CHAPTER III

MULTIVARIATE MODELS

3.1 INTRODUCTION

This chapter represents an extension to the calibration problem in the simple univariate case. We now examine calibration procedures for an unknown q-vector X when a p-vector Y has been observed. Whilst the simple calibration problem has a long history, the same is not true of the multivariate extension. In 1982 Brown presented a number of important results for the multivariate calibration problem for both the classical approach and also from a Bayesian perspective. Other papers have since appeared although the multivariate aspect of the calibration problem has been relatively slow to develop.

Oman and Wax (1984) gave an example of how multivariate calibration could be used to improve the quality of estimation of the gestational age of an unborn child. Standard practice is to compare the rate of fetal development with published charts and tables. Such quantities as bone lengths (femur length, F and biparietal diameter ,BPD) are usually used. However, the comparison of each measurement separately with its respective 'standard' value ignores the relationship between all three factors (age, F, and BPD). The authors demonstrated that a model which took cognizance of the covariance structure of all three variables resulted in an index of gestational age which was significantly more

accurate than the use of either alone. Notice that in this example both F and BPD are dependent, although our future interest centers on determining age from measurements on F and BPD.

Other applications of multivariate calibration have included a method by Spezzaferri (1985) for choosing among K different calibration experiments associated with K different instruments and applications in remote sensing [O'Muircheartaich and Gaver (1986)] and chemistry [Sjostrom et.al.(1983)]. The problem of detecting influential observations in a calibration experiment has been touched on (e.g. Oman (1984) and Spiegelman (1984)) although most of this work has concentrated on the univariate case. Applications to the multivariate setting are examined in section 3.6

We commence our treatment of multivariate calibration with the development of fundamental equations for both the classical (m.l.e.) and inverse approaches. The orthogonal estimator developed for the univariate model of Chapter I is also extended to the cover multivariate analyses. Monte-Carlo methods are employed to compare properties of the classical,inverse, and orthogonal estimators. The problem of conditional calibration has hitherto not been examined. Development of the necessary theory and illustrative examples are provided. Discrete multivariate calibration has similarly received little or no attention in the literature and in section 3.7 a procedure is developed for factorial experiments. Implications for fractional factorial designs are also examined. Finally, the problem of calibrating when there is only one dependent variable but more than one potential regressor is available is reviewed. This model is a hybrid of the univariate and multivariate

approaches and has received little attention in the literature.

3.2 THE THEORY OF MULTIVARIATE CALIBRATION.

We have as our assumed model :

$$Y = 1\,\alpha^{T} + X\beta + \xi \qquad (3.1)$$

where Y is $(n \times p)$; 1 is $(n \times 1)$ vector of ones;
α is $(p \times 1)$; X is $(n \times q)$;
β is $(q \times p)$; ξ is $(n \times p)$.

Furthermore, assume $\xi \sim N_p(0,\Sigma_y)$ .In the subsequent development we will assume that both the X and Y data have been centered on their respective means and hence equation (3.1) reduces to $Y = X\beta + \xi$.

3.2.1 THE INVERSE ESTIMATOR.

Inverse estimation is perhaps the easiest and most straightforward approach. It is mentioned in passing that if the q components of X cannot be considered random then the same criticisms arise as for the univariate case. Brown (1982) notes that if both X and Y are random then there is no preferred method and both regression of X on Y and Y on X are equally valid. For the inverse regression we have

$$X = Y\gamma + \zeta \qquad (3.2)$$

where γ is a (p x q) matrix of parameters for the regression of X on Y. The GLS estimator for γ is thus

$$\hat{\gamma} = (Y^T \Omega_x^{-1} Y)^{-1} Y^T \Omega_x^{-1} X \tag{3.3}$$

where Ω_x is a positive definite, symmetric matrix with $Cov[\zeta_i \zeta_j] = \delta_{ij} \Omega_x$.

The covariance matrix of $\hat{\gamma}$ is

$$Cov[\hat{\gamma}_i, \hat{\gamma}_j] = \delta_{ij} (Y^T \Omega_x^{-1} Y)^{-1} \tag{3.4}$$

Note, under OLS conditions $\Omega_x = I$,and equation (3.4) reduces to the more familiar $\delta_{ij} (Y^T Y)^{-1}$.

In calibrating for an unknown X-vector we assume we have a p-vector of responses $y_0 = [y_{01}, y_{02}, \ldots, y_{0p}]^T$ and thus

$$\hat{X}_0 = y_0^T \hat{\gamma}$$

$$= [y_0^T \hat{\gamma}_1 \mid y_0^T \hat{\gamma}_2 \mid \ldots \mid y_0^T \hat{\gamma}_q] \tag{3.5}$$

with

$$Cov[\hat{X}_{0i}\ \hat{X}_{0j}] = \delta_{ij}\ y_0^T (Y^T \Omega_x^{-1} Y)^{-1} y_0 \tag{3.6}$$

3.2.1.1 Confidence Intervals.

Under the assumption of multivariate normality the calibrated $\hat{X}_0$ is

distributed as :

$$N_q[\gamma^T y_0, y_0^T (Y^T \Omega_x^{-1} Y)^{-1} y_0 \Sigma_x]$$

where

$$\Sigma_x = \{\delta_{ij}\}$$

and

$n\hat{\Sigma}_x$ is distributed independently with Wishart distribution $W_{n-p}(\cdot|\Sigma_x)$ where $\hat{\Sigma}_x = \frac{1}{n}\hat{\zeta}^T\hat{\zeta}$.

Hotelling's T^2 statistic is therefore

$$T^2 = \left[\frac{\hat{\gamma}^T y_0 - \gamma^T y_0}{\{y_0^T[(Y^T \Omega_x Y)^{-1}] y_0\}^{\frac{1}{2}}} \right]^T \left[\frac{n\Sigma_x}{n-p} \right]^{-1} \left[\frac{\hat{\gamma}^T y_0 - \gamma^T y_0}{\{y_0 (Y^T \Omega_x Y)^{-1} y_0\}^{\frac{1}{2}}} \right]$$

and the $(1-\alpha)100\%$ confidence ellipsoid for $\hat{\gamma}^T y_0$ is

$$(\hat{\gamma}^T y_0 - \gamma^T y_0)^T \left[\frac{n\,\Sigma_x}{n-p} \right]^{-1} (\hat{\gamma}^T y_0 - \gamma^T y_0)$$

$$\leq y_0^T (Y^T \Omega_x^{-1} Y)^{-1} y_0 \left[\frac{q(n-p)}{n-p-q+1} F_{q, n-p-q+1}(\alpha) \right]$$

where $F_{q, n-p-q}(\alpha)$ is the upper (100α)th percentile of the F-distribution with q and n-p-q degrees of freedom.

The $(1-\alpha)100\%$ simultaneous confidence intervals for $y_0^T \gamma_i$ are

$$y_0^T \hat{\gamma}_i \pm \left[\frac{q(n-p)}{n-p-q+1} F_{q, n-p-q+1}(\alpha) \right]^{\frac{1}{2}} \left[y_0^T (Y^T \Omega_x^{-1} Y)^{-1} y_0 \left(\frac{n\delta_{ii}}{n-p} \right) \right]^{\frac{1}{2}}$$

$$i = 1, \ldots, q \; . \qquad (3.7)$$

3.2.2 THE CLASSICAL ESTIMATOR.

For the model and assumptions given by equation (3.1) we can express the likelihood function in terms of the unknowns β and Σ_y. Let x_i' be the i^{th} *row* of X and y_i' be the i^{th} *row* of Y (i.e. x_i' is a 1xq vector and y_i' is a 1xp vector). Then the likelihood function is

$$L(\beta, \Sigma_y) = \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\Sigma_y|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_i' - x_i'\beta)\Sigma_y^{-1}(y_i' - x_i'\beta)^T} \qquad (3.8)$$

Maximizing equation (3.8) (or its logarithm) produces the maximum likelihood estimates :

$$\hat{\beta} = (X^TX)^{-1}X^TY \qquad (3.9)$$

and

$$\hat{\Sigma}_y = \frac{1}{n}(Y - X\hat{\beta})^T(Y - X\hat{\beta}) \qquad (3.10)$$

The m.l.e. for an x_0 vector is obtained as follows :
Suppose x_0 is an unknown (1xq) vector corresponding to a new (1xp) vector y_0 . Assuming for the time being that Σ_y and β are known then the likelihood function of equation (3.8) can be written as a function of x_0. Thus

$$L(x_0) = (2\pi)^{-\frac{p}{2}} |\Sigma_y|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_0 - x_0\beta)\Sigma_y^{-1}(y_0 - x_0\beta)^T} \qquad (3.11)$$

Now equation (3.11) is a maximum when the exponent is a minimum. Minimization of $(y_0 - x_0\beta)\Sigma_y^{-1}(y_0 - x_0\beta)^T$ is equivalent to a weighted

least-squares regression for unknown $\mathbf{x}_0$. Consider the model

$$\mathbf{y}_0^T = \beta^T \mathbf{x}_0^T + \zeta \tag{3.12}$$

in which β is assumed known. The least-squares estimate for $\mathbf{x}_0$ is thus

$$\hat{\mathbf{x}}_0^T = (\beta\beta^T)^{-1}\beta\, \mathbf{y}_0^T \tag{3.13}$$

Now there exists a non-singular matrix P such that $PP^T=\Sigma_y^{-1}$. Premultiplying both sides of equation (3.12) by P^T gives

$$P^T\mathbf{y}_0^T = P^T\beta^T\mathbf{x}_0^T + P^T\zeta$$

or
$$\mathbf{y}_0^{*} = \beta^{*}\mathbf{x}_0^T + \zeta^{*} \tag{3.14}$$

From equation (3.13) the least-squares solution is thus

$$\hat{\mathbf{x}}_0^T = (\beta^{*T}\beta^{*})^{-1}\beta^{*T}\mathbf{y}_0^{*}$$

$$= (\beta PP^T\beta^T)^{-1}\beta PP^T\mathbf{y}_0^T$$

Thus ,

$$\hat{\mathbf{x}}_0^T = (\beta\, \Sigma_y^{-1}\beta^T)^{-1}\beta\, \Sigma_y^{-1}\mathbf{y}_0^T \tag{3.15}$$

Also,

$$\mathrm{Cov}[\hat{\mathbf{x}}_0] = (\beta\,\Sigma_y^{-1}\beta^T)^{-1}\beta\,\Sigma_y^{-1}\beta^T(\beta\,\Sigma_y^{-1}\beta^T)^{-1} \qquad (3.16)$$

where $\Sigma_y = \{\sigma_{ij}\}$.

3.2.2.1 Confidence Intervals.

For known β and Σ_y , the quadratic form in equation (3.11) has a χ^2_p distribution. Thus a $(1-\alpha)100\%$ confidence region for $\hat{\mathbf{x}}_0$ is

$$(\mathbf{y}_0 - \mathbf{x}_0\beta)\Sigma_y^{-1}(\mathbf{y}_0 - \mathbf{x}_0\beta)^T \le \chi^2_{p,1-\alpha} \qquad (3.17)$$

For most practical purposes, the quantities β and Σ_y will not be known and in which case we would use the corresponding estimates $\hat{\beta}$ and $\hat{\Sigma}_y$ obtained from the regression of Y on X.

Let the statistic W be formed by replacing β and Σ_y in equation (3.17) by their sample estimates i.e.

$$W = (\mathbf{y}_0 - \mathbf{x}_0\hat{\beta})\hat{\Sigma}_y^{-1}(\mathbf{y}_0 - \mathbf{x}_0\hat{\beta})^T \qquad (3.18)$$

Now

$$\mathbf{y}_0 \sim N_p(\mathbf{x}_0\beta, \Sigma_y)$$

and

$$\begin{aligned} E[(\mathbf{y}_0 - \mathbf{x}_0\hat{\beta})] &= E[\mathbf{y}_0] - E[\mathbf{x}_0\hat{\beta}] \\ &= \mathbf{x}_0\beta - \mathbf{x}_0\beta \qquad \text{since } \hat{\beta} \text{ is unbiassed.} \\ &= 0 \end{aligned}$$

$$\text{Also } \mathrm{Cov}[(y_{0i} - x_0\hat{\beta}_i)\ (y_{0j} - x_0\hat{\beta}_j)] = \mathbb{E}[(y_{0i} - x_0\hat{\beta}_i)\ (y_{0j} - x_0\hat{\beta}_j)^T]$$

$$= \mathbb{E}\left[[(y_{0i}-x_0\beta_i) + (x_0\beta_i-x_0\hat{\beta}_i)]\ [(y_{0j}-x_0\beta_j) + (x_0\beta_j-x_0\hat{\beta}_j)]^T\right]$$

$$= \mathbb{E}\left[[(y_{0i}-x_0\beta_i) + (x_0\beta_i - x_0\hat{\beta}_i)][(y_{0j}-x_0\beta_j)^T + (x_0\beta_j-x_0\hat{\beta}_j)^T]\right]$$

$$= \mathbb{E}\left[(y_{0i} - x_0\beta_i)(y_{0j} - x_0\beta_j)^T\right] + \mathbb{E}\left[(y_{0i} - x_0\beta_i)(x_0\beta_j - x_0\hat{\beta}_j)^T\right]$$
$$+ \quad \mathbb{E}\left[(x_0\beta_i-x_0\hat{\beta}_i)(y_{0j}-x_0\beta_j)^T\right] + \mathbb{E}\left[(x_0\beta_i-x_0\hat{\beta}_i)(x_0\beta_j-x_0\hat{\beta}_j)\right]$$

$$= \sigma_{ij} + x_0\ \mathrm{Cov}[\hat{\beta}_i\hat{\beta}_j]\ x_0^T$$
$$= \sigma_{ij} + x_0\ [\sigma_{ij}\ (X^TX)^{-1}]\ x_0^T$$
$$= \sigma_{ij} + \sigma_{ij}\ x_0\ (X^TX)^{-1}\ x_0^T$$
$$= \sigma_{ij}\ [1 + x_0(X^TX)^{-1}x_0^T]$$

Thus

$$\mathrm{Cov}[y_0 - x_0\hat{\beta}] = \Sigma_y\ [1 + x_0(X^TX)^{-1}x_0^T] \qquad (3.19)$$

Furthermore, $n\hat{\Sigma}_y$ is distributed independently with Wishart distribution $W_{n-q}(\cdot|\Sigma_y)$ where $\hat{\Sigma}_y = \frac{1}{n}\hat{\xi}^T\hat{\xi}$.

Thus

$$(y_0 - x_0\hat{\beta}) \sim N_p\left[0, \Sigma_y\ [1 + x_0(X^TX)^{-1}x_0^T]\right]$$

$$\Rightarrow \quad \frac{(y_0-x_0\hat{\beta})}{c^{\frac{1}{2}}} \sim N_p[0, \Sigma_y]$$

where $c = [1 + x_0(X^TX)^{-1}x_0^T]$.

Also,

$$[c^{-\frac{1}{2}}(y_0 - x_0\hat{\beta})]^T[c^{-\frac{1}{2}}(y_0 - x_0\hat{\beta})] \sim T^2_{p,n-q}$$

$$\Rightarrow \quad \frac{W}{c} \sim T^2_{p,n-q} \tag{3.20}$$

where

$$T^2_{p,n-q} = \frac{(n-q)p}{n-p-q+1} F_{p,n-p-q+1} .$$

A $(1-\alpha)100\%$ confidence interval for $\hat{x}_0$ is given as all x_0 for which

$$W \leq c\, T^2_{p,n-q}$$

$$\Rightarrow \quad W \leq [1 + x_0(X^TX)^{-1}x_0^T] \frac{(n-q)p}{n-p-q+1} F_{p,n-p-q+1} \tag{3.21}$$

Additionally, an α-level test of :

$$H_0 : x_0 = b$$

$$\text{against} \quad H_1 : x_0 \neq b$$

rejects if

$$(y_0 - b\hat{\beta})^T\hat{\Sigma}_y^{-1}(y_0 - b\hat{\beta}) > \lambda \frac{(n-q)p}{n-p-q+1} F_{p,n-p-q+1}(\alpha) \tag{3.22}$$

where $\lambda = [1 + b(X^TX)^{-1}b^T]$.

3.2.3 THE ORTHOGONAL ESTIMATOR.

In §2.3.1 the orthogonal estimator was developed for the simple regression/calibration case. We now present a generalization for both

the multiple linear regression and multivariate multiple regression situations. Consider the case of a single dependent Y and multiple independent X's as is the case in multiple regression. The three-dimensional situation is depicted in figure 9. In this representation, $\underline{a}_1$ and $\underline{a}_2$ are any two independent vectors in the plane and hence the matrix $\mathbf{A} = [\underline{a}_1, \underline{a}_2]$ represents a basis for the plane. Now, each vector in the column space of A is a linear combination of the columns, with some coefficients $\alpha_1, \ldots, \alpha_n$ which we may write as $\mathbf{A}\alpha$. For all choices of α these vectors in the plane must be perpendicular to the error vector

$\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}$ (see figure 9). Thus

$$(\mathbf{A}\alpha)^T(\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) = 0$$

$$\Rightarrow \quad \alpha^T[\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}} - \mathbf{A}^T\mathbf{b}] = 0$$

therefore

$$\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}} - \mathbf{A}^T\mathbf{b} = 0$$

$$\Rightarrow \quad \bar{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \ .$$

This last equation is the usual least-squares solution. Thus the point b has projection $\mathbf{p} = \mathbf{A}\bar{\mathbf{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ and hence the corresponding projection matrix is $\mathcal{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. The generalization to higher dimensions is straightforward. First assume the data have been centered thus eliminating a constant term in the model. Next consider

$$\mathbf{Y} = \mathbf{X}\beta + \xi$$

where Y is (n x 1) , X is (n x p) and β is (p x 1) i.e.

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \xi_i$$

Now the columns of matrix A form a basis for the plane (in this case in (p+1) dimensional space). Any set of independent vectors will

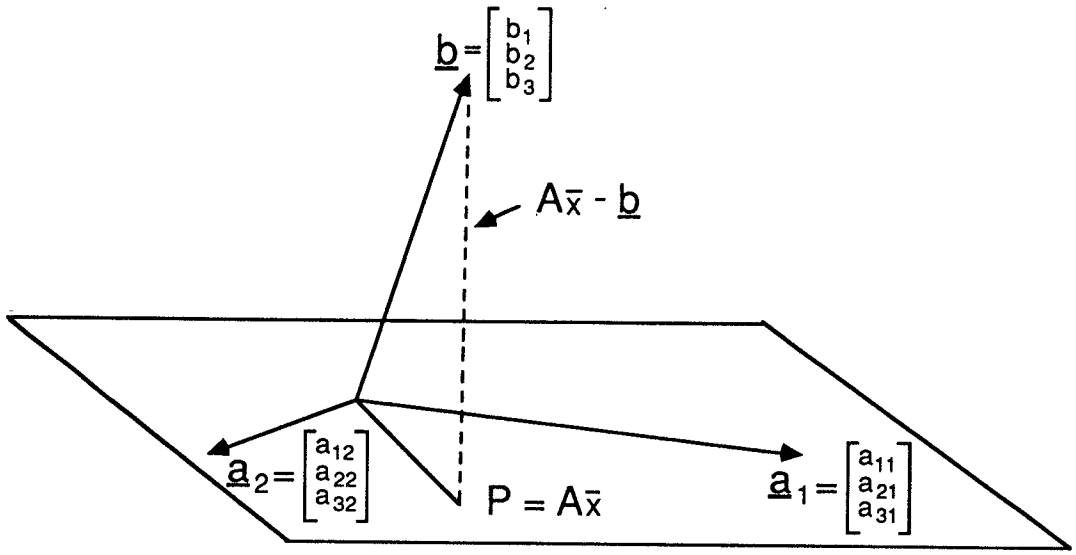

Figure 9. Orthogonal projection of arbitrary point $\underline{b}$ onto plane $A = [\underline{a}_1, \underline{a}_2]$.

do, although we choose a convenient basis as :

$$A = \left[\frac{I_p}{\beta^T}\right]$$

where

$$\beta^T = [\beta_1, \beta_2, \ldots, \beta_p]$$

and I_p is the (pxp) identity matrix.

$$\text{Now } A^TA = [I_p + \beta\beta^T]$$

$$\text{and } (A^TA)^{-1} = [I_p + \beta\beta^T]^{-1} \qquad (3.23)$$

The right-hand side of equation (3.23) can be written as

$$[I_p + \beta\beta^T]^{-1} = I_p - \beta(I + \beta^T\beta)^{-1}\beta^T$$

$$= I_p - \beta(1 + \beta^T\beta)^{-1}\beta^T$$

$$= I_p - \frac{\beta\beta^T}{1+\beta^T\beta} \qquad (3.24)$$

Thus the projection matrix is

$$\mathcal{P} = A(A^TA)^{-1}A^T$$

$$= \left[\frac{I_p}{\beta^T}\right]\left[I_p - \frac{\beta\beta^T}{1+\beta^T\beta}\right]\left[I_p \mid \beta\right]$$

Let $Q = I_p - \frac{\beta\beta^T}{1+\beta^T\beta}$ and therefore

$$\mathcal{P} = \left[\frac{I_p}{\beta^T}\right] Q \left[I_p \mid \beta\right]$$

$$= \left[\frac{Q}{\beta^T Q}\right] \left[I_p \mid \beta\right]$$

thus

$$\underset{(p+1)x(p+1)}{\mathcal{P}} = \left[\begin{array}{c|c} Q & Q\beta \\ \hline \beta^T Q & \beta^T Q \beta \end{array}\right] \tag{3.25}$$

where Q is $(p \times p)$

$Q\beta$ is $(p \times 1)$

and $\beta^T Q \beta$ is (1×1)

The elements of Q are

$$\frac{1}{1 + \sum_{i=1}^{p} \beta_i^2} \begin{bmatrix} 1+\underline{\beta}_{(1)}^T \underline{\beta}_{(1)} & -\beta_1\beta_2 & -\beta_1\beta_3 & \cdots & -\beta_1\beta_p \\ -\beta_2\beta_1 & 1+\underline{\beta}_{(2)}^T \underline{\beta}_{(2)} & -\beta_2\beta_3 & \cdots & -\beta_2\beta_p \\ \vdots & \vdots & \vdots & & \vdots \\ -\beta_p\beta_1 & -\beta_p\beta_2 & -\beta_p\beta_3 & \cdots & 1+\underline{\beta}_{(p)}^T \underline{\beta}_{(p)} \end{bmatrix} \tag{3.27}$$

where $\underline{\beta}_{(i)}$ is β with β_i removed .

Thus, for the situation depicted in figure 9 we have

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \xi_i$$

for which

$$Q = \frac{1}{1+\beta_1^2+\beta_2^2}\begin{bmatrix} 1+\beta_2^2 & -\beta_1\beta_2 \\ -\beta_1\beta_2 & 1+\beta_1^2 \end{bmatrix}$$

$$Q\beta = \frac{1}{1+\beta_1^2+\beta_2^2}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

and

$$\beta^T Q\beta = \frac{1}{1+\beta_1^2+\beta_2^2}(\beta_1^2 + \beta_2^2)$$

Therefore the corresponding projection matrix is

$$\frac{1}{1+\beta_1^2+\beta_2^2}\begin{bmatrix} 1+\beta_2^2 & -\beta_1\beta_2 & \beta_1 \\ -\beta_1\beta_2 & 1+\beta_1^2 & \beta_2 \\ \beta_1 & \beta_2 & \beta_1^2+\beta_2^2 \end{bmatrix}$$

3.2.3.1 Orthogonal Least-Squares estimation.

We now describe a procedure for estimating the coefficients of β such that the sum of the projected vector norms squared is minimized.

Let y_i^* be the orthogonal projection of y_i and x_{ji}^* be the orthogonal projection of x_{ji} where x_{ji} is the i^{th} value for the j^{th} x-variable. It can be shown (see Appendix F) that

$$(y_i^* - y_i) = \frac{\sum_{j=1}^{p} \beta_j x_{ji} - y_i}{1 + \sum_{k=1}^{p} \beta_k^2}$$

and

$$(x^{*}_{ji} - x_{ji}) = \frac{-\sum_{k=1}^{p} \beta_j \beta_k x_{ki}}{1 + \sum_{k=1}^{p} \beta_k^2}$$

Our criterion is to select the β's such that

$$\mathbb{Z} = \sum_{i=1}^{n} \left[\left[\frac{\sum_{j=1}^{p} \beta_j x_{ji} - y_i}{1 + \sum_{k=1}^{p} \beta_k^2} \right]^2 + \left[\frac{-\sum_{k=1}^{p} \beta_j \beta_k x_{ki}}{1 + \sum_{k=1}^{p} \beta_k^2} \right]^2 \right] \qquad (3.28)$$

is a minimum.

The minimization of equation (3.28) requires differentiating $\mathbb{Z}$ with respect to β_j , setting the result to zero and solving the resulting equation for β_j . This process is tedious and will not be presented here although the derivation is given in Appendix G. It turns out that the equations that need to be solved for the β_j's have no closed form solution and one is necessarily forced to use some iterative procedure. An equivalent, although computationally simpler, procedure has been developed and is now described.

In the most general situation Y is (n x p) , X is (n x q) , and β is (q x p). The iterative procedure for the orthogonal least-squares

estimation of β relies on the preceding results applied to a single column of Y at a time to estimate at each stage a column of β. Specifically, the steps involved are as follows :

Step 1 : Obtain an initial estimate using OLS i.e.

$$\hat{\beta}_0 = (X^T X)^{-1} X^T Y$$

Step 2 : For the j^{th} column of $\hat{\beta}_0$:

Compute the projection matrix P_j , where

$$P_j = \left[\begin{array}{c|c} Q_j & Q_j\hat{\beta}_j \\ \hline \hat{\beta}_j^T Q_j & \hat{\beta}_j^T Q_j \hat{\beta}_j \end{array}\right]$$

$$Q_j = \left[I_q - \frac{\hat{\beta}_j \hat{\beta}_j^T}{1 + \hat{\beta}_j^T \hat{\beta}_j}\right]$$

and

$\hat{\beta}_j$ is the j^{th} column of $\hat{\beta}$.

Step 3 : Let $Z = \left[X \mid Y_j \right]$ where Y_j is the j^{th} column of Y. Thus Z is (q+1)x n . The matrix of projected values is then $Z^* = Z P_j$.

Step 4 : Extract the X and Y projections from Z^* as :

$Y_j^* = Z_{(q+1)}^*$ [i.e the $(q+1)^{th}$ column of Z^*]

$X_j^* = [Z_1^* \mid Z_2^* \mid . . . \mid Z_q^*]$

Step 5 : Compute the criterion $\mathbb{Z}$ as :

$$\mathbb{Z} = (Y_j^* - Y)^T(Y_j^* - Y) + \sum_{k=1}^{q} (X_k^* - X_k)^T(X_k^* - X_k)$$
$$= (Y_j^* - Y)^T(Y_j^* - Y) + tr[(X^* - X)^T(X^* - X)]$$

Step 6 : Update $\hat{\beta}$ using Newton-Raphson iteration :

$$\hat{\beta}_j^{(m+1)} = \hat{\beta}_j^{(m)} - \left[\nabla^2 \mathbb{Z}[\hat{\beta}_j^{(m)}] \right]^{-1} \nabla \mathbb{Z}[\hat{\beta}_j^{(m)}]$$

Step 7 : Repeat steps 1 - 6 until convergence established.

Step 8 : Repeat steps 1 - 7 for j = 1,2,...,p components of β and Y.

The complete iterative procedure has been programmed using the matrix-based language GAUSS. A listing is given in Appendix H. We now illustrate the use of all three procedures discussed so far with the use of an illustrative example.

3.2.4 AN EXAMPLE.

The data in table 11 are reproduced from Brown (1982). There are four Y variables representing infrared measurements taken on twenty-one samples of hard wheat. The two X variables are laboratory determinations of %water and %protein for each of the samples. Thus in this example n=21, p=4, and q=2. Assuming no particular ordering existed in the presentation of the data, we have treated the first 15 observations as

Table 11. Twenty-one samples of hard wheat. Y-variables represent infrared reflectance measurements , X_1 is % water and X_2 is % protein.

Observation	Y_1	Y_2	Y_3	Y_4	X_1 %water	X_2 %protein
1	361	108	96	243	9.00	10.73
2	361	107	98	245	8.94	11.05
3	362	110	94	241	9.12	9.86
4	362	105	94	246	9.06	11.41
5	362	104	70	221	10.02	11.57
6	367	113	75	221	10.06	9.42
7	366	108	82	233	9.52	10.93
8	360	104	86	236	9.32	11.61
9	362	113	85	229	9.56	8.82
10	360	103	90	242	9.10	11.81
11	351	97	88	238	9.14	12.33
12	353	95	73	227	9.70	12.93
13	352	97	77	228	9.60	12.69
14	355	96	52	206	10.62	13.13
15	357	106	69	216	10.04	10.41
16	351	93	69	222	10.00	13.57
17	363	113	88	231	9.46	9.26
18	363	110	101	248	8.86	9.82
19	366	96	85	235	9.34	12.85
20	350	96	85	235	9.34	12.85
21	355	97	63	216	10.12	12.81

representing the calibration test data. The last six observations will then be used for calibrating purposes and the methods compared in terms of MSE.

The mean and standard deviation data are as follows :

n=15

	Y_1	Y_2	Y_3	Y_4	X_1	X_2
mean	359.40	104.40	81.93	231.47	9.520	11.247
S.D.	4.82	5.89	12.63	11.77	0.493	1.267

3.2.4.1 The Inverse estimate.

For the inverse regression of X on Y the following matrix of parameter estimates was obtained (N.B. data was centered prior to analysis) :

$$\hat{\gamma} = \begin{bmatrix} 0.0185 & 0.1928 \\ -0.0052 & -0.3945 \\ -0.0131 & 0.1087 \\ -0.0287 & -0.0983 \end{bmatrix}$$

Calibration of observations 16-21 :

Observation #	X_1	$\hat{X}_1$	X_2	$\hat{X}_2$
16	10.00	9.8635	13.57	13.6513
17	9.46	9.4765	9.26	9.2542
18	8.86	8.8350	9.82	10.1783
19	9.78	9.8451	9.46	9.1484
20	9.34	9.2479	12.85	12.7352
21	10.12	10.1673	12.81	12.7825

MSE = 0.0234

3.2.4.2 The Classical estimate.

The matrix of parameter estimates obtained for the regression of Y on X is

$$\hat{\beta} = \begin{bmatrix} 0.0057 & -0.7869 & -24.1135 & -23.5830 \\ -2.7611 & -4.4276 & -2.1461 & -0.0471 \end{bmatrix}$$

Calibration of observations 16-21 :

Observation #	X_1	$\hat{X}_1$	X_2	$\hat{X}_2$
16	10.00	9.8626	13.57	13.6822
17	9.46	9.4784	9.26	9.2259
18	8.86	8.8327	9.82	10.1703
19	9.78	9.8490	9.46	9.1147
20	9.34	9.2449	12.85	12.7595
21	10.12	10.1689	12.81	12.7976

MSE = 0.0250

3.2.4.3 The Orthogonal estimate.

Parameter estimates obtained using the iterative procedure described in §3.2.3.1 are

$$\hat{\beta}' = \begin{bmatrix} -0.4263 & -1.2088 & -24.3542 & -24.1554 \\ -5.0579 & -4.7169 & -2.1356 & -0.0147 \end{bmatrix}$$

Calibration of observations 16-21 :

Observation #	X_1	$\hat{X}_1$	X_2	$\hat{X}_2$
16	10.00	9.8531	13.57	13.5453
17	9.46	9.4433	9.26	8.9732
18	8.86	8.8272	9.82	10.1473
19	9.78	9.8272	9.46	8.9815
20	9.34	9.2482	12.85	12.7452
21	10.12	10.1719	12.81	12.7897

MSE = 0.0389

The above results show little difference between the three methods employed although no inferences should be drawn about the relative merits of each from such a small example. To this end we next present the results of a simulation study similar to that used for the univariate case in Chapter 2.

3.3 A COMPARATIVE STUDY OF POINT ESTIMATORS IN MULTIVARIATE CALIBRATION.

The procedure for producing a simulation to compare the inverse, classical, and orthogonal estimators in terns of bias and MSE is considerably more complex for the multivariate case than for the univariate case. Experimentation with the simulation model has been limited due to two factors : (i) there are many more parameters in the multivariate case which can be varied giving rise to excessively long runs if a full factorial design is implemented ; (ii) each simulation run is very computationally intensive since the orthogonal estimate has to be obtained by iterative methods. Nevertheless a small resolution three fractional factorial experiment has been conducted to examine main effects for a number of factors thought to influence the performance of the various estimators.

The simulation process (for the *multivariate measurement-error model*) is described by the following sequence of steps (NB. Y is a 4-component vector and X is 2-component) :

Step 1 : Data entry - user must specify β matrix coefficients in addition to the following parameters

means and variances for X_1 and X_2 ; measurement error variances $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$; coefficient of variation for Y_1, Y_2, Y_3, and Y_4 ; correlation coefficients for errors in X's , errors in Y's and error in X with error in Y.

Step 2 : Construct covariance matrix Σ_X for the X_1 and X_2 data , where

$$\Sigma_X = \begin{bmatrix} \sigma^2_{X_1} & \rho_{X_1X_2}\sigma_{X_1}\sigma_{X_2} \\ \rho_{X_1X_2}\sigma_{X_1}\sigma_{X_2} & \sigma^2_{X_2} \end{bmatrix}$$

Step 3 : Perform the spectral decomposition of Σ_X :

$\Sigma_X = PD_1P^T$ and let $A = PD_1^{\frac{1}{2}}$.

Step 4 : Generate (n+1) observations from $\underset{(1x2)}{Z} \sim N(0,I)$.

Step 5 : Form matrix of *true* (unobserved) x-values :

$$x = ZA^T + \left[\mu_{X_1} \mid \mu_{X_2} \right] .$$

Step 6 : Compute the vector of expected Y's :

$\mu_Y = \mu_X\beta$ (β is fixed).

Step 7 : Assemble combined error covariance matrix :

$$\Sigma = \left[\begin{array}{c|c} \Sigma_e & \Sigma_{eu} \\ \hline \Sigma_{eu}^{\top} & \Sigma_u \end{array}\right]$$

where Σ_e is the error covariance matrix for the ξ terms in the multivariate model $Y = X\beta + \xi$, Σ_u is the covariance matrix for the measurement-error terms (u), and Σ_{eu} is the cross-covariance matrix between the ξ and u errors.

The $\{i,j\}^{th}$ element of Σ_e is $\rho_{e_i e_j}(c_i\mu_{Y_i})(c_j\mu_{Y_j})$ where $\rho_{e_i e_j}$ is the correlation between the errors in the Y's, c_i and c_j are respectively the coefficients of variation of Y_i and Y_j. The diagonal elements of Σ_u are $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$. The off-diagonal elements are $\rho_{u_1u_2}\sigma_{u_1}\sigma_{u_2}$ where $\rho_{u_1u_2}$ is the correlation between measurement errors for the X_1 and X_2 variables. The $\{i,j\}^{th}$ element of Σ_{eu} is $\rho_{eu}\sigma_{e_i}\sigma_{u_j}$ where ρ_{eu} is the correlation between the errors in Y and the measurement errors in X and the error in Y.

Step 8 : Obtain the spectral decomposition of Σ :

$$\Sigma = QD_2Q^{\top} \quad \text{and let} \quad B = QD_2^{\frac{1}{2}} .$$

Step 9 : Generate (n+1) observations from $Z_{2(1\times 6)} \sim N(0,I)$.

Step 10 : Obtain matrix of errors : $\mathbf{E} = \mathbf{Z}_2\mathbf{B}^T$.

Step 11 : Extract ξ (model error) and U (measurement error) components from E.

Step 12 : Compute the matrix of *observable* x's :

$$\mathbf{X}_i = \mathbf{x}_i + \mathbf{u}_i \qquad i = 1,\ldots,q \ .$$

Step 13 : Compute the matrix of *observable* y's :

$$\mathbf{Y} = \mathbf{x}\beta + \xi \ .$$

Step 14 : Using the first n observations on X and Y, compute all three estimates : inverse ; classical ; and orthogonal.

Step 15 : Using each of the estimated models from step 14 calibrate for $\mathbf{X}_{(n+1)}$ using observation $\mathbf{Y}_{(n+1)}$ and examine difference $[\mathbf{X}_{(n+1)} - \hat{\mathbf{X}}_{(n+1)}]$.

Step 16 : Repeat steps 1-15 for length of simulation run.

3.3.1 RESULTS.

As previously mentioned, comprehensive experimentation with the simulation model is made difficult by virtue of the time and complexity of the model. We present here the results of a small trial in which

various combinations of six factors were investigated. Each of the six factors (described below) assumed either a high or low value for a particular simulation run. Thus a complete factorial experiment would involve 64 treatment combinations. In view of the fact that *one* treatment combination requires N simulations (where N is reasonably large $\geq$ 50 say) and one simulation requires the generation of 6n (n is typically 100 or more) random variates having a prescribed covariance structure and that the orthogonal estimate requires iteration on *each column* of $\hat{\beta}$, the need for fractional replication becomes apparent. To this end a one-eighth fractional factorial design has been employed. Since this is a resolution three design only unbiased estimates of main effects can be examined. It is well recognized that interaction effects are likely to be significant determinants of the efficiency and bias of the estimators considered. This is an aspect of the study that will require future investigation.

Summary of simulation input data :

Fixed parameters :

$$\beta = \begin{bmatrix} 0.5 & 2.0 & 0.1 & 1.0 \\ 0.5 & -1.0 & 0.9 & -2.0 \end{bmatrix}$$

$$\mu_{X_1} = \mu_{X_2} = 10.0$$

$$\Rightarrow \quad \mu_Y = \mu_X \beta = [10.0 \quad 10.0 \quad 10.0 \quad 10.0]$$

Variable parameters (factors) :

Factor #	Description	High value	Low value
1	Coefficient of variation for Y data.	0.20	0.01
2	Measurement error variances.	1.00	0.10
3	Variance of x data.	2.00	0.50
4	Correlation between measurement errors.	0.90	0.20
5	Correlation between residual error terms.	0.90	0.20
6	Correlation between measurement error and residual error.	0.50	-0.50

Design matrix :

Factor ->	1	2	3	4	5	6
Run						
1	0	0	0	0	0	0
2	0	0	1	1	1	1
3	1	1	0	0	1	1
4	1	1	1	1	0	0
5	0	1	0	1	0	1
6	0	1	1	0	1	0
7	1	0	0	1	1	0
8	1	0	1	0	0	1

Where a zero indicates corresponding factor is at low level and a one indicates the high level.

For each of the eight runs fifty simulations were performed. For each simulation 101 observations for the X and Y data were generated - the first one-hundred being used for parameter estimation and the last

observation used for calibration. A listing of the complete GAUSS program may be found in Appendix I. Table 12 lists the bias and MSE data that was obtained for this simulation experiment. With reference to table 12 the following comments can be made :

(i) with respect to bias , the orthogonal estimator performed best overall with the classical estimator having the worst bias. It is readily apparent that the classical estimator consistently underestimates by about 1.5% of the true calibrated value. The inverse and orthogonal estimators have biases of about 1/10th of this.

(ii) The inverse estimator is demonstrably better in terms of MSE. Again the classical estimator fared worst of all. Particularly high MSE's were observed for the classical and (to a lesser extent) the orthogonal estimators for the second treatment combination. Relatively high MSE's were also observed for the last treatment combination. Both of these treatment combinations are characterized by small measurement error, highly correlated residual errors, and a high correlation between measurement error and residual error.

Since the above experimental design leaves only one degree of freedom (if the mean is included) for the error term, it would be

Table 12. Comparison of the inverse, classical, and orthogonal estimators in terms of bias and MSE for the discrete calibration example.

	Bias			MSE		
	Inverse	Classical	Orthogonal	Inverse	Classical	Orthogona
1	-0.0060	-0.0094	-0.0048	0.0250	0.0263	0.0246
2	-0.1079	-0.6832	0.1109	2.3233	68.0956	30.3368
3	0.0040	-0.0148	0.0003	0.0197	0.0914	0.0217
4	0.0089	-0.1491	-0.0385	0.5088	1.8884	1.2300
5	-0.0180	-0.0017	0.0005	0.0376	0.0423	0.0358
6	-0.0180	-0.0026	0.2497	0.4451	1.5781	7.4686
7	0.0027	0.0028	0.0029	0.0016	0.0015	0.0015
8	-0.2279	-0.3532	-0.2341	2.0321	2.7239	2.9503

Overall :	Inverse	Classical	Orthogonal
mean bias	-0.0175	-0.1514	0.0109
av. MSE	0.6742	9.3059	5.2587

tenuous to try and draw any statistical inference from these results. However the magnitude of estimated effects does provide some insight as to possibly significant factors. It would appear that the inverse estimator is the most stable and is probably affected mostly by measurement-error variance ,residual error variance ,and the correlation between the two. On the other hand there is evidence to suggest that the classical and orthogonal estimators are affected by all six factors considered in this experiment. Interactions are probably likely to play a significant role, although such information on such effects is not admissible from this design.

3.4 CONDITIONAL MULTIVARIATE CALIBRATION.

The idea now is that when calibrating for X_0 it may be that some of the components of X_0 are fixed in advance and we thus wish to modify the procedure so that not only will $\hat{X}_0$ have the required fixed or known values, but also that the remaining (free) components of $\hat{X}_0$ be such that $\hat{X}_0$ as a whole has the correct covariance structure. Development of this aspect of statistical calibration was motivated by procedures developed for conditional simulations by Professor L. Borgman at the University of Wyoming [For example, see Borgman (1982).

Before addressing the conditional calibration problem we first present some standard results for the conditional multivariate normal distribution.

3.4.1 THE CONDITIONAL NORMAL P.D.F.

Let $\mathbf{X}$ and $\mathbf{Y}$ be two random vectors such that $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim N(\mu, \mathbb{C})$ where

$$\mu = \mathbb{E}\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{and} \quad \mathbb{C} = \text{Cov}\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{bmatrix} .$$

Let $\tilde{\mathbf{Y}}$ be the random vector $\mathbf{Y}|\mathbf{X}=\mathbf{x}$ where $\mathbf{x}$ is a vector of deterministic scalars.

It is shown in Appendix J that

$$\mathbb{E}[\tilde{\mathbf{Y}}] = C_{12}^T C_{11}^{-1}(\mathbf{x} - \mu_x) + \mu_y \tag{3.29}$$

and

$$\text{Cov}[\tilde{\mathbf{Y}}] = C_{22} - C_{12}^T C_{11}^{-1} C_{12} \tag{3.30}$$

We now utilize these results to calibrate conditionally in the case where the data is multivariate normal.

Let $\hat{\mathbf{X}}_0^T$ be partitioned as :

$$\hat{\mathbf{X}}_0^T = \begin{bmatrix} \hat{\mathbf{S}} \\ \hline \hat{\mathbf{T}} \end{bmatrix}$$

where $\hat{\mathbf{S}}$ has ν_1 components and $\hat{\mathbf{T}}$ has ν_2 components and $\nu_1+\nu_2=q$.

Furthermore, let $\mathbb{E}[\hat{\mathbf{X}}_0^T] = \begin{bmatrix} \mu_s \\ \mu_t \end{bmatrix}$ and $\text{Cov}[\hat{\mathbf{X}}_0^T] = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{bmatrix}$ and let $\tilde{\mathbf{T}}$ be the vector $\hat{\mathbf{T}}|\hat{\mathbf{S}}=\mathbf{s}$.

Thus, using equations (3.29) and (3.30) we have that

$$\mathbb{E}[\tilde{T}] = C_{12}^{T} C_{11}^{-1} (s - \mu_s) + \mu_t \qquad (3.31)$$

and

$$\mathrm{Cov}[\tilde{T}] = C_{22} - C_{12}^{T} C_{11}^{-1} C_{12} \qquad (3.32)$$

We are now in a position to identify the steps involved in calibrating conditionally.

3.4.2 THE CONDITIONAL CALIBRATION PROCEDURE.

Step 1:

Conduct the *unconditional* calibration to obtain $\hat{X}_0^T$ using either equation (3.5) or (3.15). For the MLE (classical) estimator we have

$$\hat{X}_0^T = (\hat{\beta}\, \hat{\Sigma}_y^{-1}\, \hat{\beta}^T)^{-1} \hat{\beta}\, \hat{\Sigma}_y^{-1} y_0^T \qquad (3.33)$$

(where $\hat{\Sigma}_y$ denotes the estimated residual error-covariance matrix for the regression of Y on X).

Partition $\hat{X}_0^T$ as : $\hat{X}_0^T = \begin{bmatrix} \hat{S} \\ \hat{T} \end{bmatrix}$.

Step 2:

For given S=s compute :

$$\tilde{T} = C_{12}^{T} C_{11}^{-1} (s - \hat{S}) + \hat{T} \qquad (3.34)$$

where $\hat{S}$ and $\hat{T}$ are obtained from the *unconditional* calibration.

Observe that our ability to calibrate conditionally is dependent upon us having knowledge of $Cov[\hat{X}_0]$ .In order that we may develop an expression for $Cov[\hat{X}_0]$ it will be necessary for us to digress momentarily to consider the regression of X on Y.

3.4.2.1 The regression of X on Y.

As noted by Brown (1982) and others, when both X and Y are multivariate normally distributed the use of classical and inverse estimators are equally valid. Using the inverse estimator of equation (3.5) we obtain the calibrated vector $\hat{X}_0^*$ directly from the estimated regression model. Furthermore, after suitably transforming both $\hat{X}_0^*$ and $\hat{X}_0$, the resulting quantities are in fact equivalent up to a constant of proportionality and that the constants are the cannonical correlations between X and Y. We now develop these ideas further.

In what follows we assume that X and Y have been centered.
Recall, the inverse regression model is :

$$X = Y\gamma + \zeta$$

where the dimensions of X,Y, and γ are respectively (n x q), (n x p), and (p x q). The estimator, $\hat{\gamma}$ is obtained via OLS or GLS as in equation (3.3).

Thus, given a new (1 x p) vector y_0, we obtain $\hat{X}_0^*$ as :

$$\hat{X}_0^{*T} = X^T Y (Y^T Y)^{-1} y_0^T \qquad (3.35)$$

After some algebraic manipulation it can be shown that :

$$\hat{X}_0^{*T} = [(X^TX)^{-1} + \hat{\beta}\,\hat{\Sigma}^{-1}\,\hat{\beta}^T]^{-1}(\hat{\beta}\,\hat{\Sigma}_y^{-1}\,\hat{\beta}^T)\hat{X}_0^T \qquad (3.36)$$

where $\hat{\beta}$ is obtained from the regression of Y on X using equation (3.9) and

$$\hat{\Sigma}_y = Y^TY - \hat{\beta}^TX^TX\hat{\beta} \qquad (3.37)$$

Next, let $W = (X^TX)^{\frac{1}{2}}\,\hat{\beta}\,\hat{\Sigma}_y^{-\frac{1}{2}}$

$$\Rightarrow W\,W^T = (X^TX)^{\frac{1}{2}}\,\hat{\beta}\,\hat{\Sigma}_y\,\hat{\beta}^T(X^TX)^{\frac{1}{2}} \qquad (3.38)$$

Using the Binomial Inverse Theorem [Press (1972),p.23] it can also be shown that

$$(X^TX)^{-\frac{1}{2}}\hat{X}_0^{*T} = [I_q + W\,W^T]^{-1}W\,W^T(X^TX)^{-\frac{1}{2}}\hat{X}_0^T \qquad (3.39)$$

Let $\Omega = [I_q + W\,W^T]^{-1}$

$\Rightarrow \Omega\,[I_q + W\,W^T] = I$

and $\Omega + \Omega(W\,W^T) = I$

hence $\Omega(W\,W^T)^{-1} + \Omega = (W\,W^T)^{-1} \quad \Rightarrow \quad \Omega\,[(W\,W^T)^{-1} + I_q] = (W\,W^T)^{-1}$

$$\Rightarrow \Omega = (W\,W^T)^{-1}\,[(W\,W^T)^{-1} + I_q]^{-1} \qquad (3.40)$$

However, since $(W\,W^T)$ and I_q are both symmetric the product in equation (3.40) is commutative and hence we may write :

$$\Omega = [I_q + (W\,W^T)^{-1}](W\,W^T)^{-1} \qquad (3.41)$$

Now from equation (3.39) we have

$$(X^TX)^{-\frac{1}{2}}\hat{X}_0^{*T} = [I_q + W W^T]^{-1} W W^T (X^TX)^{-\frac{1}{2}}\hat{X}_0^T$$

$$= [I_q + (W W^T)^{-1}]^{-1}(W W^T)^{-1} W W^T (X^TX)^{-\frac{1}{2}}\hat{X}_0^T$$

$$= [I_q + (W W^T)^{-1}]^{-1}(X^TX)^{-\frac{1}{2}}\hat{X}_0^T \qquad (3.42)$$

Let $U = (X^TX)^{-\frac{1}{2}}X^TY(Y^TY)^{-\frac{1}{2}}$

$$= \Sigma_x^{-\frac{1}{2}}\Sigma_{xy}\,\Sigma_y^{-\frac{1}{2}} \qquad (3.43)$$

Thus the positive eigenvalues of U are the cannonical correlations between X and Y.

Now, $\hat{\Sigma}_y = Y^TY - Y^TX(X^TX)^{-1}X^TY$ (3.44)

Letting $V = (Y^TY)^{-\frac{1}{2}}\hat{\Sigma}_y\,(Y^TY)^{-\frac{1}{2}}$ we have

$$V^{-1} = (Y^TY)^{\frac{1}{2}}\hat{\Sigma}_y^{-1}(Y^TY)^{\frac{1}{2}} \qquad (3.45)$$

Furthermore, it is relatively easy to show that

$$V = I - U^TU \qquad (3.46)$$

Next consider $(I + W W^T)^{-1} W W^T$. In the most general case where X is (n x q) and Y is (n x p), U will be a (q x p) matrix and V is a (p x p) matrix.

Now

$$(I + W W^T)^{-1} W W^T = (I + UV^{-1}U^T)^{-1} UV^{-1}U^T$$

Using the Binomial Inverse theorem we have

$$(I + UV^{-1}U^T)^{-1} = [I - UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}V^{-1}U^T]$$

and hence

$$\begin{aligned}
(I + UV^{-1}U^T)^{-1}UV^{-1}U^T &= [I - UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}V^{-1}U^T]UV^{-1}U^T \\
&= UV^{-1}U^T - UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}V^{-1}U^TUV^{-1}U^T \\
&= UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}(V^{-1} + V^{-1}U^TUV^{-1})U^T \\
&\quad - UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}V^{-1}U^TUV^{-1}U^T \\
&= UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}[(V^{-1} + V^{-1}U^TUV^{-1})U^T - V^{-1}U^TUV^{-1}U^T] \\
&= UV^{-1}(V^{-1} + V^{-1}U^TUV^{-1})^{-1}V^{-1}U^T \\
&= UV^{-1}[V^{-1}(I + U^TUV^{-1})]^{-1}V^{-1}U^T \\
&= UV^{-1}(I + U^TUV^{-1})^{-1}U^T \\
&= U(V + U^TU)^{-1}U^T
\end{aligned}$$

but by equation (3.46), $V + U^TU = I$. Thus substituting into the last expression we finally have that

$$(I + W W^T)^{-1} W W^T = (I + UV^{-1}U^T)^{-1}UV^{-1}U^T = UI^{-1}U^T = UU^T \qquad (3.47)$$

Thus the eigenvalues of $(I + W W^T)^{-1} W W^T$ are the same as those of UU^T. However, as previously noted, the eigenvalues of U are the cannonical correlations between X and Y and hence the eigenvalues of UU^T are the corresponding correlations squared.

From equation (3.39) we have :

$$(X^T X)^{-\frac{1}{2}} \hat{X}_0^{*T} = [I_q + (W W^T)^{-1}]^{-1} (X^T X)^{-\frac{1}{2}} \hat{X}_0^T$$

$$\Rightarrow \quad [I_q + (W W^T)^{-1}] (X^T X)^{-\frac{1}{2}} \hat{X}_0^{*T} = (X^T X)^{-\frac{1}{2}} \hat{X}_0^T \qquad (3.48)$$

Let $[I_q + (W W^T)^{-1}]$ have the spectral decomposition QDQ^T where the columns of Q are the normalized eigenvectors of $[I_q + (W W^T)^{-1}]$ and $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$ be a diagonal matrix of eigenvectors. Thus equation (3.48) can be written as :

$$QDQ^T (X^T X)^{-\frac{1}{2}} \hat{X}_0^{*T} = (X^T X)^{-\frac{1}{2}} \hat{X}_0^T$$

$$\Rightarrow \quad DQ^T (X^T X)^{-\frac{1}{2}} \hat{X}_0^{*T} = Q^T (X^T X)^{-\frac{1}{2}} \hat{X}_0^T \qquad (3.49)$$

If we let $\Delta = Q^T (X^T X)^{-\frac{1}{2}}$ we obtain :

$$D \Delta \hat{X}_0^{*T} = \Delta X_0^T \qquad (3.50)$$

In otherwords, after applying the same transformation (namely Δ) to both $\hat{X}_0^*$ and $\hat{X}_0$, we find they only differ by constants $(\lambda_1, \lambda_2, \ldots, \lambda_q)$ which, as we have shown, are the squared cannonical correlations between X and Y. We now utilize this fact to derive the covariance information for $\hat{X}_0$.

3.4.2.2 The covariance of $\hat{X}_0$.

For a given y_0 vector we can use the previous results to obtain $\hat{X}_0 = [\hat{x}_1, \hat{x}_2, \ldots \hat{x}_q]$. Observe that $\hat{X}_0$ is a random vector having some mean vector μ_{X_0} and covariance matrix Σ_{X_0}. In general, it will not be possible to estimate Σ_{X_0} using a sample covariance matrix since there will only be one q-vector $\hat{X}_0$. As an alternative, we use the relationship between $\hat{X}_0$ and $\hat{X}_0^*$ established in the previous section.

Returning to the regression of X on Y :

$$X = Y\gamma + \zeta$$

thus

$$\hat{X}_0^* = y_0\hat{\gamma}$$

where

$$\hat{\gamma} = (Y^TY)^{-1}Y^TX$$

and

$$\text{Cov}[\hat{X}_0^*] = y_0(Y^TY)^{-1}y_0^T\,\Sigma_x \qquad (3.51)$$

where Σ_x is the residual error-covariance matrix for the regression of X on Y. Now from equation (3.50) we have :

$$D\,\Delta\,\hat{X}_0^{*T} = \Delta\,X_0^T$$

$$\Rightarrow \quad \hat{X}_0 = \hat{X}_0^*\,\Omega \qquad \text{where} \quad \Omega = \Delta\,D\,\Delta^{-1}.$$

and thus :

$$\text{Cov}[\hat{X}_0] = \hat{\Sigma}_{X_0} = \Omega\,\hat{\Sigma}_{X_0}\,\Omega^T \qquad (3.51)$$

Thus, by performing the regression of X on Y we can obtain an estimate of $Cov[\hat{X}_0]$. Knowledge of $Cov[\hat{X}_0]$ allows us to calibrate conditionally using the procedure outlined in section 3.4.2.

It should be pointed out that we are not necessarily advocating the use of the classical estimator over the seemingly more direct method of regressing X on Y (which as we have seen, needs to be performed anyway if estimates of $Cov[\hat{X}_0]$ are to be obtained). All we are attempting to do is to provide a reasonable mechanism for calibrating conditionally *given* that one wishes to use the maximum likelihood estimate given by equation (3.15). We now illustrate the procedure using the previously considered wheat data. A comparison of the methods will also be provided.

3.4.2.3 An example.

The results of the previous sections are now applied to the wheat data reported in table 11.

Note : (i) only the first 15 observations in table 11 will be used for the purpose of model fitting. The remaining data will be used for calibration.

(ii) all data is centered prior to analysis.

Preliminary data analysis

A cannonical correlation analysis of the data (n=15) gave the following (standardized) cannonical variates :

Reflectivity : $\phi_1 = -0.3541Y_1 + 0.5195Y_2 + 0.0397Y_3 + 0.8691Y_4$

$\phi_2 = -0.6524Y_1 + 1.7849Y_2 - 1.2006Y_3 + 0.6315Y_4$

Composition : $\tau_1 = -0.9320X_1 - 0.2513X_2$

$\tau_2 = 0.3914X_1 - 0.9791X_2$

The cannonical correlations are : r_1=0.998067 and r_2=0.992321. The correlations squared (also the eigenvalues of equation 3.15). are λ_1=0.996138 and λ_2=0.984702.

For the regression of Y on X we obtain :

$$\hat{\alpha}^T = (359.40,\ 104.40,\ 81.933,\ 231.467)$$

$$\hat{\beta} = \begin{bmatrix} 0.0027 & -0.3747 & -11.4830 & -11.2304 \\ -3.3795 & -5.4193 & -2.6268 & -0.0576 \end{bmatrix}$$

$$\hat{\Sigma}_y = \begin{bmatrix} 154.323 & 68.159 & 30.857 & 71.739 \\ 68.159 & 34.082 & 16.089 & 28.259 \\ 30.857 & 16.089 & 19.606 & 19.853 \\ 71.739 & 28.259 & 19.853 & 45.024 \end{bmatrix}$$

$$W = \begin{bmatrix} 6.0425 & -2.2184 & -9.1563 & -10.6294 \\ 2.7771 & -8.5708 & -1.4343 & -0.8891 \end{bmatrix}$$

$$Q = \begin{bmatrix} -0.318778 & 0.947829 \\ 0.947829 & 0.318778 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.9847 & 0.0000 \\ 0.0000 & 0.9961 \end{bmatrix} \qquad \Delta = \begin{bmatrix} -0.101054 & 0.252790 \\ 0.240629 & 0.064895 \end{bmatrix}$$

$$\Omega^{T} = \begin{bmatrix} 1.005046 & -0.004193 \\ -0.002829 & 1.014407 \end{bmatrix}$$

Regression of X on Y

$$\hat{\gamma} = \begin{bmatrix} 0.038935 & 0.157483 \\ -0.010858 & -0.322290. \\ -0.027428 & 0.088771. \\ -0.060185 & -0.080345. \end{bmatrix}$$

$$\hat{\Sigma}_{x} = \begin{bmatrix} 0.068538 & -0.030882 \\ -0.030882 & 0.203757 \end{bmatrix}$$

Calibration of observation #16.

For this observation we have :

$$y_0 = [351, 93, 69, 222]$$

From the regression of X on Y we obtain :

$\hat{\bar{X}}_0^* = [0.72123, 1.96381]$ (NB: these are centered values)

The calibration data yield : $\bar{X}_1 = 9.520$ and $\bar{X}_2 = 11.247$. Adding these to the respective components of $\hat{\bar{X}}_o^*$ we obtain

$\hat{X}_o^* = [10.241, 13.211]$.

Applying equation (3.50) gives $\hat{\bar{X}}_o = [0.73991, 1.198538]$ from which we obtain

$\hat{X}_o = [10.260, 13.232]$ (NB: these are the unconditional estimates).

$$\text{Now, } \hat{\Sigma}_{X^*_o} = \begin{bmatrix} 0.0308109 & -0.0138828 \\ -0.0138828 & 0.0915984 \end{bmatrix}$$

from which we compute :

$$\hat{\Sigma}_{X_o} = \begin{bmatrix} 0.0312553 & -0.0138411 \\ -0.0138411 & 0.0937138 \end{bmatrix}$$

Finally, *given* that % water for this observation is 10.00 we obtain the conditionally calibrated value of % protein as 13.3498. This represents a 40% reduction in error when compared with the actual value of 13.57 and the value of 13.2105 obtained from the regression of X on Y.

The procedure has been applied to all six observations *not* used as part of the model fitting exercise. These results are summarized in table 13. Overall, the conditional calibration of %protein using the procedure outlined in this paper has resulted in an approximate 30% reduction of mean square error when compared with the unconditional results.

3.5 CALIBRATION IN UNIVARIATE MULTIPLE REGRESSION MODELS

The problem of calibrating in a univariate multiple regression setting perhaps arises less frequently than the other cases considered thus far, although is no less deserving of formal investigation. The situation presently referred to is when there is a *single* Y vector and two or more regressor or explanatory variables. This would appear to be a special case of the multivariate methods already considered thus requiring no further discussion.

Table 13. Comparison of the various calibration methods for the multivariate conditional calibration of % protein.

Calibration of %Protein | %water.

Observation.	Unconditional (inverse)	Conditional (classical)	Unconditional (classical)	Actual
16	13.2320	13.3498	13.2105	13.57
17	9.6031	9.5842	9.6180	9.26
18	10.3569	9.9951	10.3730	9.82
19	9.5205	9.7108	9.5316	9.46
20	12.4693	12.2924	12.4620	12.85
21	12.5197	12.8733	12.5007	12.81

Summary

	Average error	mean square error
inverse	-0.0114	0.1255
conditional	0.0059	0.0937
classical	-0.0124	0.1358

However, this case presents special problems of its own and as such is given separate treatment in this section.

Consider the model :

$$Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_q x_{qi} + \xi_i \qquad (3.53)$$

Without loss of generality, assume we wish to calibrate for X_{q0} on the basis of a future observed value of Y , y_0 . A number of possibilities exist :

(i) perform the simple regression of Y on X_q and use the classical estimator of equation (1.2).

(ii) perform the simple regression of X_q on Y and use the inverse estimator of equation (1.3).

(iii) perform the *multiple* regression of Y on $X_1, X_2, \ldots, X_q$ and use the classical estimator (i.e. algebraically manipulate the estimated regression equation so as to express $\hat{X}_q$ as a function of the other variables.

(iv) perform the *multiple inverse regression* :

$$X_q = \lambda_0 + \lambda_1 X_1 + \ldots + \lambda_{q-1} X_{q-1} + \text{error}.$$

(v) use all the available information and the multivariate methods presented in this chapter.

Each of the suggestions (i) - (v) has inherent difficulties. These are now explored in turn. Suggestions (i) and (ii) use only one regressor and ignore the information contained in the remaining variables. Such approaches should only be entertained when no additional information is gained by incorporating all q variables - in other words when the covariance matrix of $[Y \mid X_1, X_2, \ldots, X_{q-1}]$ is a diagonal matrix. This is unlikely to arise in practice. Suggestion (iii), whilst utilizing all the available information also poses problems. If the classical estimator is to be used then the calibrated X_{q0} will be obtained as

$$\hat{X}_{q0} = \frac{y_0 - \sum_{j=1}^{q-1} \hat{\beta}_j x_{j0}}{\hat{\beta}_q} \qquad (3.54)$$

If there is some degree of multicollinearity present among the regressor variables then $\hat{\beta}_q$ may assume values very close to zero in which case the above estimate (and its variance) are grossly inflated. In any event, the estimator of equation (3.54) suffers from the same problems of undefined expectation and infinite variance as for the simple case discussed in Chapter I. As an alternative to the use of equation (3.54) one may use the multivatiate inverse estimator of equation (3.2). However this also leads to inefficient calibration for the following reason. In calibrating for X_{q0} we presumably have observations on the remaining X's as well as y_0 . In using the inverse estimator we effectively calibrate for the entire vector of X's and then

discard all but $\hat{X}_{q0}$. It is intuitively obvious that the quality of calibration could be improved by examining how well we have calibrated for the X's not of interest and on the basis of this comparison "readjust" the calibrated $\hat{X}_{q0}$. This procedure forms the basis of the estimator proposed in the following section. The classical multivariate estimator of equation (3.15) is not particularly useful in the present context. In this case β is a (q x 1) vector and hence the inverse $(\beta \Sigma_y^{-1} \beta^T)^{-1}$ in equation (3.15) does not exist since the (q x q) matrix inside the parentheses is rank one. A g-inverse could be used although problems of uniqueness then arise. Suggestion (iv) is mentioned by Madddula (1988) although Goldberger (1984) notes that all of the λ's are biased and recommends against the use of this approach.

We now examine a means of calibration for the present situation which utilizes all of the available calibration data and avoids the problems mentioned above.

3.5.1 THE CONDITIONAL INVERSE ESTIMATOR FOR MULTIPLE REGRESSION

We commence with the inverse regression of X on Y

$$X = Y\gamma + \zeta$$

where the dimensions of X are (n x q) , Y is (n x 1) and γ is (q x 1). Furthermore, it will be assumed that the ζ are distributed with zero mean and covariance matrix Σ_x . Now the OLS estimate of γ is

$\hat{\gamma} = (Y^T Y)^{-1} Y^T X$ with $\mathrm{Cov}[\hat{\gamma}_i, \hat{\gamma}_j] = \zeta_{ij} (Y^T Y)^{-1} = \zeta_{ij} \left[\sum_{i=1}^{n} y_i^2 \right]^{-1}$ where ζ_{ij} is the $(i,j)^{th}$ element of Σ_x.

Now,

$$\mathrm{Cov}[\hat{X}_{i0}, \hat{X}_{j0}] = \mathrm{Cov}[y_0 \hat{\gamma}_i, y_0 \hat{\gamma}_j]$$

$$= y_0^2 \left[\sum_{i=1}^{n} y_i^2 \right]^{-1} \zeta_{ij} \qquad (3.55)$$

Thus,

$$\mathrm{Cov}[\hat{X}_0] = \frac{y_0^2}{\sum_{i=1}^{n} y_i^2} \Sigma_x \qquad (3.56)$$

We next apply the reults presented in §3.4 for conditional calibration. First partition $\hat{X}_0$ as

$$\hat{X}_0' = \left[\begin{array}{c} \hat{X}_{0(q)}^T \\ \hline \hat{x}_{0q} \end{array} \right]$$

where $\hat{X}_{0(q)}^T$ is a (q-1) x 1 vector obtained by deleting $\hat{x}_{0q}$.

Let $\mathrm{Cov}[\hat{X}_0'] = \mathbb{C} = \left[\begin{array}{c|c} C_{11} & C_{12} \\ \hline C_{12}^T & C_{22} \end{array} \right]$

where C_{11} is the (q-1) x (q-1) covariance matrix of $[\hat{X}_{01},\hat{X}_{02},\ldots,\hat{X}_{0q-1}]$; C_{12} is the (q-1) x 1 covariance matrix of $\hat{X}_{i0}$ and $\hat{x}_{0q}$; and C_{22} is the (1 x 1) covariance matrix of $\hat{x}_{0q}$ i.e the variance of $\hat{x}_{0q}$. Next, let $\tilde{X}_0$ be $\hat{x}_{0q} \Big| \hat{X}^T_{0(q)} = x^T_{0(q)}$ where $x^T_{0(q)}$ is a vector of deterministic scalars. From the results of §3.4.1 we have

$$\mathbb{E}[\tilde{X}_0] = C^T_{12}C^{-1}_{11}(x^T_{0(q)} - \hat{X}^T_{0(q)}) + \mathbb{E}[\hat{x}_{0q}] \qquad (3.57)$$

and

$$\mathrm{Cov}[\tilde{X}_0] = C_{22} - C^T_{12}C^{-1}_{11}C_{12} = \mathrm{Var}[\tilde{X}_0] \qquad (3.58)$$

Thus the procedure uses equation (3.57) as follows :

(i) perform the regression of X on Y and estimate γ

(ii) using the residuals from the fitted model, estimate Σ_x and hence estimate $\mathrm{Cov}[\hat{X}_0]$ using equation (3.56) with $\hat{\Sigma}_x$ replacing Σ_x .

(iii) partition $\hat{\Sigma}_x$ and identify C_{11}, C_{12}, and C_{22}.

(iv) given new post-calibration observations on $y_0, x_{01}, x_{02}, \ldots, x_{0q-1}$,obtain the *unconditional* calibrated vector $\hat{X}_0$.

(v) partition $\hat{X}_0$ as indicated earlier.

(vi) the *conditionally* calibrated $\tilde{\mathbf{x}}_0$ is obtained using equation (3.57) with $\hat{\mathbf{x}}_{0(q)}^T$ replaced by the actual post-calibration data, $\hat{\mathbf{X}}_{0(q)}^T$ replaced by the *unconditional* estimate obtained from step (iv), and $\mathbb{E}[\hat{\mathbf{x}}_{0q}]$ replaced by the *unconditional* estimate.

Observe that the variance of the conditionally calibrated value is equal to $C_{22} - C_{12}^T C_{11}^{-1} C_{12}$. The quantity C_{22} is in fact the variance that would result if one performed the simple inverse regression of X_q on Y. The present method utilizes all the available information and "adjusts" the calibrated x_{0q} on the basis of a comparison of the calibrated $X_{01}, X_{02}, \ldots, X_{0q-1}$ with the actual post-calibration data. The actual reduction in variance is thus equal to $C_{12}^T C_{11}^{-1} C_{12}$ which is always positive. At *worst* the conditional method will yield an estimator whose variance is *no larger* than that obtained by performing the simple inverse regression of X_q on Y. This situation would arise when the X's are mutually independent in which case $C_{12} = 0$ and hence the reduction in variance is zero. Thus, compared to simple inverse regression, the conditional procedure has an efficiency of

$$\text{efficiency} = \frac{C_{12}^T C_{11}^{-1} C_{12}}{C_{22}} \qquad (3.59)$$

[NB: C_{22} is a scalar]

We now illustrate the preceding methods with the use of a small example.

3.5.1.2 An example : univariate multiple calibration

The data in table 14 are taken from Fuller (1987,p65) and report on the hectares of corn for 37 area segments (approximately 250 hectares) in north-central Iowa obtained using three different methods by the U.S. Department of Agriculture. The three methods were digitized aerial photography, satellite imagery, and personal interview with the farm operator. Observations 1 to 27 will be used for model estimation while the remaining 10 observations will be used for calibration purposes.

For observations 1 to 27 in table 14 we compute the means and standard deviations :

	X_1	X_2	Y
mean	122.96	133.68	119.79
Std.Dev.	34.56	32.14	31.36

Each of these means was subtracted from the respective X and Y data (for both the calibration and post-calibration data) prior to analysis.

For the inverse regression of X on Y we obtain :

$$\hat{\gamma} = [1.0867 \quad 0.8199]$$

and

$$\hat{\Sigma}_x = \begin{bmatrix} 31.7349 & 6.7937 \\ 6.7937 & 358.0830 \end{bmatrix}$$

Table 14. Hectares of corn determined by three methods.

Segment	Photograph (X_1)	Satellite (X_2)	Interview (Y)
1	167.14	168.30	165.76
2	159.04	162.45	162.08
3	161.06	129.60	152.04
4	163.49	166.05	161.75
5	97.12	94.05	96.32
6	123.02	140.85	114.12
7	111.29	110.70	100.60
8	132.33	158.85	127.88
9	116.95	121.95	116.90
10	89.84	106.65	87.41
11	84.17	99.00	88.59
12	88.22	153.00	88.59
13	161.87	159.75	165.35
14	106.03	117.45	104.00
15	87.01	84.15	88.63
16	159.85	157.50	153.70
17	209.63	194.40	185.35
18	122.62	165.15	116.43
19	93.08	99.45	93.48
20	120.19	166.05	121.00
21	115.74	154.35	109.91
22	125.45	153.90	122.66
23	99.96	132.30	104.21
24	99.55	92.70	92.88
25	163.09	142.20	149.94
26	60.30	65.25	64.75
27	101.98	113.40	99.96
28	138.40	131.85	140.43
29	94.70	92.70	98.95
30	129.50	135.90	131.04
31	132.74	159.75	127.07
32	133.55	132.75	133.55
33	83.37	100.35	77.70
34	78.51	113.85	76.08
35	205.98	206.55	206.39
36	110.07	130.50	108.33
37	134.36	138.15	118.17

Using the above results we can calibrate for the unused portion of the data. For comparative purposes both the simple inverse estimate and the conditional inverse estimate will be computed. These results are summarized below .

Calibration of X_2 (satellite data) :

Segment	actual	simple inverse	conditional
28	131.85	150.6034	149.1069
29	92.70	116.5926	115.3911
30	135.90	142.9042	141.6870
31	159.75	139.6491	140.0491
32	132.75	144.9623	144.0281
33	100.35	99.1690	100.4858
34	113.85	97.8407	98.4940
35	206.55	204.6862	202.3116
36	130.50	124.2836	124.1903
37	138.15	132.3517	135.1691
		MSE = 185.8221	MSE = 166.3987

In this case the mean squared error in calibrating X_2 has been reduced by approximately 10%. Whilst not spectacular, it nevertheless highlights the fact that the proposed method is better than the use of the simple inverse estimator. As already pointed out the efficiency of the conditional estimator is very much dependent on the nature of the covariance matrix Σ_x . To indicate the improvements that are possible a synthetic data set, similar to that of table 14 but with a different covariance structure, has been generated. These data are presented in table 15.

Table 15. Synthetic data for hectares of corn determined by three methods.

Segment	Photograph (X_1)	Satellite (X_2)	Interview (Y)
1	129.64	158.12	165.76
2	122.89	124.12	162.08
3	126.86	153.96	152.04
4	134.18	165.09	161.75
5	124.24	126.12	96.32
6	119.39	129.24	114.12
7	123.92	137.16	100.60
8	126.26	142.46	127.88
9	119.07	135.45	116.90
10	123.53	135.93	87.41
11	124.16	134.64	88.59
12	123.52	143.92	88.59
13	115.81	111.25	165.35
14	115.37	102.17	104.00
15	119.58	120.75	88.63
16	118.15	125.33	153.70
17	131.71	167.37	185.35
18	123.61	142.25	116.43
19	121.85	119.57	93.48
20	129.68	159.37	121.00
21	133.98	164.40	109.91
22	126.90	151.75	122.66
23	122.88	143.59	104.21
24	122.00	133.37	92.88
25	121.75	122.02	149.94
26	124.12	147.90	64.75
27	114.25	106.86	99.96
28	123.02	130.62	140.43
29	114.37	104.71	98.95
30	130.45	160.57	131.04
31	122.84	129.95	127.07
32	122.14	126.13	133.55
33	122.00	131.81	77.70
34	118.69	123.99	76.08
35	122.18	136.19	206.39
36	122.88	149.42	108.33
37	122.94	135.40	118.17

After performing the inverse regression the estimated covariance matrix is

$$\hat{\Sigma}_x = \begin{bmatrix} 950.0649 & 1016.8997 \\ 1016.8997 & 1231.8416 \end{bmatrix}$$

A simple calculation indicates that for this structure the variance of the simple inverse estimator can be reduced by 88% through the use of the conditional procedure.

This reduction is evident when a comparison of the two estimates is made.

Calibration of X_2 (synthetic satellite data) :

Segment	actual	simple inverse	conditional
28	130.62	149.3985	136.9042
29	104.71	105.2694	126.0930
30	160.57	139.4088	144.5024
31	129.95	135.1853	136.2147
32	126.13	142.0791	135.7060
33	131.81	82.6623	133.4740
34	123.99	80.9388	129.8706
35	136.19	219.5710	138.4536
36	149.42	115.2484	135.5579
37	135.40	125.7169	135.9909
		MSE = 1356.6344	MSE = 112.0507

In this case the reduction in MSE has been dramatic. Using equation (3.59) it was computed that an 88% reduction was possible, the actual figure is 92% .

3.6 DIAGNOSTIC QUANTITIES

We now examine various diagnostic quantities similar in spirit to those used in ordinary regression although now couched in the framework of (multivariate) calibration.

Considerable attention has been given in recent years to the examination of the adequacy of the fitted regression model and in particular the identification of "influential" observations [for example, see Andrews and Pregibon (1978) ; Belsley, Kuh, and Welsch (1980); Cook (1977) and Atkinson (1985)].

Since calibration is a variant of multiple (multivariate) regression we should similarly concern ourselves with the detection of single observations or groups of observations which appear to be abnormal in some respect. These atypical observations can arise in several ways, for example gross errors in either the response or explanatory variables or the use of an improper scale etc. The potential difficulty in using these data for modeling is that the fitted model may try to accommodate points which are in some way "strange" , leading to residual differences between the data and the fitted model which are not sufficiently large as to be noteworthy. This section is concerned with the derivation of diagnostic quantities useful in detecting abnormal observations in a multivariate calibration setting.

3.6.1 PRELIMINARIES : OLS AND THE HAT MATRIX

Consider the multivariate regression model

$$Y = X\beta + \xi$$

for which the least-squares estimate of β is

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

Now the matrix of predicted Y's is thus $X\hat{\beta} = X(X^TX)^{-1}X^TY = HY$. The matrix H is referred to as the "hat" matrix (presumably because it puts the hats on the Y's). Geometrically, H projects Y onto the column space of X . Furthermore, the error matrix $Y - \hat{Y} = Y - HY = Y(I - H)$ is the component of Y in the left-null space of X (the orthogonal compliment of the column space of X) and thus I - H is similarly a projection matrix.

The hat matrix appears often in the derivation of regression diagnostics. Of particular interest are the diagonal elements of H . The i^{th} diagonal of H is

$$h_i = \underline{X}_i(X^TX)^{-1}\underline{X}_i^T$$

where $\underline{X}_i$ (1 x q) is the i^{th} row of X . h_i is referred to as the *leverage* and is a measure of the remoteness of the i^{th} observation from the remaining (n-1) in the X-space. We now examine the effect on parameter estimates, residual sum of squares and predicted (calibrated) values

arising from the deletion of one or more observations. Before doing so, the introduction of some extra notation is necessary.

Definitions and notation :

The effect of deleting the i^{th} observation on the estimate $\hat{\beta}$ is to yield an estimate $\hat{\beta}_{(i)}$. If the number of observations deleted is $m > 1$, belonging to the set {I}, the rows of X corresponding to the deleted observations form an m x p matrix denoted by X_I . The resulting X matrix (after deletion) is denoted by $X_{(I)}$. $Y_{(I)}$ and Y_I are similarly defined.

3.6.2 REVISED PARAMETER ESTIMATES

We first define the hat matrix for X_I as

$$H_I = X_I(X^TX)^{-1}X_I \ .$$

The matrix of revised parameter estimates is

$$\beta_{(I)} = [X_{(I)}^TX_{(I)}]^{-1}X_{(I)}^TY_{(I)} \ . \qquad (3.60)$$

Recalculation of the revised parameter estimates using equation (3.60) is not necessary as it can be shown (see Appendix K) that

$$\hat{\beta}_{(I)} - \hat{\beta} = (X^TX)^{-1}X_I^T(I - H_I)^{-1}R_I \qquad (3.61)$$

where

$$R_I = \hat{Y}_I - Y_I \ .$$

3.6.3 REVISED RESIDUAL SUM OF SQUARES

For the multivariate regression model we assume

$$\mathbb{E}[\xi] = 0 \qquad \text{and} \qquad \text{Cov}[\xi] = \Sigma \ .$$

Prior to deletion, our estimate of Σ is

$$(n-q)\hat{\Sigma} = Y^T Y - \hat{\beta}^T X^T Y \tag{3.62}$$

After deletion this becomes

$$(n-m-q)\hat{\Sigma}_{(I)} = Y^T Y - Y_I^T Y_I - \hat{\beta}_{(I)}^T (X^T Y - X_I^T Y_I) \tag{3.63}$$

It is shown in Appendix K that equation (3.63) can be written as

$$(n-m-q)\hat{\Sigma}_{(I)} = (n-q)\hat{\Sigma} - R_I^T (I - H_I)^{-1} R_I \tag{3.64}$$

We now have the necessary results in order to investigate the effect on calibrated values resulting from the deletion of one or more observations.

3.6.4 DELETION OF A SINGLE OBSERVATION

The effect of deleting the i^{th} observation is now examined. We have previously discussed a number of procedures for calibrating, namely the classical estimator, the inverse estimator, and the orthogonal estimator. Since the orthogonal estimates are obtained iteratively, expressions for moments are not easily obtained and hence this form of estimation will not be considered for the purposes of calibration diagnostics. An expression for the variance of the classical estimator was obtained in §3.4.2.2 . Since the derivation of this variance necessitated the use of inverse regression it is suggested that for the purposes of detecting influential observations the computationally simpler and more direct method of inverse regression be employed. In this case the preceding developments for deletion statistics still apply - all that changes is the roles of X and Y are reversed.

Thus, for the inverse regression

$$\begin{aligned} H_I &= Y_I (Y^T Y)^{-1} Y_I \\ &= \underline{Y}_i^T (Y^T Y)^{-1} \underline{Y}_i^T \\ &= h_i \end{aligned}$$

and

$$\gamma_{(i)} - \gamma = (Y^T Y)^{-1} \underline{Y}_i^T (1-h_i)^{-1} \underline{r}_i$$

$$= \frac{(Y^T Y)^{-1} y_i^T r_i}{(1-h_i)}$$

and

$$(n-p-1)\hat{\Sigma}_{(i)} = (n-p)\hat{\Sigma} - \frac{r_i^T r_i}{(1-h_i)}$$

Now $\underline{x}_i - \hat{\underline{x}}_{(i)} = \underline{x}_i - \underline{y}_i\hat{\gamma}_{(i)} \sim N_q\left[\left[\, 0 \,,\, \Sigma_x[1+y_i(Y_{(i)}^T Y_{(i)})^{-1}y_i^T] \,\right]\right]$.

Let $c = 1 + y_i(Y_{(i)}^T Y_{(i)})^{-1}y_i^T$ (scalar) , then

$$\frac{1}{c^{\frac{1}{2}}}\,[\underline{x}_i - \hat{\underline{x}}_{(i)}]\,\frac{1}{c^{\frac{1}{2}}}\,\hat{\Sigma}_{(i)}^{-1}\,[\underline{x}_i - \hat{\underline{x}}_{(i)}]^T \sim T^2_{q,n-p-1}$$

where T^2 is Hotelling's T^2 distribution.

Next consider $c = 1 + y_i(Y_{(i)}^T Y_{(i)})^{-1}y_i^T$. The quantity $(Y_{(i)}^T Y_{(i)})^{-1}$ can be written as

$$(Y_{(i)}^T Y_{(i)})^{-1} = (Y^T Y)^{-1} + (Y^T Y)^{-1}Y_I^T(I - H_I)^{-1}Y_I(Y^T Y)^{-1}$$

thus

$$y_i(Y_{(i)}^T Y_{(i)})^{-1}y_i^T = y_i(Y^T Y)^{-1}y_i^T + y_i(Y^T Y)^{-1}y_i^T(1-h_i)^{-1}y_i(Y^T Y)^{-1}y_i^T$$

$$= y_i(Y^T Y)^{-1}y_i^T + \frac{y_i(Y^T Y)^{-1}y_i^T y_i(Y^T Y)^{-1}y_i^T}{(1-h_i)}$$

$$= h_i + \frac{h_i^2}{(1-h_i)} = \frac{h_i}{(1-h_i)}$$

Therefore $$c = 1 + \frac{h_i}{(1-h_i)} = \frac{1}{(1-h_i)}$$

Thus, an observation is deemed "influential" if :

$$\mathcal{Z}_i = \frac{r_i \hat{\Sigma}_{(i)}^{-1} r_i^T}{(1-h_i)} \qquad (3.65)$$

exceeds the critical value of $T^2_{q,n-p-1}$ at some chosen level of significance. Alternatively, and equivalently, observation i is influential if

$$\mathcal{Z}_i > \frac{(n-p-1)q}{(n-p-q)} F_{q,n-p-q,(\alpha)} \qquad (3.66)$$

3.6.5 A MEASURE OF INFLUENCE FOR THE UNIVARIATE MULTIPLE REGRESSION MODEL

In §3.5.1 the conditional estimator for the univariate multiple regression case was introduced. We now incorporate the present methods of deletion statistics to assist in identifying influential observations in such calibration problems.

In the following development we consider the effect of deleting a single observation. An observation in the calibration data set will be deemed to be influential if the difference between the calibrated values for X_{0i} obtained when the i^{th} observation is included and when it is removed are "significantly" different.

As before, the vector of calibrated values $\hat{X}_0$ is partitioned as

$$\hat{X}_0' = \left[\begin{array}{c} \hat{X}_{0(q)}^T \\ \hline \hat{x}_{0q} \end{array}\right] .$$

In addition we need to distinguish calibrated values for which the i^{th} observation was deleted, we denote such a vector as

$$\hat{X}_{0(i)}' = \left[\begin{array}{c} {}_{(i)}\hat{X}_{0(q)}^T \\ \hline {}_{(i)}\hat{x}_{0q} \end{array}\right]$$

The notation is somewhat clumsy, although is necessary to fully indicate the processes taking place. In the above representation, ${}_{(i)}\hat{X}_{0(q)}^T$ is the transpose of the vector of calibrated values obtained from a model which was estimated using all but the i^{th} observations. The parenthesized q is used to denote the partitioning of the vector and represents the (q-1) vector resulting from the deletion of variable $\hat{x}_{0q}$. Using equations (3.61) and (3.63) we can obtain expressions for the revised vector of parameter estimates and the revised error-covariance matrix. These are

$$\hat{\gamma}_{(i)} = \hat{\gamma} + \frac{y_i}{(1-h_i)\sum_{i=1}^{n} y_i^2} R_I \qquad (3.67)$$

and

$$\hat{\Sigma}_{(i)} = \frac{n-1}{n-2}\hat{\Sigma} - \frac{R_I^T R_I}{(n-2)(1-h_i)} \qquad (3.68)$$

Also,

$$\hat{X}_{0(i)} = y_0 \hat{\gamma}_{(i)} \tag{3.69}$$

and

$$\mathbb{C}^* = \mathrm{Cov}[\hat{X}_{0(i)}] = y_0^2 (Y_{(I)}^T Y_{(I)})^{-1} \hat{\Sigma}_{(i)} \tag{3.70}$$

Furthermore, $\mathbb{C}^*$ is partitioned as

$$\mathbb{C}^* = \left[\begin{array}{c|c} C_{11}^* & C_{12}^* \\ \hline C_{12}^{*T} & C_{22}^* \end{array}\right] \tag{3.71}$$

By analogy with the definition of $\tilde{T}$ in §3.4.1, we let $\tilde{X}_{0(i)}$ be the conditional *variable* :

$$ {}_{(i)}\hat{x}_{0q} \Big| {}_{(i)}\hat{X}_{0(q)}^T = x_{0(q)}^T$$

where $x_{0(q)}^T$ is a vector of *given values* for the variables $X_1, \ldots, X_{q-1}$. Thus

$$\tilde{X}_{0(i)} = C_{12}^{*T} C_{11}^{*-1} [\hat{x}_{0(q)}^T - {}_{(i)}\hat{X}_{0(q)}^T] + {}_{(i)}\hat{X}_{0q} \tag{3.72}$$

The discrepancy between the *actual* Y_0 and $\tilde{Y}_{0(i)}$ is a measure of the influence of the i^{th} observation . A formal test can be constructed around the quantity

$$t = \frac{Y_0 - \tilde{X}_{0(i)}}{SE[\tilde{X}_{0(i)}]} \qquad (3.73)$$

where

$$Var[\tilde{X}_{0(i)}] = C^*_{22} - C^{*T}_{12} C^{*-1}_{11} C^*_{12}$$

Therefore, an observation is influential if

$$t = \frac{| X_0 - \tilde{X}_{0(i)} |}{\left[C^*_{22} - C^{*T}_{12} C^{*-1}_{11} C^*_{12} \right]^{\frac{1}{2}}} \qquad (3.74)$$

exceeds $T_{n-q-2,(\frac{\alpha}{2})}$, where T has the Student's t-distribution.

The advantage of equation (3.74) is that in deciding if an observation is influential or not, all of the available information on both the X's and Y is utilized. When performing the simple inverse regression of X_q on Y it is possible that an influential observation will not be detected by the usual diagnostic quantities. Such a situation is now presented by the following example.

3.6.5.1 <u>An example</u>

The data in table 16 are from Atkinson (1985,p53) and were originally given by Prater and later used by Hader and Grandage (1958) to illustrate multiple regression calculations. The dependent variable (Y) is the percentage of crude oil which, after distillation and

Table 16. Prater's gasoline data.

Obsn.	Crude oil gravity x_1	Crude oil vapour pressure x_2	Crude oil ASTM 10% point x_3	Gasoline end point X_4	Gasoline yield y
1	38.4	6.1	220	235	6.9
2	40.3	4.8	231	307	14.4
3	40.0	6.1	217	212	7.4
4	31.8	0.2	316	365	8.5
5	40.8	3.5	210	218	8.0
6	41.3	1.8	267	235	2.8
7	38.1	1.2	274	285	5.0
8	50.8	8.6	190	205	12.2
9	32.2	5.2	236	267	10.0
10	38.4	6.1	220	300	15.2
11	40.3	4.8	231	‹267›	26.8
12	32.2	2.4	284	351	14.0
13	31.8	0.2	316	379	14.7
14	41.3	1.8	267	275	6.4
15	38.1	1.2	274	365	17.6
16	50.8	8.6	190	275	22.3
17	32.2	5.2	236	360	24.8
18	38.4	6.1	220	365	26.0
19	40.3	4.8	231	395	34.9
20	40.0	6.1	217	272	18.2
21	32.2	2.4	284	424	23.2
22	31.8	0.2	316	428	18.0
23	40.8	3.5	210	273	13.1
24	41.3	1.8	267	358	16.1
25	38.1	1.2	274	444	32.1
26	50.8	8.6	190	345	34.7
27	32.2	5.2	236	402	31.7
28	38.4	6.1	220	410	33.6
29	40.0	6.1	217	340	30.4
30	40.8	3.5	210	347	26.6
31	41.3	1.8	267	416	27.8
32	50.8	8.6	190	407	45.7

fractionation, ends up as gasoline. The four explanatory variables are :

x_1 crude oil gravity, ^{o}API

x_2 crude oil vapour pressure, lbf/in^2

x_3 crude oil ASTM 10 percent, ^{o}F

x_4 gasoline end point, ^{o}F .

For observation #11, x_4 was erroneously recorded by Prater as 267 - it should have been 367. Such recording errors are not easily detected and in this case cannot be identified by a plot of the x_4 values. More importantly, it is possible that this error would go unnoticed in a calibration of X_4 using a simple inverse regression of X_4 on Y as we will now demonstrate.

The data is first centered prior to analysis by subtracting from each observation the sample mean of the respective variable.

For the regression of X on Y we obtain

$\hat{\gamma}$ = [0.1295 0.0938 -1.1030 4.4286] using all observations.

and

$\hat{\gamma}_{(11)}$ = [0.1292 0.0939 -1.0975 4.6250] with observation #11 removed.

Calibration of observation #11.

The actual (centered) values for observation #11 are :

x_1 = 1.0500 ; x_2 = 0.6188 ; x_3 = -10.500 ; x_4 = -61.9688 ; y = 7.1406

Using the conditional calibration procedure of §3.5.1 we obtain

$$C_{11}^{*} = \begin{bmatrix} 0.4332 & 0.1128 & -1.9123 \\ 0.1128 & 0.0850 & -1.1215 \\ -1.9123 & -1.1215 & 18.4305 \end{bmatrix} \qquad C_{12}^{*} = \begin{bmatrix} -2.8368 \\ -1.5159 \\ 24.2018 \end{bmatrix}$$

$$C_{22}^{*} = 34.8834$$

with

$$\tilde{X}_{0(11)} = 30.3607 \qquad \text{and} \qquad \text{Var}[\tilde{X}_{0(11)}] = 2.5120 \ .$$

Using the test for influence [equation (3.74)] we have

$$t = \frac{30.3607 - (-61.9688)}{[2.5120]^{\frac{1}{2}}} = 59.9357$$

This result is highly significant and we therefore conclude that observation #11 is indeed influential or suspect. For comparison the statistic given by equation (3.65) is computed to be $\mathbb{I} = 48.5735$. The $\alpha = 0.05$ critical value for the test given by equation (3.66) is 12.2362 and thus the computed result is clearly significant. In comparison, an examination of the regression diagnostics obtained from a simple regression of X_4 on Y does not reveal the extent of observation #11's abnormality. The studentized residual for observation #11 is reported by MINITAB's regression program to be -1.89 (called a TRESID in MINITAB). This is *not significant* at the 5% level ($t_{crit} = \pm 2.045$). The utility of the proposed approach is apparent.

3.7 DISCRETE MULTIVARIATE CALIBRATION

We now consider briefly the problem of discrete calibration i.e. where one or more of the X-variables is discrete. This situation is most likely to arise in experimental design problems where one is interested in identifying the treatment combination that gave rise to an observed response Y_0 . This type of calibration is similar to discriminant analysis although the main difference being in the calibration context we first estimate the response generating model. The multivariate procedures already discussed in this chapter are not appropriate as it is tacitly assumed that the X-variables are measured on an interval or ratio scale. Rounding the calibrated values to the nearest integer offers no solution to the discrete calibration, what is required is a modified approach. A variant of the classical calibration estimator applicable for discrete calibration problems is now proposed and the method illustrated with the use of an example.

3.7.1 A MODIFIED CLASSICAL ESTIMATOR

We assume the response-generating model is of the usual form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$

In the present context, X is a design matrix . The matrix of parameter estimates (or 'effects') is obtained in the usual manner as

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad .$$

and $$\hat{\Sigma} = \frac{1}{n}(Y-\hat{Y})^T(Y-\hat{Y}) .$$

The proposed criterion for calibration of X is based on a minimum distance concept (in this case we use the regular Euclidian distance). Specifically, *given* a new value of the response Y_0 , the calibrated $\hat{X}_0$ is that treatment combination for which

$$(Y_0 - \hat{Y}_0)\hat{\Sigma}^{-1} (Y_0 - \hat{Y}_0)^T \qquad (3.75)$$

is a minimum, where $\hat{Y}_0 = X_0\hat{\beta}$.

The concept is quite simple and has intuitive appeal. We now apply this technique to previously published data.

3.7.2 AN EXAMPLE

The data presented in table 17 are taken from Brown (1982,p302) and were obtained from an experiment in which information was sought on the effect of two factors on the reflectivity of paint used for road marking. A single patch of paint was tinted with pigment at three levels (0% , 0.15% , 0.30%), and the viscosity of the sample was adjusted before spraying to one of three levels (30,33,36 seconds in an efflux cup); each of the resulting paints was replicated four times, giving a total of 36 panels. The optical properties of each panel was measure in three ways :

Table 17. 3^2 factorial experiment with four replicates and six response variables for Brown's paint data.

	P	V	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
x	0	0	1.9	35.0	75.0	40.9	101.0	20.0
x	0	0	1.9	35.3	75.8	40.7	114.0	17.5
	0	0	1.9	36.1	77.0	40.6	101.0	19.8
	0	0	1.9	36.8	77.0	40.6	100.5	17.5
	0	1	1.8	32.4	73.2	39.8	95.0	20.6
	0	1	1.8	31.9	72.9	39.7	107.0	19.5
x	0	1	1.7	29.6	72.1	39.2	93.5	21.0
x	0	1	1.7	30.6	72.5	39.4	93.0	18.0
	0	2	1.6	25.7	66.7	37.2	93.5	22.0
x	0	2	1.7	26.8	67.9	37.9	86.0	21.3
x	0	2	1.6	23.8	62.9	37.4	84.0	21.3
	0	2	1.7	27.2	67.2	38.2	84.5	19.0
	1	0	1.8	33.5	76.0	39.1	102.0	21.0
x	1	0	1.8	31.3	71.8	39.1	105.0	19.8
x	1	0	1.8	31.8	71.8	39.3	103.0	20.0
	1	0	1.8	31.8	72.5	38.7	101.0	20.8
	1	1	1.7	30.5	71.5	39.3	103.0	20.1
x	1	1	1.7	28.7	71.1	38.6	98.5	20.9
x	1	1	1.7	29.6	71.1	38.5	99.0	21.0
	1	1	1.7	29.5	70.0	39.1	101.0	20.8
x	1	2	1.5	21.0	63.0	35.8	85.0	21.1
x	1	2	1.5	21.9	63.9	35.7	85.0	21.8
	1	2	1.5	21.2	63.0	35.7	84.0	22.5
	1	2	1.5	20.6	61.6	35.8	85.0	22.2
x	2	0	1.9	35.8	74.0	38.0	101.0	19.5
	2	0	1.9	33.9	72.0	38.1	101.0	20.1
	2	0	1.9	35.0	73.0	37.9	92.5	19.5
x	2	0	1.9	33.7	70.5	38.2	83.0	21.8
	2	1	1.9	33.0	71.0	37.2	96.0	21.0
	2	1	1.9	31.5	68.5	37.2	91.5	22.5
x	2	1	1.9	33.0	70.0	37.8	99.0	19.5
x	2	1	1.9	31.5	68.0	37.3	95.0	20.1
x	2	2	1.8	27.8	64.8	34.4	82.5	20.8
	2	2	1.6	24.6	67.4	34.1	71.0	21.0
x	2	2	1.6	24.0	63.0	34.2	80.0	21.4
	2	2	1.6	23.0	60.0	34.2	86.0	20.7

(i) spectrometer measurements of incident light. Measurements at two different inclinations were used to create three responses $,Y_1,Y_2,Y_3$ of table 16.

(ii) integrated reflectance with normal incident light, Y_4 of table 17.

(iii) peak-height and band-width on a recording goniophotometer, Y_5,Y_6 of table 16.

It is pointed out that Brown's analysis of this data was different to the one presently being considered. Rather than attempting to identify the levels of pigmentation and viscosity from post-calibration Y data, Brown's objective was to determine an optimal subset of the Y's that best modeled the data.

The observations in table 17 preceded by an asterisk have been used for model estimation (these were randomly selected). The remaining observations will be used for calibration purposes. A preliminary inspection of the data is provided by figure 10 in which each of the response variables has been plotted against the levels of each treatment factor. A reasonable degree of curvature is indicated by these plots and we therefore elect to use an orthogonal polynomial coding for the design matrix X.

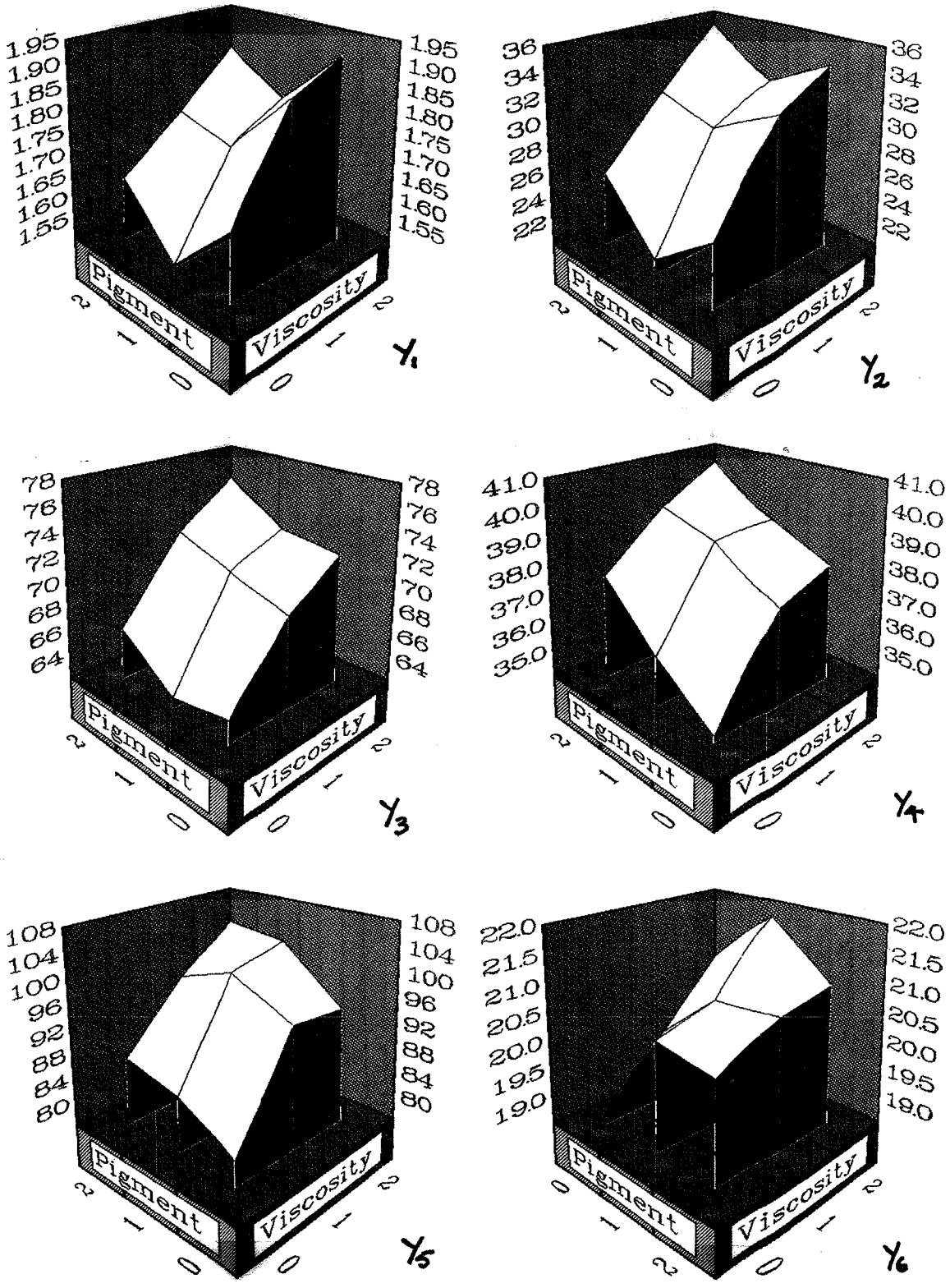

Figure 10. Profile plots for each of the response variables used in the wheat analysis data.

The matrix of estimated parameters is :

$$\hat{\beta} = \begin{bmatrix} 1\cdot743 & 29\cdot450 & 69\cdot333 & 37\cdot870 & 94\cdot167 & 20\cdot417 \\ -0\cdot047 & -0\cdot483 & 1\cdot225 & 1\cdot247 & 3\cdot208 & -0\cdot275 \\ 0\cdot040 & 1\cdot033 & 0\cdot275 & 0\cdot015 & -0\cdot875 & -0\cdot175 \\ 0\cdot129 & 4\cdot892 & 4\cdot550 & 1\cdot804 & 8\cdot083 & -0\cdot817 \\ -0\cdot009 & -0\cdot525 & -0\cdot733 & -0\cdot304 & -1\cdot083 & 0\cdot167 \\ 0\cdot003 & 0\cdot387 & 0\cdot562 & -0\cdot076 & 2\cdot000 & -0\cdot613 \\ 0\cdot017 & 0\cdot296 & -0\cdot212 & 0\cdot192 & 2\cdot542 & -0\cdot062 \\ -0\cdot004 & -0\cdot079 & 0\cdot188 & 0\cdot032 & -0\cdot708 & -0\cdot021 \\ 0\cdot003 & 0\cdot179 & 0\cdot212 & 0\cdot030 & 0\cdot167 & 0\cdot129 \end{bmatrix}$$

Using the above parameter estimates we can calibrate for the observations which were not previously used. For each Y-vector, a row vector $\hat{Y}$ is computed using a particular treatment combination (i.e a particular row of the design matrix). The quantity $(Y - \hat{Y})\hat{\Sigma}^{-1} (Y - \hat{Y})^{T}$ is then computed and the procedure repeated for all treatment combinations (rows of X). The treatment combination yielding the smallest value of $(Y - \hat{Y})\hat{\Sigma}^{-1} (Y - \hat{Y})^{T}$ becomes the calibrated value for X. This has been carried out for all 18 observations and the results displayed in table 18.

It can been seen from table 18 that the procedure resulted in 10 out of the 18 observations being correctly classified. Such a result has an infinitesimally small probability of occurring due to chance alone.

3.8 SEEMINGLY UNRELATED CALIBRATION

We conclude this discussion on multivariate calibration techniques with a brief look at the seemingly unrelated calibration model (this terminology has been coined by the present author).

Table 18. Values of calibration criterion for 18 observations of table 17 not used for model fitting. Underlined entries correspond to minimum value of the criterion. Entries enclosed with ⟨ ⟩ indicate misclassified observation.

Treatment combination								
00	01	02	10	11	12	20	21	22
4.28	26.61	70.67	13.36	33.55	247.92	124.07	117.24	466.15
34.18	80.68	137.32	64.59	93.26	361.01	224.33	214.27	603.68
11.85	12.01	43.43	⟨5.10⟩	16.50	195.58	92.85	86.11	395.58
9.40	13.06	44.41	⟨2.19⟩	15.48	192.79	88.01	79.73	395.14
115.33	23.53	3.31	39.51	15.20	39.06	12.27	8.92	151.59
95.64	12.90	1.54	30.96	11.08	54.96	18.59	15.84	178.48
41.98	2.82	14.73	6.43	⟨1.99⟩	16.98	43.30	40.19	277.89
40.74	3.06	12.44	4.24	⟨1.07⟩	114.19	40.14	35.14	273.96
13.32	10.02	37.39	⟨1.81⟩	12.00	181.04	84.05	75.69	378.20
18.00	7.93	30.57	⟨0.46⟩	8.26	163.37	71.26	63.01	353.13
278.21	111.32	53.22	150.08	96.43	0.25	34.45	35.63	45.26
222.66	77.78	29.29	109.91	64.77	5.31	24.99	23.34	71.17
103.97	25.26	10.80	35.12	17.62	55.38	4.57	⟨2.87⟩	164.97
165.47	51.77	22.17	74.44	43.55	28.66	1.42	3.84	104.37
183.18	62.82	26.42	84.66	50.78	22.27	⟨3.14⟩	4.43	90.50
150.54	46.88	16.71	62.79	35.30	33.02	2.65	1.83	116.73
900.36	569.71	434.58	659.41	543.19	182.94	329.69	347.56	65.92
396.76	194.21	110.53	239.85	173.62	16.79	74.30	75.55	11.81

The use of seemingly unrelated regression models is widespread in econometrics and was first suggested by Zellner in 1962. This section, whilst not presenting any new contribution to the body of literature on SUR models, does however indicate their potential application in the calibration context. We now give a brief overview of the SUR model and associated estimation procedures.

3.8.1 SUR MODELS AND GLS ESTIMATION

The type of situation to which SUR models apply is probably best illustrated by way of example. Consider the wheat data of table 11 and the associated analysis of §3.2.4. For this case there was *one* multivariate model describing the relationship between the laboratory measurements Y and the physical properties of the wheat X. Suppose now that the same analyses had been conducted at m different laboraties, in which case we end up with m multivariate models. If one assumes independence between *all observations* then we could utilize standard MANOVA or MANCOVA techniques and which are offered as options in many statistical packages (e.g. SPSS-X and SAS). However in many cases in econometrics the assumption of independence is not tenable and thus the standard procedures need to be modified. In the present example such lack of independence may result from sending portions of the *same* sample to each laboratory for analysis. Formally, the m models may be written as follows :

$$\begin{bmatrix} Y_1 \\ \hline Y_2 \\ \hline \vdots \\ \hline Y_m \end{bmatrix} = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \hline \beta_2 \\ \hline \vdots \\ \hline \beta_m \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \hline \xi_2 \\ \hline \vdots \\ \hline \xi_m \end{bmatrix} \qquad (3.76)$$

where

$$Y_i \text{ is } (n \times p) \; ; \quad \beta_i \text{ is } (q_i \times p) \; ;$$

$$X_i \text{ is } (n \times q_i) \; ; \quad \xi_i \text{ is } (n \times p) \; .$$

Equation (3.76) can be written more conveniently as

$$Y = Z\,\gamma + \varphi \qquad (3.77)$$

where

$$Y \text{ is } (mn \times p) \; ; \quad \gamma \text{ is } (k \times p) \; ;$$

$$Z \text{ is } (mn \times k) \; ; \quad \varphi \text{ is } (mn \times p)$$

$$\text{where } k = \sum_{i=1}^{m} q_i \; .$$

Furthermore, assume $\mathbb{E}[\xi_i] = 0$, $\mathbb{E}[\xi_i^T \xi_j] = \underset{(p \times p)}{\Sigma_{ij}}$ and that

$$\text{Cov}[\varphi] = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{bmatrix} \otimes I_n = \Gamma \; .$$

The BAN estimator of γ is the GLS estimator :

$$\hat{\gamma} = (Z^T\Gamma^{-1}Z)^{-1}Z^T\Gamma^{-1}Y \qquad (3.78)$$

When OLS conditions are met equation (3.78) is equivalent to performing each of the m regressions separately to obtain $\hat{\beta}_i = (X_i^TX_i)^{-1}X_i^TY_i$. However, when OLS conditions are not satisfied, the GLS estimator of equation (3.78) is superior since it takes cognizance of the covariance structure between the ξ_i and because it uses information on explanatory variables that are included in the system, but which do not appear in the i^{th} equation. The potential gain in efficiency through the use of the GLS estimator prompted Zellner (1962) to use the term "seemingly unrelated regression equations". In the present context we apply the term seemingly unrelated calibration to emphasize that the primary goal is that of calibration and not simply regression.

From a practical viewpoint, equation (3.78) cannot be implemented without knowledge of Γ , and this will rarely be the case. In such situations Zellner suggests that Γ be replaced with its sample estimate $\hat{\Gamma}$, where $\hat{\Gamma}$ is computed from OLS residuals. Thus, the modified GLS estimator becomes

$$\tilde{\gamma} = (Z^T\hat{\Gamma}^{-1}Z)^{-1}Z^T\hat{\Gamma}^{-1}Y \qquad (3.79)$$

where

$$\hat{\Gamma} = \mathrm{Cov}[\hat{\varphi}]$$

and the components of $\hat{\varphi}$ are obtained as $\hat{\varphi}_i = Y_i - X_i\hat{\beta}_i$.

Equation (3.79) may be applied repeatedly with $\hat{\varphi}$ successively updated from the previous fit. This iterative process is continued until $\tilde{\gamma}$ converges. For normally distributed ξ , the resulting $\tilde{\gamma}$ is in fact the m.l.e. [Dhrymes (1971), Oberhofer and Kmenta (1974), Magnus (1978)]. We now illustrate the procedure by applying it to an artificial data set.

3.8.2 AN EXAMPLE

The data appearing in tables 19 to 22 were computer-generated and are intended to reproduce the kind of responses that might be obtained for the wheat data of table 11 when a number of different laboratories each perform the same analyses.

The OLS parameter estimates are

$$\hat{\beta} = \begin{bmatrix} -0.351 & 1.998 & 3.192 & 5.692 \\ 0.047 & -1.474 & -3.145 & 5.464 \\ 5.386 & 13.737 & 4.977 & 6.437 \\ -0.556 & -4.212 & -1.463 & 2.079 \\ -1.088 & 0.433 & -0.901 & 0.350 \\ 0.485 & 0.351 & 1.668 & -15.968 \\ 1.993 & 0.937 & 16.073 & 12.859 \\ -0.336 & -1.037 & -5.687 & -9.810 \end{bmatrix} \begin{matrix} \\ \\ \hat{\beta}_1 \\ \\ \hat{\beta}_2 \\ \\ \hat{\beta}_3 \\ \\ \hat{\beta}_4 \end{matrix}$$

The estimates obtained by the iterative GLS procedure are

$$\tilde{\gamma} = \begin{bmatrix} -0.554 & 2.336 & 3.670 & 5.449 \\ 0.187 & -1.721 & -3.532 & -5.365 \\ 7.235 & 12.347 & 0.381 & 9.629 \\ -1.026 & -3.735 & 0.267 & -4.837 \\ -1.525 & 0.914 & -0.550 & -0.785 \\ 1.322 & -0.841 & -0.994 & -19.805 \\ -0.293 & 5.502 & 11.740 & -3.829 \\ 1.062 & -1.895 & -3.538 & -4.200 \end{bmatrix} \begin{matrix} \\ \\ \tilde{\gamma}_1 \\ \\ \tilde{\gamma}_2 \\ \\ \tilde{\gamma}_3 \\ \\ \tilde{\gamma}_4 \end{matrix}$$

Table 19. Laboratory #1 data.

%water	%protein	Y_1	Y_2	Y_3	Y_4
9.093	9.447	360.498	106.711	86.604	250.18
8.857	10.241	368.325	103.684	102.201	234.32
9.837	13.021	361.880	102.193	105.941	256.31
9.340	10.264	357.174	104.316	89.506	219.23
9.146	9.865	357.027	102.885	97.458	248.62
10.537	13.113	359.418	104.757	79.353	217.98
9.407	12.669	362.326	103.359	66.340	210.98
8.977	10.621	368.908	101.260	90.997	222.54
9.364	11.891	363.424	96.745	77.027	228.28
10.127	12.468	363.922	111.402	95.952	243.59
10.781	14.280	352.792	112.377	95.281	223.49
9.130	9.780	361.478	118.354	103.656	252.03
9.708	11.747	359.335	103.329	89.487	232.59
9.446	11.605	358.200	102.511	104.697	241.70
9.077	11.146	363.462	102.053	74.662	225.39
8.886	10.226	369.316	108.990	85.704	231.68
8.574	8.749	359.900	101.332	108.902	259.45
8.958	7.642	352.073	97.396	87.345	234.25
8.766	9.241	363.902	108.824	102.591	243.67
9.752	11.743	367.445	117.185	79.150	215.68
8.958	9.182	357.931	96.729	93.240	231.15
8.798	8.159	361.994	113.864	98.754	255.26
9.858	12.465	349.763	87.807	55.225	204.10
10.022	13.489	360.875	98.025	64.118	217.88
9.013	9.017	359.562	106.343	87.996	220.59

Table 20. Laboratory #2.

%water	%protein	Y_1	Y_2	Y_3	Y_4
9.111	10.226	354.784	113.612	92.318	254.46
9.175	9.182	353.878	106.525	87.195	225.05
9.839	11.982	354.607	111.402	96.777	218.95
9.925	12.510	358.804	113.590	107.042	232.26
9.521	10.925	359.330	104.677	89.734	257.27
9.144	11.104	350.377	102.336	68.028	209.37
9.611	11.112	368.370	98.982	63.216	197.01
10.092	12.225	368.283	115.560	102.269	251.54
10.041	12.422	364.874	99.645	65.508	231.30
8.132	6.327	353.619	107.871	89.427	231.56
9.150	11.043	361.946	96.225	83.340	233.26
10.419	13.578	354.451	106.423	76.785	226.84
9.105	10.533	351.274	104.595	91.158	221.95
8.547	9.599	361.526	94.048	94.846	238.59
8.702	9.134	349.917	108.185	61.324	217.19
9.040	10.390	369.271	99.723	72.684	240.20
8.879	8.535	352.309	98.522	92.071	242.52
8.833	9.486	356.968	98.465	95.043	236.36
10.072	13.468	363.255	103.181	94.416	233.35
9.697	11.463	361.953	114.253	73.783	210.33
9.024	8.153	362.183	93.980	96.166	227.76
9.458	12.041	355.342	109.003	97.088	249.58
9.880	14.562	354.305	90.812	78.063	227.68
9.417	12.145	355.237	103.415	74.306	228.71
9.596	9.927	371.651	104.736	82.895	231.44

Table 21. Laboratory #3.

%water	%protein	Y_1	Y_2	Y_3	Y_4
9.793	12.197	359.439	98.666	95.603	240.69
9.041	9.433	359.236	112.188	112.362	246.76
9.642	12.399	359.862	124.695	98.239	236.45
9.291	11.131	352.710	108.528	79.877	228.83
8.702	8.492	357.065	101.320	77.980	496.87
9.239	10.642	355.274	99.001	69.051	215.53
8.240	7.719	362.849	101.807	77.732	218.87
9.392	10.855	358.833	100.804	90.176	235.35
9.757	11.483	352.496	99.106	91.552	240.36
8.770	8.756	356.282	112.386	93.343	235.05
19.922	11.349	347.454	112.122	79.333	234.73
10.023	11.896	356.244	110.522	93.618	248.09
8.617	10.486	361.419	114.844	86.829	233.07
9.940	13.114	365.629	103.893	86.827	232.38
9.606	11.123	359.022	101.773	89.237	232.52
9.710	11.875	362.321	105.516	83.745	224.03
10.189	12.751	356.947	108.377	105.593	259.60
9.195	10.807	353.980	113.649	86.741	232.47
9.223	9.667	367.566	118.326	103.347	256.29
9.290	10.967	368.111	102.887	85.605	212.73
10.205	12.972	351.661	111.332	87.211	241.80
8.978	9.689	366.551	109.743	93.100	248.29
9.750	11.754	363.444	93.864	58.638	223.66
9.337	11.343	353.996	101.804	59.107	208.15
8.520	9.168	353.342	103.924	86.631	232.86

Table 22. Laboratory #4.

%water	%protein	Y_1	Y_2	Y_3	Y_4
9.876	11.350	361.071	98.751	92.966	234.93
7.982	7.043	357.446	102.212	101.313	247.44
9.418	11.213	371.641	115.846	95.013	205.60
9.312	11.544	362.329	108.342	108.691	221.69
9.336	10.203	356.611	104.971	88.191	238.11
9.911	12.331	362.733	103.933	109.624	211.02
8.901	9.180	364.714	95.691	69.253	222.48
10.610	15.823	362.227	97.951	77.481	196.23
10.222	14.026	360.708	92.043	82.649	213.55
9.122	9.800	360.809	103.675	86.720	243.81
9.274	11.909	355.541	97.750	64.096	222.68
9.370	10.396	360.044	109.418	107.305	246.57
9.503	11.284	363.197	107.813	72.439	220.73
9.413	10.803	355.723	95.075	81.208	227.14
8.841	9.418	363.989	104.684	81.694	247.47
8.985	9.283	366.016	114.356	87.268	255.32
9.422	9.138	356.419	111.076	87.246	239.48
9.705	12.496	359.176	107.020	82.956	224.48
9.482	9.752	361.304	110.318	88.489	228.38
9.797	11.219	366.073	124.109	96.293	232.91
9.364	14.271	364.471	102.425	112.280	246.33
9.127	9.560	357.619	116.248	110.155	268.68
9.070	9.233	347.887	110.714	74.788	225.55
8.389	7.822	355.191	94.460	82.267	215.41
9.979	12.975	356.408	98.409	87.766	227.93

Using the GLS parameter estimates the cross-covariance matrix may be estimated (this indicates not only the covariance structure *within* a single laboratory, but also the dependencies *between* laboratories). This matrix will not be reproduced her due to size limitations (matrix is 16 x 16). Having estimated the cross-covariance matrix all the previous results of this chapter may be applied. A number of interesting calibration scenarios are possible. For example, the conditional calibration procedure may be applied to the situation where one of the components (say %water) is already known for the samples sent to a *different* laboratory. If measurements among laboratories are *not independent* then we can utilize the covariance structure to obtain a more precise calibrated value for the laboratory under consideration.