Fox, David R.,
Statistical Calibration : Theory and Applications, Ph.D.,
Department of Statistics, University of Wyoming, May 1989.

This dissertation provides a comprehensive treatment of statistical calibration theory and methodology. In addition to the adaptation and refinement of existing theory, a number of new results are reported. These include :

(i) development of orthogonal estimation procedures in both the univariate and multivariate situations.

(ii) calibration procedures for measurement-error models and methods for extracting variance components when measurement-error variance is confounded with residual-error variance.

(iii) development of the theory associated with conditional multivariate calibration

(iv) derivation of procedures for calibrating in the univariate, multiple-regression model.

(v) derivation of a new measure of influence for conditional calibration procedures.

(vi) development of a procedure for discrete calibration - for example, in experimental design situations where a dummy-variable coding of the X-matrix is used.

(vii) development of theory and methods for calibrating in a non-stationary field where the model parameters are

functions of the position in a multi-dimensional space-time framework.

Chapter I provides a review of existing literature in the areas of univariate and multivariate calibration, tracing briefly the development of calibration as a statistical technique and the controversy surrounding the so-called "inverse" and "classical" approaches.

Chapter II deals with univariate calibration problems and examines the role of standard estimation procedures in addition to introducing alternative forms of estimation. Both point and interval estimation procedures are examined. Questions of optimal design in controlled-calibration experiments are also addressed. Results of small-scale simulation studies are presented which indicate the relative utility of the methods considered. The performance of the various procedures is also studied for measurement-error models and some recommendations provided. Many aspects of univariate calibration are succinctly illustrated with the use of a detailed application arising from legal proceedings brought against a speeding driver.

Chapter III expands upon the ideas introduced in Chapter II and considers multivariate applications of calibration. The multivariate analogs for inverse and classical estimation are developed as are confidence interval procedures. The orthogonal estimator introduced in Chapter II is extended to cater for multivariate problems and computer software developed to perform the iterative calculations necessary to

implement this means of calibration in practice. A significant portion of this chapter is devoted to the development of the necessary theory for conditional multivariate calibration. New procedures are developed which enhance the performance of calibration when one can utilize knowledge of some of the components of the vector being calibrated for. These methods are also applied with great effectiveness to the univariate multiple linear regression situation.

In Chapter IV methods are developed for spatial calibration where the usual assumptions of independence do not hold and the parameters of the model are linearly related to the position within a random field. Various procedures for estimation in this varying-parameters model are devised together with the development of accompanying computer routines.

Chapter V concludes the dissertation with suggestions for future research.

STATISTICAL CALIBRATION : THEORY AND APPLICATIONS

by

David R. Fox

A dissertation

submitted to the Department of Statistics

and The Graduate School of The University of Wyoming

in partial fulfillment of Requirements

for the Degree of

DOCTOR OF PHILOSOHPY

in

STATISTICS

Laramie, Wyoming

May, 1989

To Jennifer who endured it all.

ACKNOWLEDGEMENTS

The research reported in these pages is not the result of a single-handed effort on the part of the author, but rather is the culmination of the efforts and talents of many people. I should first like to express my sincere thanks to Curtin University for granting me a rather protracted leave of absence to complete this work and in particular Mr. W.S. Perriman and Mr. Brian White. Special thanks are also due to my Committee : Professors Bieber, Bonett, Borgman, Bridges, and McDonald for their time, assistance and understanding. I am particularly indebted to Dr. Bonnet whose ready availability and willingness to participate at every stage of this research was invaluable.

To all my friends and family who acted as a support group over the years I shall remain indebted. In Australia thanks to Molloy for managing my affairs. In the U.S. thanks to Daniel, Darlene, Thom, Joanne, Gus, and Barb for helping me keep a sensible perspective on life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

REVIEW OF LITERATURE

1.1 INTRODUCTION

The history of statistical calibration is long and checkered. One of the earliest calibration examples was provided by Wolf (1908). An examination of the literature suggests little interest in the problem until the late 1930's and early 1940's with the publication of papers by Eisenhart (1939) and Fieller (1940). There was another gap in the literature from the 1940's to the late 1960's when Krutchkoff (1967) provided results of a simulation study which brought into question standard statistical practice in calibration problems. Interest in the calibration controversy was intense for a few years thereafter but again waned and went into recession. Only recently has there been a renewed interest in the subject, which one may speculate is partly due to the associated developments in regression theory, for example, detection of influential observations and outliers,and multivariate regression.

Whilst much work has already been completed, more needs to be done to increase the body of theory on this important statistical problem. It is true to say that the calibration problem in its infancy was a difficult child that demanded the careful attention of the best statisticians. Thankfully, today the so-called calibration controversy has been largely resolved and attention is now correctly being given to

more diverse applications as well as refining existing theory. This present chapter is intended to provide a concise review of the major developments that have taken place over the last fifty years.

1.2 CALIBRATION DEFINED

The most common form of calibration is that of calibrating a quick and easy (and usually inexpensive) method of obtaining measurements against a tedious and/or expensive, but highly accurate method. In calibrating some instrument we take readings Y on some physical process X and use the empirical relationship between Y and X to 'predict' the value of X given some future reading y_0. Examples of such procedures abound in the literature. Carroll et. al. (1981) consider the calibration of a mass spctrometer while Knafl et. al. (1984) have used calibration as part of the nuclear safeguards program. A more novel, but nonetheless serious application of calibration theory was recently provided by Smith and Corbett (1987) who looked at the determination of the length of an olympic marathon course based on the counter readings taken from a bicycle wheel.

Williams (1969) has stressed the existence of two types of calibration problem. "Absolute calibration" is used in reference to the situation just described in which an alternative procedure for obtaining measurements is calibrated against a known or 'true' method. "Comparative calibration" on the other hand is used to describe the process of comparing one instrument or measurement technique against another. Rosenblatt and Spiegelman (1981) define the following

categories for the variety of possible uses of calibration.

(i) single use : where a point or interval estimate is required for a single future reading.

(ii) multiple uses : when the process being calibrated can be assumed to be sufficiently stable to permit the repeated use of the calibration model.

(iii) single and multiple use in combination with other measurements : when results from the calibration experiment are to be used with other data to estimate a quantity that is expressed as a function of several variables.

Aitchison and Dunsmore (1975) distinguish between designed and natural calibration experiments. By their definition a designed experiment is one which is set up under controlled (labatory) conditions and is intended to span the range of expected X_0 values. A natural calibration experiment on the other hand, is one in which the data simply 'presents' itself as is often the case in biological or clinical research where experimentation is not possible. Rosenblatt and Spiegelman (1981) also report on an increasing requirement and use of non-linear calibration which the authors claim is often brought about by the non-linearities introduced by automated recording procedures.

1.3 THE CLASSICAL VERSUS INVERSE CONTROVERSY

In many instances it is assumed that the relationship between Y and X is linear, and thus estimation of parameters β_0 and β_1 in the (simple) regression of Y on X is readily achieved via OLS to give :

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad (1.1)$$

In contrast to the 'normal' use of equation (1.1) where some new value of Y is to be predicted for a given value of $X=x_0$, the requirement here is to estimate x_0 having observed $Y=y_0$. This in turn has led to the development of two distinct approaches, hereafter referred to the classical and inverse regression methods.

In the so-called "classical" calibration method an estimate of X_0, $\hat{X}_0$, is obtained by a simple rearrangement of the terms in equation (1.1) to give :

$$\hat{X}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \qquad (1.2)$$

An alternative, and equally appealing approach is to treat X as dependent and regress X on Y. This procedure is known as the "inverse" calibration method for which we obtain $\hat{\gamma}_0$ and $\hat{\gamma}_1$ as our parameter estimates in the regression

$$X_i = \gamma_0 + \gamma_1 y_i + \zeta_i \qquad (1.3)$$

The problem of deciding between these two methods is not new. Eisenhart (1939) suggests that both methods were in common use up to the time of his paper, although firmly rejects procedures based on equation (1.3) arguing that the least-squares line should be fitted to the variable which is observed with error (the case of errors in *both*

variables complicates this assertion and is examined in Chapter two).

Krutchkoff (1967,1969) on the other hand advocates the use of inverse calibration and presents the results of simulation studies in which the relative merits of each method were asessed.

A number of papers critical of Krutchkoff's work have since appeared, although general agreement on the 'best' approach has not been forthcoming. The problem, it seems, stems from the fact that there is no universally accepted properties of an optimal estimator in the calibration experiment.

Williams (1969) pointed out that in the case of normally distributed errors, the classical estimator has an undefined expectation and infinite variance and as such any comparison based on mean squared error (MSE) is rendered meaningless.

Berkson (1969) advocates the concept of Pitman closeness as a means of comparison although notes that estimators obtained by the inverse method are not consistent nor asymptotically unbiased. This lack of consistency was also observed by Madansky (1959). Shukla (1972) showed that the asymptotic MSE for the inverse estimator is given by

$$\lim_{n\to\infty} \text{MSE}(\hat{X}_0^*) = \frac{\sigma^4}{\beta_1^4 S_{xx}^2 \theta^2} (\bar{x} - x_0)^2$$

where $\theta = 1 + \dfrac{\sigma^2}{\beta_1^2 S_{xx}}$ and $\hat{X}_0^*$ denotes the calibrated value from the inverse regression. Oman (1985) showed that the exact MSE for the inverse estimator is given as

$$MSE[\hat{X}_0^*] = \delta\{1+\frac{\lambda}{2}[(n-6)\phi(n-1,\lambda)-(n-4)\phi(n-3,\lambda)]\} + (1+\frac{1}{n})\frac{S_{xx}}{2}[(n-2)\phi(n-1,\lambda)-(n-4)\phi(n-3,\lambda)]$$

where S_{xx} represents the usual sum of squares, $\lambda = \frac{\beta_1^2 S_{xx}}{\sigma^2}$, $\delta = (x_0 - \bar{x})$, and $\phi(j,\lambda) = E[\frac{1}{j+2W}]$ $,j>0$ and W has a Poisson$(\frac{\lambda}{2})$ distribution.

Martinelle (1970) shows that the MSE for the inverse estimator is less than that of the classical estimator provided

$$(X_0 - \bar{X})^2 < S_{xx}[2 + \frac{1}{\gamma^2 S_{xx}}]$$

where $\bar{X}$ and S_x^2 are the sample mean and (biased) sample variance respectively and $\gamma = \frac{\beta_1}{\sigma}$. Furthermore, Matrinelle (1970) suggests that when $\gamma^2 s_{xx}$ is large, then there is little advantage in using the inverse method. The general relation between MSE for the inverse and classical estimators is depicted in figure 1.

1.4 INTERVAL ESTIMATION

Having obtained a point estimate for X_0 whether it be by classical or inverse regression, a natural extension is to develop interval estimates. Many authors have devoted their attention to this problem. Probably one of the earliest references to interval estimation in the calibration context was by Fieller (1940,1954) who constructed a $(1-\alpha)100\%$ confidence interval for X_0. However the procedure had inherent difficulties, most notably that the $(1-\alpha)$ confidence region sometimes included the whole real line or two disjoint semi-infinite lines. This phenomenon occurred when β_1 was close to zero (in which case one may

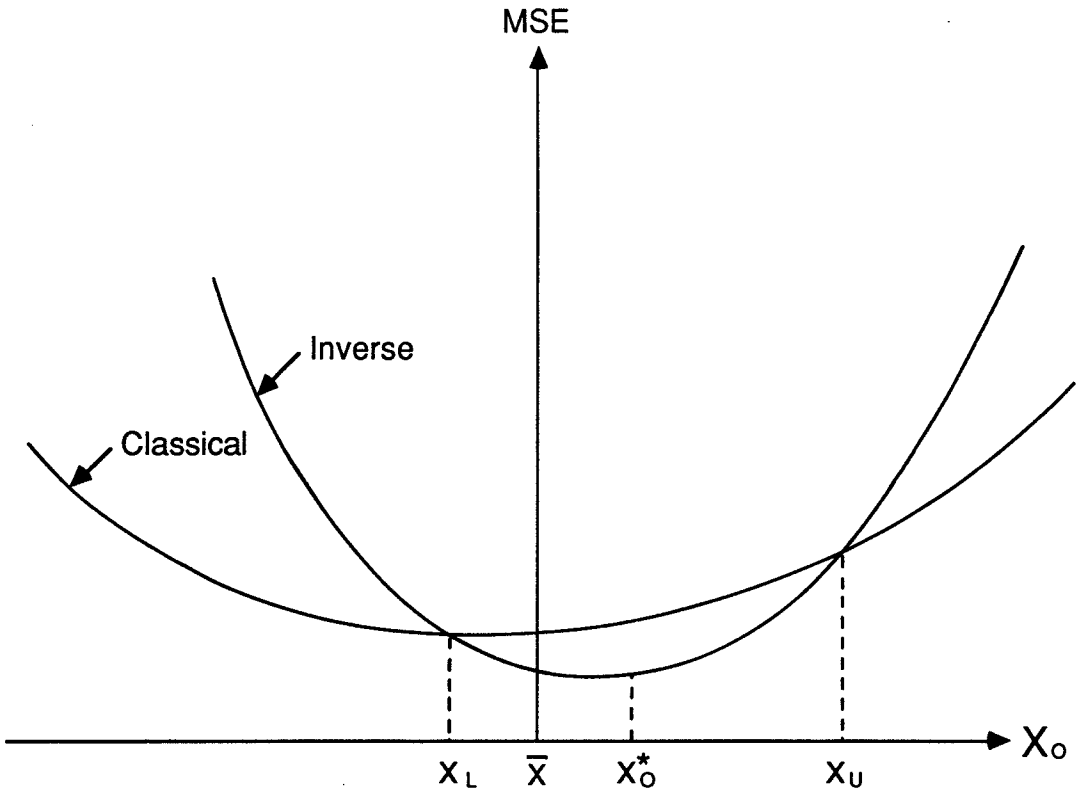

Figure 1. Comparison of mean square error for the inverse and classical estimators.

argue that there is little value in conducting the calibration experiment at all). Dobrigal, Fraser, and Gebotys (1987) provide a conditional confidence interval for X_0 which avoids the paradox of Fieller's method although the level of confidence is only approximate.

Sheffé (1973) provided confidence bands for the classical estimator although his method required the use of his own set of tables for determining the values of certain coefficients. Carrol, Sacks, and Spiegleman (1988) modified Sheffe's approach so as to be easier to implement in practice. The resulting intervals were also shown to be substantially shorter than the corresponding Sheffe interval. Oden (1973) has considered the problem of finding simultaneous confidence intervals for the inverse estimator. Further aspects of interval estimation techniques are taken up in chapter two.

1.5 OTHER DEVELOPMENTS

Many alternatives to the classical and inverse estimators have been proposed, each being designed to overcome certain inherent difficulties associated with the standard procedures. Conditioning, compound estimation, and Bayesian methods usually form the basis of these approaches. In this section we examine some of these alternatives.

Lwin and Maritz (1980) proposed the "non-linear" predictor of X_0 which was shown to have desirable properties not shared by other methods. In a subsequent development Lwin and Maritz (1982) considered a more general class of estimators and in particular demonstrated that

$$X^*(y_0) = (1 + \theta^2 S_{xx})^{-1}\bar{x} + \theta^2 S_{xx}(1 + \theta^2 S_{xx})^{-1}[(y_0-\beta_0)/\beta_1]$$

is optimal in the sense that it results in smallest MSE when applied to previous y_i's. This estimator makes use of the current observation y as well as previous x_i values of the calibration experiment.

Ali and Singh (1981) proposed an estimator which was a weighted average of the inverse and classical estimators while Srivastava and Singh (1987) provide a class of estimators by replacing β_1^{-1} in the classical estimator with

$$T(k) = \frac{\hat{\beta}_1}{\hat{\beta}_1^2 + k\sigma^2/n} \quad ; \quad k > 0$$ and σ^2 is the residual error variance. The p.d.f. for the resulting class is obtained and the bias and MSE obtained by numerical integration. The authors conduct simulations to establish that the inverse estimator was superior in terms of MSE thus supporting Krutchkoff's (1967) conclusion.

Techniques based on conditioning have been widely used in order to obtain estimators which are well-behaved. Problems usually arise when β_1 is close to zero and so many authors prefer to 'protect' themselves from this situation by first testing the hypothesis $H_0:\beta_1=0$. This is anlagous to Fisher's protected LSD in an analysis of variance. Graybill (1976) advocates this method as standard procedure. Shukla and Datta (1985) also use this form of conditioning to investigate properties of the classical estimator under "random truncation about zero". Their suggestion is that a level of significance of at least 0.01 be used. The authors then found that the bias and MSE depend only on the ratio $\frac{\beta_1}{\sigma}$.

Specifically, |bias| decreases for increasing β_1 for both the inverse and classical estimators although |bias| for the classical estimator is always smaller than that of the inverse estimator. Furthermore, for *small* β_1 the MSE for the classical estimator is greater than the MSE for the inverse estimator for interpolation but is smaller for extrapolation. There is little difference between the two (in terms of MSE)for large β_1. Lwin (1981) also advocates the use of conditioning although prefers to be more specific, claiming that the hypothesis to be first tested should be $H_0:\beta_1>0$ since the condition of $\beta_1<0$ has no validity in a calibration context. He suggests a class of estimators linear in the current stage data mean $\bar{y}$: $\hat{x}_0(\bar{y}) = k_0 + k_1\bar{y}$ and uses the criterion $n^{-1}\{\mathbb{E}[\hat{x}_0(y_i) - x_i]^2\}$, called the compound MSE as a basis of comparison. It is shown that the unconditional minimization of the compound MSE leads to the inverse estimator whereas if the constraint of unbiasedness is imposed, the minimization results in the classical estimator. Shukla (1972) investigated the classical estimator for this conditioning situation in the case of normally distributed errors and obtained expressions [to $O(n^{-1})$] for both the inverse and classical estimators. Lwin (1981) generalized these results by considering a wider class of error distributions also to terms $O(n^{-1})$ and concludes that there is negligible effect of non-normality on the performance of the classical estimator. Furthermore, the bias of the inverse estimator is affected by both the skewness and kurtosis of the error distribution although this effect can be reduced by increasing n. Lwin (1981) suggests that in general an excessively peaked error distribution will reduce the efficiency of the inverse estimator, while a flat-peaked

error distribution will enhance its performance.

Bayesian approaches have also been applied to the calibration problem although the techniques are perhaps not as widespread as those based on more traditional frequentist methods. Hunter and Lamboy (1981) provided a Bayesian analysis of the calibration problem. Interestingly, their approach resulted in a posterior distribution for the calibrated value that also has infinite variance. The authors remark that this fact is not disturbing and go on to state :

> "Theoretically, if one's model is adequate, all of the relevant information is contained in the appropriate posterior distribution, whether its variance happens to be finite or not . . . our view is that for standard calibration problems, arguments about which estimator is best that are based on MSE or related criteria are simply irrelevant".

It should be pointed out that the duality of the classical and inverse regression methods is not resolved by a Bayesian approach since either is supported depending on the choice of the prior distribution for X_0 .

Much work has also been devoted to other aspects of the calibration problem including extensions to the multivariate setting. Spiegelman (1984) has explored the use of calibration curves in quality-control situations while Brown (1982), Wood (1982), Spezzaferri (1985), Oman and Wax (1984) and others have examined the use of multivariate calibration techniques. Oman (1984) has derived a statistic, similar in nature to Cook's distance to measure the influence of a particular observation on future estimates from the calibration curve. Spiegelman (1984) has similarly considered the role of regression diagnostics in the

calibration problem.

While much discussion continues on the relative merits of the inverse and classical approaches, a further complexity is introduced when one considers situations in which both X and Y are measured with error. Berkson (1950) formally raised the question of the existence of two separate regression lines in such cases. Small-sample properties of $\hat{\beta}_1$ were investigated by Halperin and Gurian (1971) and under certain prescribed conditions, results for $E[\hat{\beta}_1]$ and $MSE[\hat{\beta}_1]$ were derived. Clutton-Brock (1967) argues that there is "no paradox of two regression lines" and suggests that in the case of errors in both X and Y, the maximum likelihood estimates lie between the two separate regressions of Y on X and X on Y. Carrol and Spiegelman (1986) have examined the effect on confidence interval estimates of X_0 when measurement error is ignored. They showed that the ratio of (interval length ignoring measurement error) : (interval length when no measurement error present) is equal to $(1 + \beta_1^2\sigma_m^2/\sigma^2)^{\frac{1}{2}}$ where σ_m^2 is the variance of measurement error. Thus, for example when β_1 is close to 1.0, the effect of neglecting a measurement error variance which is comparable to the residual variance will cause the resulting interval estimate to be approximately 40% too large. The measurement error model will be examined further in chapter two.