*International Statistical Review* (1991), **59**, 3, pp. 309–336. Printed in Great Britain


# Statistical Calibration: A Review

**Christine Osborne**

*School of Mathematical Sciences, University of Bath, Bath BA2 7AY, England*

## Summary

**This paper reviews the wide variety of approaches to both univariate and multivariate calibration. In particular, in univariate calibration, it considers the classical, inverse, Bayesian and non-parametric approaches together with the approaches via tolerance regions. In multivariate calibration, it considers the ridge regression, multiple linear regression, partial least squares regression and principal components regression approaches together with the Bayesian and profile likelihood approaches. An extensive bibliography is provided.**

*Key words:* Absolute calibration; Bayes estimator; Classical estimator; Comparative calibration; Cross-validation; Inverse estimator; Linear model; Marginal likelihood; Mean squared error; Multivariate regression; Non-linear model; Non-parametric regression; Optimal design; Profile likelihood; Simultaneous confidence intervals.

## 1 The Nature of Calibration

### *1.1 Introduction*

Statistical calibration has some similarities with scientific calibration, which is the process whereby the scale of a measuring instrument is determined or adjusted on the basis of an informative or calibration experiment, but it has a more complicated form. A statistical calibration problem is a kind of inverse prediction, a problem of retrospection and some authors call it inverse regression rather than calibration. It is probably best explained by considering a typical univariate calibration problem.

Consider the problem of a chemist wishing to establish a calibration curve to use in measuring the amount of a certain chemical $A$ in samples sent to an analytical laboratory. There are two stages in the calibration process. In the first stage, for each of $n$ samples with *known* amounts of chemical $A$, one or more measurements are made with the relatively quick, inexpensive *test method* being calibrated ($Y$). The *known* amounts of chemical $A$ have been determined by an extremely accurate *standard method* that is slow and expensive ($X$). The resulting data constitutes the calibration experiment and is used to estimate the calibration curve $f$. This calibration curve is now ready for use in the second stage of the calibration process which involves prediction. In the second stage, samples with *unknown* amounts of chemical $A$ are analysed with the test method and the amount of chemical $A$ predicted for each new sample. For a given sample, one or more measurements using the test method may be made.

### *1.2 Absolute and Comparative Calibration*

It is important in any discussion of calibration to distinguish between absolute and comparative calibration. Williams (1969a) stressed that although these two activities are both called calibration, they are conceptually different and lead to different issues in statistical modelling.

In absolute calibration a quick or non-standard measurement technique in measurement is either known or made with negligible error. The chemist's problem outlined above is an example of absolute calibration. A high percentage of statistical papers on calibration over the past twenty years have been about absolute rather than comparative calibration.

With comparative calibration one instrument or measurement technique is calibrated against another with neither one being inherently a standard so that there is no standard measurement $X$. Williams (1969a), Barnett (1966, 1969) and Theobald & Mallinson (1978) discuss comparative calibration. Cochran (1943) proposed methods of multivariate analysis for application in comparative calibration experiments where the values obtained on two different measurement scales are correlated. Rosenblatt & Spiegelman (1981) give several references in the field of comparative calibration.

### *1.3 Mathematical Formulation of the Univariate Calibration Problem*

The chemist's problem outlined above illustrates a common calibration procedure. Let us consider the mathematical formulation of this calibration procedure. Let the true values associated with the standard and test methods be designated by $\xi$ and $\eta$ respectively. We assume $\eta = f(\xi)$. There have been many papers devoted to the linear calibration problem i.e. it is assumed that $f(\xi) = \beta_0 + \beta_1 \xi$, where $\beta_0$ and $\beta_1$ are the intercept and slope parameters respectively. However, Tallis (1969), Scheffé (1973), Lwin & Maritz (1980), Clark (1980), Lundberg & De Maré (1980) and Lwin & Spiegelman (1986) have assumed a general form for $f(\xi)$ to derive their theoretical results (see §§ 2, 3, 4 and 5). The majority of them consider the linear model as a special case of interest. Knafl et al. (1984) assumes a non-linear form for $f(\xi)$ in response to the particular demands of their pressure-volume calibration problem and Osborne (1990) assumes $f(\xi)$ is a cubic spline with knots at the data points (see § 5).

In the first stage of the calibration process, the *calibration experiment*, $n$ pairs of observations $(X_i, Y_i)$ are obtained where $X_i$ and $Y_i$ are the observed values of $\xi_i$ and $\eta_i$ respectively,

$$\begin{aligned} Y_i &= \eta_i + \varepsilon_i \quad i = 1, 2, \ldots, n, \\ X_i &= \xi_i + \delta_i \quad i = 1, 2, \ldots, n, \end{aligned} \tag{1.1}$$

where $\varepsilon_i$ and $\delta_i$ are experimental errors. In all absolute calibration problems $\delta_i = 0$ for all $i$. Lwin & Spiegelman (1986) consider a case close to absolute calibration where $\delta_i \neq 0$ for any $i$ but the $\delta_i$ have a known finite bound. Mandel (1984) and Carroll & Spiegelman (1986) also consider calibration when both $X$ and $Y$ have error. Making the assumption that $\delta_i = 0$ for all $i$, produces the following model

$$Y_i = \eta_i + \varepsilon_i = f(x_i) + \varepsilon_i \quad i = 1, 2, \ldots, n. \tag{1.2a}$$

In the case of the linear calibration problem this becomes

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \ldots, n. \tag{1.2b}$$

The next assumption which is made is that the $\varepsilon_i$'s are independent normal random variables with mean 0 and variance $\sigma_1^2$. Mandel (1957) considers a calibration problem where the assumption of independence is not valid.

Having established the calibration curve/line we proceed to the second stage of the calibration process. A sample is presented with a specific *unknown* value $\eta$ and one or more measurements are made using the test method from which are obtained the

observations

$$Y'_j = \eta + \varepsilon'_j = f(\xi) + \varepsilon'_j \quad j = 1, 2, \ldots, m, \tag{1.3a}$$

$$= \beta_0 + \beta_1 \xi + \varepsilon'_j \qquad j = 1, 2, \ldots, m, \tag{1.3b}$$

in the linear calibration problem, where $\varepsilon'_1, \varepsilon'_2, \ldots, \varepsilon'_m$ are independent normal random variables with mean 0 and variance $\sigma_2^2$. Berkson (1969) argued that $\sigma_1^2$ will ordinarily be substantially smaller than $\sigma_2^2$, since the calibration experiment is usually performed under relatively highly controlled conditions. Perng & Tong (1974) in their sequential approach to calibration assume that $\sigma_1^2 \neq \sigma_2^2$ but the majority of authors assume that these variances are equal, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Given the data from the first and second stages, inferences are now made about the unknown $\xi$ that corresponds to $\eta$ for the sample being measured. For the linear model $\xi$ is given by

$$\xi = \frac{(\eta - \beta_0)}{\beta_1}. \tag{1.4}$$

There are many approaches to this inference problem and these will be outlined in the following sections (§§ 2–7).

In comparative calibration, one is typically concerned with the calibration of $p$ measuring instruments when each is used to make measurements of the same property on every member of a group of specimens. There is usually no standard measurement to which they may be referred so there is no question of estimating the measurement expected on one instrument corresponding to that observed on another as in absolute calibration for example. Instead one is concerned in assessing the relative calibrations and relative precisions of the set of $p$ instruments. In the details given below, the term *structural* relationship is used to denote a relationship between random variables whilst the term *functional* relationship is used to denote a relationship between fixed variables.

Barnett (1969) points out that each of the $p$ instruments may not be equally consistent in its readings (on repeated measurements) for a given specimen but it is often reasonable to assume equal consistency from one specimen to another for a given instrument. Furthermore measurement on a given specimen may vary from one instrument to another quite apart from *within-specimen* sampling fluctuations, reflecting differences in relative calibrations of the instruments. Suppose $\xi_i$ represents the unobserved 'true' reading of the $i$th instrument ($i = 1, 2, \ldots, p$) and $Y_i$ is the observed measurement of the $i$th instrument so that

$$Y_i = \xi_i + \varepsilon_i \quad i = 1, 2, \ldots, p, \tag{1.5}$$

where the $\varepsilon$'s represent the errors of measurement. Assume that $\varepsilon_i$ is independent of $\xi_i$ and the $\varepsilon_i$ are independent random variables with mean zero and variance $\sigma_i^2$ ($i = 1, 2, \ldots, p$). The variances $\sigma_i^2$ assess the consistency of measurement for the different instruments in repeated measurements on the same specimen and are assumed to be constant from specimen to specimen. Barnett put forward, as a first approximation, a model which expresses a *linear* structural relationship between the values $\xi_i$ given by a pair of instruments for a particular specimen (one of the instruments is the 'reference' instrument). In particular

$$\xi_i = \alpha_i + \beta_i \xi_1 \quad i = 1, 2, \ldots, p.$$

He assumed that $\xi_1$ and hence $\xi_2, \xi_3, \ldots, \xi_p$ are normally distributed random variables, the mean and variance of $\xi_1$ being $\mu$ and $\sigma^2$ respectively. The intercept and slope

parameters $\alpha_i$ and $\beta_i$ characterise the relationship between the values given by instrument $i$ and those given by the 'reference' instrument 1, as the physical property under investigation varies from one specimen to another. By convention $\alpha_1$ and $\beta_1$ are taken to be 0 and 1 respectively. Barnett presented some consistent moment estimators of the parameters of the model and analysed a data set concerned with lung function.

The choice of 'reference' instrument is obvious on technical grounds, when there is a single control instrument and a set of experimental alternatives. Theobald & Mallinson (1978) pointed out that if there is no particular 'reference' instrument, then there is a symmetry about the problem which is not reflected in Barnett's notation. They suggested that the relationship between the true responses might be better expressed by supposing that

$$\xi_i = \mu_i + \lambda_i F \quad i = 1, 2, \ldots, p, \tag{1.6}$$

where $F$ represents a hypothetical standard measurement having zero mean and unit variance over the population of possible specimens. The $\lambda_i$ are called *calibration factors.*

Substituting (1.6) into (1.5) gives the model

$$Y_i = \mu_i + \lambda_i F + \varepsilon_i \quad i = 1, 2, \ldots, p. \tag{1.7}$$

If $F$ and $\varepsilon$ have normal distributions, model (1.7) is the standard model for factor analysis restricted to a single common factor so the parameters $\lambda_i$ and $\mu_i$ $(i = 1, 2, \ldots, p)$ can be estimated using a program for carrying out factor analysis by maximum likelihood (ML). The calibration equation for any two instruments $h$ and $l$ is

$$\xi_h = \mu_h + (\xi_l - \mu_l)\frac{\lambda_h}{\lambda_l} \quad h, l = 1, 2, \ldots, p.$$

When $\lambda_h = \lambda_l$, the true measurements differ by an additive bias so several authors (Grubbs, 1948, 1973; Theobald & Mallinson, 1978) have considered the estimation of the parameters of the model (1.7) subject to the $\lambda_i$'s all being the same. It is also possible to estimate the precision of an instrument. Let $\pi_i$ represent the precision of the $i$th instrument where

$$\pi_i = \frac{\text{St. dev}\,(\xi_i)}{\text{St. dev}\,(\varepsilon_i)} = \frac{\lambda_i}{\sigma_i} \quad i = 1, 2, \ldots, p.$$

The parameters $\pi_i$ can be estimated using ML estimation (Theobald & Mallinson, 1978).

Williams (1969a) uses a different model from model (1.7). Suppose that there are $n$ specimens to be measured by the $p$ instruments. Let $\xi_{ij}$ represent the true measurement of the $j$th specimen tested on the $i$th instrument. Then Williams assumes that

$$\xi_{ij} = \mu_i + \lambda_i \varphi_j \quad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, n,$$

where $\varphi_j$ are scalar parameters rather than values taken by the random variable $F$ as in model (1.7). This leads to the model

$$Y_{ij} = \mu_i + \lambda_i \varphi_j + \varepsilon_{ij}, \tag{1.8}$$

where it is assumed that the $\varepsilon_{ij}$'s are independent random variables with mean zero and variance $\sigma_i^2$ $(i = 1, 2, \ldots, p)$. William's model differs from Theobald and Mallinson's model (model (1.7)) in that his model leads to the calibration equations being regarded as a set of $(p-1)$ functional relationships rather than $(p-1)$ structural relationships.

Theobald & Mallinson (1978) strongly criticise William's use of the ML method to estimate the parameters of model (1.8). They state that the ML method cannot be applied satisfactorily when there is no prior information on the error variances and replicate

observations are not available. They suggest a consistent method of estimation for model (1.8).

### *1.4 Controlled and Random/Natural Calibration*

In controlled calibration the values of $X$ are fixed or pre-chosen. The calibration experiment is often designed so that chosen values of $X$ span the range of possible values of $X$. The chemist's problem described above is an example of a controlled calibration problem. In random or natural calibration the $X$-values are not chosen by the experimenter but $X$ is a random variable as well as $Y$. Often $X$ will correspond to random true values accurately determined. Brown (1982) emphasises the importance of distinguishing between controlled and random calibration. Aitchison & Dunsmore (1975) give examples of controlled and random calibration as does Brown (1982). Naes (1985a) comments that when using Beer's model in chemical spectroscopy, the basic distinction between natural and controlled calibration is unimportant if the residuals, after fitting the calibration model to the calibration data, are small, but if appreciable errors exist, the distinction is important (see § 6).

### *1.5 Milestones*

Rosenblatt & Spiegelman (1981) offer a detailed general discussion of calibration. The papers which have acted as milestones along the development road of statistical calibration theory are Eisenhart (1939), Krutchkoff (1967), Hoadley (1970), Scheffé (1973) and Brown (1982). These will be considered in detail below. There are many approaches to this inference problem and these will be outlined in the sections which follow.

To carry out this review, the author firstly used the *Current Index to Statistics* and the *Science Citation Indexes* (SCI). The above five *milestone* papers were used as citing papers when using the SCI. As both of these indexes provide only minimal information beyond what is in the title and the keywords, the *Statistical Theory and Method Abstracts* was used to give *summaries* of relevant papers.

## 2. The Classical and Inverse Approaches to Calibration

### *2.1 The Classical Estimator*

Eisenhart (1939) set the stage for statistical investigations of absolute calibration problems. His analysis and solution of the inverse estimation problem has come to be called *classical.* Eisenhart obtained his estimate of $\xi$ by considering the regression of $Y$ on $X$,

$$E(Y \mid X = x) = \beta_0 + \beta_1 x.$$

The estimated regression line of $Y$ on $X$ is given by

$$\begin{aligned} \hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 X \\ &= \bar{Y} + \frac{S_{xy}}{S_{xx}}(X - \bar{x}), \end{aligned} \tag{2.1}$$

where

$$S_{xy} = \sum_i (x_i - \bar{x})(Y_i - \bar{Y}), \quad S_{xx} = \sum_i (x_i - \bar{x})^2.$$

Eisenhart then inverted equation (2.1) to give an estimator of $\xi$, the unknown $X$, which has since become known as the *classical estimator*. Let it be denoted by $\hat{\xi}_c$. Then

$$\hat{\xi}_c = \bar{x} + \frac{S_{xx}}{S_{xy}}(\bar{Y}' - \bar{Y}),$$

where $\bar{Y}'$ is the mean of the $m$ observations at the prediction stage. If one makes the assumption of normal errors in models (1.2b) and (1.3b), then $\hat{\xi}_c$ is the maximum likelihood estimator of $\xi$. Eisenhart also produced an interval estimate for $\xi$ based on the $t$-distribution with $(n-2)$ degrees of freedom.

Subsequent texts and journal articles used the same *classical* approach as favoured by Eisenhart, i.e. regressing $Y$ on $X$. Examples are Mandel & Linnig (1957), Mandel (1958), Williams (1959) and Linnig & Mandel (1964). Mandel & Linnig (1957) adopted a joint inference approach to the problem of interval estimation of $\xi$ and produced conservative intervals when compared with Eisenhart's.

Fieller (1954) produced interval estimates for $\xi$ identical to those of Eisenhart using a fiducial argument. Fieller showed that the calibration problem could be reduced to considering the ratio of the means of two normally distributed random variables (see Hoadley, 1970). Creasy (1954) also used a fiducial argument to obtain interval estimates.

The *classical* approach to interval estimation has caused considerable consternation over the years because if the slope parameter $\beta_1$ is not significantly different from zero the interval is either the whole real line or even two disjoint semi-infinite lines. As a result of this problem, Berkson (1969) and Shukla (1972) obtained asymptotic expressions for the bias and mean square error (M.S.E.) of $\hat{\xi}_c$ conditional on the event $|\hat{\beta}_1| > 0$. Shukla & Datta (1985) studied the bias and M.S.E. of $\hat{\xi}_c$ under the truncation procedure that $H_0: \beta_1 = 0$ is rejected with some specified probability $\alpha$. The argument behind this conditioning is that in practice, it is unlikely that one would proceed to estimate $\xi$ unless one is convinced that $\beta_1 \neq 0$. One way is to test the hypothesis $H_0: \beta_1 = 0$ and estimate $\xi$ only if $H_0$ is rejected. This suggests a reasonable truncation procedure of the distribution of $\hat{\beta}_1$ around zero. Perng & Tong's (1974) two-stage sequential approach is in this spirit (see § 7).

The main support of the classical estimator $\hat{\xi}_c$, has been the consistency of its conditional mean within the assumed model (Berkson, 1969) and the fact that if one assumes normal errors, $\hat{\xi}_c$ is the maximum likelihood estimator. The main disadvantage of the classical estimator has been the fact that it has an undefined mean and infinite M.S.E. for fixed $(x_1, x_2, \ldots, x_n)$ and finite $n$ (Williams, 1969b). However, Miwa (1985) showed that $\hat{\xi}_c$ has lower moments conditional on $H_0: \beta_1 = 0$ being rejected. Consider linear functions of $Y$ in the class $\varphi(Y) = k_0 + k_1 Y$. If $k_0$ and $k_1$ are chosen to minimise

$$\sum_{i=1}^{n} E_F(\varphi(Y_i) - x_i)^2,$$

where the expectation is with respect to $\mathbf{Y}$ for a fixed value of $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $F$ is the distribution function of $Y$ for fixed $X$, the summation is called the total mean squared error or *compound error* of the calibration experiment. Finding constants $k_0$ and $k_1$ that minimise

$$\sum_{i=1}^{n} E_F(\varphi(Y_i) - x_i)^2$$

is the problem of compound estimation and the resulting optimal estimator is called a *linear compound estimator*. Lwin & Maritz (1982) established the asymptotic unbiased-

ness of the classical estimator on the basis of compound estimation without specific distributional assumptions on the errors in the model. They also showed that in the class of consistent estimators linear in $Y$, $\hat{\xi}_c$ has asymptotically minimum variance.

### *2.2 The Inverse Estimator*

Krutchkoff (1967) derived an estimator known as the *inverse estimator* by considering the regression of $X$ on $Y$ rather than $Y$ on $X$ for a controlled calibration problem. The estimated regression line is given by

$$\hat{X} = \bar{x} + \frac{S_{xy}}{S_{yy}}(Y - \bar{Y}),$$

where

$$S_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

Let the inverse estimator of $\xi$ be denoted as $\hat{\xi}_I$. This formulation of the problem results in

$$\hat{\xi}_I = \bar{x} + \frac{S_{xy}}{S_{yy}}(\bar{Y}' - \bar{Y}),$$

where $\bar{Y}'$ is the mean of the $m$ observations taken at the prediction stage of the calibration process. Krutchkoff concluded on the basis of a Monte Carlo investigation that the M.S.E. of $\hat{\xi}_I$ was uniformly less than that of the classical estimator and so the estimator $\hat{\xi}_I$ was preferable. The Monte Carlo work involved 10 000 repetitions and considered both normal and non-normal error distributions with $n \leqslant 20$. His paper caused considerable controversy when published because his suggestion went against protocols established as far back as Eisenhart (1939), since the $n$ $X$-values are fixed in a controlled calibration problem. Williams (1969b) criticised $\hat{\xi}_I$ on the grounds that it was derived on the false assumption that the errors $\zeta_i$ $(i = 1, 2, \ldots, n)$ are independent of $Y_i$ in the inverse regression of $X$ on $Y$ given by

$$X_i = \gamma + \delta Y_i + \zeta_i.$$

However, Lwin & Maritz (1982) argued that their derivation of $\hat{\xi}_I$ does not require this assumption so William's criticism of $\hat{\xi}_I$ is not justified. Perng (1987) derived the inverse estimator by the method of cross-validation (Stone, 1974). The derivation does not require any specific distributional assumptions.

Berkson (1969), Martinelle (1970), Halperin (1970) and Shukla (1972) pointed out that Krutchkoff's conclusion held only in very restrictive circumstances. Krutchkoff, in a subsequent study (1969) compared the classical and inverse approaches and demonstrated that for a sufficiently large number of observations, the classical method produces a lower M.S.E. outside the range of calibration, however the conclusions drawn in his 1967 paper remain unchanged for $X$ values within the calibration range (i.e. $0 \leqslant X \leqslant 1$). Apart from Lwin & Maritz, Hoadley (1970) appears to be the only author to give qualified support to the inverse estimator in that he argued that if one assumed a prior for $\xi$, a non-central Student density centred at $\bar{x}$ (see equation (3.1)) and one assumed a non-informative joint prior for $(\beta_0, \beta_1, \sigma^2)$, the posterior distribution of $\xi$ had mean $\hat{\xi}_I$ and so the inverse estimator was a Bayes estimator in the case of $m = 1$ in model (1.3b). Brown (1982) maintains that these results imply, in sampling theory terms, that the inverse method will do well if the unknown $\xi$ is reasonably central to the set of pre-chosen $X$-values of a

controlled calibration experiment but not perform so well if $\xi$ happened to be outside this pre-chosen range.

Exact expressions for the mean and M.S.E. of $\hat{\xi}_I$ have been derived recently (Shukla & Datta, 1985; Oman 1985). Prior to these results, Williams (1969b) indicated that the mean, variance and M.S.E. of $\hat{\xi}_I$ are finite for $n \geq 4$. However, $\hat{\xi}_I$ is biased even as $n \to \infty$; the bias is affected mainly by the kurtosis of the error distribution of the model and this effect can be reduced by increasing $n$ (Lwin, 1981). The inverse estimator is unbiased only if $\xi = \bar{x} = \sum_i x_i/n$. The inverse estimator is also an inconsistent estimator (Berkson, 1969).

### 2.3 *Comparison of* $\hat{\xi}_c$ *and* $\hat{\xi}_I$

In reply to Krutchkoff's 1967 paper, Berkson (1969) showed that in practice when $|\sigma/\beta_1|$ is small, the asymptotic M.S.E. of $\hat{\xi}_c$ is smaller than the M.S.E. of $\hat{\xi}_I$ except when $\xi$ lies very near to $\bar{x}$. Moreover, the inverse method provides an inconsistent estimator whilst the classical method/approach provides a consistent estimator. Martinelle (1970) obtained the expression for relative efficiency for large $n$ and gives results similar to those of Berkson. Halperin (1970) compared the two estimators on the basis of *Pitman's closeness* criterion, consistency and the M.S.E. of the relevant asymptotic distributions. Suppose $\hat{\xi}_1$ and $\hat{\xi}_2$ are two estimators of $\xi$, then the Pitman's nearness (PN) of $\hat{\xi}_1$ relative to $\hat{\xi}_2$ is given as

$$PN = \Pr\{|\hat{\xi}_1 - \xi| < |\hat{\xi}_2 - \xi|\}.$$

He found that for large $n$, $\hat{\xi}_I$ was superior to $\hat{\xi}_c$ in the sense of *closeness* only in a closed interval round $\bar{x}$ and inferior elsewhere. In practice this interval seems to be very small. He put forward a family of *modified* inverse estimators $\hat{\xi}(r)$ where $r = 1$ gives $\hat{\xi}_I$ but preferred the classical estimator to his modified estimators because $\hat{\xi}_c$ gives an exact interval estimate of $\xi$ whereas the modified estimators do not allow an interval estimate. Krutchkoff (1971) carried out further simulations using the Pitman closeness criterion but these did not lead to any clear-cut decisions. Rothman (1968) and Saw (1970) were also very critical of Krutchkoff's 1967 paper but perhaps the most incisive reply came from Williams (1969b). As previously stated, the main objection to using $\hat{\xi}_c$ is that its M.S.E. is infinite for fixed $(x_1, x_2, \ldots, x_n)$ and finite $n$. Williams showed that if $\sigma^2$ is assumed known and normal errors are assumed in model (1.2b) a unique unbiased estimator of $\xi$ that is a function of the minimal sufficient statistics can be found. Suppose this is denoted by $\hat{\xi}_M$ and suppose further that $\tilde{\xi}$ is any other unbiased estimator of $\xi$. Then by the Rao–Blackwell theorem

$$\text{Var}(\hat{\xi}_M) \leq \text{Var}(\tilde{\xi}).$$

Williams showed that $\hat{\xi}_M$ had infinite variance so establishing that no unbiased estimator of $\xi$ would have finite variance. He concluded that the minimisation of M.S.E. as a criterion for choosing between estimators of $\xi$ was not sensible in the calibration situation. His arguments may be taken as a reply to the objection of using the classical estimator.

A second reply to this objection is that one can always apply a truncation procedure to make both the expectation and M.S.E. of $\hat{\xi}_c$ finite. Several authors have suggested that the performance of the classical estimator should be assessed on conditioning on the event $|\hat{\beta}_1| > 0$. Using Tchebycheff's inequality

$$P(|\hat{\beta}_1 - \beta_1| \geq k) \leq \frac{\sigma^2}{k^2 \beta_1^2 S_{xx}} \quad k > 0.$$

By making $S_{xx}$ large and providing $|\sigma/\beta_1|$ is not large, the probability that $\hat{\beta}_1$ lies in an interval which contains very small values including zero can be made very small. So it is possible, by increasing $n$ and choosing values of $X$ which are not very close to each other, to truncate the distribution of $\hat{\beta}_1$ in such a way as to exclude values of $\hat{\beta}_1$ very close to zero. Krutchkoff used a truncated version of the classical estimator, which replaced $\hat{\beta}_1$ with 0·001 whenever $|\hat{\beta}_1| < 0{\cdot}001$, to obtain the variance and bias of $\hat{\xi}_c$ by means of convergent expansions (1967, 1968). As mentioned in § 2.1, Shukla (1972) and Shukla & Datta (1985) used a *conditional set-up*. Lwin & Maritz (1982) derived a linear compound estimator conditional on the event $|\hat{\beta}_1| > 0$.

When considering M.S.E., the advantage of $\hat{\xi}_I$ over the truncated version of the classical estimator is most pronounced when $n$ is small, $|\sigma/\beta_1|$ is large and when estimation of $\xi$ is restricted to the calibration range i.e.

$$\min_i X_i \leqslant \xi \leqslant \max_i X_i$$

(Krutchkoff, 1967, 1968, 1969; Berkson, 1969; Martinelle, 1970; Shukla, 1972). Lwin (1981) extended Shukla's results (1972) to a more general location and scale family distribution for the errors in model (1.2b). Lwin gives a detailed comparison of $\hat{\xi}_c$ and $\hat{\xi}_I$ with graphs of their approximate M.S.E.'s as functions of $\xi$. He showed that the inverse estimator would be worse than the classical estimator as $\xi$ moves further away from $\bar{x}$ for symmetric error distributions and except in the case of extremely peaked error distributions, the inverse estimator would be more efficient at $\xi = \bar{x}$ than the classical estimator (compare Berkson, 1969 and Halperin, 1970). Lwin argued that the M.S.E. criterion could not be dismissed outright without further consideration of the *conditioning* imposed on it.

The conditional or truncated classical estimator derived by Shukla & Datta (1985) compares very favourably with the inverse estimator. Although an earlier conclusion about $\hat{\xi}_I$ and $\hat{\xi}_c$ still holds good for their conditional classical estimator (i.e. $\hat{\xi}_I$ has a smaller M.S.E. than $\hat{\xi}_c$ when $\xi$ is close to the design mean $\bar{x}$) their conditional estimator has a small bias, is consistent and has a finite M.S.E.

Ali & Singh (1981) derived an alternative estimator $\hat{\xi}_a$ which they claimed is uniformly better than either $\hat{\xi}_c$ or $\hat{\xi}_I$. It is given by $\hat{\xi}_a = \lambda\hat{\xi}_c + (1-\lambda)\bar{x}$, where $0 \leqslant \lambda \leqslant 1$. Turiel et al. (1982) performed a simulation study in which they compared M.S.E., Pitman's closeness and probability of over-estimation (i.e. $P(\hat{\xi} > \xi)$) for the classical estimator, the inverse estimator and a modified version of Naszodi's estimator (Naszodi 1978, see § 7). They considered both the usual inverse regression problem and the inverse median estimation problem which is less familiar. The latter arises when one wishes to estimate $X$ given a specified median value of $Y$. For example, $X$ might be the stress on a component and $Y$ the time to failure of the component and it is wished to estimate the stress associated with a specified median lifetime for that component. They considered the probability of over-estimation to be an important criterion by which to compare estimators because in some practical problems, particularly the inverse median problem, the undesirable consequences of over-estimation are more severe than those of under-estimation or vice-versa. For example, if an estimate of $\xi$ is then used for designing an aircraft engine, an over-estimate of stress associated with a specified median lifetime might lead to a serious accident while an under-estimate would result only in undue conservation.

Chow & Shao (1990) studied the difference between $\hat{\xi}_I$ and $\hat{\xi}_c$, in particular they examined the probability that the ratio of $\hat{\xi}_I$ to $\hat{\xi}_c$ differs from unity by more than a specified small constant. Their results showed that this probability increases as $|\beta_1/\sigma|$ decreases. They proposed methods of estimating this probability.

## 3 The Bayesian Approach to Calibration

Hoadley's (1970) paper was a milestone along the development road of statistical calibration theory because it clearly defined a Bayesian approach to linear calibration. Let $\mathbf{z} = \{(x_i, Y_i) : i = 1, 2, \ldots, n\}$.

Hoadley argued in his paper that the classical estimator $\hat{\xi}_c$ was unsatisfactory from a point of view independent of M.S.E. considerations. The usual $F$-statistic for testing $H_0 : \beta_1 = 0$ is given by $F = \hat{\beta}_1^2 S_{xx}/\hat{\sigma}_1^2$, where

$$\hat{\sigma}_1^2 = \left\{ \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 + \sum_{j=1}^{m} (Y_j' - \bar{Y}')^2 \right\} \Big/ (n + m - 3).$$

If $F$ is much larger than $F_{\alpha;1,(n+m-3)}$ (the upper $100(1-\alpha)\%$ point of the $F$ distribution with 1 and $(n + m - 3)$ degrees of freedom), then $\hat{\xi}_c$ is fairly precise an estimator but if $\beta_1$ is close to zero and $H_0$ cannot be rejected, then $\hat{\xi}_c$ is very imprecise, i.e. the data of the calibration experiment, $\mathbf{z}$, contain information about the precision of $\hat{\xi}_c$. It would seem reasonable to have some way of giving less weight to $\hat{\xi}_c$ when it is known to be unreliable. This is precisely what a Bayes estimator does. Hoadley argued that conditioning on the data of the calibration experiment was the right way to approach the calibration problem.

Hoadley took a general form of prior density

$$p(\beta_0, \beta_1, \sigma^2, \xi) \propto p(\beta_0, \beta_1, \sigma^2) p(\xi)$$

and obtained explicit results using the conventional non-informative prior distribution for $(\beta_0, \beta_1, \sigma^2)$ i.e. $p(\beta_0, \beta_1, \sigma^2) \propto \sigma^{-2}$. He assumed normal errors in models (1.2b) and (1.3b) and showed that the posterior density of $\xi$ is given by

$$p(\xi \mid Y', \mathbf{z}) \propto p(\xi) L(\xi),$$

where $L(\xi)$ is the predictive density function of $\bar{Y}'$. Hoadley remarked that $L(\xi)$ is a kind of likelihood function representing the information about $\xi$ from all sources except the prior distribution of $\xi$. However $L(\xi)$ is not integrable so it is necessary that $p(\xi)$ is a proper density function so that overall integrability is achieved, enabling evaluation of a proper posterior density for $\xi$.

Halperin (1970) recognised the inverse estimator $\hat{\xi}_I$ as a Bayes estimator corresponding to a normal linear model for the calibration experiment and a normal prior density for $\xi$. Dunsmore (1968) derived $\hat{\xi}_I$ as the Bayes estimator of $\xi$ corresponding to a bivariate normal distribution of $(X, Y)$ and a non-informative prior density function

$$p(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \propto \sigma_1^{\nu-4} \sigma_2^{\nu-4} (1 - \rho^2)^{\frac{1}{2}\nu - 3} \quad \nu \leqslant n$$

for the set of parameters involved in the bivariate normal model. He considered the case of $m = 1$ in model (1.3b), i.e. one observation at the prediction stage. Hoadley also showed that if one assumes a prior for $\xi$ of the form

$$p(\xi) = \mathrm{St}\left(n - 3, \bar{x}, \left(1 + \frac{1}{n}\right) \frac{S_{xx}}{n - 3}\right) \tag{3.1}$$

which is the density function of a Student distribution based on $(n - 3)$ degrees of freedom, centred at $\bar{x}$ and with scale factor

$$\left(1 + \frac{1}{n}\right) \frac{S_{xx}}{n - 3},$$

then if $m = 1$ in model (1.3b), the posterior mean is $\hat{\xi}_I$ and the scale factor depends on $F$ defined above. In Bayesian terms, $\hat{\xi}_I$ can be thought of as a shift of $\hat{\xi}_c$, the classical

estimator, towards the prior mean, $\bar{x}$. The more informative the data (i.e. the larger $F$) the smaller the shift. Using Hoadley's relationship,

$$\hat{\xi}_I - \bar{x} = R[\hat{\xi}_c - \bar{x}],$$

where $R = \{1 + (n-2)/F\}^{-1}$ and $F$ is as defined above, it can be seen that in all situations

$$|\hat{\xi}_c - \bar{x}| \geqslant |\hat{\xi}_I - \bar{x}|$$

with equality only when $R = 1$ i.e. there is no error in model (1.2b). So the more informative the data, the more one moves from the prior mean towards the estimate $\hat{\xi}_c$. Since, subject to very weak conditions, Bayes estimators are consistent and the inverse estimator, $\hat{\xi}_I$, is inconsistent (Berkson, 1969), Hill (1981) argued that the inverse estimator was only Bayes for $m = 1$. If one uses the prior for $\xi$ given by equation (3.1) and Hoadley's likelihood function $L(\xi)$ for the case of $m > 1$, the resulting Bayes estimator (i.e. the posterior mean/mode) is consistent as $n, m$ both tend to infinity (Brown, 1982). Brown remarks that the posterior mean or mode may be regarded as the correct generalisation of the Krutchkoff (inverse) estimator to $m > 1$. Hill (1981) conjectures that this Bayes estimator will behave asymptotically like the maximum likelihood estimator as $m$ and $n$ go to infinity at appropriate rates. Brown (1982) obtained multivariate extensions of both Hoadley's theorems (see § 6). Here the term multivariate is intended to mean multivariate in both the $X$'s and the $Y$'s.

Aitchison & Dunsmore (1975) assumed normal errors and a non-informative prior for $(\beta_0 + \beta_1\xi, \sigma^2)$. Through a predictive distribution approach, they obtained the posterior distribution of $\xi$ which they call the calibrative distribution. Their results agree with those of Hoadley when $m = 1$ but for $m > 1$ the predictive density function, $L(\xi)$, is a Student–Siegel density function. The form of $L(\xi)$ is as follows

$$L(\xi) \propto \frac{V_2^{(m-3)/2}}{[V_1 c]^{\frac{1}{2}} V_1^{(m-1)/2}(1 + (\bar{Y}' - \hat{\beta}_0 - \hat{\beta}_1\xi)^2/(V_1 c) + V_2/V_1)^{(v+1)/2}}, \tag{3.2}$$

where

$$V_1 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2, \quad V_2 = \sum_{j=1}^{m} (Y_j' - \bar{Y}')^2, \quad c = \left(\frac{1}{m} + \frac{1}{n} + \frac{(\xi - \bar{x})^2}{S_{xx}}\right), \quad v = n + m - 3.$$

This should be compared with $L(\xi)$, using Hoadley's choice of prior for $(\beta_0, \beta_1, \sigma^2)$ which is as follows

$$L(\xi) \propto \frac{1}{[(V_1 + V_2)c]^{\frac{1}{2}}(1 + (\bar{Y}' - \hat{\beta}_0 - \hat{\beta}_1\xi)^2/\{(V_1 + V_2)c\})^{(v+1)/2}}. \tag{3.3}$$

In many cases, the calibrative density function is not of standard form because of the prior density $p(\xi)$ and has to be constructed using numerical integration techniques (Reilly, 1976). However, Aitchison & Dunsmore showed that, by choosing suitable (tractable) priors for $\xi$, the calibrative density functions are non-central Student density functions with the following means:

$$\hat{\xi}_I \qquad \text{when } m = 1,$$

$$\hat{\xi}_B = \bar{x} + S_{xy}(\bar{Y}' - \bar{Y}) \Big/ \left\{S_{yy} + \sum_{j=1}^{m} (Y_j' - \bar{Y}')^2\right\} \qquad \text{when } m > 1.$$

The calibrative density function for the case $m = 1$ is identical to that derived by Hoadley (1970).

Aitchison & Dunsmore compare (for the case of $m > 1$) the Bayes estimator, $\hat{\xi}_B$, the classical estimator, $\hat{\xi}_c$, the inverse estimator, $\hat{\xi}_I$ and Halperin's modified inverse estimator. They note that $\hat{\xi}_B$, unlike the three other estimators, takes account of the variation in the $m$ future observations $Y'_1, Y'_2, Y'_3, \ldots, Y'_m$ in such a way that the more variation there is in these observations, the nearer the predictive mean is to the prior mean, $\bar{x}$. Aitchison & Dunsmore (1975) argue that this is a good property for an estimator to have because one should be reluctant to change one's prior views if additional data are very variable.

Hunter & Lamboy (1981a) derived another approach to the calibration problem. For the case of $\sigma^2$ known, they assumed normal errors and non-informative priors for $(\beta_0, \beta_1)$ and for $\eta$ where $\eta = \beta_0 + \beta_1 \xi$. In particular, they assumed the priors $p(\beta_0, \beta_1)$ and $p(\eta)$ were locally uniform. For the case of $\sigma^2$ unknown, they again assumed normal errors and a non-informative prior for $(\beta_0, \beta_1, \sigma^2, \eta)$ of the form,

$$p(\beta_0, \beta_1, \sigma^2, \eta) \propto \sigma^{-2}.$$

They further postulated *a priori* independence between $(\beta_0, \beta_1)$ and $\eta$. The first stage of their analysis was to derive the posterior distribution of $(\beta_0, \beta_1, \eta)$ and then using equation (1.4) they derived the posterior distribution of $\xi$. The posterior densities of $\xi$ are the posterior densities of a ratio of bivariate normal random variables (in the $\sigma^2$ known case) and of bivariate $t$ random variables (in the $\sigma^2$ unknown case). The posterior distribution of $\xi$, thus derived, has infinite variance but according to the authors this presents no problems in practice. The authors obtained Bayesian highest posterior density intervals for $\xi$ in a particular practical problem and they laid down conditions under which these Bayesian intervals could be approximated by Fieller's (1954) confidence intervals.

The posterior distribution derived by Hunter & Lamboy differs from the Bayesian posterior distribution found by Hoadley (1970) because Hoadley does not treat $\xi$ as an explicit function of $\eta, \beta_0, \beta_1$ as Hunter & Lamboy do, but directly uses a prior distribution for $\xi$. In Hunter & Lamboy's analysis, the prior for $\xi$ is implicitly given by the priors for $\eta, \beta_0, \beta_1$. Another fundamental difference in the two Bayesian approaches is that Hoadley assumes an *a priori* independence between $\xi$ and $(\beta_0, \beta_1)$ whereas Hunter & Lamboy assume an *a priori* independence between $\eta$ and $(\beta_0, \beta_1)$. Several authors (Hill, 1981; Lawless, 1981; Lwin, 1981; Orban, 1981) have argued that Hoadley's independence assumption is more natural and preferable, especially when there is non-negligible prior information about $\xi$.

Hunter & Lamboy's paper stimulated six discussion papers (Hill, 1981; Lwin, 1981; Rosenblatt & Spiegelman, 1981; Easterling, 1981; Orban, 1981; Lawless, 1981). Hill (1981) was very critical of Hunter & Lamboy's paper. He criticised, amongst other things, their choice of priors for $\eta$ and $(\beta_0, \beta_1)$, their assumption of *a priori* independence between $\eta$ and $(\beta_0, \beta_1)$ and the lack of information on the location, shape and behaviour of the posterior distribution of $\xi$. Hill commented that Hunter & Lamboy's analysis is a special case of Hoadley's (1970). He argued that their posterior distribution of $\xi$ is the Hoadley posterior distribution that would arise from using the modified prior density $p(\beta_0, \beta_1, \sigma^2, \xi)$ proportional to $|\beta_1|/\sigma^2$ in the case of $\sigma^2$ unknown and $|\beta_1|$ if $\sigma^2$ is known. Hill (1981) and Lawless (1981) argued for a family of proper prior distributions which included Hunter & Lamboy's choice of priors as a limiting case. Hill felt that the choice of a family of gamma ($\text{Ga}(\alpha, \lambda)$, $\alpha, \lambda > 0$) prior distributions for $\beta_1$, with $\alpha$ and $\lambda$ chosen to reflect prior knowledge, might yield a more satisfactory analysis, including a consistent estimator of $\xi$ with finite M.S.E. Lawless (1981) made the comment that it was not clear how to incorporate prior information about $\xi$ using Hunter & Lamboy's approach. Lwin

(1981) argued that any Bayesian approach to calibration should incorporate a *conditioning* of the parameter space in the same way that Shukla (1972) conditioned on the event $|\hat{\beta}_1| > 0$. A criticism of Hunter & Lamboy's prior on $\eta$ which came in a later paper (Brown, 1982) was that there was no natural generalisation of this prior to several future $Y'$ observations corresponding to different unknown $\xi$'s.

In the practical example considered by Hunter & Lamboy, the midpoint of the $100(1-\alpha)\%$ highest posterior density intervals for $\xi$ is the maximum likelihood estimate of $\xi$. This occurs for $\alpha = 0{\cdot}1$, $0{\cdot}05$ and $0{\cdot}01$ and was noted by Hill (1981). Hunter & Lamboy (1981b) stated that in the limit as $n \to \infty$, the posterior distribution of $\xi$ would concentrate at (or near) the maximum likelihood estimate. In this paper, they give a detailed response to all the discussion papers together with a Bayesian interpretation of Fieller's theorem (1954).

It will be recalled that $\hat{\xi}_c$, the classical estimator, is the maximum likelihood estimator if one makes the usual assumption of normal errors in models (1.2b) and (1.3b). Hill (1981) and Orban (1981) suggested that the Hunter & Lamboy form of prior distribution for $\eta$ (and implicitly for $\xi$) tended to provide Bayesian support for the classical estimator, $\hat{\xi}_c$, whereas Hoadley's prior distribution for $\xi$, outlined above in equation (3.1), tended to provide Bayesian support for the inverse estimator, $\hat{\xi}_I$. This suggestion has been confirmed by Brown (1982) in his paper on multivariate calibration. In particular he stated that he saw the classical solution as valid only if the distribution of $\boldsymbol{\xi}$ is thought to be rather flat and wider than the designed distribution of $\mathbf{X}$ and Hunter & Lamboy's implied prior for $\xi$ is vague and flat.

In bioassays or enzyme assays, the calibration curves (i.e. concentration response curves) are usually non-linear. Racine-Poon (1988) used a Bayesian approach to a non-linear calibration problem arising from agro-chemical bioassays. If $\xi$ represents the unknown concentration of a new sample presented at the prediction stage, the posterior distribution of $\xi$ can be calculated by using an efficient numerical integration technique such as Naylor & Smith's (1982). Racine-Poon pointed out that in practice, when concentrations of hundreds of samples have to be determined routinely, the numerical effort can be prohibitive and therefore proposed an approximation method to reduce the calculation. The combination of two or more bioassays is also discussed.

Smith & Corbett (1987) applied Bayesian and maximum likelihood methods to the estimation of the length of a marathon course. The data came from a detailed report on the course measurement of the 1984 Olympic Marathon. Apart from considering a multivariate linear regression model, they consider a dynamic model where the calibration constants change with time. Finally, Spezzaferri (1985) used a Bayesian approach to develop a method of choosing among $K$ different multivariate calibration experiments associated with $K$ different instruments.

## 4 Multiple Use of the Calibration Curve

Rosenblatt & Spiegelman (1981) in their discussion on calibration consider it important to distinguish the case where the calibration curve is used only once and the case where it is used repeatedly and interval estimates are reported for a series of determinations. In bioassay a *standard curve* is constructed on which all future assays (calibrations) are to be run.

If we refer back to the chemist's problem in § 1.1, suppose we are presented not with one new sample at the second stage of the calibration process but with $K$ new samples and suppose we make measurements $Y_1^*, Y_2^*, Y_3^*, \ldots, Y_K^*$, one on each sample, and we require interval estimates for the unknown levels of chemical A $(\xi_1, \xi_2, \xi_3, \ldots, \xi_K)$ in

each of the $K$ samples presented. This problem involves multiple use of the calibration curve which has been derived at the first stage of the calibration problem.

The problem was first treated by Mandel (1958) and another solution was given by Miller (1966). When $K$ is unknown and possibly arbitrarily large, the results of Mandel and Miller do not apply. Lieberman et al. (1967) considered the problem of $K$ unknown or very large. They constructed simultaneous confidence intervals called *unlimited simultaneous discrimination* intervals. (The classical discriminant problem *could* be regarded as a special case of calibration with $X$ taking one of a finite number of values. However, here we are using calibration to mean that $X$ is a continuous variable so it is possibly confusing to use the word *discrimination*). Lieberman et al. (1967) assumed a linear model as given in equation (1.2b) where the $\varepsilon_i$ are independent $N(0, \sigma^2)$. For each set of constructed intervals for $\xi_1, \xi_2, \xi_3, \ldots, \xi_K$, there are two probabilities $P$ and $\alpha$. This is because there are two sources of uncertainty, there is the uncertainty associated with the outcome of the calibration experiment and there is uncertainty that can be attributed to errors in the future measurements $Y_1^*, Y_2^*, \ldots, Y_K^*$. Consider for example the case of $P = 0{\cdot}90$ and $\alpha = 0{\cdot}05$. Suppose one performs a calibration experiment, estimates a regression line and then constructs $K$ intervals for $\xi_1, \xi_2, \ldots, \xi_K$ based on this regression line and measurements $Y_1^*, Y_2^*, \ldots, Y_K^*$. The $K$ intervals have the property that at least 90% of the intervals will contain their respective $\xi$'s with confidence 0·95. So if the same calibration procedure is repeated many times, producing different estimated regression lines and each time $K$ intervals for $\xi_1, \xi_2, \ldots, \xi_K$ are constructed, then for 95% of the calibration procedures at least 90% of the intervals will contain their correct $\xi$'s. For each of the remaining 5% of calibration procedures, the percentage of intervals enclosing their $\xi$'s may be greater or less than 90%. Lieberman et al. used two methods to obtain simultaneous intervals. The generally more efficient method of the two was based on application of the Bonferroni inequality to a confidence band for the regression line and a confidence interval for the unknown standard deviation $\sigma$.

In the same spirit as Lieberman et al (1967), Clark (1979) used Bonferroni inequalities to obtain simultaneous calibration intervals for $\xi_1, \xi_2, \ldots, \xi_K$ given observations $Y_1^*, Y_2^*, \ldots, Y_K^*$ at the prediction stage and the model

$$Y_j^* = F(\xi_j) + \varepsilon_j' \quad j = 1, 2, \ldots, K.$$

It was assumed that the $\varepsilon_j'$ were independent $N(0, \sigma_j^2)\, j = 1, 2, \ldots, K$ and $F$ was some smooth but unknown function, estimated by estimators of the form (see § 5)

$$\hat{F}(x) = \mathbf{u}(x)^{\mathrm{T}}\mathbf{Y}.$$

Scheffé (1973) greatly extended this approach to calibration via simultaneous intervals. Associated with these simultaneous intervals there are two probabilities $\alpha$ and $\delta$. By the Scheffé procedure the probability is $\geqslant 1 - \delta$ that at least $(1 - \alpha)$ of the $K$ intervals constructed will contain the correct $\xi$'s (i.e. $\xi_1, \xi_2, \xi_3, \ldots, \xi_K$). Lechner et al. (1982) reviews Scheffé's procedure in some detail. Scheffé explained the use of his procedure in detail for a linear calibration curve and compared his intervals with the Bonferroni intervals of Lieberman et al. (1967).

In 1973, Oden also wrote a paper on simultaneous confidence intervals in inverse linear regression. Like Lieberman et al. (1967) and Scheffé (1973) his treatment involved the use of two specified probabilities which he called $p$ and $\delta$. Carroll & Spiegelman (1988) put forward a method based on a modification of the Scheffé (1973) confidence statements which they maintain is easier to implement than Scheffé's and generally leads to shorter intervals. They applied their method to a calibration problem in atomic absorption spectroscopy and compared their results with those obtained using Scheffé's procedure.

Scheffé's approach to calibration via tolerance regions has been criticised by Lindley (1972). Lindley argues that they violate the likelihood principle and are therefore not acceptable. He commented that such an approach produces tortuous statements which are difficult to comprehend and are replaced in the Bayesian analysis, by a single statement that is much simpler. The Bayesian approach involves predictive distributions.

Brown (1982) is in accord with Lindley's views. In response to the discussion of his paper on multivariate calibration, he commented that an interval for an unknown $\xi$ corresponding to an observed $\mathbf{Y}'$ at the second stage of the calibration process should not depend on other as yet unknown and unobserved $(\xi, \mathbf{Y}')$. A set of observed $\mathbf{Y}'$ at different unknown $\xi$ provide information on the distribution of future $\xi$. As soon as two or more $\mathbf{Y}'$ are observed corresponding to different but unknown $\xi$, it is possible to update the posterior distribution of $\xi$ (as defined by Hoadley (1970) in the univariate case) by using Bayes theorem. The updating involves calculating posterior predictive distributions. Obviously if one rejects Scheffé's approach to calibration via simultaneous calibration intervals for the unknown $\xi$'s, one must update the calibration (either continuously or at regular intervals as is practically expedient) as $Y'$ are observed corresponding to different $\xi$'s. Copas (1982) put forward a method for obtaining an updated point estimate of $\xi$, in random calibration, given a set of observed $Y'$ corresponding to different $\xi$'s. However, it should be pointed out that sometimes updating is not always possible.

## 5 Non-Parametric Approaches to Calibration

In recent years at least seven papers have taken a non-parametric approach to calibration. Three of these papers involve T. Lwin, namely Lwin & Maritz (1980), Lwin (1981), Lwin & Maritz (1982). Lwin & Maritz (1980) consider a random calibration experiment in which only the bivariate random variables $(X, Y)$ can be observed (compare Tallis, 1969). They derived the distribution of $X$ given $Y$ basing estimation of the marginal distribution of $X$ on the sample distribution function. For point estimation, if $f(y_i \mid x_i, \boldsymbol{\beta})$ is the probability density function of $Y_i$ conditional on $X_i = x_i$ $(i = 1, 2, \ldots, n)$ then the predictor of $X$ when $Y = Y'$ is observed (at the prediction stage) is $\sum w_i x_i$, where

$$w_i = f(Y' \mid x_i, \boldsymbol{\beta}) \Big/ \sum_{j=1}^{n} f(Y' \mid x_j, \boldsymbol{\beta}).$$

Their *non-linear* predictor as they call it is a special member of the following class of predictors:

$$\hat{\xi}(Y') = \sum_{i=1}^{n} x_i w_i(Y', \mathbf{z})$$

where $w_i(Y', \mathbf{z})$ are *weights* attached to the $x_i$'s. This class has been studied in detail in terms of consistency properties by Stone (1977). Comparing this with the inverse estimator, $\hat{\xi}_I$, the latter is a predictor linear in $Y'$ whereas Lwin & Maritz do not restrict themselves to the class of predictors linear in the observation $Y'$. It has the advantage that it is applicable for non-normal error distributions; the error distribution need only be a member of the location and scale family. It is also very flexible as the calibration curve $m(X, \boldsymbol{\theta})$ can have many forms, not just linear in $X$ or linear in $\boldsymbol{\theta}$.

Lwin & Maritz (1982) compared the classical and inverse estimators $\hat{\xi}_c$ and $\hat{\xi}_I$ respectively by using a compound estimation approach (see § 2.1). With this approach, $\hat{\xi}_c$, is a linear compound estimator satisfying the criterion of asymptotic unbiasedness whilst $\hat{\xi}_I$ is a linear compound estimator without the unbiasedness constraint. Their

formulation required no specific distributional assumptions. Their approach showed that $\hat{\xi}_I$ was superior to $\hat{\xi}_c$ only if the current $\xi$ value was sampled from the same population as previous $X$ values i.e. $x_1, x_2, \ldots, x_n$ of the calibration experiment. This paper should be compared with Brown (1979) who obtained an optimal linear predictor $\lambda_0 + \lambda Y$, where the $\lambda_0$ and $\lambda$ were chosen to minimise the integrated mean square error rather than the compound error as with Lwin & Maritz (1982). The integrated mean square error is defined over a range $(L, U)$ with respect to a weight function $w(x)$ as

$$\text{IMSE} = \int_U^L \text{MSE}(x) w(x)\, dx.$$

Clark has produced two papers on calibration with particular reference to radiocarbon dating (1979, 1980). His 1980 paper assumes the model

$$Y_{ij} = F(x_i) + e_{ij} \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m_i, \quad \sum m_i = N,$$

where $\{x_i\}$ are known constants, $F$ is some unknown but *smooth* function and $\{e_{ij}\}$ are uncorrelated random variables with zero mean but constant variance $\sigma^2$. The $\{x_i\}$ are assumed to be distinct with $x_1 < x_2 \ldots < x_n$. The estimate of $F$ was chosen by the method of cross-validation (Stone, 1974) from a class of estimators defined by

$$\hat{F}(x) = \mathbf{u}(x)^{\mathrm{T}} \mathbf{Y}, \tag{5.1}$$

where $\mathbf{u}(x) = (u_1(x), u_2(x), \ldots, u_n(x))^{\mathrm{T}}$ is a $n \times 1$ vector of known functions, possibly indexed by some index parameter $\beta$, $Y_{i\cdot} = \sum_j Y_{ij}/m_i$ $(i = 1, 2, \ldots, n)$ and $\mathbf{Y} = (Y_{1\cdot}, Y_{2\cdot}, Y_{3\cdot}, \ldots, Y_{n\cdot})^{\mathrm{T}}$. This class of estimators includes as special cases least-squares polynomials and splines, approximating and interpolating spines and the CS-estimators (Convolution-Smoothing). The CS-estimators are of particular interest to Clark and the generalised CS-estimator is defined as

$$\hat{F}(x) = \sum_{i=1}^{n} u_i(x) Y_{i\cdot},$$

where

$$u_i(x) = \int_{-\infty}^{\infty} v_i(t) \frac{1}{b(x)} W\left(\frac{x-t}{b(x)}\right) dt.$$

Here $v_i(.)$ is the piecewise linear natural spline interpolating data $\{(x_i, Y_{i\cdot})\}$, $W$ is an arbitrary density function and $b$, the bandwidth, is a suitable bounded non-negative function. The cross-validation algorithm used is defined as follows; let $P_1, P_2, \ldots, P_k$ be a partition of the index set $U = \{1, 2, \ldots, n\}$ and let $n_j$ be the number of elements in $P_j$. Corresponding to each subset $P_j$, one conceptually sub-divides the complete sample into an *estimation sample* $\{(x_i, Y_{i\cdot}) i \in \bar{P}_j\}$ and a *validation sample* $\{(x_i, Y_{i\cdot}) i \in P_j\}$ and for various trial values of $\beta$ computes

$$C_j(\beta) = \frac{1}{n_j} \sum_{i \in P_j} w_i \{Y_{i\cdot} - f^*(x_i \mid j, \beta)\}^2,$$

where $f^*(. \mid j, \beta)$ denotes the estimator defined by $\beta$ computed from the $j$th estimation sample *only*, and $w_1, w_2, \ldots, w_n$ are suitable non-negative weights. The cross-validation mean-square error (CVMSE) is defined as

$$\bar{C}(\beta) = \frac{1}{k} \sum_j C_j(\beta).$$

The 'best' choice of $\beta$ is then taken to be that minimising $\bar{C}(\beta)$. This CVMSE is also a device for estimating the average bias of the chosen estimator and this is discussed in detail by Clark (1980). Whereas Clark considered the construction of simultaneous calibration intervals in his 1979 paper under the assumption that the estimator $\hat{F}$ (as defined by equation (5.1)) has negligible bias, in his 1980 paper he considered the *equivalent* problem of construction of prediction intervals (i.e. intervals for $Y$ given $X = x$) under the assumption that $\hat{F}$ has a non-negligible bias. The CVMSE was used to estimate this bias and an adjustment proposed to both prediction and confidence intervals to compensate, to some extent, for the bias in $\hat{F}$.

Lechner et al. (1982) combined Scheffé's (1973) calibration approach with linear splines to produce simultaneous calibration intervals for the liquid volume ($v$) in large nuclear material processing tanks, given the differential pressure ($p$). The authors assumed the upper part of the tank to be composed of $(k+1)$ distinct and known regions. By considering the cross-sectional area, $A(x)$, at a given height $x$, and its relationship to $p$ and $v$, the authors concluded that within each of the $(k+1)$ regions, it was reasonable to assume a linear relationship between $p$ and $v$. The pressure-volume relationship could therefore be modelled as a linear spline with $k$ knots (Eubank, 1988). The locations of the knots between line segments were determined empirically and confirmed by engineering analysis. Since recalibration would not be feasible after the tank was in use at a nuclear materials processing plant, the Scheffé procedure was used for describing the uncertainty of volume measurements. The authors chose to use a $B$-spline basis made up of linear $B$-splines (De Boor, 1978). If $\{v_i\}$ $i = 1, \ldots, k$ are the interior knot locations, then the $B$-spline basis consists of the functions $\beta_i(v)$, where $\beta_i(v)$ is the piecewise linear natural spline interpolating the points $\{(v_j, \delta_{ij})\, j = 1, 2, \ldots, k\}$ where $\delta_{ij} = 1$ if $i = j$, or 0 if $i \neq j$. The authors applied their theoretical results to 172 observations from calibration runs on a processing tank and compared their results with those obtained using propagation-of-error techniques (Naszodi, 1978).

Knafl et al. (1984) also considered pressure-volume calibration of a nuclear tank and adopted Scheffé's procedure (1973), i.e. resulting calibration intervals had two associated probabilities $\alpha$ and $\delta$. Their non-parametric approach is more general than Lechner et al. (1982). They proposed a model which states that at each point $x \in [S_0, S_1]$ and for any real $t$, $m_1$, $m_2$ $(m_1 < m_2)$ and real $f$, the unknown calibration curve $f$ can be written as $f(t) = f(x) + r(t, x)$ with

$$m_1(t-x) \leq r(t, x) \leq m_2(t-x) \quad \text{if } t > x,$$

$$m_2(t-x) \leq r(t, x) \leq m_1(t-x) \quad \text{if } t < x,$$

where $m_1$ and $m_2$ are specified. If $m_1 > 0$ this model implies that $f$ is increasing at a rate that is bounded above and below. This does not require differentiability of $f$ and thus reflects the abrupt change in the tank. Let

$$\eta_i = Y_i - \left(\frac{m_1 + m_2}{2}\right)(x_i - x) \quad i = 1, 2, \ldots, n,$$

$$\rho(x_i, x) = r(x_i, x) - \left(\frac{m_1 + m_2}{2}\right)(x_i - x) \quad i = 1, 2, \ldots, n.$$

The model for the calibration experiment then becomes

$$\eta_i = f(x) + \rho(x_i, x) + \sigma\varepsilon_i \quad i = 1, 2, \ldots, n.$$

Knafl et al. sought a linear estimate $\sum c_i(x)\eta_i$ of $f(x)$. The $c_i$ were chosen to minimise the maximum M.S.E. subject to $\sum c_i(x) = 1$, which ensures that the bias term of the M.S.E. is

bounded. It was noted that the estimate of $f(x)$ i.e. $\sum \hat{c}_i(x)\eta_i$ depends only on averages of $\eta$'s at common values of the $x_i$'s. Thus replicated data (repeated $x_i$'s) provided an obvious estimate $\hat{\sigma}^2$ of $\sigma^2$ based on the replicated data only. It was further assumed that $\varepsilon_i \sim N(0, 1)$ thus $\hat{f}(x)$, conditioned on $\hat{\sigma}$, is normal with mean $f(x) + \sum \hat{c}_i(x)\rho(x_i, x)$ and variance $\hat{\sigma}^2 \sum \hat{c}_i^2(x)$. The upper and lower calibration curves of Scheffé's method are given by

$$\hat{U}(x) = \hat{f}(x) + B(x) + \hat{a}D(x) + \hat{q}\hat{\sigma},$$
$$\hat{L}(x) = \hat{f}(x) - B(x) - \hat{a}D(x) - \hat{q}\hat{\sigma},$$

where $B(x)$ is the maximum bias of $\hat{f}(x)$ and $D(x)$ is the standard deviation of $\hat{f}(x)$. Although the calculation of $\hat{f}(x)$ and $\hat{c}(x)$ is an easy process on a computer, the calculation of $\hat{a}$ and $\hat{q}$ is long and complicated involving evaluation of double integrals and complicated equations involving integrals. The authors gave a detailed seven-point procedure for the derivation of the calibration intervals as well as an extension of the method to suit *smoother* $f$'s. If $Y^*$ is a new observation corresponding to an unknown $\xi$, the calibration interval for $\xi$ is given by

$$\hat{U}^{-1}(Y^*) \leqslant \xi \leqslant \hat{L}^{-1}(Y^*).$$

Scheffé's simultaneous intervals are rather conservative (i.e. wide) because they arise from a joint inferential approach to calibration. The more general approach of Knafl et al. (1984) produces simultaneous intervals which are wider than those of Lechner et al. (1982) and which are more complicated to calculate, so it must be very questionable whether it is worthwhile adopting a more general model.

Osborne (1990) developed a variety of Bayesian non-parametric approaches to calibration. It will be noted from equation (1.3a) that $\xi = f^{-1}(\eta)$. Two of Osborne's approaches viewed $\xi$ as a non-linear functional ($f^{-1}(\cdot)$) of the calibration curve $f$ where $f$ was assumed to be a monotonic natural cubic spline with knots $\{x_i\}$ $i = 1, 2, \ldots, n$. Assume the errors $\varepsilon_i$ and $\varepsilon_j'$ in models (1.2a) and (1.3a) are independent $N(0, \sigma^2/w_i)$ and $N(0, \sigma^2)$ random variables respectively (the weights $w_i$ assumed known). Silverman (1985) showed that if one assumes $f$ to be a natural cubic spline with knots at the data points $\{x_i\}$, the posterior distribution of $f$ is a multivariate normal distribution and the posterior mean is a natural cubic smoothing spline with knots $\{x_i\}$. For the case of $f$ being a *monotonic* cubic spline, the posterior distribution of $f$ is a truncated multivariate normal distribution, truncated so as to remove any non-monotonic splines. Osborne obtained the posterior distribution of $\xi$ by firstly simulating from the posterior distribution of $f$, then simulating from the posterior distribution of $\eta$ and finally calculating values of $f^{-1}(\eta)$.

Let us consider one of these approaches in a little more detail. Suppose a posterior realisation of $f$ is denoted by $\tilde{f}_s$ and a posterior realisation of $\eta$ is denoted by $\eta_v$. Suppose also that $\xi$ has a prior density $\pi(\xi)$ and that $\pi(\xi)$ is unimodal within the calibration range $[x_1, x_n]$. Let

$$\pi_{\text{MAX}} = \operatorname*{Max}_{x_1 \leqslant \xi \leqslant x_n} \pi(\xi).$$

Osborne showed that if $p(\eta \mid \mathbf{Y}')$ is the posterior density of $\eta$ corresponding to the prior density $\pi(\xi)$ for $\xi$, then realisations of $p(\eta \mid \mathbf{Y}')$ can be obtained by:

(i) simulating $\eta_v^*$ from the posterior density $p_u(\eta \mid \mathbf{Y}')$, which is the posterior density of $\eta$ corresponding to a uniform prior density for $\xi$;

(ii) accepting $\eta_v^*$ as a realisation of a random variable with density function $p(\eta \mid \mathbf{Y}')$ iff

$$U\pi_{\text{MAX}} \leqslant \pi(\tilde{f}_s^{-1}(\eta_v^*)),$$

where $U$ is an independent $U(0, 1)$ random variable.

A switching algorithm was used to simulate $\eta_v^*$ from the posterior density $p_u(\eta \mid \mathbf{Y}')$. The posterior distribution of $\xi$ is obtained by calculating $\tilde{f}_s^{-1}(\eta_v^*)$ for as many values of $s$ and $v$ as are required (a *different* set of values $\eta_v^*$ being simulated for *each* $\tilde{f}_s$). Osborne tested out all her non-parametric methods on a variety of simulated and real data sets.

## 6 Multivariate Calibration

Naes & Martens (1984) state that multivariate calibration is a young discipline and Brown's (1982) paper contributed a great deal to this discipline in its early stages. Brown's (1982) paper is an important paper on multivariate calibration not only because of its contents but because it stimulated great interest in this area of calibration.

Brown (1982) adopted both a *classical* approach akin to that of Eisenhart (1939), Williams (1959) and Fieller (1954) and a *Bayesian approach* akin to that of Hoadley (1970). He assumed for both the calibration experiment and the prediction experiment (i.e. the second stage of the calibration process) multivariate linear models. Suppose there are $n$ observations in the calibration experiment, $q$ response variables $Y_1, Y_2, \ldots, Y_q$ and $p$ explanatory variables $X_1, X_2, \ldots, X_p$ and $q \geqslant p$.

$$Y = \mathbf{1}\boldsymbol{\alpha}^{\mathrm{T}} + XB + E, \tag{6.1}$$

where $Y$ $(n \times q)$ and $E$ $(n \times q)$ are random matrices, $X$ is a $n \times p$ matrix of fixed constants and $\mathbf{1}$ is a $(n \times 1)$ vector of units. $B$ is a $p \times q$ matrix of unknown parameters, $\boldsymbol{\alpha}$ is a $q \times 1$ vector of unknown parameters. The model for the prediction experiment is given by

$$Y' = \mathbf{1}\boldsymbol{\alpha}^{\mathrm{T}} + \mathbf{1}\boldsymbol{\xi}^{\mathrm{T}}B + E', \tag{6.2}$$

where $Y'$ $(m \times q)$, $E'$ $(m \times q)$ are random matrices, $\boldsymbol{\xi}$ is a $p \times 1$ vector of unknowns and $\mathbf{1}$ is a $(m \times 1)$ vector of units. It is wished to draw inferences about $\boldsymbol{\xi}$. As defined the calibration experiment is a controlled calibration experiment. If $\mathbf{e}_i$ is the $i$th row of $E$, it is assumed that $E(\mathbf{e}_i) = 0$, $E(\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}) = \Gamma$ and $\mathbf{e}_i \sim N(\mathbf{0}, \Gamma)$ for $i = 1, 2, \ldots, n$. If $\mathbf{e}_j'$ is the $j$th row of $E'$, the $\mathbf{e}_j'$ satisfy the above also and it is assumed that they are independent of the $\mathbf{e}_i$. Brown pointed out that $\mathbf{X}$ might consist of $p$ variables derived from a smaller set as in polynomial regression. In this case, $\boldsymbol{\xi}$ is a vector function of the same reduced number of unknowns.

Let $S$ be the $(q \times q)$ residual sum of products matrix, pooled from the calibration and prediction experiments when $m > 1$. Further let

$$\sigma^2(\boldsymbol{\xi}) = \frac{1}{m} + \frac{1}{n} + \boldsymbol{\xi}^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\boldsymbol{\xi}.$$

Brown showed by using *classical* arguments that a $100(1-\gamma)\%$ confidence region for $\boldsymbol{\xi}$ is all $\boldsymbol{\xi}$ such that

$$(\bar{\mathbf{Y}}' - \hat{\boldsymbol{\alpha}} - \hat{B}^{\mathrm{T}}\boldsymbol{\xi})^{\mathrm{T}}S^{-1}(\bar{\mathbf{Y}}' - \hat{\boldsymbol{\alpha}} - \hat{B}^{\mathrm{T}}\boldsymbol{\xi})/\sigma^2(\boldsymbol{\xi}) \leqslant \frac{q}{\nu}F_{\gamma,q,\nu}, \tag{6.3}$$

where $F_{\gamma,q,\nu}$ is the upper $100(1-\gamma)\%$ point of the standard $F$ distribution on $q$ and $\nu$ degrees of freedom $(\nu = n - p + m - q - 1)$. This is the multivariate form of the confidence region obtained by Fieller (1954) and others who considered $p = q = 1$. In standard multivariate linear regression the region corresponds to fiducial limits of Williams (1959). When $p = q$ the above inequality can be written as

$$\|\boldsymbol{\xi} - C^{-1}D\|_C^2 - \text{constant} \leqslant 0,$$

where

$$C = \hat{B}S^{-1}\hat{B}^{\mathrm{T}} - k(X^{\mathrm{T}}X)^{-1} = \hat{B}S^{-1}\hat{B}^{\mathrm{T}} - \left(\frac{q}{\nu}\right)F_{\gamma,q,\nu}(X^{\mathrm{T}}X)^{-1}$$

and $C$ is assumed positive definite. So the confidence region defined by inequality (6.3) is a closed ellipsoid. As stated in § 2.1, in the univariate case (i.e. $p = q = 1$) the confidence region sometimes degenerates into two disjoint semi-infinite lines or the whole real line. This occurs if

$$\frac{\hat{\beta}_1^2 S_{xx}}{\hat{\sigma}^2} < F_{\gamma,1,n+m-3}.$$

The condition that $C$ is positive definite is the direct multivariate extension of the univariate condition. If one lets $\gamma \to 1$ in inequality (6.3), then $k \to 0$ and this condition on $C$ will be satisfied if $B \neq 0$. The resulting confidence region degenerates to the point

$$\hat{\boldsymbol{\xi}} = (\hat{B}S^{-1}\hat{B}^{\mathrm{T}})^{-1}\hat{B}S^{-1}(\bar{\mathbf{Y}}' - \hat{\boldsymbol{\alpha}}). \tag{6.4}$$

If $p = q = 1$, this gives the classical estimator $\hat{\xi}_c$ defined in § 2.1. As equation (6.4) arises when considering the regression of $\mathbf{Y}$ on $\mathbf{X}$, $\hat{\boldsymbol{\xi}}$ is usually called the classical estimator of $\boldsymbol{\xi}$. Brown & Sundberg (1987) have shown that for $q = p$, the estimator $\hat{\boldsymbol{\xi}}$ defined by equation (6.4) is the maximum likelihood estimator of $\boldsymbol{\xi}$. When $q > p$, however, it is not the maximum likelihood estimator, the latter being shifted from $\hat{\boldsymbol{\xi}}$ by an amount depending on the inconsistency diagnostic

$$R = (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\hat{\boldsymbol{\xi}})^{\mathrm{T}}S^{-1}(\mathbf{Y}' - \hat{B}^{\mathrm{T}}\hat{\boldsymbol{\xi}}). \tag{6.5}$$

The inconsistency diagnostic $R$ measures the relative lack of consistency of a $q$-variate observation $\mathbf{Y}'$.

Lieftinck-Koeijers (1988) considered the classical estimator given by equation (6.4) in the case of $p = 1$ and $\Gamma$ known. He showed that this estimator has a finite mean if $q > 2$ and a finite M.S.E. if $q > 4$. He gave exact expressions for the mean and M.S.E. in terms of expectations of Poisson variables. Nishii & Krishnaiah (1988) also considered the classical estimator and showed that when $\Gamma$ is unknown, the mean of $\hat{\boldsymbol{\xi}}$ is finite iff $q \geq p + 1$ and the M.S.E. is finite iff $q \geq p + 2$. They give the exact moments of $\hat{\boldsymbol{\xi}}$ using expectations of Poisson variables and compare their results with those of Lieftinck-Koeijers.

Oman (1988) considered the case of where $\boldsymbol{\xi}$ in model (6.2) is given by $\boldsymbol{\xi}^{\mathrm{T}} = [h_1(\boldsymbol{\zeta}), h_2(\boldsymbol{\zeta}), \ldots, h_p(\boldsymbol{\zeta})]$ for *known* functions $h_j$ (e.g. squares or logarithms of components of $\boldsymbol{\zeta}$) and one wishes to construct a confidence region for the *unknown* $\boldsymbol{\xi}$ corresponding to a future $\mathbf{Y}'$. One approach to the problem is as follows; obtain a $100(1-\gamma)\%$ confidence region for $\boldsymbol{\xi}$ by using expression (6.3). Suppose this region is called $R$. The confidence region $C$ for $\boldsymbol{\zeta}$ is then given by $C = \{\boldsymbol{\zeta} : \boldsymbol{\xi}(\boldsymbol{\zeta}) \in R\}$. Oman points out that a disadvantage of this approach is that although the region $R$ might be nicely behaved, the region $C$ need not be. He put forward an alternative confidence region for $\boldsymbol{\zeta}$ and compared his method with asymptotic results obtained by Fujikoshi & Nishii (1984) and Brown & Sundberg (1987). An application to the estimation of gestational age using ultrasound foetal bone measurements is given in the paper.

Brown's Bayesian approach was similar to that of Hoadley. Hoadley (1970) assumed a general form of prior density $p(\beta_0, \beta_1, \sigma^2, \xi) \propto p(\beta_0, \beta_1, \sigma^2)p(\xi)$. Brown assumed

$$\pi(B, \boldsymbol{\alpha}, \Gamma, \boldsymbol{\xi}) = \pi(B, \boldsymbol{\alpha}, \Gamma)\pi(\boldsymbol{\xi}).$$

Brown also assumed $\pi(\boldsymbol{\xi} \mid X) = \pi(\boldsymbol{\xi})$ i.e. the controlled $X$ values provide no information on $\xi$ and he assumed a Jeffrey's invariant prior

$$\pi(B, \boldsymbol{\alpha}, \Gamma) \propto |\Gamma|^{-(q+1)},$$

which is the multivariate analogue of Hoadley's non-informative prior for $(\beta_0, \beta_1, \sigma^2)$ i.e.

$p(\beta_0, \beta_1, \sigma^2) \propto \sigma^{-2}$. Brown obtained multivariate extensions to both of Hoadley's theorems which provide extra insight into Hoadley's results.

Brown derived a multivariate generalisation ($\hat{\boldsymbol{\xi}}$) of the Krutchkoff inverse estimator by regressing $\mathbf{X}$ on $\mathbf{Y}$, which he compared with the classical estimator $\hat{\boldsymbol{\xi}}$ given by equation (6.4). Brown's multivariate methods, including one which is a multivariate extension of Lwin & Maritz (1980) are compared on data from a random calibration experiment and data from a controlled calibration experiment. Some important conclusions are drawn. In particular, Brown suggested that it was a good idea to consider each of the $p$ characteristics, $X_1, X_2, X_3, \ldots, X_p$, one at time, forgetting the existence of the other $(p-1)$ characteristics. Sundberg (1985) justifies this in some circumstances. Brown's paper is a discussion paper and there is a considerable amount of discussion of his results with further references.

A recent paper by Brown & Sundberg (1987) has considered an approach to multivariate calibration which involves the profile or maximum relative likelihood (Kalbfleisch & Sprott, 1970). Suppose that the model for the calibration experiment is given by model (6.1) i.e.

$$Y = \mathbf{1}\boldsymbol{\alpha}^{\mathrm{T}} + XB + E.$$

If $\mathbf{e}_i$ is the $i$th row of $E$, it is assumed that $E(\mathbf{e}_i) = 0$, $E(\mathbf{e}_i\mathbf{e}_i^{\mathrm{T}}) = \Gamma$ and $\mathbf{e}_i \sim N(\mathbf{0}, \Gamma)$ for $i = 1, 2, \ldots, n$. This profile likelihood approach entails forming the maximised likelihood 'as if $\boldsymbol{\xi}$ were known' and normalising by the likelihood maximised over all unknown parameters $\boldsymbol{\alpha}$, $B$, $\Gamma$, $\boldsymbol{\xi}$. Suppose the model for the prediction stage is given by model (6.2) with $m = 1$. This approach is offered as an alternative to the Bayesian approach (Brown, 1982), avoiding the need for specification of a prior distribution. The profile likelihood is a function of $\boldsymbol{\xi}$ which has a maximum value of one at a maximum likelihood estimate of $\boldsymbol{\xi}$ and the profile log-likelihood is given (up to an additive constant) by

$$-\tfrac{1}{2}(n+1)\ln(|\hat{\Gamma}(\boldsymbol{\xi})|), \tag{6.6}$$

where

$$|(n+1)\hat{\Gamma}(\boldsymbol{\xi})| \propto \frac{n_0(n+1)}{n} + \frac{[R + (\boldsymbol{\xi} - \hat{\boldsymbol{\xi}})^{\mathrm{T}}\hat{B}S^{-1}\hat{B}^{\mathrm{T}}(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}})]}{r(\boldsymbol{\xi})},$$

with

$$r(\boldsymbol{\xi}) = 1 + \frac{n}{(n+1)}\boldsymbol{\xi}^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\boldsymbol{\xi}$$

and $R$ the inconsistency diagnostic given in equation (6.5), $\hat{B}$ the least squares estimator of $B$ and $S$ the residual sum of products matrix from the calibration experiment.

The profile likelihood for $m \geqslant 1$ is given by

$$C\left(\frac{\sigma^2(\boldsymbol{\xi})}{\sigma^2(\boldsymbol{\xi}) + (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\boldsymbol{\xi})^{\mathrm{T}}S^{-1}(\mathbf{Y}' - \hat{B}^{\mathrm{T}}\boldsymbol{\xi})}\right)^{\frac{1}{2}(n+1)}, \tag{6.7}$$

where

$$\sigma^2(\boldsymbol{\xi}) = \frac{1}{m} + \frac{1}{n} + \boldsymbol{\xi}^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}\boldsymbol{\xi}$$

and $S$ is the residual sum of products matrix, pooled from the calibration experiment and the prediction stage. The normalising constant is such that $C^{-1}$ is the value of expression (6.7) at its maximum. The Bayes posterior density, $\pi(\boldsymbol{\xi} \mid \mathbf{Y}', \mathbf{z})$, obtained when using a

vague or non-informative prior for $\xi$ is given by

$$\pi(\xi \mid \mathbf{Y}', \mathbf{z}) \propto \frac{[\sigma^2(\xi)]^{\frac{1}{2}\nu}}{[\sigma^2(\xi) + (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\xi)^{\mathrm{T}} S^{-1} (\mathbf{Y}' - \hat{B}^{\mathrm{T}}\xi)]^{\frac{1}{2}(\nu+q)}},$$

where $\nu = n - p + m - q - 1$. If one compares the profile likelihood and the posterior density, one can see that they are closely related. The powers of numerator and denominator of Bayes posterior are $\frac{1}{2}\nu$ and $\frac{1}{2}(\nu + q)$ compared with $\frac{1}{2}(n+1)$ and $\frac{1}{2}(n+1)$ in the profile likelihood. Brown & Sundberg obtained likelihood-based confidence regions for $\xi$ which have the intuitively desirable property of expansion with increasing values of $R$. This should be contrasted with the unnatural behaviour of the classical confidence regions (see inequality (6.3)) when $q > p$, which expand for decreasing $R$ and get narrower with increasing $R$ sometimes shrinking to a point (see Oman & Wax, 1984, for a practical example of this).

Brown & Sundberg (1989) examine the case of there being more variables than observations ($n < p + q + 1$) when assuming a standard multivariate linear regression model given by equation (6.1) with no derived variables. They considered the case of $\mathbf{X}_i$ being regarded as fixed (controlled calibration) and $\mathbf{X}_i$ being regarded as random (random calibration). They showed that if $n > q$, the generalised least squares estimator of $\xi$ and the estimated best linear predictor ($\hat{\xi}$) are both unique. By way of an example, they used the NIR (near infra-red) data of Fearn (1983).

There are a wide variety of approaches to multivariate calibration. Naes et al. (1986) compare the multiple linear regression (MLR), ridge regression (RR), principal component regression (PCR) and partial least squares regression (PLSR) approaches with particular reference to the calibration of near infra-red (NIR) instruments. In a NIR calibration problem, the instrument response $\mathbf{Y} = (Y_1 Y_2 \ldots Y_q)^{\mathrm{T}}$ is called the spectrum. In the NIR problem considered by Fearn (1983), $q = 6$ and $Y_1, Y_2, \ldots, Y_6$ are measurements of the reflectance of NIR radiation at six different wavelengths. There is only one dependent variable $X$ which is the protein content of wheat samples. The NIR data matrix is highly multicollinear and the data set have been used by several authors since 1983 (Farebrother, 1984; Hoerl et al. 1985; Naes et al. 1986; Brown & Sundberg, 1989). With reference to the PCR and PLSR methods, both methods project the NIR spectrum into a space determined by vectors and use coordinates in this space as regressors with $\mathbf{X}$ as regressand. However, while PCR projects into a space estimated by the spectral variables $Y_1, Y_2, \ldots, Y_q$, PLSR projects into a space determined by both spectrum and $\mathbf{X}$. According to Naes et al. (1986) PLSR is a serious competitor to PCR and RR for multicollinear data.

Ridge regression (Hoerl et al. 1970) is particularly recommended for cases in which the explanatory variables have high intercorrelations. Fearn (1983) compared the MLR and RR approaches. The paper provoked quite a bit of discussion about whether prediction methods such as RR should be used in the calibration of NIR instruments (Farebrother, 1984; Hoerl et al. 1985). Naes et al. (1986) performed computations on Fearn's data set and found that when the ratio of the number of calibration samples and the number of wavelengths in the NIR spectrum is low i.e. $n/q$ is small, the RR, PLSR and PCR methods which are biased regression methods, gave much better results than MLR.

Naes (1985a) compared the generalised least squares estimator of $\xi$, $\hat{\xi}_{\mathrm{GLS}}$ and the best linear predictor of $\xi$, $\hat{\xi}_{\mathrm{BLP}}$ which arise from the controlled and random calibration situations respectively. In this paper he assumed that Beer's model held; this states that the spectrum, $\mathbf{Y}$, is a linear function of the concentrations $(X_1, X_2, X_3, \ldots, X_p)$ of one or more chemical constituents ($p$ constituents). If the parameters of the model in both the controlled and random situations are replaced by estimates, let $\hat{\xi}_{\mathrm{GLS}}$ and $\hat{\xi}_{\mathrm{BLP}}$ be the

resulting estimators. If the estimates of the parameters of the model are maximum likelihood estimates, $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ coincide with the estimators $\hat{\xi}$ and $\hat{\hat{\xi}}$ in Brown (1982). Naes showed that, if the estimates of the parameters of the model were consistent, then $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ behave more and more like $\hat{\hat{\xi}}_{GLS}$ and $\hat{\hat{\xi}}_{BLP}$ respectively as $n \to \infty$. He showed that the set of $\xi$'s where $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ have the same M.S.E. is an ellipsoid given by

$$F = \{\xi : E[(\hat{\xi}_{BLP} - \xi)^T(\hat{\xi}_{BLP} - \xi) \mid \xi] = \operatorname{tr}(B^T\Gamma^{-1}B)^{-1}\}$$

using the notation of model (6.1). If $F'$ denotes $F$ and the region inside $F$, then if $\xi \notin F'$, $\hat{\xi}_{GLS}$ is best otherwise $\hat{\xi}_{BLP}$ is best. Finally he suggested a new predictor which is a combination of $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ involving indicator functions.

The problem of estimating $\hat{\xi}_{GLS}$ and $\hat{\xi}_{BLP}$ in the best way is treated in Naes (1985b, 1986). Methods based on different assumptions on $\Sigma$, the error covariance matrix of the multivariate linear model, are proposed and analysed in these two papers. It is known from practical experiments (for example NIR measurements) that the covariance matrix of multivariate measurements from spectrophotometers is often highly multicollinear. This means that the number of variation sources for the spectrum is smaller than the number of variables in the spectrum. It is therefore assumed in Naes (1985b) that the measurement errors have a linear factor structure, i.e.

$$\mathbf{e}_i = P\mathbf{t}_i + \mathbf{e}_i' \quad i = 1, 2, \ldots, n,$$

where $P$ is a fixed matrix and $\mathbf{e}_i$, $\mathbf{e}_i'$ and $\mathbf{t}_i$ are random vectors. These 'rank reduction' models have proved to be very suitable for multicollinear data and have given good results in NIR spectroscopy (see Naes, 1985b). The model (6.1) becomes

$$Y = \mathbf{1}\alpha^T + XB + TP + E',$$

where $X$, $Y$, $T$ and $E'$ are random matrices, $B$ and $P$ are fixed matrices (the rows of $P$ may be regarded as spectra of unknown constituents, while $B$ contains spectra of the known constituents). As this is a random calibration problem, Naes obtained expressions for the best linear predictor of $\xi$ i.e. $\hat{\xi}_{BLP}$ and the estimated best linear predictor, $\hat{\hat{\xi}}_{BLP}$. A principal component analysis is used to estimate $\Sigma$. Naes examined the properties of $\hat{\hat{\xi}}_{BLP}$ and discussed the links between

(a) $\hat{\hat{\xi}}_{BLP}$ and the predictor based on multiple linear regression of $\mathbf{X}$ on $\mathbf{Y}$,
(b) $\hat{\hat{\xi}}_{BLP}$ and the PCR predictor.

Naes pointed out that his calibration procedure using $\hat{\hat{\xi}}_{BLP}$ allows calibration of two or more constituents in a mixture simultaneously whereas when random calibration is approached by MLR, stepwise MLR or PCR, each constituent must be calibrated separately, i.e. calibration of different constituents in a mixture requires completely separate computations.

## 7 Other Approaches to Calibration

Two papers, Kalotay (1971) and Minder & Whitney (1975) approach calibration using structural inference. He gives the structural distribution of $(\beta_0, \beta_1, \beta_0 + \beta_1\xi, \sigma)$ as

$$|\beta_1|\left\{\prod_{i=1}^{n} f\left(\frac{Y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\right\}\left\{\prod_{j=1}^{m} f\left(\frac{Y_j' - \beta_0 - \beta_1\xi}{\sigma}\right)\right\} d\beta_0\, d\beta_1\, d\xi\, \frac{d\sigma}{\sigma},$$

where

$$f(x) = \frac{1}{\sqrt{(2\pi\sigma)}} \exp(-\tfrac{1}{2}x^2).$$

The structural distribution of $\xi$ is obtained by integrating over $\beta_0$, $\beta_1$ and $\sigma$ and finding the constant of proportionality. Comparing this with a Bayesian analysis, Kalotay's approach is equivalent to using a prior likelihood $p(\beta_0, \beta_1, \sigma, \xi)$ proportional to $|\beta_1|/\sigma$ and integrating over the nuisance parameters $\beta_0$, $\beta_1$ and $\sigma$. It is therefore not surprising that the posterior distribution of $\xi$ derived by Hunter & Lamboy (1981a) is mathematically identical to the structural distribution of $\xi$ derived by Kalotay. Kalotay compared his results with those of Hoadley (1970). Minder & Whitney (1975) derived a *marginal likelihood* function (MLF) for $\xi$ and then used various approximations to the RMLF (relative marginal likelihood function obtained by normalising the MLF to have a maximum value of one at the maximum likelihood estimate of $\xi$) to derive confidence regions for $\xi$. There are mathematical similarities between their MLF for $\xi$ and the profile likelihood function for $\xi$ (see Kalbfleisch & Sprott, 1970, for the latter) but Minder & Whitney maintained that despite these similarities, the two functions should lead to considerably different inferences for small $(n+m)$. Minder & Whitney comment that their MLF for $\xi$ is rather similar to the predictive density function, $L(\xi)$, of Hoadley (1970) and the approximate structural density function of Kalotay (1971).

With the approach of Dobrigal et al. (1987), the linear calibration problem is transformed so that the primary parameter is an angle, the nuisance parameter is a radial distance and the density is rotationally symmetric. They point out that were the nuisance parameter known, exact location confidence intervals could be constructed using location or structural arguments. Dobrigal et al. use a confidence distribution to average out the nuisance parameter, yielding an approximate confidence interval for $\xi$ that involves a measure of precision derived from the radial distance.

Tallis (1969) considered the problem of calibration with supplementary information. In addition to the information of a controlled calibration experiment which gives knowledge of the conditional distribution of $Y$ given $X = x$ for selected values of $x$, Tallis made extra observations on the marginal distribution of $Y$ alone. The combined information is then used to estimate the conditional distribution of $X$ given $Y$. The attainment of a viable solution depends on finding a unique solution to a Fredholm integral equation of the first kind.

Naszodi (1978) used traditional propagation-of-error techniques (i.e. estimates of the first two moments obtained from a Taylor expansion) with a bias correction to derive a new estimator of $\xi$ which is approximately unbiased, is more efficient than the classical estimator $\hat{\xi}_c$ and is consistent. His expression for the approximate bias of $\hat{\xi}_c$ agrees with that of Shukla (1972). Whereas Shukla assumed that the errors $\varepsilon_i$ in model (1.2b) were normally distributed, Naszodi assumed only that $Y$ was symmetrically distributed about $E(Y)$. Naszodi's estimator is given by

$$\bar{\xi} = \bar{x} + \frac{(Y' - \bar{Y})}{(\hat{\beta}_1 + \mathrm{Var}\,(\hat{\beta}_1)/\hat{\beta}_1)},$$

where $\hat{\beta}_1 = S_{xy}/S_{xx}$.

A two-stage sequential approach is considered in two papers by Perng & Tong (1974, 1977). In the first stage (1974 paper) the sequence $\{Y_{1i}\}$ is observed sequentially for the estimation of $\beta_0$ and $\beta_1$ the intercept and slope parameters respectively. If $\hat{\beta}_1$ is not significantly different from zero, one does not proceed to the second stage and it is concluded that the model is not suitable for estimation of $\xi$. Otherwise one proceeds to the second stage and observes $\{Y_{2j}\}$ sequentially. When the experiment terminates a fixed width confidence interval can be constructed for $\xi$.

The final approach to be considered is that of Vecchia et al. (1989). Changes in calibration curves from one time period to the next, caused by drift, often require

measuring devices to be recalibrated at frequent intervals. This, however, is not always practical. Suppose we take $T$ calibration periods and suppose the $T$ models for the calibration experiments and prediction stages are given respectively by

$$Y_{ij} = f(x_{ij}; \boldsymbol{\beta}_i) + \text{error} \quad i = 1, 2, \ldots, T, \quad j = 1, 2, \ldots, n_i,$$

$$Y'_{ik} = f(\xi_{ik}; \boldsymbol{\beta}_i) + \text{error} \quad i = 1, 2, \ldots, T, \quad k = 1, 2, \ldots, r_i.$$

Vecchia et al. considered a random coefficient regression model where the $\boldsymbol{\beta}_i$'s are regarded as random variables, varying from period to period, but assuming $f$ to be unchanging throughout. The usual practice is to estimate $\xi_{ik}$ only using data from the $i$th calibration period. Vecchia et al. showed that it was more efficient to combine the data from all calibration periods.

## 8 Design Aspects of Calibration

Papers devoted solely to the design aspects of calibration have been written by Ott & Myers (1968), Thomas & Myers (1973), Spiegelman & Studden (1980) and Buonaccorsi (1986). Ott & Myers developed the optimal designs for estimating $\xi$ based on the *classical* approach whilst Thomas & Myers (1973) studied the same problem using the *inverse* approach. The design criterion used by both Ott & Myers and by Thomas & Myers was the minimum integrated or average mean squared error. Both papers used a linear approximation when the true model was linear or quadratic. The results of Ott & Myers on the classical estimator agree with those of Shukla (1972) who briefly looked at design in his paper.

Buonaccorsi also took a *classical* approach examining the distribution of $\hat{\xi}_c$ via modes, medians and moments. He estimated the asymptotic variance of $\hat{\xi}_c$ and looked at $AV$ optimality, $V_x$ optimality and $M$ optimality; $AV$ optimality results from an average asymptotic variance being minimised with respect to some prior distribution for $\xi$ and Ott & Myer's design criterion is closest to Buonaccorsi's $AV$ optimality. Buonaccorsi presented, in one of the sections of his paper, an overview of some problems associated with multiple $\xi$'s (see § 4) and Bayesian inference (see § 3). His paper is a review paper of calibration design. Spiegelman & Studden (1980) studied the optimal design of calibration experiments considered by Lechner et al. (1982) and Knafl et al. (1984) involving nuclear accountability tanks (see § 5). Naszodi (1978) devoted a section of his paper to T-optimum designs. T-optimality is achieved by minimising the average absolute bias of $\xi$ on $[a, b]$, where $a \leqslant X_i \leqslant b$, $i = 1, 2, \ldots, n$. Buonaccorsi (1986) obtained similar results to those of Naszodi using different arguments.

## 9 Conclusion

This review of calibration has shown that there are a wide variety of statistical methods proposed for the calibration problem . Orban (1981) comments that there are at least as many philosophical issues and possibly the only way these issues can be resolved is to focus our attention on more specific problems. Hunter & Lamboy (1981b) comment that it is rare for persons familiar with important real problems to have the skill and the time to present such valuable information in understandable form. They go on to say that it would benefit all if clear voices could be persuaded to provide similar descriptions in other problem areas and that such descriptions could be published. I would like to go a stage further and suggest that it would benefit all in the calibration field if a wide variety of methods could be applied to an appropriate practical calibration problem so that these

methods could be compared and assessed. In the field of multivariate calibration such comparisons have taken place (Brown (1982), Naes (1985b), Naes et al. (1986)) but there does not seem to have been an extensive comparative study of methods in the field of univariate calibration.

This review would seem to indicate that a very high proportion of univariate calibration papers have been devoted to linear calibration. There must however be many practical problems where a linear model is inappropriate. One possible solution would be to consider optimal transformations of both $X$ and $Y$ so that $g(Y)$ was a *linear* function of $s(X)$. Breiman & Friedman (1985) proposed a powerful method called 'ACE' (Alternating Conditional Expectation) which finds transformations $g(Y)$ and $s(X)$ that maximise the correlation $\text{corr}(g(Y), s(X))$ subject to $\text{Var}(g(Y)) = 1$. Tibshirani (1988) proposed as powerful a method which uses a non-parametric variance stabilising transformation for the response variable and according to the author the technique alleviates many of the anomalies that 'ACE' suffers with regression data including the inability of 'ACE' to reproduce model transformations and its sensitivity to the marginal distribution of the predictors. It would be good to see applications of both of these methods to non-linear calibration problems and also to see more research papers tackling the non-linear calibration problem directly (i.e. without transformation of the data or the model to produce linearity).

## Acknowledgements

I should like to thank Professor B.W. Silverman and Dr C. Chatfield for their helpful comments.

## References

Aitchison, J. & Dunsmore, I.R. (1975). *Statistical Prediction Analysis*. Cambridge University Press.
Ali, M.A. & Singh, N. (1981). An alternative estimator in inverse linear regression. *J. Statist. Comp.* **14,** 1–15.
Barnett, V.D. (1966). Discussion of a paper by P. Sprent. *J. R. Statist. Soc.* B **28,** 291–292.
Barnett, V.D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics* **25,** 129–142.
Berkson, J. (1969). Estimation of a linear function for a calibration line: consideration of a recent proposal. *Technometrics* **11,** 649–660.
Breiman, L. & Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Am. Statist. Assoc.* **80,** 580–597.
Brown, G.H. (1979). An optimisation criterion for linear inverse estimation. *Technometrics* **21,** 575–579.
Brown, P.J. (1982). Multivariate calibration (with discussion). *J. R. Statist. Soc.* B **44,** 287–321.
Brown, P.J. & Sundberg, R. (1987) Confidence and conflict in multivariate calibration. *J. R. Statist. Soc.* B **49,** 46–57.
Brown, P.J. & Sundberg, R. (1989). Multivariate calibration with more variables than observations. *Technometrics* **31,** 365–371.
Buonaccorsi, J.P. (1986). Design considerations for calibration. *Technometrics* **28,** 149–155.
Carroll, R.J., & Spiegelman, C.H. (1986). The effect of ignoring small measurement errors in precision instrument calibration. *J. Qual. Tech.* **18,** 170–173.
Carroll, R.J., & Spiegelman, C.H. (1988). A quick and easy multiple-use calibration-curve procedure. *Technometrics* **30,** 137–141.
Chow, S.-C. & Shao, J. (1990). On the difference between the classical and inverse methods of calibration. *Appl. Statist.* **39,** 219–228.
Clark, R.M. (1979). Calibration, Cross-validation and carbon-14. I. *J. R. Statist. Soc.* A **142,** 47–62.
Clark, R.M. (1980). Calibration, Cross-validation and carbon—14.II. *J. R. Statist. Soc.* A **143,** 177–194.
Cochran, W. C. (1943). The comparison of different scales of measurement for experimental results. *Ann. Math. Statist.* **14,** 205–216.
Copas, J.B. (1982). Discussion of P.J. Brown's paper. *J. R. Statist. Soc.* B **44,** 312–313.
Creasy, M.A. (1954). Limits for the ratio of means. *J. R. Statist. Soc.* B **16,** 186–194.
De Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag.
Dobrigal, A., Fraser, D.A.S. & Gebotys, R. (1987). Linear calibration and conditional inference *Comm. Statist.* A **16,** 1037–1048.
Dunsmore, I.R. (1968). A Bayesian approach to calibration. *J. R. Statist. Soc.* B **30,** 396–405.
Easterling, R.G. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics* **23,** 343–344.
Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Statist.* **10,** 162–186.
Eubank, R.L. (1988). *Spline Smoothing and Non-parametric Regression*. New York: Marcel Dekker.
Farebrother, R.W. (1984). A note on Fearn's "Misuse of ridge regression". *Appl. Statist.* **33,** 74–75.

Fearn, T. (1983). Misues of ridge regression in the calibration of a near infrared reflectance instrument. *Appl. Statist.* **32,** 73–79.
Fieller, E.C. (1954). Some problems in interval estimation. *J. R. Statist. Soc.* B **16,** 175–185.
Fujikoshi, Y. & Nishii, R. (1984). On the distribution of a statistic in multivariate inverse regression analysis. *Hiroshima Math. J.* **14,** 215–225.
Grubbs, F.E. (1948). On estimating precision of measuring instruments and product variability. *J. Am. Statist. Assoc.* **43,** 243–264.
Grubbs, F.E. (1973). Errors of measurement, precision, accuracy, and the statistical comparison of measuring instruments. *Technometrics* **15,** 53–66.
Halperin, M. (1970). On inverse estimation in linear regression. *Technometrics* **12,** 727–736.
Hill, B.M. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics* **23,** 335–338.
Hoadley, B. (1970). A Bayesian look at inverse linear regression. *J. Am. Statist. Assoc.* **65,** 356–369.
Hoerl, A.E., Kennard, R.W. & Hoerl, R.W. (1970). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* **12,** 55–67.
Hoerl, A.E., Kennard, R.W. & Hoerl, R.W. (1985). Practical use of ridge regression: a challenge met. *Appl. Statist.* **34,** 114–120.
Hunter, W.G. & Lamboy, W.F. (1981a). A Bayesian analysis of the linear calibration problem. *Technometrics* **23,** 323–328.
Hunter, W.G. & Lamboy, W.F. (1981b). Response to discussion of their 1981 paper. *Technometrics* **23,** 344–350.
Kalbfleisch, J.D. & 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.
Knafl, G., Spiegelman, C., Sacks, J. & Ylvisaker, D. (1984). Nonparametric calibration. *Technometrics* **26,** 233–241.
Krutchkoff, R.G. (1967). Classical and inverse regression methods of calibration. *Technometrics* **9,** 425–439.
Krutchkoff, R.G. (1968). Letter to the editor. *Technometrics* **10,** 430–431.
Krutchkoff, R.G. (1969). Classical and inverse regression methods of calibration in extrapolation. *Technometrics* **11,** 605–608.
Krutchkoff, R.G. (1971). The calibration problem and closeness. *J. Statist. Comput. Simul.* **1,** 87–95.
Lawless, J.F. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics* **23,** 334–335.
Lechner, J., Reeve, C. & Spiegelman, C. (1982). An implementation of the Scheffé approach to calibration using spline functions, illustrated by a pressure-volume calibration. *Technometrics* **24,** 229–234.
Lieberman, G.J., Miller, R.G. & Hamilton, M.A. (1967). Unlimited simultaneous discrimination intervals in regression. *Biometrika* **54,** 133–145.
Lieftinck-Koeijers, C.A.J. (1988). Multivariate calibration: a generalisation of the classical estimator. *J. Mult. Anal.* **25,** 31–44.
Lindley, D.V. (1972). *Bayesian Statistics: A Review.* Philadelphia: SIAM.
Linnig, F.J. & Mandel, J. (1964). Which measure of precision? *Analy. Chem.* **36,** 25A–32A.
Lundberg, E. & De Mare, J. (1980). Interval estimates in the spectroscopy calibration problem. *Scand. J. Statist.* **7,** 40–42.
Lwin, T. (1981). Discussion of Hunter & Lamboy's paper (1981). *Technometrics* **23,** 339–341.
Lwin, T. & Maritz, J.S. (1980). A note on the problem of statistical calibration. *Appl. Statist.* **29,** 135–141.
Lwin, T. & Maritz, J.S. (1982). An analysis of the linear calibration controversy from the perspective of compound estimation. *Technometrics* **24,** 235–242.
Lwin, T. & Spiegelman, C.H. (1986). Calibration with working standards. *Appl. Statist.* **35,** 256–261.
Mandel, J. (1957). Fitting a straight line to certain types of cumulative data. *J. Am. Statist. Assoc.* **52,** 552–566.
Mandel, J. (1958). A note on confidence intervals in regression problems. *Ann. Math. Statist.* **29,** 903–907.
Mandel, J. (1984). Fitting straight lines when both variables are subject to error. *J. Qual. Tech.* **16,** 1–4.
Mandel, J. & Linnig, F.J. (1957). Study of accuracy in chemical analysis using linear calibration curves. *Anal. Chem.* **29,** 743–749.
Martinelle, S. (1970). On the choice of regression in linear calibration. *Technometrics* **12,** 157–161.
Miller, R.G. (1966). *Simultaneous Statistical Inference.* New York: McGraw-Hill.
Minder, Ch. E. & Whitney, J.B. (1975). A likelihood analysis of the linear calibration problem. *Technometrics* **17,** 463–471.
Miwa, T. (1985). Comparison among point estimators in linear calibration in terms of mean squared error. *Jp. J. Appl. Statist.* **14,** 83–93.
Naes, T. (1985a). Comparison of approaches to multivariate linear calibration. *Biom. J.* **27,** 265–275.
Naes, T. (1985b). Multivariate calibration when error covariance-matrix is structured. *Technometrics* **27,** 301–311.
Naes, T. (1986). Multivariate calibration using covariance adjustment. *Biom. J.* **28,** 99–107.
Naes, T., Irgens, C. & Martens, H. (1986). Comparison of linear statistical methods for calibration of NIR instruments. *Appl. Statist.* **35,** 195–206.
Naes, T. & Martens, H. (1984). Multivariate calibration. II. Chemometric Methods. *Trends. Anal. Chem.* **3,** 266–271.
Naszodi, L.J. (1978). Elimination of the bias in the course of calibration. *Technometrics* **20,** 201–206.
Naylor, J.C. & Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distributions. *Appl. Statist.* **31,** 214–225.

Nishii, R. & Krishnaiah, P.R. (1988). On the moments of classical estimates of explanatory variables under a multivariate calibration model. *Sankhyā* **50,** 137–148.
Oden, A. (1973). Simultaneous confidence intervals in inverse linear regression. *Biometrika* **60,** 339–343.
Oman, S.D. (1985). An exact formula for the M.S.E. of the inverse estimator in linear calibration problem. *J. Statist. Plan. Infer.* **11,** 189–196.
Oman, S.D. (1988). Confidence regions in multivariate calibration. *Ann. Statist.* **16,** 174–187.
Oman, S. & Wax, Y. (1984). Estimating foetal age by ultrasound measurements: an example of multivariate calibration. *Biometrics* **40,** 947–960.
Orban, J.E. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics* **23,** 342–343.
Osborne, C. (1990). Non-parametric calibration. Ph.D. Thesis, University of Bath.
Ott, R.L. & Myers, R.H. (1968). Optimal experimental designs of estimating the independent variable in regression. *Technometrics* **10,** 811–823.
Perng, S.K. (1987). A note on the inverse estimator for the linear calibration problem. *Comm. Statist.* A **16,** 1743–1747.
Perng, S.K. & Tong, Y.L. (1974). Sequential solution to inverse linear regression problem. *Ann. Statist.* **2,** 535–9.
Perng, S.K. & Tong, Y.L. (1977). Optimal allocation of observations in inverse regression—note. *Ann. Statist.* **5,** 191–6.
Racine-Poon, A. (1988). A Bayesian approach to non-linear calibration problems. *J. Am. Statist. Assoc.* **83,** 650–656.
Reilly, P.M. (1976). The numerical computation of posterior distributions. *Appl. Statist.* **25,** 201–209.
Rosenblatt, J.R. & Spiegelman, C.H. (1981). Discussion of Hunter and Lamboy's 1981 paper. *Technometrics* **23,** 329–333.
Rothman, D. (1968). Letter to the editor. *Technometrics* **10,** 429–430.
Saw, J.G. (1970). Letter to the editor. *Technometrics* **12,** 937.
Scheffé, H. (1973). A statistical theory of calibration. *Ann. Statist.* **1,** 1–37.
Shukla, G.K. (1972). On the problem of calibration. *Technometrics* **14,** 547–553.
Shukla, G.K. & Datta, P. (1985). Comparison of the inverse estimator with the classical estimator subject to a preliminary test in linear calibration. *J. Statist. Plan. Infer.* **12,** 93–102.
Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. R. Statist. Soc.* B **47,** 1–52.
Smith, R.L. & Corbett, M. (1987). Measuring marathon courses: an application of statistical calibration theory. *Appl. Statist.* **36,** 283–295.
Spezzaferri, F. (1985). A note on multivariate calibration experiments. *Biometrics* **14,** 267–272.
Spiegelman, C.H. & Studden, W.J. (1980). Design aspects of Scheffé's theory using linear splines. *J. Res. Nat. Bureau of Standards* **85,** 295–304.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc.* B **36,** 111–147.
Stone, M. (1977). Consistent non-parametric regression. *Ann. Statist.* **5,** 595–645.
Sundberg, R. (1985). When is the inverse regression estimator M.S.E. superior to the standard regression estimator in multivariate controlled calibration situations? *Statist. Prob. Letters* **3,** 75–79.
Tallis, G.M. (1969). Note on a calibration problem. *Biometrika* **56,** 505–508.
Tibshirani, R.J. (1988). Estimating transformations for regression via additivity and variance stabilisation. *J. Am. Statist. Assoc* **83,** 394–405.
Theobald, C.M. & Mallinson, J.R. (1978). Comparative calibration, linear structural relationships and congeneric measurements. *Biometrics* **34,** 39–45.
Thomas, M.A. & Myers, R.H. (1973). Optimal designs for the inverse regression method of calibration. *Comm. Statist* **2,** 419–433.
Turiel, T.P., Hahn, G.J. & Tucker, W.T. (1982). New simulation results for the calibration and inverse median estimation problems. *Comm. Statist.* B **11,** 677–713.
Vecchia, D.F., Iyer, H.K. & Chapman, P.L. (1989). Calibration with randomly changing standard curves. *Technometrics* **31,** 83–90.
Williams, E.J. (1959). *Regression Analysis.* New York: Wiley.
Williams, E.J. (1969a). Regression methods in calibration problems. *Bull. Int. Statist. Inst.* **43,** 17–28.
Williams, E.J. (1969b). A note on regression methods in calibration. *Technometrics* **11,** 189–192.

## Résumé

Cet article revise les divers approches à la calibration, soit en une variable ou en plusieurs. Dans le cas de calibration en une variable, on considère les approches: classique, inverse, Bayesienne, non-paramétrique ainsi que les approches par régions de tolérance. Dans le cas de calibration en plusieurs variables, on considère diverses approches de régressions ainsi que les approches Bayesiennes et de profile de vraisemblance. Une bibliographie étendue est donnée.

[*Received February* 1988, *accepted April* 1991]