CHAPTER II

UNIVARIATE MODELS

2.1 INTRODUCTION

Various procedures associated with statistical calibration in the univariate case are examined in this chapter. Our discussion commences with the problem of 'optimally' selecting the x-values in a controlled calibration experiment followed by a review of simple linear regression methods from a geometrical perspective. The usual notion of a vertical projection onto the regression line will be extended to accommodate other types of projections. This situation is most relevant to the so-called errors-in-variables model in which both the X and Y data are subject to measurement error. Some of the results presented in this section have been previously derived, for example Cramer (1954) mentions briefly the possibility of non-vertical projections onto the regression line. A comprehensive and unifying treatment of the geometry of least-squares and its relation to the spectral decomposition of the covariance matrix is presented here. Fuller (1987) gives an excellent account of measurement error models although little attention is given to the calibration problem.

Procedures for obtaining interval estimates will be examined and various methods compared. Results of simulation studies are presented and some conclusions/recommendations for the practitioner given. An application to the problem of vehicle speed determination from airborne observation is given for illustrative purposes.

2.2 DESIGN CONSIDERATIONS IN CALIBRATION EXPERIMENTS.

In this section we review the problem of choosing the location and number of X-values in a controlled calibration experiment. Aspects of this design problem have been addressed by Andrews and Herzberg (1973), Naszodi (1978), Ott and Myers (1968) and more recently by Buonaccorsi (1986). The following development follows closely that given by Buonaccorsi (1986).

The design problem in calibration is associated with the selection of X-values for the calibration experiment. This usually entails determination of location, spacing, and degree of replication.

For the simple regression equation given in equation (1.1) we have

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (2.1)$$

Now $$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} (X_i - \bar{X})[\beta_1 (X_i - \bar{X}) + (\xi_i - \bar{\xi})]$$

$$= \beta_1 \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (X_i - \bar{X})(\xi_i - \bar{\xi})$$

where $$\bar{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i$$

Thus $$\hat{\beta}_1 = \beta_1 + \left[\sum_{i=1}^{n} (X_i - \bar{X})(\xi_i - \bar{\xi}) \right] S_{xx}^{-1}$$

For sufficiently large n, $\bar{\xi} \doteq 0$ and we write

$$\hat{\beta}_1 = \beta_1 + \left[\sum_{i=1}^{n} (X_i - \bar{X})\xi_i \right] S_{xx}^{-1}$$

We will assume that the errors are bounded with support on the interval $[-\delta, \delta]$ and that the error distribution is continuous. These assumptions imply that $\hat{\beta}_1$ has a continuous distribution with support on the interval

$$\beta_1 \pm \frac{\delta \sum_{i=1}^{n} |X_i - \bar{X}|}{S_{xx}} \quad .$$

As mentioned in §1.3, a major criticism of the use of the classical estimator is the problem of non-existent moments. This situation arises whenever the distribution of $\hat{\beta}_1$ has sufficient probability around zero.

A sufficient condition for the distribution of $\hat{\beta}_1$ to be bounded away from zero is $|\beta_1 \pm k| > \zeta$ or $|\beta_1| - k > \zeta$ where

$$k = \frac{\delta \sum_{i=1}^{n} |X_i - \bar{X}|}{S_{xx}}$$

and ζ is some non-negative constant.

Letting $\zeta = 0$, we have

$$|\beta_1| - \frac{\delta \sum_{i=1}^{n} |X_i - \bar{X}|}{S_{xx}} > 0$$

$$\Rightarrow \quad \frac{|\beta_1|}{\delta} > \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{S_{xx}} \qquad (2.2)$$

Furthermore, for $a \leq x_i \leq b$ it can readily be shown [e.g. Buonaccorsi (1986),p155] that

$$\frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{S_{xx}} \geq \frac{2}{|b-a|} \qquad (2.3)$$

Now, moments for $\hat{X}_0$ will exist provided

$$\frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{\sum_{i=1}^{n} (X_i - \bar{X})^2} < \frac{|\beta_1|}{\delta}$$

However, from equation (2.3) it is established that there is no design for which these moments exist if $\frac{|\beta_1|}{\delta} < \frac{2}{|b-a|}$.

We now examine designs for which the lower bound in equation (2.3) is attained (this choice not only ensures the existence of moments but minimizes the chances of obtaining large deviations for $|\hat{X}_0 - X_0|$). We consider designs having equally-spaced X-values with possible multiple observations at each X. To this end, we divide the interval [a,b] into k equal parts and let the number of observations at the j^{th} position be n_j with $\sum_{j=1}^{k+1} n_j = n$.

Thus

$$x_j = a + (j-1)\left(\frac{b-a}{k}\right) \tag{2.4}$$

Now,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{j=1}^{k+1} n_j X_j$$

hence

$$\bar{X} = a + \frac{b-a}{nk}\left[\sum_{m=1}^{k+1} n_m(m-1)\right] \tag{2.5}$$

Let s be defined such that $X_j - \bar{X} < 0$ for j=1,2,...,s-1 and $X_j - \bar{X} > 0$ for j=s,s+1,...,k+1.

Therefore

$$\sum_{i=1}^{n} |X_i - \bar{X}| = \sum_{j=1}^{k+1} n_j |X_j - \bar{X}|$$

$$= \sum_{j=1}^{s-1} n_j |X_j - \bar{X}| + \sum_{j=s}^{k+1} n_j |X_j - \bar{X}| \qquad (2.6)$$

Also,

$$\sum_{j=1}^{k+1} n_j (X_j - \bar{X}) = 0$$

$$\Rightarrow \quad \sum_{j=1}^{s-1} n_j (X_j - \bar{X}) + \sum_{j=s}^{k+1} n_j (X_j - \bar{X}) = 0 \qquad (2.7)$$

Equation (2.6) can be written as

$$\sum_{j=1}^{k+1} n_j |X_j - \bar{X}| = \sum_{j=s}^{k+1} n_j (X_j - \bar{X}) - \sum_{j=1}^{s-1} n_j (X_j - \bar{X})$$

and combining with equation (2.5) we obtain

$$\sum_{j=1}^{k+1} n_j |X_j - \bar{X}| = 2 \sum_{j=s}^{k+1} n_j (X_j - \bar{X}) \qquad (2.8)$$

Next, consider $\sum_{i=1}^{n} (X_i - \bar{X})^2$

$$= \sum_{j=1}^{k+1} n_j (X_j - \bar{X})^2$$

$$= \sum_{j=1}^{s-1} n_j (X_j - \bar{X})^2 + \sum_{j=s}^{k+1} n_j (X_j - \bar{X})^2 .$$

$$\text{Now} \quad X_j - \bar{X} = \frac{b-a}{k}\left[(j-1) - \frac{1}{n}\sum_{m=1}^{k+1} n_m(m-1)\right] \tag{2.9}$$

Our criterion is to select a design (or designs) for which

$$\frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{2}{(b-a)} \qquad \text{for} \quad a,b > 0\ ,\ b>a\ .$$

After some algebraic manipulation this reduces to

$$\frac{\sum_{j=1}^{k+1} n_j\left[(j-1) - \frac{1}{n}\sum_{m=1}^{k+1} n_m(m-1)\right]^2}{\sum_{j=s}^{k+1} n_j\left[(j-1) - \frac{1}{n}\sum_{m=1}^{k+1} n_m(m-1)\right]} = k \tag{2.10}$$

where s is the *smallest integer exceeding*

$$1 + \frac{1}{n}\sum_{j=1}^{k+1} n_j(j-1) \tag{2.11}$$

To remove the dependency on n let $p_j = \frac{n_j}{n}$. Thus equations (2.10) and (2.11) become

$$\frac{\sum_{j=1}^{k+1} p_j[(j-1) - \frac{1}{n}\sum_{m=1}^{k+1} p_m(m-1)]^2}{\sum_{j=s}^{k+1} p_j[(j-1) - \frac{1}{n}\sum_{m=1}^{k+1} p_m(m-1)]} = k \qquad (2.12)$$

$$1 + \frac{1}{n}\sum_{j=1}^{k+1} p_j(j-1) \qquad (2.13)$$

Designs for which the criterion of equation (2.12) are satisfied are three-point designs for which

$$p_1 = \frac{\alpha}{2} \; ; \quad p_{k+1} = \frac{\alpha}{2} \; ; \quad p_{\frac{k+2}{2}} = (1-\alpha) \qquad \text{where } 0 < \alpha \leq 1,$$

assuming k is even (odd k pose no problem).

To see this we first determine s using equation (2.13). Thus

$$1 + \frac{1}{n}\sum_{j=1}^{k+1} p_j(j-1) = 1 + (0)\frac{\alpha}{2} + (\frac{k+2}{2} - 1)(1-\alpha) + k\frac{\alpha}{2}$$
$$= \frac{k}{2} + 1$$

Therefore s is $2 + \frac{k}{2} = \frac{k+4}{2}$.

Now, the left-hand side of equation (2.12) is

$$\frac{\frac{\alpha}{2}[0-\frac{k}{2}]^2 + (1-\alpha)[\frac{k}{2} - \frac{k}{2}]^2 + \frac{\alpha}{2}[k - \frac{k}{2}]^2}{\frac{\alpha}{2}[k-\frac{k}{2}]}$$

$$= k \qquad \text{(QED)}.$$

Note, this three-point design can be reduced to a two-point design if we let $\alpha = 1$. Thus, in its simplest form the 'optimal' design for a controlled calibration experiment is to take $\frac{n}{2}$ readings at x=a and $\frac{n}{2}$ readings at x=b. For completeness we demonstrate that an equally-weighted (k+1) point design does not possess this 'optimality'. For this case

$$n_j = \frac{n}{k+1} \quad \forall\, j.$$

$$\Rightarrow \quad p_j = \frac{1}{k+1} \quad \forall\, j.$$

and thus

$$\sum_{j=1}^{k+1} p_j(j-1) = \frac{k}{2}$$

with

$$s = \begin{cases} 1 + \frac{k}{2} & k \text{ even} \\ 1 + \frac{k+1}{2} & k \text{ odd} \end{cases}$$

Without loss of generality we assume k to be even and hence $s = \frac{k+2}{2}$.

Now

$$\sum_{j=1}^{k+1} [(j-1) - \frac{k}{2}]^2$$

$$= \sum_{j=1}^{k+1} [(j-1)^2 - 2k(j-1) + \frac{k^2}{4}]$$

$$= \frac{k^3 + 3k^2 + 2k}{12} \qquad (2.14)$$

and

$$\sum_{j=\frac{k+2}{2}}^{k+1} [(j-1) - \frac{k}{2}] = \frac{k^2 + 6k + 8}{8} \qquad (2.15)$$

Hence the ratio in equation (2.12) is

$$\frac{\sum_{j=1}^{k+1} (\frac{1}{k+1}) [(j-1) - \frac{k}{2}]^2}{\sum_{j=s}^{k+1} (\frac{1}{k+1}) [(j-1) - \frac{k}{2}]} = \frac{\sum_{j=1}^{k+1} [(j-1) - \frac{k}{2}]^2}{\sum_{j=s}^{k+1} [(j-1) - \frac{k}{2}]}$$

Substituting equations (2.14) and (2.15) into the above we find that this ratio equals

$$\frac{2}{3} \frac{k^3 + 3k^2 + 2k}{k^2 + 6k + 8}$$

which is always less than k.

2.3 THE GEOMETRY OF LEAST SQUARES

The usual requirements of ordinary least-squares (OLS) curve-fitting is that the independent variable (invariably identified as X) is measured without error and, unlike Y, is not assumed to be subject to random variation. In other words X is a deterministic variable while Y is a random variable.

Furthermore we usually assume Y to follow a normal distribution when inferences about the estimated regression line are to be made.

It is not inconceivable that there are many practical situations in which the above assumptions simply do not hold. In this section the problem of simple linear regression when both variables are subject to error is examined. It is noted that OLS is just a special case of a more general regression methodology in which we allow the X,Y data to be projected at any angle onto the regression line.

When the OLS conditions are met then the x,y data are projected vertically and the least-squares criterion seeks that regression line which minimizes the sum of the vertical distances squared. If we take the other extreme where all the error is in X and Y is measured without error then we can simply interchange the roles of X and Y in our original regression formulation. This is equivalent to minimizing the sum of the horizontal distances squared. In the case where there is error in both X and Y it would seem appropriate to project the x,y data at an angle somewhere between 0 and $\frac{\pi}{2}$ - the actual value depending in some way on the ratio of the error variances.

For the present situation we assume the regression model :

$$Y_i = \beta_0 + \beta_1 X_i + \xi_i \qquad (2.16)$$

At this stage we make no assumptions about the distributional properties of ξ since our treatment (at least initially) will be purely data-analytic and not inferential.

Consider a typical data point (x_i,y_i). Figure 2 shows the point (x_i,y_i), its relation to the regression line and the various angles involved.

In the following development we will assume that the X,Y data have been centered at their respective means i.e. X^* and Y^* will be used to denote the centered values $X-\bar{X}$ and $Y-\bar{Y}$ respectively. After such a transformation the regression equation in equation (2.16) reduces to

$$Y_i^* = \beta_1^* x_i + \xi_i^* \tag{2.17}$$

$$\text{where} \quad \xi_i^* = \xi_i - \bar{\xi}_i \quad \text{and} \quad \bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i \ .$$

With reference to figure 2, $(\tilde{x}_i^*,\tilde{y}_i^*)$ denotes the projection of the point (x_i,y_i) onto the regression line, θ is the angle of projection, γ is the angle the regression line makes with the line segment joining (x_i^*,y_i^*) and $(\tilde{x}_i^*,\tilde{y}_i^*)$ and ϕ is the angle the regression line makes with the Y^* axis.

Now, from elementary mensuration we have that the angle ψ formed by the intersection of two lines having gradients ν_1 and ν_2 is given by the following equation

$$\tan(\psi) = \frac{\nu_1 - \nu_2}{1+\nu_1\nu_2} \ . \tag{2.18}$$

Since the Y-axis has slope $\nu_2=\infty$ we use the limiting form of equation (2.18)

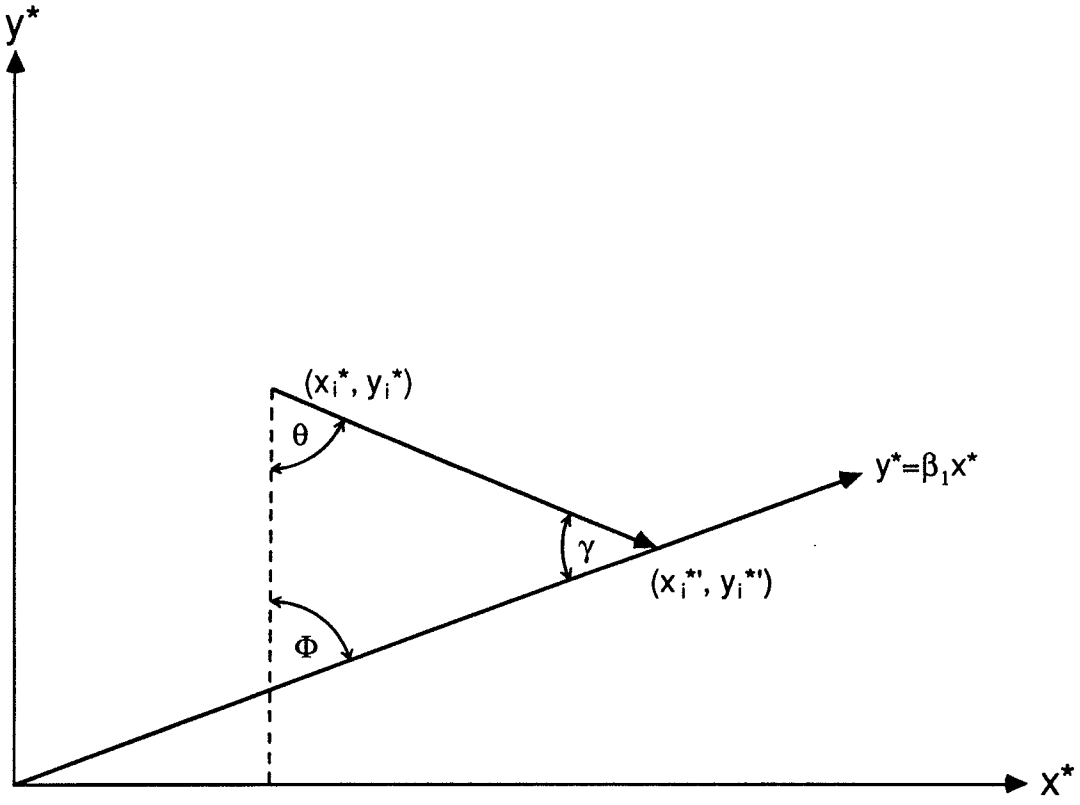

Figure 2. A typical data point and its projection onto the regression line.

$$\lim_{\nu_2 \to \infty} \left[\frac{\nu_2 - \nu_1}{1 + \nu_1 \nu_2} \right] = \lim_{\nu_2 \to \infty} \left[\frac{1 - \frac{\nu_1}{\nu_2}}{\frac{1}{\nu_2} + \nu_1} \right] = \frac{1}{\nu_1} \tag{2.19}$$

Thus,

$$\phi = \tan^{-1}\left(\frac{1}{\nu_1} \right) \tag{2.20}$$

and

$$\gamma = \pi - \theta - \phi$$

$$\Rightarrow \quad \gamma = \pi - \theta - \tan^{-1}\left(\frac{1}{\nu_1} \right) \tag{2.21}$$

By equation (2.18) we have that $\tan \gamma = \frac{\nu_2 - \beta_1}{1 + \nu_2 \beta_1}$ where β_1 is the slope of the regression line. But $\tan \gamma = \tan[\pi - \theta + \tan^{-1}(\frac{1}{\beta_1})]$. Using elementary trigonometric relations we have

$$\tan[\pi - \theta + \tan^{-1}\left(\frac{1}{\beta_1} \right)] = -\tan \theta + \tan^{-1}\left(\frac{1}{\beta_1} \right)$$

and

$$\tan \theta + \tan^{-1}\left(\frac{1}{\beta_1} \right) = \frac{\tan(\theta) + \tan[\tan^{-1}(\frac{1}{\beta_1})]}{1 - \tan(\theta)\tan[\tan^{-1}(\frac{1}{\beta_1})]}$$

$$= \frac{\tan(\theta) + \frac{1}{\beta_1}}{1 - \frac{1}{\beta_1}\tan(\theta)}$$

Therefore

$$\tan(\gamma) = \frac{1 + \beta_1 \tan(\theta)}{\beta_1 - \tan(\theta)} \tag{2.22}$$

However, using equation (2.18) we also have that $\tan(\gamma) = \frac{\nu_2 - \beta_1}{\beta_1 - \tan(\theta)}$ where ν_2 is the slope of the line segment joining $(\tilde{x}_i^*, \tilde{y}_i^*)$ and $(\tilde{x}_i, \tilde{y}_i)$. Thus equating this expression for $\tan(\gamma)$ with equation (2.20) we obtain the following relation

$$\frac{\nu_2 - \beta_1}{1 + \beta_1 \nu_2} = \frac{1 + \beta_1 \tan(\theta)}{\beta_1 - \tan(\theta)} \tag{2.23}$$

Substituting $\nu_2 = \frac{\tilde{y}_i^* - y_i^*}{\tilde{x}_i^* - x_i^*}$ and $\tilde{y}_i^* = \beta_1 \tilde{x}_i^*$ into equation (2.23) gives

$$\nu_2 = -\cot(\theta) \tag{2.24}$$

Now $\nu_2 = \frac{\tilde{y}_i^* - y_i^*}{\tilde{x}_i^* - x_i^*} = \frac{\beta_1 \tilde{x}_i^* - y_i^*}{\tilde{x}_i^* - x_i^*} = -\cot(\theta)$ from which we obtain the transformation equations

$$\tilde{x}_i^* = \frac{y_i^* + x_i^* \cot(\theta)}{\beta_1 + \cot(\theta)} \tag{2.25a}$$

and

$$\tilde{y}_i^* = \frac{\beta_1 y_i^* + x_i^* \beta_1 \cot(\theta)}{\beta_1 + \cot(\theta)} \tag{2.25b}$$

The transformation represented by equations (2.25a) and (2.25b) can be conveniently written in terms of the corresponding projection matrix:

$$\mathcal{P} = \frac{1}{\beta_1+\cot(\theta)} \begin{bmatrix} \cot(\theta) & 1 \\ \beta_1\cot(\theta) & \beta_1 \end{bmatrix} \qquad (2.26)$$

2.3.1 ORTHOGONAL PROJECTIONS

The special case where the data are projected orthogonally onto the regression line is considered next. This situation is depicted in figure 3. The point p in figure 3 is given as $p = \frac{a^T b}{a^T a} a$ and substituting $a = (x^*, \beta_1 x^*)$, and $b = (x_i^*, y_i^*)$ we obtain

$$p = \frac{x^* x_i^* + \beta_1 x^* y_i^*}{x^{*2}(1 + \beta_1^2)} \begin{bmatrix} x^* \\ \beta_1 x^* \end{bmatrix}$$

$$= \frac{x_i^* + \beta_1 y_i^*}{(1 + \beta_1^2)} \begin{bmatrix} 1 \\ \beta_1 \end{bmatrix}$$

and hence

$$\begin{bmatrix} x_i^* \\ y_i^* \end{bmatrix} = \frac{1}{(1 + \beta_1^2)} \begin{bmatrix} x_i^* + \beta_1 y_i^* \\ \beta_1 x_i^* + \beta_1^2 y_i^* \end{bmatrix}.$$

The corresponding projection matrix is easily seen to be

$$\mathcal{P} = \frac{1}{(1 + \beta_1^2)} \begin{bmatrix} 1 & \beta_1 \\ \beta_1 & \beta_1^2 \end{bmatrix}.$$

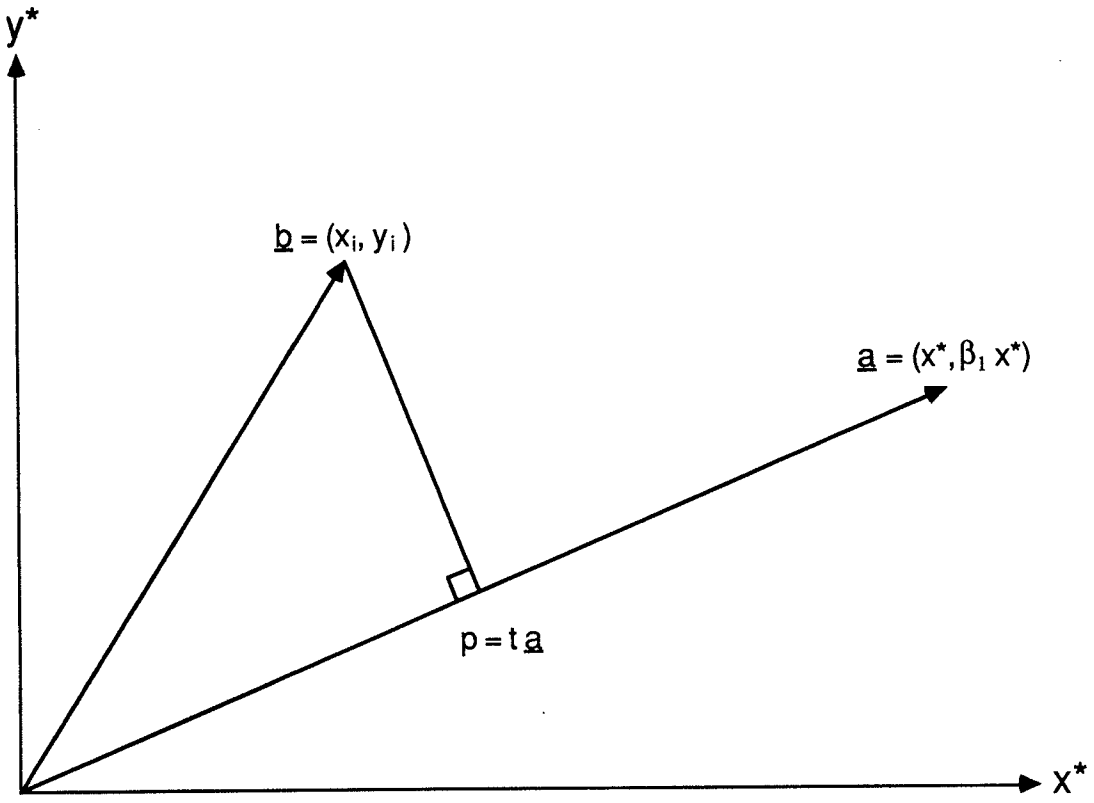

Figure 3. Orthogonal projection of a point onto the regression line.

Notice that the same result is obtained by letting $\gamma = \frac{\pi}{2}$ (see figure 2) : $\gamma = \pi - \theta - \phi$ and with $\gamma = \frac{\pi}{2}$ we have $\theta = \frac{\pi}{2} - \phi$

$$\Rightarrow \quad \tan(\theta) = \tan(\frac{\pi}{2} - \phi) = \cot(\phi)$$

but by equation (2.20)

$$\phi = \tan^{-1}(\frac{1}{\beta_1})$$

hence $$\cot(\phi) = \cot \tan^{-1}(\frac{1}{\beta_1}) = \beta_1$$

and $$\tan(\theta) = \beta_1 \Rightarrow \cot(\theta) = (\frac{1}{\beta_1}) \qquad (2.27)$$

Substituting equation (2.27) into the projection matrix given in equation (2.26) we obtain

$$\mathcal{P} = \frac{1}{(\beta_1 + \frac{1}{\beta_1})} \begin{bmatrix} \frac{1}{\beta_1} & 1 \\ 1 & \beta_1 \end{bmatrix}$$

$$\mathcal{P} = \frac{1}{(1 + \beta_1^2)} \begin{bmatrix} 1 & \beta_1 \\ \beta_1 & \beta_1^2 \end{bmatrix} \qquad \text{as before.}$$

2.3.2 THE 'OPTIMUM' PROJECTION ANGLE

As has already been alluded to there are infinitely many ways of arriving at a so-called least-squares fit to the experimental data with

any particular regression line being determined by the angle of projection. Vertical and orthogonal projections are simply two possibilities. Thus if one ignores the usual statistical considerations which dictate the fitting via OLS and allow θ to assume any value then we may reasonably seek an 'optimal' angle of projection. Not suprisingly, it will be shown that this situation corresponds to the orthogonal projection just discussed.

The length of the vector connecting the point (x_i^*, y_i^*) with its projection $(\tilde{x}_i^*, \tilde{y}_i^*)$ is $(\tilde{x}_i^* - x_i^*)^2 + (\tilde{y}_i^* - y_i^*)^2$. Thus our least-squares criterion becomes

$$\text{minimize } Q = \sum_{i=1}^{n} (\tilde{x}_i^* - x_i^*)^2 + (\beta_1 \tilde{x}_i^* - \tilde{y}_i^*)^2 \tag{2.28}$$

Now,
$$(\tilde{x}_i^* - x_i^*) = \frac{\tilde{y}_i^* - \beta_1 \tilde{x}_i^*}{\beta_1 + \cot(\theta)} \tag{2.29a}$$

and

$$(\tilde{y}_i^* - y_i^*) = \frac{(\beta_1 \tilde{x}_i^* - \tilde{y}_i^*)}{\beta_1 \tan(\theta) + 1} \tag{2.29b}$$

Thus
$$Q = \sum_{i=1}^{n} \left[\left[\frac{y_i^* - \beta_1 x_i^*}{\beta_1 + \cot(\theta)} \right]^2 + \left[\frac{\beta_1 x_i^* - y_i^*}{\beta_1 \tan(\theta) + 1} \right]^2 \right] \tag{2.30}$$

We now take partial derivatives of Q with respect to both θ and β_1.

First with respect to θ :

$$\frac{\delta Q}{\delta \theta} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \left[\frac{(y_i^* - \beta_1 x_i^*)^2}{\beta_1 + \cot(\theta)} + \frac{\beta_1 x_i^* - y_i}{\beta_1 \tan(\theta) + 1} \right]$$

from which we obtain

$$\beta_1 \tan^2(\theta) \left[\frac{\beta_1 + \cot(\theta)}{1 + \beta_1 \tan(\theta)} \right]^3 = 1$$

or

$$[1 + \beta_1 \tan(\theta)]^3 = \beta_1 \tan^2(\theta) [\beta_1 + \cot(\theta)]^3.$$

Letting $\tan(\theta) = z$ this last expression becomes

$$(1 + \beta_1 z)^3 = \beta_1 z^2 (\beta_1 + \frac{1}{z})^3$$

$$\Rightarrow \qquad (1 + \beta_1 z)^3 = \frac{\beta_1}{z} (\beta_1 z + 1)^3$$

Choosing the root which makes physical sense for this problem we find that

$$z = \beta_1 \quad \Rightarrow \quad \tan(\theta) = \beta_1 \tag{2.31}$$

which, as was shown in the previous section corresponds to the orthogonal projection case. Note, the other real solution to equation (2.31) is $z = \tan(\theta) = -1/\beta_1$ which represents a line perpendicular to the orthogonal projection (as is shown later, these two lines correspond to the eigenvectors of the X,Y covariance matrix).

The relationship between Q, β_1, and γ (angle projection makes with the regression line) is graphically illustrated in figure 4. It is apparent from figure 4 that the minimum for Q is attained at $\gamma = \pm 1.57$ radians (i.e. $\pm \frac{\pi}{2}$) as has just been demonstrated.

We next take the derivative of Q with respect to β_1.

First observe that Q can be rewritten as follows :

$$Q = \sum_{i=1}^{n} \left[\left[\frac{y_i^* - \beta_1 x_i^*}{\beta_1 + \cot(\theta)} \right]^2 + \left[\frac{\beta_1 x_i^* - y_i^*}{\beta_1 \tan(\theta) + 1} \right]^2 \right]$$

$$= \frac{1}{[1 + \beta_1 \tan(\theta)]^2} \sum_{i=1}^{n} [y_i^* - \beta_1 x_i^*]^2 [1 + \tan^2(\theta)]$$

$$= \frac{\sec^2(\theta)}{[1 + \beta_1 \tan(\theta)]^2} \sum_{i=1}^{n} [y_i^* - \beta_1 x_i^*]^2 \qquad (2.32)$$

Incidentally, when $\theta = 0$ equation (2.32) reduces to

$$Q' = \sum_{i=1}^{n} [y_i^* - \beta_1 x_i^*]^2$$

which is the usual least-squares criterion.

Returning now to equation (2.32) we have

$$\frac{\delta Q}{\delta \beta_1} = 0 \Rightarrow$$

$$\frac{-2\sec^2(\theta)\tan(\theta)}{[1 + \beta_1 \tan(\theta)]^3} \sum_{i=1}^{n} [y_i^* - \beta_1 x_i^*]^2 + \frac{\sec^2(\theta)}{[1 + \beta_1 \tan(\theta)]^2} \sum_{i=1}^{n} 2[y_i^* - \beta_1 x_i^*]^2 (-x_i)$$

$$= 0$$

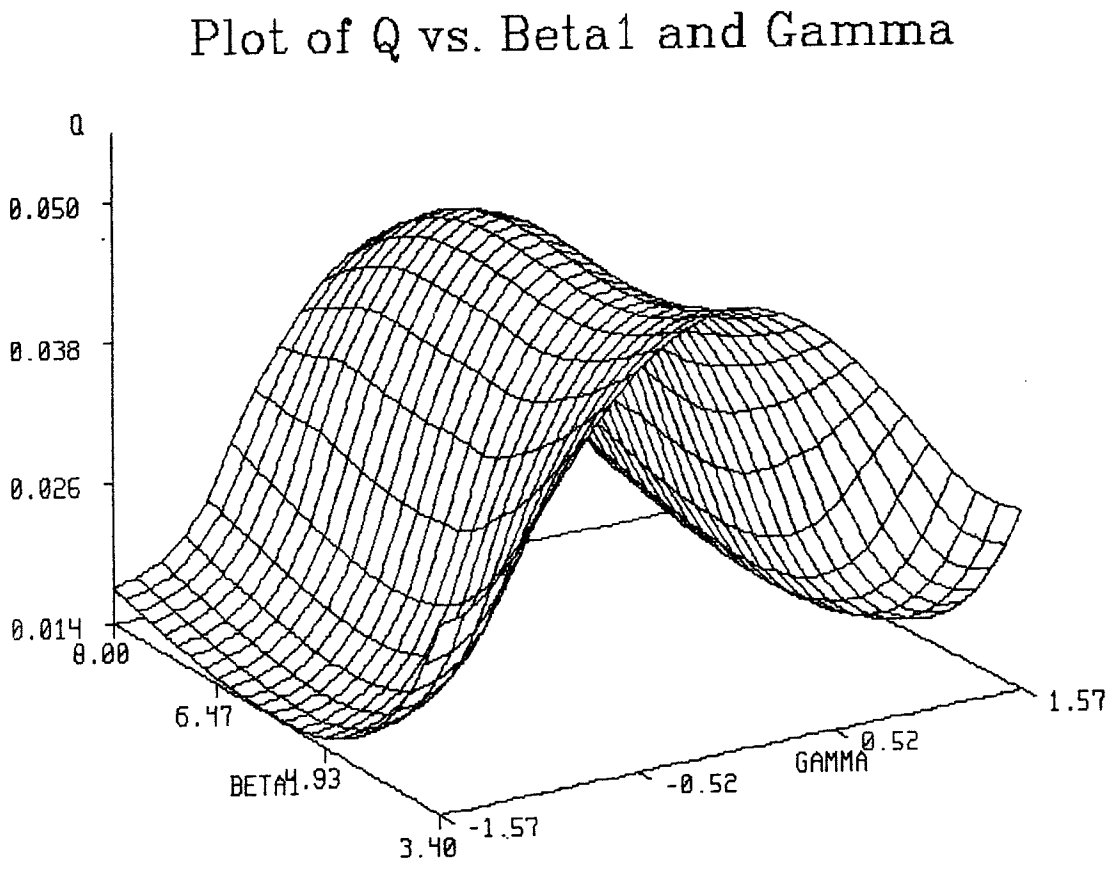

Figure 4. Relationship between error sum-of-squares (Q), the slope of the regression line (β_1), and the angle of projection (γ).

$$\Rightarrow \quad \frac{\sin(\theta)\sum_{i=1}^{n}(y_i^* - \beta_1 x_i^*)^2}{[\cos(\theta) + \beta_1\sin(\theta)]} + \sum_{i=1}^{n}(x_i^* y_i^*) - \beta_1\sum_{i=1}^{n} x_i^{*2} = 0$$

$$\Rightarrow \quad \beta_1[\sin(\theta)x_i^* y_i^* + \cos(\theta)x_i^{*2}] + \sum_{i=1}^{n}[\sin(\theta)y_i^{*2} + \cos(\theta)x_i^* y_i^*] = 0$$

Thus,

$$\hat{\beta}_1 = \frac{\sin(\theta)\sum_{i=1}^{n} y_i^{*2} + \cos(\theta)\sum_{i=1}^{n} x_i^* y_i^*}{\cos(\theta)\sum_{i=1}^{n} x_i^{*2} + \sin(\theta)\sum_{i=1}^{n} x_i^* y_i^*} \qquad (2.33)$$

Note, when $\theta = 0$ equation (2.33) reduces to

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i^* y_i^{*2}}{\sum_{i=1}^{n} x_i^{*2}}$$ which is the OLS estimate.

2.3.3 RELATIONSHIP TO THE SPECTRAL DECOMPOSITION OF THE COVARIANCE MATRIX.

Let $\mathbb{C}$ denote the sample covariance matrix i.e.

$$\mathbb{C} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}$$

Furthermore let $\underline{e} = [e_1, e_2]^T$ be an eigenvector of $\mathbb{C}$ and hence $\mathbb{C}\underline{e} = \lambda\underline{e}$ where λ is the corresponding eigenvalue. Now $\lambda\underline{e} = [\lambda e_1, \lambda e_2]^T$ and the gradient of the line from the origin to the point $(\lambda e_1, \lambda e_2)$ is equal to $\beta_1 = \frac{\lambda e_2}{\lambda e_1} = \frac{e_2}{e_1}$.

But,

$$\mathbb{C}\underline{e} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} e_1 S_{xx} + e_2 S_{xy} \\ e_1 S_{xy} + e_2 S_{yy} \end{bmatrix} \qquad \text{and so}$$

$$\beta_1 = \frac{e_2 S_{yy} + e_1 S_{xy}}{e_1 S_{xx} + e_2 S_{xy}} \qquad (2.34)$$

Thus equating coefficients in equations (2.33) and (2.34) we see that

$$e_2 = \sin(\theta) \qquad \text{and} \qquad e_1 = \cos(\theta)$$

i.e.

$$\underline{e} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$$

from which we observe $\beta_1 = \frac{\sin(\theta)}{\cos(\theta)} = \tan(\theta)$.

However, in section 2.3.1 it was shown (equation (2.27)) that for an orthogonal projection $\beta_1 = \tan(\theta)$. We are therefore led to conclude that the eigenvector corresponds to a least-squares fit when the data

are projected *orthogonally* onto the regression line. Rather than perform the spectral decomposition of $\mathbb{C}$ we can obtain an expression for $\hat{\beta}_1$ from previous equations. Equation (2.33) is reproduced below :

$$\hat{\beta}_1 = \frac{\sin(\theta)\sum_{i=1}^{n} y_i^{*2} + \cos(\theta)\sum_{i=1}^{n} x_i^* y_i^*}{\cos(\theta)\sum_{i=1}^{n} x_i^{*2} + \sin(\theta)\sum_{i=1}^{n} x_i^* y_i^*}$$

Dividing through by $\cos(\theta)$ we obtain

$$\hat{\beta}_1 = \frac{\tan(\theta)S_{yy} + S_{xy}}{S_{xx} + \tan(\theta)S_{xy}}$$

but, as has already been established for a perpendicular projection $\hat{\beta}_1 = \tan(\theta)$. Substituting into the expression above for $\tan(\theta)$ we have

$$\hat{\beta}_1 = \frac{\hat{\beta}_1 S_{yy} + S_{xy}}{S_{xx} + \hat{\beta}_1 S_{xy}}$$

This is a quadratic in $\hat{\beta}_1$ for which the solutions are given as

$$\hat{\beta}_1 = \frac{(S_{yy} - S_{xx}) \pm [(S_{yy} - S_{xx})^2 + 4S_{xy}^2]^{\frac{1}{2}}}{2S_{xy}} \qquad (2.35)$$

2.4 CALIBRATION AND MEASUREMENT ERROR MODELS.

We now examine the use of various calibration techniques in the presence of measurement error in X. The basic measurement error model as defined by Fuller (1987) is

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + e_i \\ X_i &= x_i + u_i \end{aligned} \tag{2.36}$$

with

$$\begin{bmatrix} x_i \\ e_i \\ u_i \end{bmatrix} \sim N(\mu, \Sigma) \qquad \text{and} \qquad \mu = \begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix}$$

where $\Sigma = \text{diag}(\sigma_x^2, \sigma_e^2, \sigma_u^2)$

In this representation x_i is a true (unobserved) random variable and X_i is the observed measure of x_i. Some properties of this model are now developed.

Let σ_Y^2 be the variance of the Y_i's , $\sigma_{XY}^2 = \text{Cov}[X_i, Y_i]$ and $\sigma_X^2 = \text{Var}[X_i]$. Furthermore, $\mu_X = \mathbb{E}[X_i] = \mathbb{E}[x_i + u_i] = \mu_x$ and $\mu_Y = \mathbb{E}[\beta_0 + \beta_1 x_i + e_i] = \beta_0 + \beta_1 \mu_x$.

Now,

$$\begin{aligned} \text{Cov}[X_i, Y_i] &= \mathbb{E}[(X_i - \mu_x)(\beta_0 + \beta_1 x_i + e_i - \beta_0 - \beta_1 \mu_x)] \\ &= \mathbb{E}\{[X_i - \mu_x][\beta_1 (x_i - \mu_x) + e_i]\} \\ &= \beta_1 \mathbb{E}[(X_i - \mu_x)(x_i - \mu_x)] + \mathbb{E}[(X_i - \mu_x) e_i] \end{aligned}$$

the last expectation is zero by the (assumed) independence of X_i and e_i thus

$$\mathrm{Cov}[X_i, Y_i] = \beta_1 \mathbb{E}[(X_i - \mu_x)(x_i - \mu_x)]$$

but $\mu_X = \mu_x$ and $X_i = x_i + u_i$ therefore

$$\begin{aligned}\mathrm{Cov}[X_i, Y_i] &= \beta_1 \mathbb{E}[(x_i + u_i - \mu_x)(x_i - \mu_x)] \\ &= \beta_1 \mathbb{E}[(x_i - \mu_x)^2 + u_i(x_i - \mu_x)] \\ &= \beta_1 \sigma_x^2 + \beta_1 \mathbb{E}[u_i(x_i - \mu_x)]\end{aligned}$$

again, by the independence of u_i and x_i the last expectation is zero, and hence

$$\mathrm{Cov}[X_i, Y_i] = \beta_1 \sigma_x^2 \qquad (2.37)$$

Moment estimators for the parameters of this model are readily obtained if it is assumed that σ_u^2 (measurement error variance) is known. The defining equations are

$$\begin{aligned}\mu_Y &= \beta_0 + \beta_1 \mu_x \\ \mu_X &= \mu_x \\ \sigma_Y^2 &= \beta_1^2 \sigma_x^2 + \sigma_e^2 \\ \sigma_{XY} &= \beta_1 \sigma_x^2 \\ \sigma_X^2 &= \sigma_x^2 + \sigma_u^2\end{aligned}$$

Starting with the last equation first and working backwards we can successively solve the above system replacing the the population values with their corresponding sample estimates. The following moment estimators are obtained :

$$\hat{\sigma}_x^2 = S_{XX} - \sigma_u^2$$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX} - \sigma_u^2}$$

$$\hat{\sigma}_e^2 = S_{YY} - \hat{\beta}_1 S_{XY}$$

$$\hat{\mu}_x = \bar{X}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Since it is required that $\hat{\sigma}_x^2 > 0$ and $\hat{\sigma}_e^2 > 0$ we have

$$S_{XX} - \sigma_u^2 > 0$$

and $$S_{YY} - \hat{\beta}_1 S_{XY} > 0 \Rightarrow S_{YY} > \frac{S_{XY}}{S_{XX} - \sigma_u^2} .$$

Note that the first constraint is made redundant by the second and therefore the requirement becomes

$$S_{YY}(S_{XX} - \sigma_u^2) - S_{XY} > 0 \quad .$$

Estimation of the true, unknown x may be handled using a GLS estimator.

Writing the model in matrix form we have

$$\begin{bmatrix} Y_i - \beta_0 \\ X_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ 1 \end{bmatrix} x_i + \begin{bmatrix} e_i \\ u_i \end{bmatrix}$$

or

$$Z = \beta x + \xi$$

then the GLS estimator for x is

$$\hat{x} = [\beta^T \Sigma_e^{-1} \beta]^{-1} \beta^T \Sigma_e^{-1} Z \qquad (2.38)$$

with

$$\text{Var}[\hat{x}] = [\beta^T \Sigma_e^{-1} \beta]^{-1}$$

where

Σ_e is the covariance matrix of $[e_i, u_i]$.

If Σ_e is the diagonal matrix (σ_e^2, σ_u^2) then equation (2.38) reduces to

$$\hat{x} = \frac{\frac{\beta_1}{\sigma_e^2}(Y_i - \beta_0) + \frac{X_i}{\sigma_u^2}}{\frac{\beta_1^2}{\sigma_e^2} + \frac{1}{\sigma_u^2}} \qquad (2.39)$$

The errors in both variables presents no special difficulties for the calibration of x . The techniques of classical or inverse regression apply equally well in this situation although we mention an alternative estimator for the calibration of an unknown x as proposed by Fuller (1987).

The model of equation (2.36) can be rewritten as

$$Y_i = y_i + e_i$$

where $y_i = \beta_0 + \beta_1 x_i$ is the "true" y-value corresponding to the true x_i. Thus, the m.l.e. for some unknown x_0 is

$$x_0 = \frac{y_0 - \beta_0}{\beta_1} = \gamma_0 + \gamma_1 y_0$$

where

$$\gamma_0 = -\frac{\beta_0}{\beta_1} \qquad \text{and} \qquad \gamma_1 = \frac{1}{\beta_1} \ .$$

Hence, γ_0 and γ_1 are parameters relating the regression of true y-values on the true x-values (neither of which are known). A future calibrated value of x (x_0) corresponding to a reading y_0 would be obtained as

$$\hat{x}_0 = \hat{\gamma}_0 + \hat{\gamma}_1 y_0 \ . \qquad (2.40)$$

Previous estimates for β_0 and β_1 could be used to obtain estimates for γ_0 and γ_1 using

$$\hat{\gamma}_0 = -\frac{\hat{\beta}_0}{\hat{\beta}_1} \qquad \text{and} \qquad \hat{\gamma}_1 = \frac{1}{\hat{\beta}_1}$$

however, as has already been noted, moments for these do not exist (under OLS conditions).

Fuller (1987) suggests the following modification which yields an estimator having finite mean and variance.

$$\hat{\gamma}_1 = \frac{(S_{XX} - \sigma_u^2)S_{XY} + \frac{2}{(n-1)}S_{XY}\sigma_u^2}{S_{XY} + \frac{1}{(n-1)}[S_{XX}S_{YY} - S_{XY}]} \quad \text{if} \quad \hat{\lambda} > 1$$

or

$$\hat{\gamma}_1 = \frac{S_{XY}}{S_{YY}} \quad \text{otherwise}$$

where

$$\hat{\lambda} = \frac{S_{XX} - \frac{(S_{XY})^2}{S_{YY}}}{\sigma_u^2}$$

An estimator for the variance of $\hat{\gamma}_1$ has also been provided by Fuller (1987,p178).

2.5 A COMPARATIVE STUDY OF POINT ESTIMATORS FOR THE MEASUREMENT ERROR MODEL.

We now present the results of a monte-carlo simulation in which four different methods of calibrating in the presence of measurement error are compared in terms of bias and efficiency.

The methods chosen are : (i) the classical estimator (ii) the estimator based on an orthogonal projection (iii) an estimator proposed by Mandel (1984) and (iv) the estimator suggested by Fuller (1987). Estimator (iii) has not yet been discussed. We digress momentarily to provide details.

Mandel (1984) proposed the following estimator for the measurement error model :

Define :

$$\lambda = \frac{\sigma_e^2}{\sigma_u^2}$$

and

$\theta = \rho\lambda^{\frac{1}{2}}$ where ρ is the correlation between σ_e^2 and σ_u^2 .

The proposed estimator for β_1 is :

$$\hat{\beta}_1 = \frac{(S_{YY} - \lambda S_{XX}) + [(S_{YY} - \lambda S_{XX})^2 - 4(S_{XY} - \theta S_{XX})(\theta S_{YY} - \lambda S_{XY})]^{\frac{1}{2}}}{2(S_{XY} - \theta S_{XX})} \qquad (2.41)$$

It is noted that when λ=1 and θ=0 equations (2.35) and (2.41) are identical.

For the purposes of simulation the parameters ρ, σ_e^2, σ_u^2, and σ_x^2 were each varied over three levels to yield a 3^4 factorial experiment. β_0 was held fixed at 0.5 and β_1 at 1.0 (these being reasonable choices

in a calibration setting). ρ assumed the values 0.0, 0.5, and 0.95 which represent no correlation, mild correlation, and strong correlation respectively. σ_e^2, σ_u^2, and σ_x^2 were each varied over 0.5, 4.0, and 25.0 - these representing approximately 2.5% , 20% , and 50% coefficient of variations for both the Y and X data (μ_x was held constant at 10.0). For each of the 81 treatment combinations the procedure was as follows :

Step 1 : Generate 101 values from $Z_{(1x2)} \sim N(0 , I)$. These form the basis for σ_e^2 and σ_u^2 .

Step 2 : Impose the required covariance structure on σ_e^2 and σ_u^2 as follows :

Given $\Sigma_e = \begin{bmatrix} \sigma_e^2 & \rho\sigma_e\sigma_u \\ \rho\sigma_e\sigma_u & \sigma_u^2 \end{bmatrix} = QDQ^T$ where the columns of Q contain the normalized eigenvectors of Σ_e and D is the diagonal matrix of corresponding eigenvectors , compute [e,u] as $Z A^T$ where $A = QD^{\frac{1}{2}}$.

Step 3 : Generate 101 observations for the x_i from $N(\mu_x, \sigma_x^2)$.

Step 4 : Compute $X_i = x_i + u_i$ and $Y_i = \beta_0 + \beta_1 x_i + e_i$.

Step 5 : Using the first *100* observations, estimate β_0 and β_1 using each of the four methods to be compared. The last observation (Y_{101}, X_{101}) is then used for calibration and the result $(\hat{X}_{101})$ compared with x_{101} .

This procedure gives *one* observation on $(\hat{X}_0 - x_0)$ for each of the four methods. Steps 3 to 5 are then repeated another 199 times for each treatment combination. The bias and MSE are then computed from the 200 observations for each treatment combination.

The entire simulation has been programmed using the matrix-based language GAUSS. A listing may be found in Appendix A.

2.5.1 RESULTS.

Results from the simulation run are displayed in tables 1 and 2. It is readily apparent that there is little difference in performance of Fuller's estimator and the classical estimator with respect to both bias and MSE. One is lead to the conclusion that the modification suggested by Fuller is of more theoretical rather than practical significance. We would therefore elect to use the computationally simpler classical approach in practice. Also obvious from table 2 are the two missing values reported for treatment combination ρ=0.95,σ_e^2=4.0,σ_x^2=0.25,σ_u^2=4.0 (labelled as 3,2,1,1) and ρ=0.95,σ_e^2=25.0,σ_x^2=0.25,σ_u^2=4.0 (labelled as 3,3,1,1). These observations are not in fact missing however their actual values are so large as to mask all other numbers in the table were they to be included. Interestingly both of these 'outliers' are associated with Mandel's procedure and one must therefore question the validity of this procedure under the condition of highly correlated errors and large error variance. Examination of the corresponding entries in table 1 also reveals exceptionally high bias. We now examine the bias and MSE data in more detail.

Table 1. Bias data for estimators (i) - (iv) obtained from simulation.

CONTROL: rho = 1
ROWS: var-e COLUMNS: var-x / var-u

	1			2			3		
	1	2	3	1	2	3	1	2	3
1	-0.06219	0.03179	-0.01604	0.07610	0.02214	0.05650	0.01543	-0.02199	0.00876
	-0.06278	0.03142	-0.01595	0.07150	0.02558	0.05648	0.01658	-0.01929	0.00896
	-0.0628	0.0312	-0.0159	0.0715	0.0285	0.0565	0.0166	-0.0169	0.0092
	-0.06224	0.03162	-0.01583	0.07591	0.02319	0.05646	0.01548	-0.02110	0.00919
2	-0.06087	-0.10434	-0.04820	0.25071	0.12507	-0.05756	-0.20727	0.11790	0.14463
	-0.01933	-0.10445	-0.04496	0.11510	0.11001	-0.04497	-0.21801	0.09764	0.14960
	-0.0204	-0.1044	-0.0431	0.2296	0.1100	-0.0366	-0.2087	0.0976	0.1539
	-0.03980	-0.10450	-0.04762	0.24680	0.12433	-0.05572	-0.20748	0.11697	0.14578
3	-1.00512	-0.03520	-0.18277	-0.01898	-0.34484	-0.35412	0.09743	-0.05739	0.53742
	0.02956	0.08530	-0.14026	-0.09484	-0.08895	-0.30093	0.10941	-0.02546	0.50000
	-0.2036	0.0558	-0.1403	-0.0285	-0.2062	-0.3009	0.0971	-0.0478	0.5000
	-0.07901	0.00355	-0.18085	-0.01227	-0.32431	-0.35203	0.09717	-0.05677	0.53572

CONTROL: rho = 2
ROWS: var-e COLUMNS: var-x / var-u

	1			2			3		
	1	2	3	1	2	3	1	2	3
1	-0.03977	-0.00653	-0.03390	0.00730	-0.01233	0.01002	-0.04221	-0.02982	0.01248
	-0.03199	-0.00656	-0.03386	0.00978	-0.01538	0.01089	-0.04397	-0.02946	0.01216
	-0.0301	-0.0066	-0.0338	0.0099	-0.0184	0.0118	-0.0440	-0.0290	0.0118
	-0.03936	-0.00655	-0.03383	0.00739	-0.01364	0.01169	-0.04228	-0.02975	0.01184
2	-0.04854	-0.10869	0.01515	0.18918	0.05032	0.01053	0.05686	0.06090	-0.08582
	-0.02365	-0.09195	0.01365	0.09166	0.02752	0.00150	0.02644	0.04242	-0.06741
	-0.2181	-0.0874	0.0129	0.2109	0.0216	-0.0055	0.0618	0.0403	-0.0516
	-0.04757	-0.10754	0.01500	0.18660	0.04924	0.00881	0.05619	0.06008	-0.08270
3	-0.19161	0.34094	0.09933	0.43562	-0.04430	0.51056	-0.35480	-0.00285	-0.11778
	-0.02874	0.17736	0.08562	-0.01617	-0.14885	0.40091	-0.14804	-0.03408	0.01513
	-1.1600	0.1811	0.0819	0.7074	-0.1238	0.3713	-0.3802	-0.0015	0.0487
	-0.18406	0.33384	0.09818	0.39583	-0.04791	0.50378	-0.34872	-0.00357	-0.11130

CONTROL: rho = 3
ROWS: var-e COLUMNS: var-x / var-u

	1			2			3		
	1	2	3	1	2	3	1	2	3
1	0.02378	0.00161	0.02577	0.00734	0.01476	0.00869	0.01187	0.01129	0.05084
	0.02508	0.00169	0.02576	0.00679	0.01574	0.00865	0.01265	0.01010	0.05108
	0.0277	0.0018	0.0258	0.0035	0.0170	0.0086	0.0130	0.0084	0.0513
	0.02392	0.00175	0.02576	0.00733	0.01536	0.00863	0.01190	0.01091	0.05130
2	0.08912	-0.07846	0.06349	0.02805	0.02850	-0.19231	0.02208	-0.09614	0.01874
	0.06073	-0.07621	0.06299	0.04354	0.01381	-0.19349	0.02572	-0.12639	0.02688
	5.1866	-0.0734	0.0624	0.0150	-0.0157	-0.1947	0.0197	-0.2273	0.0379
	0.08827	-0.07779	0.06274	0.02839	0.02733	-0.19321	0.02215	-0.09758	0.02126
3	0.09746	-0.02164	0.14469	0.37870	-0.12506	-0.07835	-0.33713	-0.23816	0.09189
	0.07067	-0.01730	0.14039	-0.05197	-0.10631	-0.08022	-0.12636	-0.09319	0.08845
	21.1917	-0.0100	0.1353	1.7885	0.2557	-0.0828	-0.4062	-0.5388	0.0838
	0.09664	-0.02125	0.14293	0.35730	-0.12445	-0.07854	-0.33143	-0.23468	0.09170

Table 2. Mean square error data for estimators (i) - (iv) obtained from simulation.

CONTROL: rho = 1
ROWS: var-e COLUMNS: var-x / var-u

	1			2			3		
	1	2	3	1	2	3	1	2	3
1	0.240	0.265	0.237	0.218	0.284	0.246	0.236	0.239	0.243
	0.1637	0.2497	0.2343	0.2051	0.2742	0.2445	0.2341	0.2373	0.2419
	0.164	0.238	0.232	0.205	0.268	0.243	0.234	0.236	0.241
	0.235	0.260	0.232	0.217	0.280	0.243	0.235	0.239	0.242
2	6.864	2.926	4.003	4.954	3.989	4.127	3.482	3.441	3.467
	0.2501	1.2368	3.4232	2.4742	2.5958	3.6157	3.0794	3.0125	3.2526
	1.884	1.237	3.099	4.422	2.596	3.341	3.413	3.013	3.160
	4.457	2.817	3.912	4.844	3.905	4.045	3.468	3.418	3.429
3	466.429	28.976	25.377	35.577	25.916	30.647	21.782	25.264	29.412
	0.2823	1.0118	10.1270	3.3676	4.4287	13.5089	12.2542	15.3540	18.4551
	116.738	3.717	10.127	31.426	10.304	13.509	21.381	20.437	18.455
	8.248	24.190	24.392	28.849	23.810	29.539	21.330	24.754	28.777

CONTROL: rho = 2
ROWS: var-e COLUMNS: var-x / var-u

	1			2			3		
	1	2	3	1	2	3	1	2	3
1	0.187	0.233	0.234	0.246	0.242	0.257	0.258	0.250	0.277
	0.1566	0.2253	0.2328	0.2419	0.2440	0.2562	0.2563	0.2541	0.2772
	0.151	0.219	0.231	0.242	0.247	0.255	0.256	0.260	0.278
	0.185	0.229	0.231	0.246	0.243	0.255	0.258	0.251	0.278
2	1.032	2.145	2.636	2.922	2.808	3.575	3.724	4.217	3.944
	0.2894	1.4126	2.4082	2.1076	2.2752	3.3071	3.3766	4.1453	3.8833
	8.240	1.247	2.249	3.251	2.181	3.121	3.809	4.155	3.870
	0.990	2.092	2.582	2.886	2.775	3.518	3.713	4.210	3.929
3	2.093	4.473	11.357	18.590	11.078	12.987	28.530	17.862	15.489
	0.3364	1.3823	7.3480	4.2990	4.8691	9.0245	15.2703	12.4137	12.4327
	248.668	1.538	6.439	37.514	7.796	8.139	31.240	18.629	11.924
	1.901	4.298	11.054	16.608	10.774	12.710	27.975	17.637	15.295

CONTROL: rho = 3
ROWS: var-e COLUMNS: var-x / var-u

	1			2			3		
	1	2	3	1	2	3	1	2	3
1	0.112	0.172	0.215	0.211	0.251	0.220	0.272	0.282	0.315
	0.1055	0.1712	0.2148	0.2059	0.2544	0.2201	0.2707	0.2869	0.3165
	0.102	0.170	0.215	0.217	0.259	0.220	0.271	0.295	0.318
	0.112	0.171	0.215	0.211	0.253	0.220	0.272	0.284	0.318
2	0.660	1.045	2.462	3.050	2.428	2.494	4.159	3.696	3.438
	0.4035	0.9917	2.4343	2.2216	2.2830	2.4599	3.9390	3.7648	3.4729
	--	0.929	2.405	4.908	2.156	2.423	4.466	5.034	3.542
	0.650	1.027	2.420	3.022	2.416	2.467	4.151	3.696	3.447
3	0.813	2.469	6.441	13.719	6.241	9.105	20.805	15.747	16.091
	0.3922	2.1400	6.1708	3.8173	4.4490	8.4798	12.1426	12.3019	15.6255
	--	1.698	5.856	177.990	32.960	7.721	25.805	52.421	15.492
	0.796	2.420	6.311	12.978	6.161	8.970	20.444	15.622	16.047

2.5.1.1 **BIAS**.

Broken down by ρ :

	ρ=0.0	ρ=0.5	ρ=0.95	All
classical	-0.04048	0.02519	-0.00180	-0.00570
orthogonal	0.00841	0.00722	-0.00669	0.00298
Mandel	0.0008	-0.0155	1.0142	0.3332
Fuller	-0.00287	0.02369	-0.00235	0.00616

Broken down by σ_e^2 :

	σ_e^2=0.25	σ_e^2=4.0	σ_e^2=25.0
classical	0.00488	0.00678	-0.02874
orthogonal	0.00525	-0.00376	0.00745
Mandel	0.0054	0.1842	0.8099
Fuller	0.0050	0.00735	0.00613

Broken down by σ_u^2 :

	σ_u^2=0.25	σ_u^2=4.0	σ_u^2=25.0
classical	-0.02153	-0.01935	0.02380
orthogonal	-0.00558	-0.01279	0.02730
Mandel	0.9958	-0.0244	0.0281
Fuller	0.01197	-0.1753	0.02404

Broken down by σ_x^2 :

	σ_x^2=0.25	σ_x^2=4.0	σ_x^2=25.0
classical	-0.04114	0.03638	-0.1233
orthogonal	0.00416	-0.00491	0.00969
Mandel	0.9179	0.1076	-0.0260
Fuller	-0.00458	0.03453	-0.01147

With respect to the above tabulations we note the following :

(i) overall, the orthogonal estimator has the smallest absolute bias while Mandel's procedure has the largest.

(ii) the bias associated with Mandel's procedure gets progressively worse with the severity of violation of the independence assumption.

(iii) the bias of the orthogonal procedure is relatively unaffected by the violation of the independence assumption.

(iv) the bias of Mandel's procedure increases steadily with increasing error variance (σ_e^2). The other estimators show little variation with respect to changing levels of this parameter.

(v) there appears to be no real effect on the bias due to changing levels of the measurement error variance (σ_u^2) or the variance of the unobserved random variable x.

2.5.1.2 MSE.

Broken down by ρ :

	ρ=0.0	ρ=0.5	ρ=0.95	All
classical	26.994	5.617	4.336	12.314
orthogonal	3.845	3.4343	2.3261	3.5318
Mandel	10.160	15.043	4165.0	1397.0
Fuller	9.273	5.449	4.263	6.328

Broken down by σ_e^2 :

	σ_e^2=0.25	σ_e^2=4.0	σ_e^2=25.0
classical	0.239	3.248	33.454
orthogonal	0.2324	2.5228	7.8402
Mandel	0.232	140.0	4050.0
Fuller	0.238	3.122	15.626

Broken down by σ_u^2 :

	σ_u^2=0.25	σ_u^2=4.0	σ_u^2=25.0
classical	23.747	6.183	7.011
orthogonal	2.672	3.0469	4.8766
Mandel	4180.0	6.0	5.0
Fuller	6.270	5.860	6.855

Broken down by σ_x^2 :

	σ_x^2=0.25	σ_x^2=4.0	σ_x^2=25.0
classical	21.263	7.273	8.4053
orthogonal	1.622	3.0271	5.9463
Mandel	4168.0	13.0	9.0
Fuller	3.942	6.757	8.286

The following observations are based on the above MSE data :

(i) overall, the orthogonal estimator has the smallest MSE while Mandel's procedure has the largest. It is disturbing to the practitioner to observe such extreme departures from the true calibrated value that are possible using Mandel's procedure. Again, this problem is most severe when the violation of the usual OLS assumptions is most extreme.

(ii) whilst the similarities between Fuller's method and the classical procedure have already been noted, it is nevertheless apparent that the latter results in an overall reduction in MSE of approximately 50%.

(iii) all estimators show a steadily increasing MSE as σ_e^2 increases, although the orthogonal procedure has a MSE which is consistently less than all of the others.

(iv) there is no obvious trend between MSE and changing levels of the measurement error variance (with the possible exception of the orthogonal estimator which shows a steady increase in MSE as σ_u^2 increases). The same comments apply to σ_x^2 .

The foregoing comments are perhaps better appreciated when considered with the visual displays provided in figures 5(a) to 5(c) and 6(a) to 6(c). We now subject the data to closer scrutiny and formally test the assertions made above.

2.5.1.3 ANALYSIS OF VARIANCE.

We next apply standard analysis of variance procedures to make inferences concerning the possible effects of the four factors considered on the bias and MSE of each of the four estimators. The computations associated with this aspect of the simulation study were performed using the SAS ANOVA procedure. For the sake of brevity, computer output has not been reproduced here, only final results are given.

With respect to bias, none of the estimators examined showed any main effect in the four factors (α <0.05). The MSE data was transformed prior to analysis in an attempt to remove the skewness of these distributions. The actual transformation employed was $MSE_{new} = (MSE_{old})^{\frac{1}{4}}$. The subsequent analysis of variance suggested a significant difference between calibration methods (p=0.0179) and a significant effect due to the residual error variance (σ_e^2) (p=0.0001). All remaining effects were judged to be insignificant with the exception of a possible interaction between ρ and the calibration method used (p=0.0339). Multiple comparison procedures (Duncan and SNK) indicated the following groupings for the four procedures : {orthogonal} ; {classical,Fuller} ; {Mandel}.

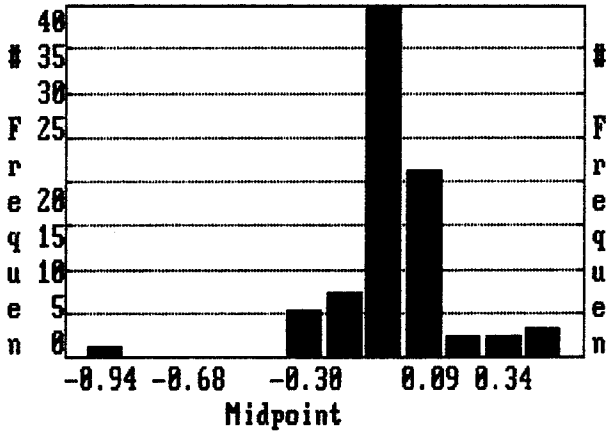

Figure 5(a). Histogram of bias for classical estimator.

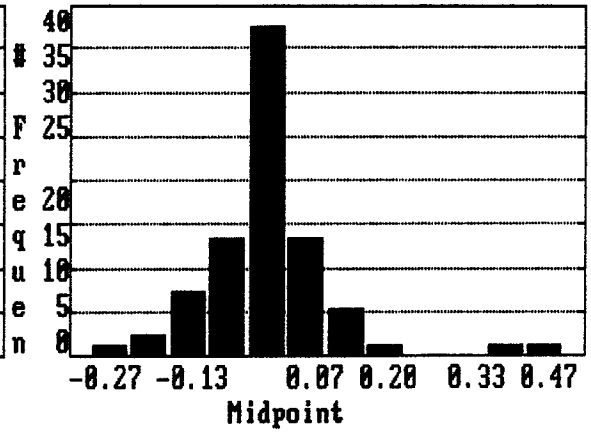

Figure 5(b). Histogram of bias for orthogonal estimator.

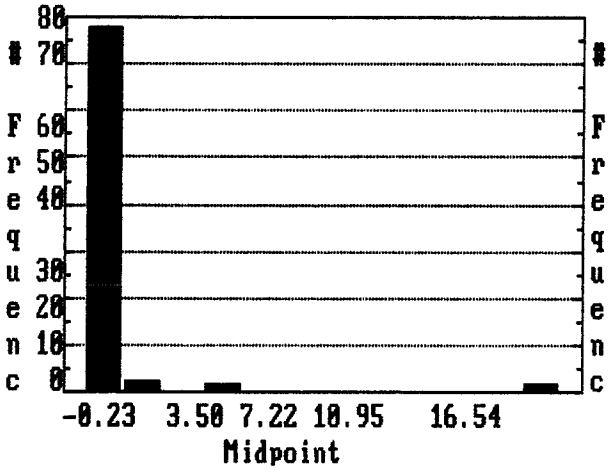

Figure 5(c). Histogram of bias for Mandel's estimator.

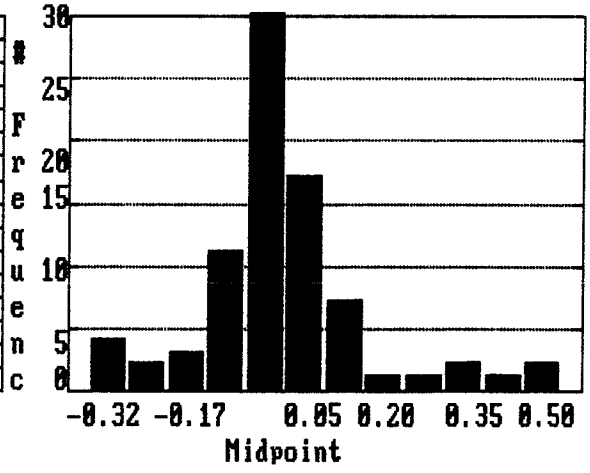

Figure 5(d). Histogram of bias for Fuller's estimator.

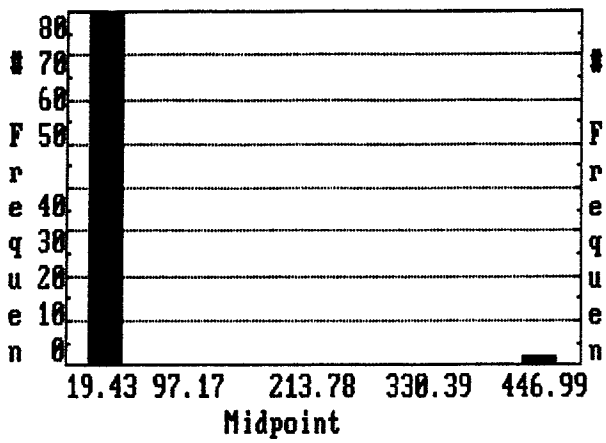

Figure 6(a). Histogram of MSE for classical estimator.

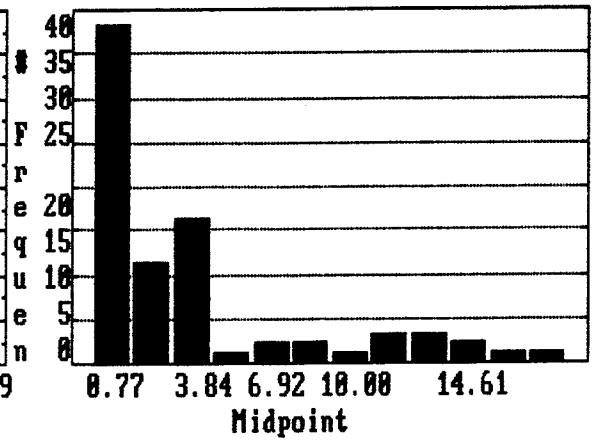

Figure 6(b). Histogram MSE for orthogonal estimator.

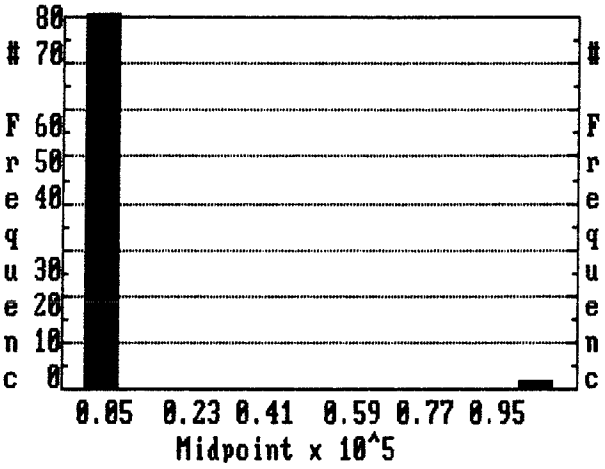

Figure 6(c). Histogram of MSE for Mandel's estimator.

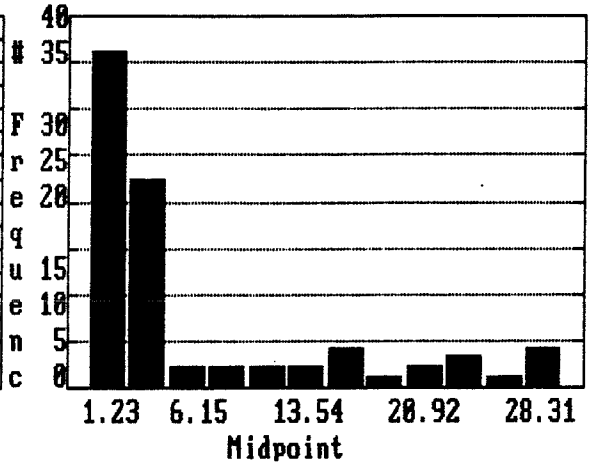

Figure 6(d). Histogram MSE for Fuller's estimator.

2.6 INTERVAL ESTIMATION.

Developments associated with interval estimation in calibration problems were discussed in chapter one. We now examine the performance of five different approaches to the interval estimation problem.

2.6.1 THE CLASSICAL INTERVAL ESTIMATE (I).

Let $J_0(x)$ be the $(1-\alpha)100\%$ prediction interval for response y i.e.

$$J_0(x) = \hat{y} \pm t_{n-2,\alpha/2}\, \hat{\sigma}_e \left[1 + \frac{1}{n} + \frac{(\hat{x}_0 - \bar{x})^2}{S_{xx}}\right]^{\frac{1}{2}} \qquad (2.42)$$

where $$\hat{\sigma}_e^2 = \frac{SSE_Y}{n-2}$$

and $$SSE_Y = (\hat{Y} - Y)^T(\hat{Y} - Y).$$

Then the confidence interval for X_0 is $K_0(y) = \{x \mid y \in J_0(x_0)\}$ (Carroll, et. al., 1988). This procedure is illustrated in figure 7.

2.6.2 THE INVERSE INTERVAL ESTIMATE.

The procedure for use with the inverse estimator is essentially the same as that for the classical estimator, except the roles of X and Y are reversed. Let the parameter estimates obtained from the regression of X on Y be denoted $\hat{\gamma}_0$ and $\hat{\gamma}_1$ for the intercept and slope respectively.

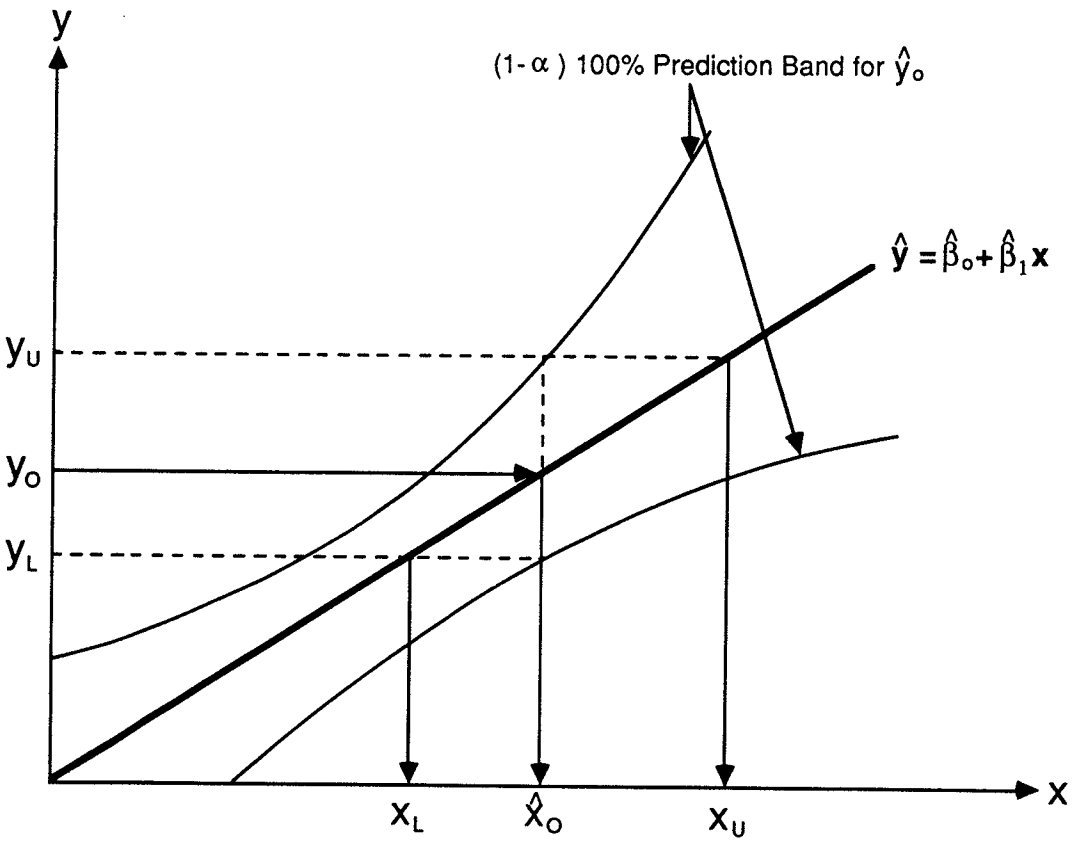

Figure 7. Graphical illustration of the procedure for obtaining confidence interval for X_0 in the classical regression case.

Then a $(1-\alpha)100\%$ confidence interval for a calibrated X_0 is :

$$\hat{x}_0 \pm t_{n-2,\alpha/2}\,\tilde{\sigma}_e\left[1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{S_{yy}}\right]^{\frac{1}{2}} \qquad (2.43)$$

where $$\tilde{\sigma}_e^2 = \frac{SSE_x}{n-2}$$

$$SSE_x = (\hat{X} - X)^T(\hat{X} - X).$$

and $$\hat{x}_0 = \hat{\gamma}_0 + \hat{\gamma}_1 y_0 \ .$$

2.6.3 THE CLASSICAL INTERVAL ESTIMATE (II).

Graybill (1976,p283) suggests the following interval estimate for use with classical regression.

$$\bar{x} + \frac{\hat{\beta}_1(y_0-\bar{y})}{a} \pm \frac{\hat{\sigma}_e}{a} t_{n-2,\alpha/2}\left[a\left(1 + \frac{1}{n}\right) + \frac{(y_0 - \bar{y})^2}{S_{xx}}\right]^{\frac{1}{2}} \qquad (2.44)$$

where

$$a = \hat{\beta}_1^2 - \frac{\hat{\sigma}_e^2\, t_{n-2,\alpha/2}}{S_{xx}} \ .$$

2.6.4 THE CSS INTERVAL ESTIMATE.

Carrol, Sacks, and Spiegelman (1988) provided an alternative to Scheffé type confidence intervals. Scheffé (1973) suggested replacing the $J_0(x)$ of equation (2.42) with prediction bands of the form

$$J_1(x) = \{y \mid \hat{y}-\hat{\sigma}_e[c_1+c_2S(x)] \leq y \leq \hat{y}-\hat{\sigma}_e[c_1+c_2S(x)]\} \qquad (2.45)$$

where S(x) is the expression inside the square brackets of equation (2.42) and c_1 and c_2 are constants chosen such that the interval has $(1-\alpha)100\%$ confidence. This gives a confidence interval of the form $K_1(y) = \{x \mid y \in J_1(x_0)\}$.

The Carrol, Sacks, and Spiegelman (CSS) method is designed to take cognizance of two sources of error - one from the calibration experiment itself, and the other from the post-calibration measurement. Data from the calibration experiment is used to estimate the model (i.e. β_0, β_1, and σ_e^2). One measure of uncertainty is expressed as the probability of the calibration experiment producing a "good" estimate of β_0, β_1, and σ_e^2. Sheffe (1973) defined a "good" set by

$$G_s = \{y \mid |\hat{y}-y| \leq c_2\hat{\sigma}_e S(x)\} \qquad \forall\, x \in I \text{ and } \frac{\hat{\sigma}_e}{\sigma_e} \geq b \ .$$

where I denotes the calibration region or set of x's of interest and b and c_2 are chosen such that the probability that the experiment results in a good outcome is $1-\delta$ i.e.

$$P[Y \in G_s] \geq 1 - \delta \ .$$

The other uncertainty is the probability of "capturing" x_0 when y_0 is observed. Scheffé requires that *given a good outcome of the calibration experiment* , the probability of the interval containing x_0 is at least $1 - \alpha$. As noted by CSS, the drawback with Scheffe's method is that his own set of tables must be consulted in order to determine c_1 and c_2 .

The quicker alternative proposed by CSS is to compute c_1 and c_2 as follows :

$$c_1 = t_{n-2,1-\alpha/2} \qquad \text{and} \qquad c_2 = [2f_{2,n-2,1-\delta}]^{\frac{1}{2}}$$

where t and f are the indicated percentiles of the T and F distributions respectively.

The difficulty with both the Scheffé method and the CSS modification is that the confidence level α is <u>conditional</u> and is therefore not directly comparable with the other $(1-\alpha)100\%$ confidence intervals discussed. A small Monte-Carlo simulation was conducted to investigate the *unconditional* level of confidence for various choices of α and δ (program listing may be found in Appendix B). Using the following parameter values , the results shown in table 3 were obtained

$$\beta_0=0.0 \;\; ; \;\; \beta_1=1.0 \;\; ; \;\; \sigma_e^2=0.02$$
$$n=100 \;\; ; \;\; N=2000.$$

where n is the number of observations generated each time for the estimation of the model parameters and N is the length of the simulation run.

2.6.5 <u>THE ORTHOGONAL INTERVAL ESTIMATE</u>.

The last estimator to be compared is the orthogonal estimator given by equation (2.35). This estimator was shown to have superior bias and

MSE characteristics. It, as well as the others, will now be compared from the perspective of interval estimation. The criteria used will be the actual level of significance attained versus the nominal level of significance employed and the average length of the intervals generated by each procedure.

Table 3. Unconditional confidence levels for the CSS procedure for various combinations of α and δ obtained from 2000 simulations.

	α			
δ	0.01	0.025	0.05	0.10
0.01	0.0045	0.0080	0.02050	0.0525
0.025	0.0020	0.0125	0.0210	0.0405
0.05	0.0050	0.0065	0.0220	0.0475
0.10	0.0040	0.0120	0.0255	0.0540

Naturally, the most desirable situation is that in which the probability of coverage is at least as large as the nominal level of significance and the average interval width is small. Details of another Monte-Carlo simulation experiment used to examine these aspects of

interval estimation follow. It should be pointed out that this is not intended to be an exhaustive investigation, but rather the results should be viewed as an indicator of possible differences between the various methods considered.

2.6.6 A MONTE CARLO SIMULATION OF INTERVAL ESTIMATION PROCEDURES.

A Gauss program was written (see Appendix C for listing) to repeatedly construct interval estimates using each of the five procedures discussed above and to then determine if the true x_0 value being calibrated was covered by the interval. One hundred observations on both X and Y were generated - those for X being uniformly distributed over the interval [0,1] and the corresponding Y values obtained from the regression model $Y=\beta_0+\beta_1 X+e$ with the errors generated from $N(0,.02^2)$. In this experiment, the values of β_0 and β_1 were fixed at 0.0 and 1.0 respectively. The nominal level of significance used was α = 0.0525 since we know from table 3 that this is attained for the CSS procedure using δ = 0.01 and α = 0.10. Thus for the CSS method we used $F_{2,98,0.01}$ = 4.828 and $T_{98,0.05}$ = 1.66051 . For all remaining procedures the critical t-value employed was $T_{98,0.02625}$ = 1.96278. The results of 5000 simulations are displayed in table 4. Inspection of table 4 indicates that the estimators differ very little in terms of both probability of coverage and average interval length. Theortically, the probability of coverage should be 0.9475. As can be seen, each of the five procedures averaged slightly higher than this with the CSS method having the highest probability of coverage (although this procedure also had the largest average interval width).

Table 4. Comparison of 5 methods of interval estimation in the calibration problem.Comparisons based on probability of coverage and average interval length from 5000 simulations,each consisting of 100 X,Y data pairs.

	METHOD				
	I	II	III	IV	V
Prob. of coverage.	0.95520	0.95620	0.95540	0.96300	0.95460
Av. length of interval	0.079138	0.078945	0.079146	0.083568	0.078855

METHOD I : Classical (I) METHOD II : Inverse
METHOD III : Classical (II) METHOD IV : CSS procedure
METHOD V : Orthogonal

The shortest average interval width was obtained with the orthogonal estimator, however the the inverse and classical procedures had average widths which were within 1% of this. Overall, the conclusion is that in terms of interval estimation there is very little to distinguish between the five methods. However based on practical considerations some recommendations are possible. As already noted, the difficulty with the CSS method is that the researcher has no way of knowing what values to use for δ and α to achieve an overall prescribed confidence level. Indeed, many combinations of δ and α may result in the same overall confidence level. Furthermore, this study has revealed that this estimation procedure (for this set of prescribed conditions) resulted in an average interval width which is approximately 6% larger than can be obtained by other methods. For these reasons the use of the CSS method is not recommended. The classical, inverse and orthogonal procedures are remarkably similar. However in view of earlier findings regarding the bias and MSE attributes of the orthogonal procedure for point estimation, this method is rated more highly than any other method considered in this dissertation.

The results of these and previous analyses are now applied to a practical problem that was considered by Fox (1987) and which became the subject of a court case in which the author was called upon to testify as an expert witness.

2.7 AN EXAMPLE : CALIBRATION AND THE LAW.

In Australia and other parts of the world, the police sometimes use aircraft to detect and prosecute speeding drivers. The standard procedure is for the airborne observer to time the vehicle below as it crosses two lines painted on the road surface. These lines are usually separated by either 500m or 1000m. It is then a simple matter to convert this recorded time into an average speed for the measured distance. The problem of assessing the accuracy of the procedure and/or varying the method of calculation falls under the umbrella of statistical calibration. To this end an experiment was conducted by the Mathematics and Statistics Department of Curtin University (formally Western Australian Institute of Technology) in conjunction with the Western Australian Police Department in late 1986 to help address the accuracy question. Results from this experiment were to be subsequently used in a court case in which the validity of the airborne procedure was brought into question.

The calibration experiment involved a simulation of actual conditions in which a driver traversed the lines at a variety of speeds while two independent observers made measurements from an overhead aircraft. Vehicle speeds over the range 80 km/hr to 140 km/hr were used with three replications made at each speed. Police vehicles whose speedometers had been previously checked for accuracy with radar devices were used in the experiment.

Given two marked lines separated by some distance d meters we can compute a vehicle's average speed as

$$X = \frac{3.6d}{Y} \qquad (2.46)$$

where X is the recorded speed (km/hr) and Y is the measured time (seconds).

This particular problem falls into the category of calibration for the measurement error model discussed in section 2.4 . A simple check on the accuracy of the airborne procedure is provided by taking time measurements on a vehicle whose speed can otherwise be accurately determined and comparing the actual speed with that calculated using (2.46). In setting up such an experiment the problem becomes one of determining to what extent discrepancies between the calculated speed from equation (2.46) and the assumed speed are attributable to the two sources :

(i) error in time readings due to

(a) observer's ability to accurately judge the crossing of the two lines.

(b) observer's reaction time.

(ii) driver error (a driver told to travel at 80 km/hr say, will have trouble in maintaining a speedometer reading of *exactly* 80 km/hr).

Thus, whilst the *true* time (y) is precisely determined by the *true* or actual speed (x), in practice the *measured* time Y is recorded against the *assumed* speed (X). Note, that the true vehicle speed x is never known-the driver is simply instructed to travel at some nominal speed X.

In terms of the measurement error model we have :

$$X_i = x_i + u_i \qquad \text{and} \qquad Y_i = y_i + e_i \tag{2.47}$$

where

X_i is some nominal (or assumed) speed .

x_i is the actual speed.

u_i is a random error reflecting the driver's inability to maintain constant speed X_i .

Y_i is the measured time made by the airborne observer.

y_i is the true time that would be required to travel distance d when travelling at speed x_i .

e_i is a measurement error associated with recording times.

Furthermore, the assumed model is :

$$Y_i = \beta_0 + \beta_1 x_i^* + e_i$$

where

$$\beta_0 + \beta_1 x_i^* = y_i$$

and

$$x_i^* = \frac{c}{x_i} \qquad \text{and} \quad c = 3.6d. \tag{2.48}$$

Validation of the airborne procedure would imply $\beta_0 = 0.0$ and $\beta_1 = 1.0$.

Another important aspect of this study is determine the relative magnitudes of σ_e^2 and σ_u^2 . Ideally, one would like to be able to establish that $\sigma_u^2 \gg \sigma_e^2$,for this situation would ascribe most of the total error in the experiment to the inability of a driver to maintain a

prescribed speed under test conditions, thereby giving credence to the accuracy of the airborne observation method. Whilst the comparison of σ_u^2 and σ_e^2 has intrinsic interest of its own it is pointed out that Mandel's procedure for the measurement error model tacitly assumes knowledge of not only the ratio σ_e^2/σ_u^2 but also of ρ (the correlation between σ_e^2 and σ_u^2) . By virtue of the replication performed in this particular calibration experiment, we are able to estimate these quantities. We digress briefly to examine how this may be achieved.

2.7.1 ESTIMATION OF THE COVARIANCE STRUCTURE OF THE COMPONENTS OF ERROR.

Given certain distributional assumptions for the errors u_i and e_i , together with other simplifying approximations, a procedure is developed for estimating the components of variability σ_e^2 , σ_u^2 , and the covariance of these quantities. We commence by writing

$$\begin{aligned} Y_i &= f(X_i) + e_i \\ &= f(x_i + u_i) + e_i \end{aligned} \tag{2.49}$$

where $f(\cdot)$ is ,in this case, the reciprocal relationship between a measured time and measured speed. Using a first-order Taylor approximation, we may write equation (2.49) as

$$Y_i = f(x_i) + u_i f'(x_i) + e_i \tag{2.50}$$

In the following development we will assume that u_i and e_i have the bivariate normal distribution :

$$h(e_i,u_i) = \frac{1}{2\pi\sigma_e\sigma_u(1-\rho^2)^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}\left(\frac{1}{1-\rho^2}\right)\left[\frac{e_i^2}{\sigma_e^2} - 2\,\frac{e_iu_i}{\sigma_e\sigma_u} + \frac{u_i}{\sigma_u^2}\right]\right] \qquad (2.51)$$

Thus :

$$\mathbb{E}[Y_i] = f(x_i)$$

and

$$\sigma_i^2 = \mathrm{Var}[Y_i] = \sigma_e^2 + \sigma_u^2[f'(x_i)]^2 + 2f'(x_i)\mathrm{Cov}[e_i,u_i]$$

$$= \sigma_e^2 + \sigma_u^2[f'(x_i)]^2 + 2f'(x_i)\rho\sigma_e\sigma_u \qquad (2.52)$$

We shall call σ_i^2 the "effective variance" at point x_i . An unbiased estimator of σ_i^2 is provided by S_i^2 where

$$S_i^2 = \sum_{j=1}^{n} \frac{(E_{ij} - \bar{E}_i)}{n-1}$$

where $\bar{E}_i = \frac{1}{n}\sum_{i=1}^{n} E_{ij}$ and E_{ij} is the j^{th} difference in a measured time and a calculated time for the assumed speed x_i . The problem is to estimate σ_e^2 , σ_u^2 , and ρ given values of S_i^2 at various values of x_i. Now, S_i^2 has distribution given by the following p.d.f.

$$g(s_i^2) = \frac{m^m}{\Gamma(m)} (s_i^2)^{m-1} \left[\frac{1}{\sigma_i^2}\right]^m e^{-ms_i^2/\sigma_i^2} \qquad (2.53)$$

where $m = \frac{(n-1)}{2}$.

We assume that at each x_i , i = 1, ... ,t there are n replications. The likelihood function is therefore given as

$$L(\sigma_i^2; s_i^2) = K^t \left[\exp\left(-m \sum_{i=1}^{t} \frac{s_i^2}{\sigma_i^2}\right) \right] \prod_{i=1}^{t} (s_i^2)^{m-1} (\sigma_i^2)^{-m} \qquad (2.54)$$

where $K = \frac{m^m}{\Gamma(m)}$ and the log-likelihood function

$$\ln L = t \ln(k) - m \sum_{i=1}^{t} \frac{s_i^2}{\sigma_i^2} - m \sum_{i=1}^{t} \ln \sigma_i^2 + (m-1) \sum_{i=1}^{t} \ln(s_i^2) \qquad (2.55)$$

Given sample data $\{s_i^2\}$, we wish to determine the unknown parameters σ_e^2 , σ_u^2 , and ρ that maximize equation (2.55). Specifically

$$\max \; F(\sigma_e^2; \sigma_u^2; \rho) = \sum_{i=1}^{t} (\ln w_i - s_i^2 w_i) \qquad (2.56)$$

where

$$w_i = \frac{1}{\sigma_i^2}$$

and

$$\sigma_i^2 = \sigma_e^2 + \sigma_u^2 [f'(x_i)]^2 + 2f'(x_i) \rho \sigma_e \sigma_u$$

subject to

$$\sigma_e^2 \geq 0$$

$$\sigma_u^2 \geq 0$$

$$-1 \leq \rho \leq 1 \quad .$$

Rather than maximizing equation (2.56) we choose to minimize

$$\phi(\sigma_e^2;\sigma_u^2;\rho) = \sum_{i=1}^{t} (s_i^2 w_i - \ln w_i) \qquad (2.57)$$

To this end we use a modified Newton algorithim as described below.

2.7.2 PARAMETER ESTIMATION.

Let $\underline{\Phi}$ denote a vector of parameters to be estimated. In this case

$$\underline{\Phi}^T = [\sigma_e^2, \sigma_u^2, \rho]$$

We wish to find values $\underline{\Phi}^*$ of $\underline{\Phi}$ for which $\phi(\underline{\Phi})$ is minimized. Starting with a given point $\underline{\theta}_1$,a sequence of points $\underline{\theta}_2, \underline{\theta}_3$, ...is generated which hopefully converges to the point $\underline{\Phi}^*$. Let $\underline{H}(\underline{\theta})$ be the Hessian matrix of the function $\phi(\underline{\theta})$ and $\underline{g}(\underline{\theta})$ the gradient vector of $\phi(\underline{\theta})$. Then the i^{th} iteration of the Newton method is

$$\underline{\theta}_{i+1} = \underline{\theta}_i - \underline{H}_i^{-1} \underline{g}_i \qquad (2.58)$$

Thus, for the present case we have :

$$\underline{g}(\underline{\theta}) = \begin{bmatrix} \frac{\partial \phi}{\partial \sigma_e^2} \\ \frac{\partial \phi}{\partial \sigma_u^2} \\ \frac{\partial \phi}{\partial \rho} \end{bmatrix} \tag{2.59}$$

and

$$\mathbf{H}(\underline{\theta}) = \begin{bmatrix} \frac{\partial^2 \phi}{\partial \sigma_e^4} & \frac{\partial^2 \phi}{\partial \sigma_u^2 \partial \sigma_e^2} & \frac{\partial^2 \phi}{\partial \rho \partial \sigma_e^2} \\ \frac{\partial^2 \phi}{\partial \sigma_u^2 \partial \sigma_e^2} & \frac{\partial^2 \phi}{\partial \sigma_u^4} & \frac{\partial^2 \phi}{\partial \sigma_u^2 \partial \rho} \\ \frac{\partial^2 \phi}{\partial \rho \partial \sigma_e^2} & \frac{\partial^2 \phi}{\partial \rho \partial \sigma_u^2} & \frac{\partial^2 \phi}{\partial \rho^2} \end{bmatrix} \tag{2.60}$$

Elements of the Hessian matrix are given in Appendix D.

2.7.3 MODEL VALIDATION.

Prior to implementation with actual data, we first investigate the validity of the foregoing methods by applying them to simulated data sets for which the population parameters are known in advance. To this end 30 observations on X and Y were generated from a bivariate normal distribution with parameters $\sigma_u^2 = 4.5$; $\sigma_e^2 = 0.85$; and $\rho = 0.8$ at each x_i. Values of S_i^2 were then obtained and the resulting {X ,S } data used as input to the Newton algorithm (see Appendix E for listing) to obtain

the estimates $\hat{\sigma}_e^2$, $\hat{\sigma}_u^2$, and $\hat{\rho}$. Results of this procedure are displayed in table 5. As can be seen from table 5, in most cases the estimates are in close agreement with the true parameter values although some evidence of bias is noted. All three parameters have been overestimated by about half a standard error. This is not totally unexpected since the expression for σ_i^2 [equation (2.52)] is only a first-order approximation. Presumably the degree of bias will be very much dependent on the exact nature of f(x) and how well it is represented by a first-order approximation. It is difficult to draw any specific conclusions on this matter from the above results. Further experimentation would be required to assess the effects of different values of n, σ_e^2 ,σ_u^2 , and ρ . Nevertheless, these results indicate that the methodology is sound and give us no reason to modify the procedure. The method is now applied to the actual experimental data taken by an airborne observer.

2.7.4 APPLICATION TO AIR SURVEILLANCE METHOD.

Data obtained from the experiment described earlier in this section is presented in table 6. Using equation (2.48) we may compute a time for each of the assumed speeds and compare these with those given in table 6. These differences or errors are shown in table 7. It is obvious from inspection of the data in table 7 that there is evidence of bias in the time readings. Seventy-three out of the 78 observations were below the median value of 0.2155 seconds. A simple sign test on such a result yields a p-value which is zero to four decimal places. It is not possible to say whether this represents a true underestimation on the

Table 5. Results from model validation. Tabulated entries are sample variances obtained from 30 observations at each x_i for each of ten replications. Model parameters used were : $\sigma_e^2 = 0.85; \sigma_u^2 = 4.5; \rho = 0.8$.

x_i	1	2	3	4	5	6	7	8	9	10
80	0.5461	0.5929	0.6037	0.8226	0.4706	0.5055	0.6529	0.8761	0.6989	1.0000
85	0.4858	0.6304	0.5041	0.4007	0.3215	0.4147	0.5715	0.3697	0.5373	0.4928
90	0.3493	0.2560	0.2490	0.3329	0.5716	0.4900	0.3387	0.4530	0.3341	0.3091
95	0.3697	0.4083	0.2323	0.4083	0.2683	0.2694	0.3047	0.2070	0.2683	0.4045
100	0.2830	0.2266	0.2172	0.2440	0.2852	0.2480	0.2948	0.4020	0.3709	0.3493
105	0.2421	0.1989	0.3493	0.2756	0.2323	0.3564	0.3564	0.2256	0.2381	0.2052
110	0.2381	0.3238	0.3364	0.1978	0.1529	0.2352	0.3260	0.1936	0.1980	0.1823
115	0.1892	0.2025	0.4135	0.3434	0.2200	0.1197	0.3215	0.2905	0.2905	0.1624
120	0.1849	0.2381	0.1706	0.2809	0.2052	0.2228	0.2070	0.1260	0.2381	0.1568
125	0.3552	0.3318	0.2275	0.1747	0.3648	0.1616	0.3226	0.3376	0.2611	0.3215
130	0.3612	0.2735	0.3058	0.2500	0.2652	0.3894	0.4007	0.1584	0.3376	0.3306
135	0.4251	0.4775	0.2981	0.3069	0.1900	0.3919	0.2862	0.3215	0.3758	0.1498
140	0.2798	0.2852	0.2798	0.2884	0.4409	0.2632	0.4147	0.3283	0.4529	0.2981
$\hat{\sigma}_v$	0.9248	0.9772	0.7460	0.8123	0.8543	0.8230	0.8946	0.8763	1.0229	0.9224
$\hat{\sigma}_u$	5.0064	5.3932	3.7315	4.6493	4.6475	4.5126	4.6761	5.1612	5.6528	5.7595
ρ	0.8423	0.8624	0.7190	0.7934	0.8232	0.8020	0.7963	0.8384	0.8743	0.8707

Mean	S.D.
0.8854	0.0817
4.9190	0.6071
0.8222	0.0470

Table 6. Data obtained from air surveillance experiment. Tabulated entries are measured times (secs) made by an observer in overhead aircraft for vehicle below to travel 500m.

COLUMNS: Speed (km/hr)

80	85	90	95	100	105	110
21.960	21.020	19.520	18.610	17.780	16.930	16.1
21.960	21.040	19.710	18.590	17.710	16.820	16.0
22.210	20.810	19.750	18.660	17.740	17.000	15.9
22.040	20.980	19.460	18.640	17.800	16.870	16.1
21.950	20.980	19.620	18.540	17.670	16.840	16.0
22.160	20.890	19.750	18.690	17.650	16.960	16.0

115	120	125	130	135	140
15.430	14.710	14.000	13.600	13.170	12.590
15.200	14.680	14.150	13.570	13.120	12.560
15.360	14.680	14.350	13.450	13.020	12.630
15.410	14.640	14.010	13.570	13.130	12.520
15.280	14.640	14.070	13.570	13.120	12.590
15.430	14.690	14.230	13.530	13.120	12.670

CELL CONTENTS --
M-time:DATA

Table 7. Differences in measured time from table 6 and time computed using equation (2.48) at each speed x_i.

COLUMNS: Speed (km/hr)

80	85	90	95	100	105	110
-0.54000	0.08977	-0.48000	-0.33737	-0.04178	-0.05113	-0.18364
-0.54000	0.10977	-0.29000	-0.35737	-0.11178	-0.16113	-0.34364
-0.29000	-0.36647	-0.03022	-0.28737	-0.26000	0.01887	-0.22622
-0.46000	0.04977	-0.54000	-0.30737	-0.02178	-0.11113	-0.24364
-0.55000	0.04977	-0.38000	-0.40737	-0.15178	-0.14113	-0.32364
-0.34000	-0.28647	-0.03022	-0.06000	-0.35000	-0.18286	-0.13622

115	120	125	130	135	140
-0.22217	-0.16603	-0.28571	-0.24615	-0.16333	-0.17596
-0.45217	-0.19603	-0.13571	-0.27615	-0.21333	-0.20596
-0.15724	-0.19603	-0.05000	-0.39615	-0.21529	-0.13596
-0.24217	-0.23603	-0.27571	-0.27615	-0.20333	-0.24596
-0.37217	-0.23603	-0.21571	-0.27615	-0.21333	-0.17596
-0.08724	-0.18603	-0.17000	-0.31615	-0.11529	-0.09596

CELL CONTENTS --
Error:DATA

part of the airborne observer or is a consequence of the driver always travelling slightly slower than required. The normality assumption for the error distribution is well supported by the histogram and boxplot of figures 8(a) and 8(b). The Q-Q plot of figure 8(c) is strongly linear with a correlation coefficient of 0.995 which is not small enough to warrant rejection of the hypothesis of normality ($p > 0.95$).

A summary of the sample variances computed at each (assumed) speed is provided in table 8. Application of the iterative Newton-Raphson procedure converged at the point $\hat{\sigma}_e^2 = 0.0072$; $\hat{\sigma}_u^2 = 1.3250$; $\hat{\rho} = 0.8981$. These results suggest that the variability (as measured by one standard deviation) in maintaining constant vehicle speed is about 1 km/hr , the variability associated with making time measurements is of the order of 1/10 th. of a second and that the two errors are strongly and positively corrrelated. These results are not surprising and agree with what one would intuitively expect. With these estimates we are in a position to calibrate using the various procedures discussed in this chapter.

For the purposes of calibration we let Y_i be a measured time for assumed speed X_i and X_i^* is a computed time using equation (2.48). The Y_i are regressed on X_i^* for the classical method, the orthogonal method, and Mandel's procedure. For the inverse method, and Fuller's method the X_i^* are regressed on the Y_i. The results are presented in table 9. The parameter estimates for each of the five calibration methods considered are very similar which is accounted for by the regression line having a slope very nearly equal to unity and and intercept which is close to the origin.

Table 8. Summary of data obtained from the air surveillance experiment. Tabulated entries are sample error variances of timing measurements at each speed x_i.

Assumed Speed X_i	Sample variance
85	0.0441126
90	0.0483296
95	0.0216649
100	0.0162971
105	0.0057608
110	0.0063680
115	0.0183088
120	0.0007868
125	0.0080210
130	0.0028164
135	0.0016314
140	0.0027468

Midpoint	Count	
-0.55	4	****
-0.50	1	*
-0.45	2	**
-0.40	3	***
-0.35	7	*******
-0.30	12	************
-0.25	8	********
-0.20	14	**************
-0.15	10	**********
-0.10	5	*****
-0.05	6	******
-0.00	2	**
0.05	2	**
0.10	2	**

Figure 8(a). Histogram of timing errors for the air surveillance method.

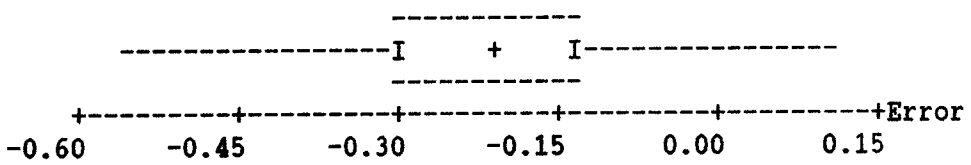

Figure 8(b). Boxplot of timing errors.

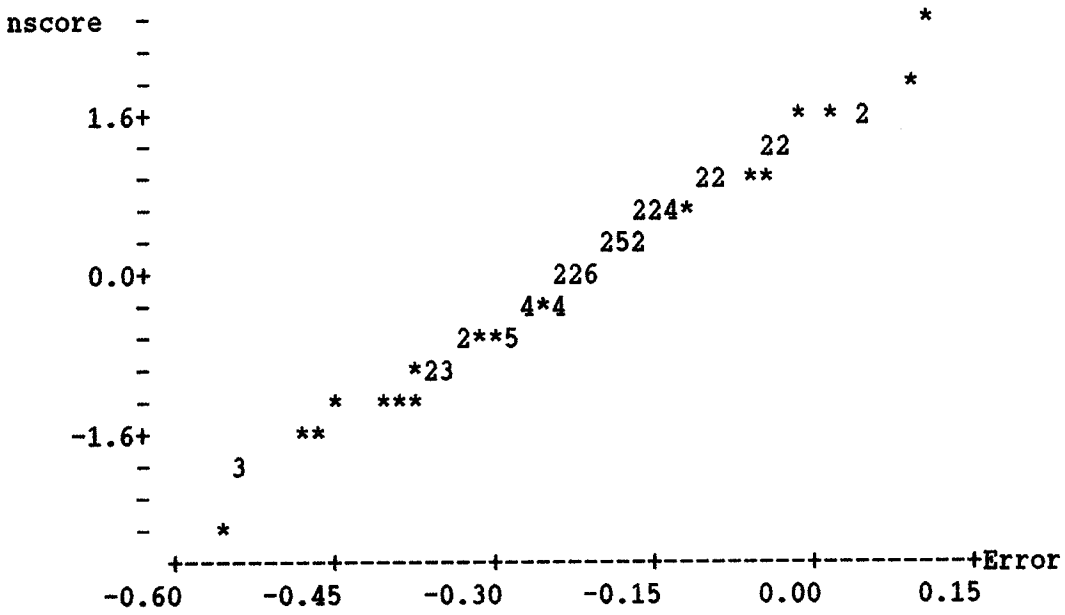

Figure 8(c). Q-Q plot of timing errors.

Table 9. Estimated parameters for five different calibration methods using the air surveillance data of table 6.

Method	slope	intercept
Classical	0.9901	-0.0572
Inverse	1.0077	0.0964
Orthogonal	0.9912	-0.0761
Mandel	0.9926	-0.0992
Fuller	1.0081	0.0891

Using each of the estimated models in table 9 we can obtain calibrated measurements for vehicle speed. This has been done for a range of *measured speeds* between 80 km/hr and 160 km/hr (approximately 50 m.p.h. to 100 m.p.h.) and results given in table 10 below.

Table 10. Calibrated speeds using each of the models from table 9.

	Calibrated Speed				
Measured Speed(km/hr)	Class.	Inverse	Orthogl	Mandel	Fuller
80	79.007	79.053	79.029	79.059	79.047
85	83.932	83.971	83.950	83.978	83.967
90	88.855	88.887	88.870	88.893	88.884
95	93.776	93.800	93.787	93.806	93.799
100	98.696	98.711	98.703	98.716	98.712
105	103.615	103.619	103.616	103.623	103.622
110	108.532	108.525	108.527	108.528	108.530
115	113.447	113.428	113.436	113.430	113.435
120	118.361	118.328	118.344	118.329	118.339
125	123.273	123.226	123.249	123.226	123.239
130	128.183	128.121	128.152	128.120	128.138
135	133.093	133.014	133.053	133.011	133.033
140	138.000	137.904	137.951	137.900	137.927
145	142.906	142.792	142.848	142.786	142.818
150	147.810	147.677	147.743	147.669	147.707
155	152.713	152.559	152.636	152.550	152.593
160	157.615	157.439	157.526	157.428	157.477

The results in table 10 show little variation among the calibration methods. This is primarily due to the very high degree of agreement between the measured times and the computed times coupled with a gradient of almost 1.0 and an intercept of approximately zero.

The entries in table 10 should be interpreted as follows : suppose a reading of 22.50 seconds was made for a 500m distance. Ordinarily, we would compute the average speed as 80 km/hr and this becomes our *measured speed*. However reference to table 10 suggests that the *true speed* is more like 79 km/hr (recall that the timing measurements consistently underestimated the true value by an average of about 1/5 th. of a second and thus overestimated the actual speed). Of course the differences here a quite small and for practical purposes inconsequential, although it is perhaps disconcerting to the driver that the bias is not working in his or her favour.