BNWL-SA-4389

CONF-720840--1

The Robustness of Tank
Calibration Slope Estimates

MASTER

Kirkland B. Stewart

Battelle Memorial Institute
Pacific Northwest Laboratories
Richland, Washington

Prepared for submission to IAEA on "Accuracy
of Nuclear Materials Accountancy and Technical
Effectiveness of Safeguards" at Vienna, Austria,
August 28 - September 1, 1972.

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

RECEIVED BY TIC SEP 22 1972

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Table of Contents

1. Introduction
2. Objectives
3. Conclusions
4. Models 1, 2 and 3
5. The Effect of the Incorrect Model
6. Results and Conclusions in Detail

1.0 Introduction

If the horizontal cross-sectional areas of a tank are the same the volume of the tank can be calibrated by the linear relationship $y(\text{volume}) = \alpha + \beta x$ (manometer height). If the cross-sectional areas are not sufficiently constant due to such things as bulging and the inclusion of equipment, higher degree polynomials may be useful in describing the volume-manometer relationship.

In the calibration of tanks it is very important to determine the correct error model for the observations, both in fitting the data and in predicting the variance of a transfer. The part of the statistical model which characterizes the expected value of y as a function of x is due to the inherent relationship between the volume and the manometer values. The variances of the estimated parameters in the model and the variances of the estimated volume transfers are functions of the design matrix of the x values in the calibration procedure and of the structure of the errors in calibration and in volume estimation. This structure of errors is important as it determines the efficiency with which the parameters and the estimable functions of the parameters are estimated as well as how the different variance estimates are determined.

In the field of safeguards it is very important to assess the error situation correctly. An underestimate of the variance of a parameter leads to an excess of type I errors, an overestimate of the variances leads to a diminished discriminatory power, i.e. an increase in type II errors. A type I error occurs, for example, when one asserts, that a MUF indicates a loss of

material when in fact the loss is due to measurement errors of the type which occur from random measurement variation. A type II error occurs when a loss of material is not indicated by the statistical significance of the MUF.

When using the model $y = \hat{\alpha} + \hat{\beta}x$, a transfer from manometer readings x_1 to x_2 leads to the volume transfer estimate :

$$y_2 - y_1 = \hat{\alpha} + \hat{\beta} x_2 - \hat{\alpha} - \hat{\beta} x_1 = \hat{\beta} (x_2 - x_1)$$

and in transfers the intercept is subtracted out. If $\hat{\beta}$, as an estimate of β , is off by $\epsilon\%$, each transfer will be off by $\epsilon\%$. Though the error in $\hat{\beta}$ may be due to random errors, its effect in practice will be to induce a constant relative bias in transfer estimates.

If in the calibration process the amounts $\Delta y_1, \Delta y_2, \dots, \Delta y_n$ are successively added into the process vessel so that the observed values are $y_1 = \Delta y_1, y_2 = \Delta y_1 + \Delta y_2, \dots, y_n = \Delta y_1 + \Delta y_2 + \dots + \Delta y_n$, are obtained, where the measurement errors originate in the incremental volumes and not in the x (manometer) values, one has the so-called cumulative error model.

If, on the other hand, the values $y_1, y_2, \dots, y_n$ are obtained by successively emptying the tank and inserting the nominal amounts $y_1, y_2, \dots, y_n$ one has the independent errors model. For both models one can also consider two additional cases: 1) the variance of the measurement error for the dependent variable is proportional to the amount added, and 2) the variance of the measurement error is independent of the amount added. The cases considered herein are model 1--cumulative, measurement variance proportional to the amount added, model 2--cumulative, measurement variance constant, independent of the amount added, and model 3--independent measurements, measurement variance constant and independent of the amount.

There is another model which sometimes occurs. It arises in the situation where the y_i volume values, are measured precisely in comparison to the x_i manometer values. Then one fits the model

$$x_i = \alpha' + \beta' y_i + e_i'$$

and solves for the relationship

$$y = (-\alpha'/\beta') + (1/\beta')x$$

$$= \hat{\alpha} + \hat{\beta}x.$$

This model is so inherently different from the others that it is difficult to include in a study of the effects of the wrong error model. It has a somewhat stormy recent history due to an article by Krutchkoff [6].

2.0 Objectives

The objective of this paper is to study the robustness of the different error models by

- a. investigating the efficiency of the β estimate and the bias in the estimate of $\sigma_{\hat{\beta}}$ when the wrong error model is assumed, and
- b. seeing what effects the data size and the x spacings have on the two indicators of lack of robustness.

3.0 Conclusions

The conclusions are given in detail in section 6. The main points are as follows.

- a. The greatest losses of efficiency and largest errors in incorrectly assessing the β estimate variance situation comes when the cumulative and independent error situations are confused. The poorest efficiencies occur when model 3 is assumed and either model

1 or 2 is correct. The biggest overstatements of the $\hat{\beta}$ variance estimates come when 3 is true but either models 1 or 2 is assumed. The variance is understated the most when model 3 is assumed but either model 1 or model 2 is correct.

- b. If the cumulative case is correct and the volume increments are quite similar, the confusion as to whether model 1 or model 2 is the correct error model does not effect the efficiency of β estimation nor the bias in the estimation of the standard deviation of the β estimate very much.
- c. The evenly-spaced case is studied to see what the effect of the number of calibration points is on the choice of the incorrect error model. The result is that the deleterious effects of confusing the cumulative and independent error models are increased as n increases.
- d. Three alternatives to the evenly-spaced case are explored. These are 1) when the big increments in x occur at either the beginning or at the end of the manometer scale, 2) when the big increments occur at the center of the scale and 3) when the increments are oscillatory such that one small increment is followed by a large one. The results in each case are described by a parameter. The conclusions are best described from the graphs which are given. The main conclusion is that the graphs do not give much assurance as to the robustness when the assumed model is at variance with the correct error model.

4.0 Models 1, 2, and 3

As indicated before the correct underlying relationship between the volume and the manometer readings $y = f(x)$ depends upon the physical situation. If the tank has constant cross-sectional areas at the different heights the model

$$y = \alpha + \beta x$$

is appropriate. In calibrating the tank this model is used but the best estimate in the sense of being minimum variance unbiased estimates (mvue) depends on the error structure. Thus to describe the observed (y_i, x_i) values an error term is added to 3.1 so that

$$y_i = \alpha + \beta x_i + e_i.$$

The mvue of α and β depend upon the nature of the error term e_i .

Model 1

Model 1 arises when the tank is filled in successive increments so that

$$y_i = \Delta y_1 + \Delta y_2 + \dots + \Delta y_i, \quad i = 1, 2, \dots, n.$$

The error in adding the amount Δy_i is ϵ_i , so that $\sigma_{y_i}^2 = \sum_{j=1}^i \sigma_{\epsilon_j}^2$ where $\sigma_{\epsilon_i}^2 = (x_i - x_{i-1})\sigma^2$ and σ^2 is the variance associated with a unit increase in x . An element of the variance-covariance matrix of the y_i values is $\sigma_{ij} = \sigma^2 \min x_i, x_j$.

It can be shown [1] [2] and [3], that the mvue of α and β are

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} (y_1 x_n - x_1 y_n) / (x_n - x_1) \\ (y_n - y_1) / (x_n - x_1) \end{pmatrix} = \begin{pmatrix} y_1 - x_1 \hat{\beta} \\ \hat{\beta} \end{pmatrix}$$

with covariance matrix

$$\Sigma_{\hat{\alpha}\hat{\beta}} = \sigma^2 \begin{pmatrix} x_1 x_n / (x_n - x_1) & -x_1 / (x_n - x_1) \\ -x_1 / (x_n - x_1) & 1 / (x_n - x_1) \end{pmatrix}$$

so that

$$\hat{\sigma}_{\hat{\alpha}}^2 = s^2 x_1 x_n / (x_n - x_1)$$

$$\hat{\sigma}_{\hat{\beta}}^2 = s^2 / (x_n - x_1)$$

$$\hat{\sigma}_{\hat{\alpha}\hat{\beta}} = -s^2 x_1 / (x_n - x_1)$$

are variance estimates where s^2 is obtained from

$$s^2 = \left(\sum_{i=2}^n (y_i - y_{i-1})^2 / (x_i - x_{i-1}) - \hat{\beta}^2 (x_n - x_1) \right) / (n-2)$$

or its algebraic equivalent

$$s^2 = \frac{\sum_{i=2}^n \left\{ y_i - y_{i-1} - \hat{\beta}(x_i - x_{i-1}) \right\}^2 / (x_i - x_{i-1})}{n-2}$$

Model 2

Model 2 is similar in its development to model 1 except that the variance of $\sigma_{\epsilon_i}^2 = \text{constant}$. This leads to the mvue's of

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} (y_1 \sum_1^n L_i^2 - x_1 \sum_1^n L_i z_i) / \Delta \\ (-y_1 x_1 + \sum_1^n L_i z_i) / \Delta \end{pmatrix} = \begin{pmatrix} y_1 - x_1 \hat{\beta} \\ \hat{\beta} \end{pmatrix}$$

where $z_1 = y_1$, $z_i = y_i - y_{i-1}$, $i > 1$; $L_1 = x_1$, $L_i = x_i - x_{i-1}$, $i > 1$

and
$$\Delta = \sum_1^n L_i^2 - x_1^2 = \sum_2^n L_i^2 = \sum_2^n (x_i - x_{i-1})^2.$$

Note that $\hat{\beta}$ can be simplified to $\hat{\beta} = \sum_2^n L_i z_i / \sum_2^n L_i^2$. The variance-covariance matrix of $\hat{\alpha}$ and $\hat{\beta}$ is

$$\Sigma_{\alpha, \beta} = \sigma^2 \begin{pmatrix} \sum_1^n L_i^2 / \sum_2^n L_i^2 & -x_1 / \Delta \\ -x_1 / \Delta & 1 / \Delta \end{pmatrix}$$

so that variance estimates are provided by

$$\hat{\sigma}_{\hat{\alpha}}^2 = s^2 \left(\sum_1^n L_i^2 / \sum_2^n L_i^2 \right) = (x_1^2 / \Delta + 1) s^2$$

$$\hat{\sigma}_{\hat{\beta}}^2 = s^2 / \Delta$$

$$\hat{\sigma}_{\hat{\alpha}\hat{\beta}} = s^2 (-x_1) / \Delta$$

where s^2 is obtained from

$$s^2 = \left(\sum_2^n z_i^2 - \hat{\beta} \sum_2^n z_i L_i \right) / (n-2)$$

or its algebraic equivalent

$$s^2 = \sum_2^n [y_i - y_{i-1} - \hat{\beta} (x_i - x_{i-1})]^2 / (n-2).$$

Model 3

Model 3 arises when the errors are in the volume values y_i , the x_i values are known exactly and e_i , the errors, are independent with zero expectation and constant variance. This is the linear regression situation commonly handled in statistical textbooks. Then the mvue's are

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \frac{\sum_1^n x_i^2 \sum_1^n y_i - \sum_1^n x_i \sum_1^n x_i y_i}{n S_{xx}} \\ S_{xy}/S_{xx} \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{\beta} \bar{x} \\ \hat{\beta} \end{pmatrix}$$

where $S_{xy} = \sum_1^n (x_i - \bar{x})(y_i - \bar{y})$, $S_{xx} = \sum_1^n (x_i - \bar{x})^2$

with

$$\hat{\sigma}_{\alpha}^2 = s^2 \sum_1^n x_i^2 / n$$

$$\hat{\sigma}_{\beta}^2 = s^2 / S_{xx}$$

$$\hat{\sigma}_{\alpha} \hat{\sigma}_{\beta} = -s^2 \bar{x} / S_{xx}$$

where

$$\begin{aligned} s^2 &= \left(\sum_1^n y_i^2 - \left(\sum_1^n y_i \right)^2 / n - \hat{\beta}^2 S_{xx} \right) / (n-2) \\ &= \left(\sum_1^n (y_i - \bar{y})^2 - \hat{\beta} S_{xy} \right) / (n-2) \\ &= \sum_1^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 / (n-2) \end{aligned}$$

5.0 The Effect of the Incorrect Error Model

If the linear model is correct for the true relationship between the volume and the manometer readings, the effect of assuming the wrong error model can be assessed in terms of

1. The efficiency of the estimate under the incorrect error model and
2. the error in the variance made by using the variance formula associated with the incorrect error model.

Since the error in β estimation is particularly pertinent as it affects transfers of part of the tank holding, the β estimates are studied. This leads to the following six cases.

Error Models

<u>Case</u>	<u>cumulative</u>		<u>Independent</u>
	<u>Model 1</u> $\sigma_{\Delta y}^2 \propto \Delta y$	<u>Model 2</u> $\sigma_{\Delta y}^2 = \text{constant}$	<u>Model 3</u> $\sigma_y^2 = \text{constant}$
(1,2)	assumed	true	
(1,3)	assumed		true
(2,1)	true	assumed	
(2,3)		assumed	true
(3,1)	true		assumed
(3,2)		true	assumed

One can write out the covariance matrices except for a scalar factor

$$\Phi_1 = \sigma_1^2 \begin{pmatrix} x_1 & x_1 & \dots & x_1 \\ x_1 & x_2 & \dots & x_2 \\ \dots & & & \\ x_1 & x_2 & \dots & x_n \end{pmatrix} = \sigma_1^2 \Phi_1^0$$

$$\Phi_2 = \sigma_2^2 \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 2 \\ \dots & & & \\ 1 & 2 & \dots & n \end{pmatrix} = \sigma_2^2 \Phi_2^0$$

$$\Phi_3 = \sigma_3^2 \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \dots & \\ & & & 1 \end{pmatrix} = \sigma_3^2 \Phi_3^0$$

Efficiencies

Let model i be the assumed model and model j the correct model. Let Φ_i and Φ_j denote the covariance matrices under models i and j respectively. Let A_j denote the column vector of coefficients such that $\hat{\beta} = A_j'Y$ where Y is the column vector of y values, and $\hat{\beta}$ is the mvue of β . Then the efficiency of the estimate of β assuming model j is correct is

$$\text{Eff} = \frac{A_i' \Sigma_j^{-1} A_i}{A_j' \Sigma_j^{-1} A_j}$$

and this calculation is straightforward and easily carried out.

Bias in Estimating the Variance of the β Estimate Under the Wrong Model

It is, however, somewhat more difficult to assess the effect of the incorrect model on the incorrect estimate of the variance of $\hat{\beta}$ because of scaling factors. For example the user would assume that the variance of the β estimate is

$$s_{\hat{\beta}_i | M_j}^2 = A_i' \Sigma_i^{-1} A_i = \sigma_i^2 A_i' \Sigma_i^{-1} A_i$$

when in fact the true variance is $A_i' \Sigma_j^{-1} A_i = \sigma_j^2 A_i' \Sigma_j^{-1} A_i$, but Σ_i and Σ_j are hard to relate since σ_i^2 and σ_j^2 have different scales and connotations. All three covariance matrices can be written in the form

$$\Sigma_i = \sigma_i^2 \Sigma_i^0$$

say where the meanings of σ_i^2 for the different cases have been indicated before. Then $s_{\hat{\beta}_i | M_j}^2$, the variance when model i is assumed but model j is correct, is calculated by

$$s_{\hat{\beta}_i | M_j}^2 = s_{i|j}^2 A_i' \Sigma_i^{-1} A_i$$

when the true variance is

$$\sigma_{\hat{\beta}_i | M_j}^2 = \sigma_j^2 A_i' \sum_j^0 A_i$$

so that

$$\frac{s_{\hat{\beta}_i | M_j}^2}{\sigma_{\hat{\beta}_i | M_j}^2} = \frac{s_{i|j}^2 A_i' \sum_i^0 A_i}{\sigma_j^2 A_i' \sum_j^0 A_i}$$

gives the ratio of the estimated and true variances of $\hat{\beta}$ under the situation (i,j). Then it is seen that errors can be introduced in the variance estimate in two ways, one by the fact that the covariance matrix assumed is incorrect and second by using $s_{i|j}^2$ an incorrect residual variance. Write $E(s_{i|j}^2)$ for the average residual variance under (i,j), the ith assumed model, jth correct model situation. In this paper R_{ij} is used to assess the extent of the bias here where

$$R_{ij}^2 = \frac{E(s_{i|j}^2) A_i' \sum_i^0 A_i}{\sigma_j^2 A_i' \sum_j^0 A_i}$$

In most cases this evaluation is not as formidable as it looks but when model 3 is assumed the formulae get somewhat complicated. This procedure leads to Table I. In cases (3,1) and (3,2) one can use the formulae as given in Table I for $E(s_{\hat{\beta}_i | M_j}^2)$ or evaluate

$$E[(S_{yy} - S_{xy}^2/S_{xx})/(n-2)/S_{xx}].$$

Another procedure is to write the individual residuals as

$$\begin{aligned} \rho_i &= y_i - \hat{y}_i = y_i - \bar{y} - \hat{\beta}(x_i - \bar{x}) \\ &= y_i - \sum_1^n y_{ij}/n - \frac{(x_i - \bar{x}) \sum_1^n y_j(x_j - \bar{x})}{S_{xx}} \\ &= \sum_1^n y_j (\delta_{ij} - 1/n - \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}) \\ &= \sum_1^n y_j u_{ij} \end{aligned}$$

$E(\rho_i) = 0$ since $\hat{\alpha}$ and $\hat{\beta}$ are unbiased even under the wrong error model. Let $U_i = (u_{i1} \ u_{i2} \ \dots \ u_{in})$ so that $\rho_i = U_i Y$, $(y_i - \hat{y}_i)^2 = \rho_i' \rho_i = Y' U_i' U_i Y = Y' W_i Y$ and

$$E(y_i - \hat{y})^2 = \text{tr } W_i \sum_j \frac{1}{n-2} + E(Y' - \hat{Y}') \sum_j \frac{1}{n-2} E(Y - \hat{Y}) = \text{tr } W_i \sum_j \frac{1}{n-2}.$$

Then

$$E(s_{i|j}^2) = \sum_1^n \text{tr } W_i \sum_j \frac{1}{n-2}$$

and

$$R_{ij}^2 = \frac{E(s_{i|j}^2) A_i' \sum_i^0 A_i}{\sigma_j^2 A_i' \sum_j A_i}$$

$E(s_{i|j}^2)$ comes out as a multiple of σ_j^2 so that the σ_j^2 values cancel out.

Table I

i	j	$\hat{\beta}_i$	$E(\text{assumed variance}) = E(\Delta_{\hat{\beta}_i}^2 M_j)$	$\sigma_{\hat{\beta}_i}^2 M_j$ Actual Variance
1	2	$\frac{y_n - y_1}{x_n - x_1}$	$\frac{\sigma_2^2}{n-2} \left[\sum_{i=2}^n \frac{1}{L_i} - \frac{n-1}{x_n - x_1} \right] \left(\frac{1}{x_n - x_1} \right)$	$\frac{(n-1) \sigma_2^2}{(x_n - x_1)^2}$
1	3		$\frac{2\sigma_3^2}{n-2} \left[\sum_{i=2}^n \frac{1}{L_i} - \frac{1}{x_n - x_1} \right] \left(\frac{1}{x_n - x_1} \right)$	$\frac{2\sigma_3^2}{(x_n - x_1)^2}$
2	1	$\frac{\sum_{j=2}^n L_j \bar{z}_j}{\sum_{j=2}^n L_j^2}$	$\frac{\sigma_1^2}{n-2} \left[x_n - x_1 - \frac{\sum_{i=2}^n L_i \hat{z}_i}{\sum_{i=2}^n L_i^2} \right] \left(\frac{1}{\sum_{j=2}^n L_j^2} \right)$	$\frac{\sum_{j=2}^n L_j^3 \sigma_1^2}{\left(\sum_{j=2}^n L_j^2 \right)^2}$
2	3		$\frac{\sigma_3^2}{n-2} \left[2(n-1) + \frac{\sum_{i=2}^n G_i^2}{\sum_{i=2}^n L_i^2} - \frac{2 \sum_{i=2}^n L_i (G_i - G_{i-1})}{\sum_{i=2}^n L_i^2} \right] \left(\frac{1}{\sum_{j=2}^n L_j^2} \right)$	$\frac{\sum_{i=2}^n G_i^2 \sigma_3^2}{\left(\sum_{i=2}^n L_i^2 \right)^2}$
3	1	$\frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$	$\frac{\sigma_1^2}{(n-2) S_{xx}} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (\min x_k, x_j) \left(\delta_{ij} - \frac{1}{n} - \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \right) \left(\delta_{ij} - \frac{1}{n} - \frac{(x_k - \bar{x})(x_i - \bar{x})}{S_{xx}} \right)$	$\frac{\sigma_1^2 \sum_{i=1}^n \sum_{j=1}^n (\min x_i, x_j) (x_i - \bar{x})(x_j - \bar{x})}{(S_{xx})^2}$
	2		$\frac{\sigma_2^2}{(n-2) S_{xx}} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (\min(i, k)) \left(\delta_{ij} - \frac{1}{n} - \frac{(x_i - \bar{x})(x_k - \bar{x})}{S_{xx}} \right) \left(\delta_{ij} - \frac{1}{n} - \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \right)$	$\frac{\sigma_2^2 \sum_{i=1}^n \sum_{j=1}^n (\min i, j) (x_i - \bar{x})(x_j - \bar{x})}{(S_{xx})^2}$

where $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$, $z_i = y_i - y_{i-1}$, $L_i = x_i - x_{i-1}$, $G_i = L_i - L_{i+1}$ ($i=1, \dots, n$) $L_i = 0, i=n+1$, $\delta_{ij} = 0, i \neq j$, $\delta_{ij} = 1, i=j$

It is apparent that the efficiency and errors made in the variance estimate of β depend upon the pattern of spacing in the x values, i.e., upon the so-called design matrix. In order to evaluate the efficiency and error in variance estimations it would be useful to use simple parameters to do this. In the four cases studied the simple parameters are n , λ , γ , and η .

In the case where the increments $L_i = x_i - x_{i-1}$ are constant it can be shown that $R_{11}^2 = 1$, $r_{13}^2 = n$, $R_{21}^2 = 1$, $R_{23}^2 = n$. In the evenly-spaced case it is difficult to get a simple expression for R_{31}^2 and R_{32}^2 which are functions of n .

The robustness of the estimates of β were studied for bias under 4 conditions:

1. The increments are evenly spaced with n allowed to vary between 5 and 40.
2. The large increments occur at the largest y values. The x_i values, $i = 1, 2, \dots, 10$ are determined by

$$x_i^! = e^{i\lambda/20} - 1 \quad ; \quad \lambda = 1, 2, \dots, 10.$$

A run is calculated for each λ . Then the x_i values are normalized by $x_i = x_i^! / x_{10}^!$. Thus larger jumps at larger x values are associated with larger λ values.

3. The large increments occur at the middle x_i values, where

$$x_i^1 = e^{i\gamma/5} - 1, \quad i = 1, 2, \dots, 5$$

$$\delta_i = x_i^1 - x_{i-1}^1$$

$$x_6^1 = x_5 + \delta_5$$

$$x_7^1 = x_6 + \delta_4$$

$$x_8^1 = x_7 + \delta_3$$

$$x_9^1 = x_8 + \delta_2$$

$$x_{10}^1 = x_9 + \delta_1$$

$$x_i = x_{i-1} + \delta_{11-i}, \quad i = 6, 7, \dots, 10.$$

Then $x_i = x_i^1/x_{10}^1$. Thus for this situation larger γ values indicate bigger jumps at the center.

4. In an attempt to portray an oscillatory effect with respect to the jumps the following formulation is used.

$$x_1^1 = 1$$

$$x_2^1 = 1 + \eta$$

$$x_3^1 = x_2^1 + 1$$

$$x_4^1 = x_3^1 + \eta$$

...

$$x_9^1 = x_8^1 + 1$$

$$x_{10}^1 = x_9^1 + \eta$$

Then $x_i = x_i^1/x_{10}^1$ and η determines the extent of the "sawtooth" effect and results are given as a function of η . Table II gives the values $L_i = x_i - x_{i-1}$ as functions of λ , γ and η .

Table II

This table gives $L_i = x_i - x_{i-1}$
for different λ , γ , and n values.

λ									
.0790	.0612	.0465	.0347	.0254	.0183	.0130	.0092	.0064	.0044
.0831	.0676	.0540	.0423	.0326	.0247	.0185	.0137	.0100	.0073
.0873	.0748	.0627	.0517	.0419	.0334	.0263	.0204	.0157	.0120
.0918	.0826	.0729	.0631	.0538	.0451	.0373	.0305	.0246	.0197
.0965	.0913	.0847	.0771	.0690	.0609	.0529	.0454	.0386	.0325
L_i .1015	.1009	.0984	.0942	.0887	.0822	.0751	.0678	.0606	.0536
.1067	.1115	.1143	.1151	.1138	.1109	.1066	.1012	.0950	.0884
.1122	.1233	.1328	.1405	.1462	.1497	.1512	.1509	.1490	.1457
.1179	.1362	.1543	.1716	.1877	.2021	.2146	.2251	.2337	.2403
.1240	.1505	.1793	.2096	.2410	.2726	.3045	.3358	.3664	.3961
γ									
.0644	.0385	.0215	.0114	.0058	.0029	.0014	.0007	.0003	.0001
.0787	.0574	.0392	.0254	.0158	.0096	.0057	.0033	.0019	.0011
.0961	.0857	.0715	.0566	.0431	.0318	.0229	.0163	.0114	.0079
.1174	.1278	.1303	.1260	.1171	.1055	.0930	.0806	.0690	.0585
L_i .1434	.1906	.2374	.2805	.3182	.3503	.3770	.3992	.4174	.4324
.1434	.1906	.2374	.2805	.3182	.3503	.3770	.3992	.4174	.4324
.1174	.1278	.1303	.1260	.1171	.1055	.0930	.0806	.0690	.0585
.0961	.0857	.0715	.0566	.0431	.0318	.0229	.0163	.0114	.0079
.0787	.0574	.0392	.0254	.0158	.0096	.0057	.0033	.0019	.0011
.0644	.0385	.0215	.0114	.0058	.0029	.0014	.0007	.0003	.0001
n									
.1000	.0667	.0500	.0400	.0333	.0286	.0250	.0222	.0200	.0182
.1000	.1333	.1500	.1600	.1667	.1714	.1750	.1778	.1800	.1818
.1000	.0667	.0500	.0400	.0333	.0286	.0250	.0222	.0200	.0182
.1000	.1333	.1500	.1600	.1667	.1714	.1750	.1778	.1800	.1818
L_i .1000	.0667	.0500	.0400	.0333	.0286	.0250	.0222	.0200	.0182
.1000	.1333	.1500	.1600	.1667	.1714	.1750	.1778	.1800	.1818
.1000	.0667	.0500	.0400	.0333	.0286	.0250	.0222	.0200	.0182
.1000	.1333	.1500	.1600	.1667	.1714	.1750	.1778	.1800	.1818
.1000	.0667	.0500	.0400	.0333	.0286	.0250	.0222	.0200	.0182
.1000	.1333	.1500	.1600	.1667	.1714	.1750	.1778	.1800	.1818

6.0 Results and Conclusions in Detail

The results are given by graphs 1-8 where the effects of choosing the wrong error model are given as efficiencies and average errors under the wrong errors models. The effects of using the wrong error model depends upon the spacing of the x values as well as on which incorrect model is used. These are studies on the robustness of different models when the assumptions are wrong where the efficiency and error bias are measures of the robustness. The x spacings which yield the design matrices of the calibrations were studied for;

- a. the effect of n using even spacings,
- b. the effect of large spacings at an end of the scale,
- c. the effect of larger spacings in the center of the x scale, and
- d. the effect of an oscillatory or sawtooth effect.

The conclusions are as follows.

1. When the x values are evenly spaced and the efficiencies and biases in estimating the standard deviation of the β estimates are functions of n, the following results are obtained (graphs 1 and 2):
 - a. The poorest efficiency arises when model 3 is assumed but either model 1 or model 2 is correct and the efficiency decreases with increasing n. In addition $\sigma_{\hat{\beta}_3}$, the standard deviation of the β estimate is underestimated more and more as n increases which increases the probability of type I errors.
 - b. When model 3 is correct the efficiency is somewhat diminished when n increases and either model 1 or model 2 is chosen. In these cases the standard deviations of the β estimate is overstated increasingly more as n increases leading to a greatly diminished test power.

- c. When models 1 and 2 are interchanged in this situation the efficiency is 1 and there is no bias in estimating the standard deviation of the β estimate.

Efficiency of the estimation of β is given as a function of the number of calibration points when the x values (manometer readings) are equally-spaced. Under equal spacing the models 1 and 2 yield the same slope so that consequently they have the same performance characteristics in this case. (i,j) means that error model i is assumed when error model j is the correct model. Efficiency is defined as the variance of the minimum variance unbiased estimator under the correct error model divided by the variance of the unbiased estimator which results from using the wrong error model.

Graph 1

These curves give the ratios $R_{ij} = \sqrt{\frac{E(s_{\beta_i}^2 | M_j)}{\sigma_{\beta_i}^2 | M_j}}$

where i is the assumed error model and j is the correct error model for the situation where the x_i values are evenly spaced, as a function of the number of observations. Under the evenly-spaced situation the models 1 and 2 behave the same so that only 3 curves are needed to represent the 6 situations. (i,j) means that error model i is assumed where error model j is the correct model.

Graph 2

2. When the spacings in the x values get larger at one or the other end of the calibration scales, i.e., increases, the following results are obtained (graphs 3 and 4):

- a. The poorest efficiencies tend to be when model 3 is chosen but either model 1 or model 2 is correct. In these cases the estimate of the standard deviation of β is greatly underestimated leading to an increased frequency of type I errors.
- b. The efficiency of using model 2 when either model 1 or model 3 is correct is good for small λ values (nearly equal spacing) but deteriorates as λ increases. If model 1 is correct, $\sigma_{\hat{\beta}_2}$ the estimate of the standard deviation of $\hat{\beta}_2$ is approximately correct for small λ but becomes more and more underestimated as λ increases. If model 3 is correct, the estimate of the standard deviation of $\hat{\beta}_2$ is too large but this bias decreases somewhat as λ increases.
- c. The efficiency of model 1 holds up well under increased λ values. However, the estimate of the standard deviation of β_1 is too large if either model 2 or model 3 holds and this bias increases as λ increases.

The efficiency of β estimation under the wrong model as a function of λ . As λ increases the x spacings become larger near the end of the x scale.

Graph 3

The curves give $\sqrt{E(s_{\beta_i}^2 | M_j) / \sigma_{\beta_i}^2 | M_j}$ as a function of λ

where λ is a parameter used to apportion the x values. For the k th point with a given λ , $x_k = \exp(\lambda k / 20) - 1$, ($k = 1, 2, \dots, 10$). The larger λ is the more relative variation there is between the x_k values. With this scheme the larger jumps $x_i = x_i - x_{i-1}$ means the model i is assumed when the correct error model is j .

Graph 4

3. When the largest increments are near the center of the range of the x 's and increasing γ increases the disparity between the jumps at the center and at the end of the x scale, one has the following conclusions (graphs 5 and 6).
 - a. The poorest efficiency arises when model 3 is assumed but model 1 or model 2 is correct. Furthermore in this case the standard deviation of the β estimate is greatly underestimated leading to an excess of type I errors.
 - b. When model 2 is assumed but either model 1 or model 3 is correct, the β estimate tends to be inefficient and particularly so as γ increases. Furthermore in this case when model 1 is correct the standard deviation of the β estimate is greatly underestimated and the frequency of type I errors would then be increased. When model 3 is correct $\sigma_{\hat{\beta}_2}$ is overestimated which results in a diminished power for the test.
 - c. When model 1 is assumed, the efficiency holds up well when either models 2 or 3 is correct. However, here $\sigma_{\hat{\beta}_1}$ is greatly overestimated when either model 2 or model 3 is correct.

The efficiency of δ estimation under the wrong model as a function of γ . As γ increases the x spacings become larger near the middle of the x scale.

Graph 5

The effect of larger x spacings near the middle of the x scale on R_{ij} are given as a function of γ . As γ increases the x spacings become increasingly large near the middle range of the x scale.

Graph 6

4. If the increments are unevenly spaced such that alternating increments are large and small and η is the ratio of the large increment to the small increment, the conclusions are as follows (graphs 7 and 8).
 - a. By far the poorest efficiency is obtained when model 3 is used but model 2 is correct. This inefficiency increases with η . Here, too, the estimate of the standard deviation of the β estimate is much too small leading to an increase of type I errors. This underestimation is almost constant as a function of η .
 - b. The efficiency of using model 3 when model 1 holds is somewhat better, ($E \sim 0.5$) and is almost independent of η . However, the standard deviation of the β estimate is biased in the direction of being much too small.
 - c. The efficiency of model 2 if either model 1 or model 2 holds, is good for small η but decreases as η increases. If model 3 is correct then using model 2 overestimates the standard deviation of the slope estimate but this bias decreases as η increases. If model 1 is correct the standard deviation of the slope estimate is more and more underestimated as η increases.
 - d. Model 1 tends to be the most robust as far as efficiency is concerned. However, if either model 2 or model 3 holds, the standard deviation of the slope estimate becomes more and more overestimated as η increases. This increase is less pronounced under model 2.

The efficiency of β estimation under the wrong model as a function of η . As η increases the oscillatory or sawtooth effect becomes more and more exaggerated. η is the ratio of the large increment in x to the small increment.

Graph 7

The effect of increasing large alternating jumps in the x scale on R_{ij} as a function of n . As n increases the oscillatory or sawtooth effect becomes more and more exaggerated and n is the ratio of the large increment in x to the small increment.

Graph 8

References

1. Mandel, John, "Fitting a Straight Line to Certain Types of Cumulative Data", Journal of the American Statistical Association, Vol. 52, 1957, pp. 552-566.
2. Jaech, J. L., "A Note on the Equivalence of Two Methods of Fitting a Straight Line Through Cumulative Data", Journal of the American Statistical Association, Sept. 1964, Vol. 59, pp. 863-866.
3. Tingey, F. H., Selected Measurement Methods, U.S.A.E.C., 1963, Chapter 5, pp. 78-83.
4. Rao, R. C., Linear Statistical Inference and its Applications, John Wiley and Sons Inc., New York, 1965, page 180.
5. Searle, S. R., Linear Models, John Wiley and Sons, Inc., New York, 1971, pages 55 and 421.
6. Krutchkoff, R. G., "Classical and Inverse Regression Methods of Calibration", Technometrics, August 1967, Vol. 9, No. 3, pp. 425-440.