A more useful type of updating might be from the accumulation of information about the marginal distribution of ξ, extending Williams (1969).

In the next subsection we examine in more detail the form of the profile likelihood in the important case $p = 1$.

3.2. *Maximum Likelihood Estimation when $p = 1$.*

The profile likelihood is the $-(n+1)/2$ power of

$$\{m + f(\mu, R)\}/\{m + f(\hat{\hat{\mu}}, R)\} \tag{3.10}$$

Thus the profile likelihood is a simple monotone transformation of $f(\mu, R)$.

When there is just one explanatory variable ($p = 1$) with q dependent variables,

$$f(\mu, R) = \{R + (\mu - \hat{\mu})^2\}/(1 + g\mu^2)$$

Furthermore note that, in this case, H given by (3.2) and G given by (2.8) are scalars and hence g in the expression for $f(\mu, R)$ above is $\{n/(n+1)\}G/H$.

For the rest of this section we exclude the case $\hat{\mu} = 0$, which is easily dealt with separately: In this special case f has the minimum at $\mu = 0$ and supremum $1/g$ if $R < 1/g$, maximum R at $\mu = 0$ and infimum $1/g$ if $R > 1/g$.

Consider $f(\mu, R) = c$. For each fixed c (3.7) is a quadratic in μ, with 0, 1 or 2 roots. The case of a single root corresponds to a maximum or a minimum of f and occurs when the constant on the right of (3.7) is zero, that is when

$$c = \{1 + g(\hat{\mu}^2 + R) \pm \sqrt{[\{1 + g(\hat{\mu}^2 + R)\}^2 - 4gR]}\}/(2g). \tag{3.11}$$

For the sequel we denote these two c-values by c_L and c_U, $c_L < c_U$. The corresponding values of μ are

$$\mu_0 = \hat{\mu}/(1 - c_U g), \quad \hat{\hat{\mu}} = \hat{\mu}/(1 - c_L g), \tag{3.12}$$

where from now on we let $\hat{\hat{\mu}}$ denote the particular value $\hat{\hat{\mu}}\ (c_L)$, the maximum likelihood estimator of μ. For reasons of symmetry we may assume $\hat{\mu} > 0$ without loss of generality. Then $\hat{\hat{\mu}}$ is to the right of $\hat{\mu}$ since $\hat{\hat{\mu}}$ is an expansion of $\hat{\mu}$. Furthermore, $f(\mu, R) \to 1/g$ as $\mu \to \pm\infty$, from below at $+\infty$, from above at $-\infty$. From (3.11), (3.12) it follows that $\hat{\hat{\mu}}\mu_0 = -1/g$ so that $\hat{\hat{\mu}}$ and μ_0 have opposite signs and consequently, from (3.12),

$$c_U > 1/g > c_L.$$

Since the right hand side of (3.7) is an increasing function of c and a decreasing function of R, it follows that the root c_L is monotonically increasing as a function of R. Hence $\hat{\hat{\mu}}$ is strictly increasing from $\hat{\mu}$ to ∞ as R goes from 0 to infinity. For large R,

$$c_L = (1/g)(1 - \hat{\mu}^2/R) + O(1/R^2)$$

and correspondingly

$$\hat{\hat{\mu}} = R/\hat{\mu} + O(1).$$

For small R,

$$c_L = R/(1 + g\hat{\mu}^2) + O(R^2)$$

and at the same time

$$\hat{\hat{\mu}} = \hat{\mu} + gR/(1 + g\hat{\mu}^2) + O(R^2).$$

Also from (3.11) it follows that $c_U c_L = R/g$, and for large R

$$c_U = R + \hat{\mu}^2 + O(1/R). \tag{3.13}$$

Figs 1a–1d give plots of $f(\mu, R)$ with $\hat{\mu} = 1$, $g = 0.1$, and $R = 0$, 4, 7, 20 respectively. The values of f at $\mu = \pm\infty$ is $1/g = 10$ and it may be noted that the minimum value of f increases from 0 for $R = 0$ when $\hat{\hat{\mu}} = \hat{\mu} = 1$ to 9.156 at $R = 20$ when $\hat{\hat{\mu}} = 11.844$. The maximum value of f, which corresponds to the least likely estimate, increases from 11.0 at $R = 0$ to 21.8 at $R = 20$, with at the same time μ_0 increasing from -10.0 to -0.8. The dotted horizontal lines give confidence regions described in the next section.

4. LIKELIHOOD-BASED CONFIDENCE REGIONS

Minus twice the log-likelihood ratio test statistic is, from (3.10),

$$W(\mu) = (n + 1)\{h(\mu, R) - h(\hat{\hat{\mu}}, R)\}$$

with

$$h(\mu; R) = \ln\{m + f(\mu, R)\}.$$

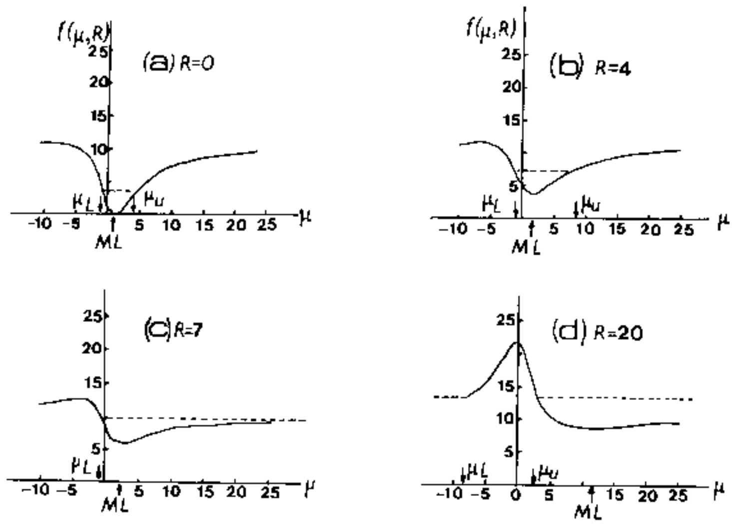

Fig. 1. Plots of $f(\mu, R)$, with maximum likelihood estimates (ML) and likelihood-based regions indicated, for $p = 1$, $1/g = 10$, $\hat{\mu} = 1$ and (a) $R = 0$; (*b*) $R = 4$; (*c*) $R = 7$; (d) $R = 20$.

A likelihood-based confidence region for μ is given by

$$\{\mu : W(\mu) \leqslant k\}, \quad k > 0, \tag{4.1}$$

or equivalently

$$\{\mu : f(\mu, R) \leqslant c^*(R)\}, \tag{4.2}$$

where

$$c^*(R) = e^{k/(n+1)}(m + f(\hat{\mu}, R)) - m = a + bc_L \tag{4.3}$$

with $a > 0$, $b > 1$.

The standard asymptotic result takes $k = k_{p,\gamma}$, the upper γ point of the chi-squared distribution on p degrees of freedom, to yield an approximate $(1-\gamma)$ 100 percent confidence level. However, usual asymptotic theory required for the limiting chi-squared distribution relies on increasing information about ξ through repetitions, whereas here only n in the training data may be considered large and predictive information about ξ remains imprecise even when n is large. Consequently a special argument is needed to justify the standard distributional assumption. For $p = q$ this is provided by letting $n \to \infty$ in the following Remark, and for $p = 1$, $q \geqslant 1$ a restricted justificaton by small Γ asymptotics is given at the end of the section.

Remark. When $p = q$ an exact likelihood ratio test readily obtains. In this case $\boldsymbol{R}$ is identically zero and (3.10) becomes $1 + f(\mu, 0)/m$, where $f(\mu, 0) = |\xi - \hat{\xi}|_H^2 / r(\xi)$, whose distribution after division by the scalar $p(n+1)/n$ is an F distribution on p and n_0 degrees of freedom, from (2.8) of Brown (1982). Harding (1986) pointed this out when $p = q = 1$. Interestingly, this gives a direct interpretation in terms of the profile likelihood of the pathological unbounded behaviour of classical confidence intervals, as described by Hoadley (1970). When $p = q = 1$ a structural solution has been given by Kalotay (1971).

Let us look at the confidence region (4.2) more closely when $p = 1$ and (say) $\hat{\mu} > 0$. If the explicit dependence of c^* on R is dropped, for a bounded interval it is necessary and sufficient that $c^* < 1/g$ and this will be aided by R small and n and γ large. The right hand side of (3.7) equated to zero, with c^* replacing c, has two roots in μ, namely $\mu_U > \mu_L$, the $\pm$values of

$$[\hat{\mu} \pm \sqrt{\{\hat{\mu}^2 - (1-gc^*)(R+\hat{\mu}^2 - c^*)\}}]/(1-gc^*). \tag{4.4}$$

To investigate the behaviour of μ_U, μ_L with increasing R, we first note that μ_U, μ_L are solutions of

$$h(\mu, R) - h(\hat{\mu}, R) = \text{const.}$$

Since the constant is independent of R, taking the derivative with respect to R gives

$$0 = \frac{\partial h(\mu, R)}{\partial \mu}\frac{d\mu}{dR} + \frac{h(\mu, R)}{\partial R} - \frac{h(\hat{\mu}, R)}{\partial \hat{\mu}}\frac{d\hat{\mu}}{dR} - \frac{\partial h(\hat{\mu}, R)}{\partial R},$$

and, since $\partial h(\hat{\mu}, R)/\partial \hat{\mu} = 0$ we get,

$$\frac{d\mu}{dR} = \left[\frac{\partial h(\hat{\mu}, R)}{\partial R} - \frac{\partial h(\mu, R)}{\partial R}\right] \Big/ \frac{\partial h(\mu, R)}{\partial \mu}$$

with $\mu = \mu_U$, μ_L. Now $\mu_U > \hat{\mu} > 0$ and from the form of h as depicted in the previous section, $\partial h/\partial \mu_U > 0$. Also

$$\frac{\partial h(\mu, R)}{\partial R} = \{h(\mu, R)(1 + g\mu^2)\}^{-1}.$$

For $\mu = \mu_U$ this is strictly less than $\partial h(\hat{\mu}, R)/\partial R$ since $\hat{\mu}$ minimises h and $1 + g\mu^2$ is an increasing function of μ. Thus $d\mu_U/dR > 0$ and the upper confidence limit is strictly increasing as a function of R.

Monotonicity with R does not apply to the lower limit however, and it is necessary to take a different approach to investigate the overall behaviour of the confidence region. Now from (4.3), $c^* = a + bc_L$ with $b > 1$ and $a > 0$. The length of the interval is from (4.4)

$$2\sqrt{\{\hat{\mu}^2 - (1-gc^*)(R + \hat{\mu}^2 - c^*)\}}/(1-gc^*). \tag{4.5}$$

As seen from (4.5) the length tends to ∞ like $2\hat{\mu}/(1-gc^*)$ when R increases in such a way that $c^* = c^*(R) \to 1/g$ from below. Thus the length increases rapidly with R for such large R. This appears to result from a rapid increase in μ_U with R and a rather slow movement of μ_L in the region of zero, where c_U increases monotonically with R. Now $c^* > 1/g$ implies that the confidence region will be unbounded, being typically doubly infinite with a centrally-excluded set of values, where by (3.13) c_U is of order R. For $R = 0, 4, 7, 20$ in Figs 1a–1d the horizontal dotted lines are at level c^* when $p = 1$, $q = 4$, $n = 50$, $\gamma = 0.05$ in addition to $\hat{\mu} = 1$, $g = 0.1$. For R larger than about 8 in this case $c^* > 1/g$ and the intervals are unbounded. The intervals are respectively: $(-0.98, 4.10)$, $(-1.02, 8.51)$, $(-0.85, 71.58)$, $(-\infty, -8.4) \cup (2.6, +\infty)$.

The dependence of this region on R is reassuring. It mimics the behaviour of the likelihood function as well as the Bayes approach, see Brown (1982), with non-informative prior distributions, although Bayes credibility intervals do not display the above discontinuous behaviour for $c^* > 1/g$. It corrects what might be interpreted as a fault in the joint sampling approach as reported in Theorem 1 of Brown (1982). The point is amplified in that author's reply to the discussion of Mr Aitchison and Professor Barnard. Perversely, in that case decreasing R expands the confidence region and increasing R shrinks it such that a very self-inconsistent observation can imply a point confidence interval for ξ! Oman and Wax (1984) have witnessed the undesirable practical effect of this. Wood (1982) has suggested basing confidence procedures on the distribution of

$$|\xi - \hat{\xi}|_H^2 \tag{4.6}$$

but this goes to the extreme of ignoring the effect of R completely. It has sampling properties developed by Fujikoshi and Nishii (1984). Oman (1985) has proposed an alternative confidence procedure, based on the best invariant region when B is known. A modification of (4.6) is obtained by exchanging $\hat{\hat{\xi}}$ for $\hat{\xi}$. This results in intervals more close to the likelihood-based regions (4.1), cf. (4.9) below. Finally, note that by Remark 3 of section 2, likelihood based confidence intervals will display similar behaviour even when Γ is known. Thus intuition that large R just suggests large Γ is faulty.

Since, as earlier mentioned, standard asymptotics do not apply to the likelihood region (4.1), we give here a brief discussion of small Γ asymptotics which may be used to justify the limiting chi-squared on p degrees of freedom in the important special case $p = 1, q \geqslant 1$. First we may note that as Γ tends to a zero matrix $\hat{\xi}$, $\hat{\hat{\xi}}$ tend in probability to ξ (regarded as fixed). A Taylor series expansion of $W(\mu)$ about the maximum likelihood estimator $\hat{\hat{\mu}}$, in terms of the ξ parameterisation and neglecting higher order terms, is

$$W(\mu) \approx \tfrac{1}{2}(n+1)f''(\hat{\hat{\xi}})(\xi - \hat{\hat{\xi}})^2/(m + c_L) \tag{4.7}$$

where

$$f(\xi) = \{R + h(\xi - \hat{\xi})^2\}/(1 + e\xi^2)$$

with $e = nG/(n+1) = O(1/n)$. Furthermore since $h = O_p(\Gamma^{-1})$ we obtain

$$f''(\hat{\hat{\xi}}) \approx 2h/(1 + e\hat{\xi}^2)$$

$$c_L \approx R/(1 + e\hat{\xi}^2)$$

and hence (4.7) further approximates to

$$W(\mu) \approx \{h/(1 + e\xi^2)\}[(n+1)/\{m + R/(1 + e\xi^2)\}](\xi - \hat{\hat{\xi}})^2 \tag{4.8}$$

Now $R = O_p(1)$ so that provided $m = n_0(n+1)/n$ is large then (4.8) becomes

$$W(\mu) \approx h(\xi - \hat{\hat{\xi}})^2. \tag{4.9}$$

Finally, $\hat{\hat{\xi}} - \hat{\xi} = O_p(\Gamma)$ and $\sqrt{h}(\hat{\xi} - \xi)$ is asymptotically $N(0, 1)$ distributed. It follows that the log-likelihood ratio test statistic has the limiting chi-squared distribution on one degree of freedom. Consequently the corresponding $(1-\gamma)100$ per cent confidence region is asymptotically given by (4.1) with $k = k_{1,\gamma}$. However, note the following restrictions.

(a) In order to conclude that the asymptotic chi-squared distribution is a good approximation for fixed values of n, m and Γ, m large and Γ small $|\xi|$ must not be too large. Hence, confidence regions derived from (exact) likelihood ratio tests and the likelihood-based regions (3.14) have been shown to agree well only when they result in bounded, reasonably narrow intervals.

(b) We have used (4.7) and subsequent approximations to find an approximate distribution of W, not an approximation of the outcome of W for atypical values of R (or of any other stochastic component of W).

Both these restrictions imply that we cannot conclude from (4.9) anything about the behavour as $R \to \infty$ of the confidence regions corresponding to likelihood ratio tests. This is of little consequence though, because a large R indicates inconsistency of Z and should lead to separate treatment as commented at the end of Section 3.1. Also, they clearly do provide intervals expanding with increasng R for moderate values of R, unlike those of (4.6).

One operationally useful alternative to likelihood-based confidence regions should be mentioned, namely, Bayes credibility intervals which, as in Brown (1982, Section 3), incorporate prior information about ξ. From a Bayesian perspective such prior information might be seen as essential given that the likelihood as described above does not tend to zero even at infinity. Copas (1982) emphasises the need for a prior for ξ.

In view of (a) and (b) and ignoring the fact that R is not exactly ancillary it may nevertheless be inferentially more appropriate to look at the sampling properties of the maximum likelihood estimator conditional on the observed value of R. Such an approach may be seen to approximate more closely the wholly conditional Bayes approach of Brown (1982).

5. AN EXAMPLE

To compare the profile likelihood and likelihood-based confidence intervals with the classical and Bayes intervals of Brown (1982) we reuse the paints data of that paper. The same subset of 27 observations was used for calibration. Considering only viscosity, for which linearity is appropriate, prediction is based on responses Y_1 and Y_4. A linear model was fitted with viscosity coded at levels -1, 0, 1. The least squares estimates are

$$\hat{\alpha} = (1.7478,\ 37.9363)'$$

$$\hat{B} = (-0.1278,\ -1.6922)$$

$$S^{-1} = \begin{pmatrix} 6.86285 & 0.03052 \\ 0.03052 & 0.02299 \end{pmatrix}$$

where S is the error sum of products from the linear model.

We give predictions at Z values of (a) (1.68, 38.64), (b) (1.86, 35.70) and (c) (1.94, 34.09). Observation (a) is in fact one of the prediction set, observation 18 of the original 36 observations and corresponds to a true X-value coded as zero. Observations (b) and (c) are constructed to be increasingly contradictory in their prediction of X. Both regression coefficients are negative and, for example, 1.94 is the highest Y_1 value in the whole original set of 36 observations whilst 34.09 is the smallest Y_4 value. Thus, the former predicts a low value of X whilst the latter a high value of X. In combination they both conspire to predict a neutral value (near zero) but with a large discrepancy diagnostic ($R = 13.14$). The discrepancy statistics for (a) and (b) are $R = 0.83$ and $R = 4.46$ respectively. Comparing with a chi-squared on one degree of freedom we see that (a) is quite typical, (b) is significant at the 5% level and (c) is significant at the 0.1 per cent level!

These three Z values give logarithms of (2.14) as plotted in Fig. 2(a) (b) and (c). The small Γ asymptotics' critical value is $k_{1,0.05} = 3.84$ for a 95% confidence interval. This encompasses all values of ξ such that the graphs are above $3.84/2 = 1.92$ below their maximum. The intervals, marked by solid horizontal lines in Fig. 2, are thus

$$\text{(a) } (-0.76,\ 1.12); \quad \text{(b) } (-1.08,\ 0.97); \quad \text{(c) } (-1.34,\ 1.16).$$

Notice how the length increases as R increases from (a) to (c). The maximum likelihood estimates are close to the generalised least squares estimates (a) 0.17, (b) -0.04, (c) -0.07 with the only perceptible difference being for case (c) where $\hat{\xi} = -0.085$.

The likelihood-based confidence intervals may be compared with the classical intervals given by (2.8) of Brown (1982). They are (a) $(-1.02, 1.04)$; (b) $(-0.82, 0.72)$; (c) empty set. Notice here how the intervals actually get shorter with increasing discrepancy R as one moves from (a) to (c). In fact, in case (c) all values of the left hand side of (2.8) are greater than 0.28, the value of the right hand side, so the interval has vanished completely! These effects corroborate the effect of R on confidence intervals as discussed in section 4 and demonstrated in Fig. 1.

6. CONCLUSION

We have examined the profile likelihood for controlled multivariate calibration, and in particular maximum likelihood estimation and likelihood-based confidence intervals.

The maximum likelihood estimator is typically close to the traditional estimator but will move some way from this if an inconsistency diagnostic is large. The paper has provided likelihood-based confidence regions together with small Γ asymptotics to specify the critical constant of a region of specified size.

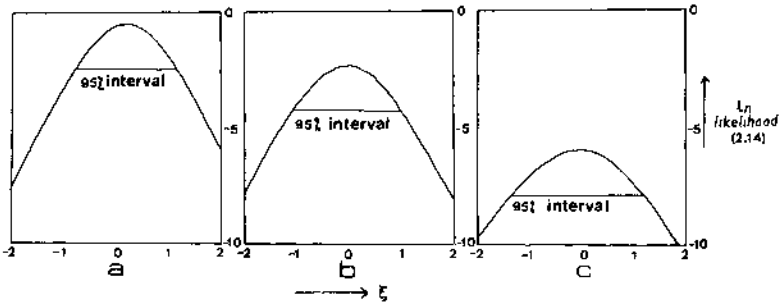

Fig. 2. Paint data: Logarithm of (2.14) for $Z =$ (a) (1.68, 38.64) (b) (1.86, 35.70) and (c) (1.94, 34.09).

In addition, likelihood-based confidence regions have been shown to possess the intuitively desirable property of expansion with increasing values of the inconsistency diagnostic.

REFERENCES

Aitchison, T. C. (1982) Discussion of paper by Brown. *J. R. Statist. Soc.* B, **44**, 309–310.

Barnard, G. A. (1982) Discussion of paper by Brown. *J. R. Statist. Soc.* B, **44**, 314.

Brown, P. J. (1982) Multivariate Calibration. (With Discussion). *J. R. Statist. Soc.* B, **44**, 287–321.

Copas, J. B. (1982) Discussion of the paper by Brown. *J. R. Statist. Soc.* B, **44**, 312–313.

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

Good, I. J. (1969) Some applications of the singular decomposition of a matrix. *Technometrics*, **11**, 823–831.

Harding, E. F. (1986) Modelling – The classical approach. *The Statistician*, **35**, 115–34.

Hoadley, B. (1970) A Bayesian look at inverse linear regression. *J. Amer. Statist. Ass.*, **65**, 356–369.

Kalbfleish, J. D. and Sprott, D. A. (1970) Applications of likelihood methods to models involving large numbers of parameters. (With Discussion). *J. R. Statist. Soc.* B, **32**, 175–208.

Kalotay, A. J. (1971) Structural solution to the linear calibration problem. *Technometrics*, **13**, 761–769.

Naes, T. (1983) Multivariate calibration. PhD. Thesis, Norwegian Food Research Institute.

Oman, S. D. (1985) Confidence regions in multivariate calibration. Dept of Stat., Hebrew Univ., Jerusalem.

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

Press, S. J. (1972) *Applied Multivariate Analysis*. New York: Holt, Rinehart and Winston.

Rosenblatt, J. R. and Spiegelman, C. H. (1981) Discussion of paper by Hunter and Lamboy. *Technometrics*, **23**, 329–333.

Sundberg, R. (1982) When does it pay to omit variables in multivariate calibration? Dept. of Maths., Royal Inst. Techn., Stockholm.

———(1985) When is the inverse regression estimator MSE-superior to the standard regression estimator in multivariate controlled calibration situations? *Statist. Prob. Letters*, **3**, 75–79.

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

———(1969) Regression methods in calibration problems. *Bull. ISI*, **43**, 17–28.

Wood, J. M. (1982) Estimating the age of an animal: an application of multivariate calibration. *Proc. 11th Int. Bio. Conf.*