UNCERTAINTY ESTIMATES FOR DERIVATIVES AND
INTERCEPTS*

Edward L. Clark
Experimental Aerodynamics Division
Sandia National Laboratories
Albuquerque, New Mexico

SAND--90-2336C

DE91 004503

ABSTRACT

Straight line least squares fits of experimental data are widely used in the analysis of test results to provide derivatives and intercepts. A method for evaluating the uncertainty in these parameters is described. The method utilizes conventional least squares results and is applicable to experiments where the independent variable is controlled, but not necessarily free of error. A Monte Carlo verification of the method is given.

INTRODUCTION

The need to estimate derivatives and intercepts occurs frequently in the analysis of aerospace test results. Examples include: (1) stability derivatives, for example, C_{N_α} , and C_{m_α} ; (2) zero-lift pitching moment coefficient, C_{m_0} , a y -intercept; and (3) trim angle, α_t , an x -intercept. Derivatives and intercepts are also encountered in transducer calibrations, for example, the sensitivity and electrical bridge locations for an internal balance. In most cases the derivatives and intercepts are evaluated from least squares straight line fits. The usual technique for estimating the uncertainty in a calculated result is by use of the Taylor's series method [1,2]. This method combines parameter uncertainties with sensitivity coefficients, obtained by differentiating the functional relationship between the result and the parameters with respect to each of the parameters, to obtain the uncertainty of the result. This approach is difficult to use for estimating the uncertainty in derivatives and intercepts. The purpose of this paper is to present an alternate method for evaluating the uncertainty in these results.

*This work was supported by the United States Department of Energy under Contract DE-AC04-76DP00789.

UNCONTROLLED AND CONTROLLED EXPERIMENTS

The type of experiment used to acquire the data is of critical importance to the analysis of derivative and intercept accuracy. Berkson [3] categorized experiments as uncontrolled or controlled and described the significant statistical differences between them. As examples of these types of experiments, consider two tests to evaluate the drag coefficient of a projectile, C_D , as a function of angle of attack, α . The assumed functional relationship is, $C_D = C_{D_0} + k\alpha^2$ where α^2 is the independent variable and C_D is the dependent variable.

In an uncontrolled experiment, both the independent variable (x) and dependent variable (y) samples are drawn from an existent population. For our present example, assume that the data are acquired from numerous range tests where the projectile is fired with a fixed muzzle velocity. The projectile is perturbed by muzzle exit conditions, thereby producing random values of α . Each test is analyzed over a time period where α is essentially constant. This is an uncontrolled experiment, since α cannot be predicted prior to the test, and the data may be considered as being drawn from an existent population. A straight line least squares fit is made to the data, (C_{D_i}, α_i^2) , to estimate the parameters C_{D_0} and k . Conventional least squares calculations are used, that is, the sum of the squared residuals for the dependent variable is minimized. If there is no error in the independent variable, the least squares fit is unbiased and will provide a true estimate of the parameters. However, if there are errors in the independent variable, the fit is biased and the parameters will not be correct. There are methods of obtaining a true estimate of the parameters even if both variables contain errors. However, these methods are complicated and lack the simplicity in estimation of parameters and variances that the conventional least squares analysis provides. The analysis of this paper does not apply to data obtained in uncontrolled experiments where there is error in the independent variable.

In contrast to the uncontrolled experiment, a controlled experiment is one in which the independent variable is set at a preassigned value, x_i , or at least as close to x_i as is feasible, before the measurement of the dependent variable, y_i , is made. In our example of the projectile drag coefficient, consider that the tests are performed in a wind tunnel. The experiment is conducted by preselecting the nominal angles of attack at which the drag will be measured. During the test, the projectile is pitched to the desired angle and data are acquired. This process is repeated until data have been obtained at the required number of angles. This type of test was defined as a controlled experiment by Berkson. He showed that for an experiment of this type, even

if there are errors in the independent variable, a conventional least squares fit will provide an unbiased estimate of the true function. As will be shown in the next section, the independent variable error is treated as an "induced" error in the dependent variable.

Many aerospace test measurements are controlled experiments. For example, in wind tunnel stability measurements, the angle of attack is usually set at prescribed values. The actual angles will vary from these values because of uncertainties in angle setting, balance deflection and flow angularity. However, this is still a controlled experiment. As another example, in balance calibrations the calibration body is loaded at specified stations. Even though there may be some uncertainty in the station value, this is a controlled experiment. The error analysis in the following sections will assume that the data were acquired in a controlled experiment and that there may be errors in both variables.

LEAST SQUARES STRAIGHT LINE FIT

It is assumed that the derivatives and intercepts will be obtained from n pairs of observations, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, by the least squares estimator

$$\hat{y} = a + bx \quad (1)$$

of the linear function

$$y = \alpha + \beta x. \quad (2)$$

In Eq.(1), a is the estimate of the y -intercept, α , and b is the estimate of the slope, β . These parameters, together with the x -intercept, x_o , are shown in Figure 1.

Figure 1 Notation

Furthermore, it will be assumed that the data used in the least squares fit matches the model. If y varies nonlinearly with x , this analysis cannot predict the error due to the poor correlation between the model and the data. In some cases, it may be known that $\alpha = 0$ and the appropriate least squares estimator is $\hat{y} = bx$. This case is analyzed in the Appendix.

For the Berkson model, the error in the controlled variable, x_i , is considered to be zero. The error in the measured variable, y_i , is composed of a "measurement" error, ϵ_{y_i} , and an "induced" error, $\beta\epsilon_{x_i}$, where ϵ_{x_i} is the difference between the nominal value of x and the actual value. Then y_i is given by [4],

$$y_i = \alpha + \beta x_i + (\epsilon_{y_i} + \beta\epsilon_{x_i}). \quad (3)$$

The variance in y_i , which will be denoted S^2 , is given by

$$S^2 = V[y_i] = V[\epsilon_{y_i}] + \beta^2 V[\epsilon_{x_i}]$$

or,

$$S = \sqrt{S_y^2 + \beta^2 S_x^2} \approx \sqrt{S_y^2 + b^2 S_x^2} \quad (4)$$

where S_x and S_y are the standard deviations of ϵ_{x_i} and ϵ_{y_i} , respectively. To estimate the uncertainty in the slope and intercepts, certain assumptions must be made about the errors, ϵ_{x_i} and ϵ_{y_i} . It will be assumed that the errors in each variable are normally distributed with mean zero and variance σ^2 , that is, $\epsilon_{x_i} \sim N(0, \sigma_x^2)$ and $\epsilon_{y_i} \sim N(0, \sigma_y^2)$. This assumption furthermore requires that the errors in each variable be independent, that is, $cov(\epsilon_{x_i}, \epsilon_{x_j}) = cov(\epsilon_{y_i}, \epsilon_{y_j}) = 0$, $i \neq j$.

In uncertainty analyses, errors are categorized as precision (random) or bias (fixed). The precision error of a measurement is estimated by the precision index which is equal to the standard deviation, or square root of the variance, of the measurement. The standard deviations S_x and S_y in Eq.(4) are the precision indices for the x - and y -variables and are evaluated by the techniques described in Ref. 1 and 2. The effect of bias errors will be discussed later. In many instances, for example, the stability derivative, $C_{m_\alpha} = dC_m/d\alpha$, evaluation of the error will require the use of propagation techniques to estimate the error in a calculated result, such as C_m in this example. Although it might seem that the standard error of estimate (SEE) of the fit, which is defined by

$$SEE = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

could be used to estimate S , this value is obtained from a very limited sample of the data and it is better to use the errors in x and y which are estimated with standard techniques.

Assuming that the x -variable was controlled in the measurement of interest, and the measurement variance is defined by Eq.(4), the uncertainty in a and b can be derived with conventional least squares techniques. The model defined in Eq.(1) is difficult to use in variance analyses because the parameters a and b are dependent and require evaluation of covariance terms. Instead, use the model suggested by Brownlee [5],

$$\hat{y} = a' + b'(x - \bar{x}) \quad (5)$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (6)$$

The parameters a' and b' are independent and are related to a and b by

$$a = a' - b'\bar{x} \quad (7)$$

and

$$b = b'. \quad (8)$$

In the following derivations, all summations are for $i = 1, 2, \dots, n$ and the index of summation will be omitted. If $\hat{y}_i$ is the value of y estimated from Eq.(5) for $x = x_i$ and y_i is the measured value of y corresponding to x_i , define the dependent variable residual, R_i , as

$$R_i = y_i - \hat{y}_i = y_i - a' - b'(x_i - \bar{x}).$$

The method of least squares provides values for the parameters a' and b' which will minimize the sum of the residuals squared. To determine a' and b' , differentiate $\sum R_i^2$ with respect to a' and b' and equate to zero. This gives two equations in a' and b' which can be solved [5] to give

$$a' = \frac{1}{n} \sum y_i = \bar{y} \quad (9)$$

and

$$b' = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}. \quad (10)$$

To obtain a and b , substitute Eqs.(9) and (8) into Eq.(7) and Eq.(10) into Eq.(8) which gives

$$a = \bar{y} - b\bar{x} \quad (11)$$

and

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}. \quad (12)$$

PRECISION INDICES

The precision index, S_u , of a variable, u , is defined as the sample standard deviation and is equal to the square root of the variance, $V[u]$, of the variable. Thus, precision indices for the derivative and intercepts can be estimated from the variances of the least squares parameters. Since a' and b' are linear functions of the y_i , which are assumed to be independent and to have a normal distribution with variance S^2 , and $V[x_i] = 0$ (Berkson's controlled x -variable case), the precision indices of a' and b' can be calculated directly from Eqs.(9) and (10),

$$S_{a'} = \sqrt{V[a']} = \sqrt{\frac{\sum V[y_i]}{n^2}} = \frac{S}{\sqrt{n}} \quad (13)$$

and

$$S_{b'} = \sqrt{V[b']} = \sqrt{\frac{\sum (x_i - \bar{x})^2 V[y_i]}{[\sum (x_i - \bar{x})^2]^2}} = \frac{S}{\sqrt{\sum (x_i - \bar{x})^2}} \quad (14)$$

where S is defined in Eq.(4). Since $a = a' - b'\bar{x}$,

$$S_a = \sqrt{V[a]} = \sqrt{V[a'] + \bar{x}^2 V[b']} = S \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} \quad (15)$$

and $b = b'$ so,

$$S_b = \sqrt{V[b]} = \sqrt{V[b']} = \frac{S}{\sqrt{\sum (x_i - \bar{x})^2}} \quad (16)$$

Equations (15) and (16) can also be obtained by the Taylor's series method for propagation of errors, for example, see Ref. 6.

The precision index of a predicted value, $\hat{y}$, at any x value is given by

$$S_{\hat{y}} = \sqrt{V[\hat{y}]} = \sqrt{V[a'] + (x - \bar{x})^2 V[b']}$$

so,

$$S_{\hat{y}} = S \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]^{1/2} \quad (17)$$

At $x = 0$, $\hat{y} = a$. Substituting $x = 0$ in Eq.(17) gives the result of Eq.(15). It is interesting to compare the standard deviations, S_a and S_b , of the parameters a and b , to the standard deviation, S , of the data. For many applications, $\bar{x} \approx 0$. Then, $a \approx \bar{y}$ and $S_a \approx S/\sqrt{n}$. This is the classical result for the standard deviation of a mean and indicates that uncertainty in the intercept is

primarily a function of the number of samples in the fit. On the other hand, $S_b \approx S/\sqrt{\sum x_i^2}$, and the uncertainty in the derivative is related to the number of samples and their x -range.

For a given value of y , the predicted value of the independent variable, $\hat{x}$, is given by

$$\hat{x} = (y - a)/b. \quad (18)$$

The precision index for this prediction is given by (e.g., see Ref.7),

$$S_{\hat{x}} = \frac{(\hat{x} - \bar{x})g/t_{95} \pm S/b[(1 - g)/n + (\hat{x} - \bar{x})^2/\sum(x_i - \bar{x})^2]^{1/2}}{1 - g}$$

where $g = t_{95}^2 S^2 / b^2 \sum(x_i - \bar{x})^2$ and t_{95} is the 95th percentile point for the Student's t -distribution with appropriate degrees of freedom. In most cases, $g \ll 1$. For example, with $t_{95} = 2$ and errors in both the x and y data equal to 5% of full scale, $g < 0.01$ if the data are symmetrical about $x = 0$. Then, setting $g = 0$, the above expression reduces to

$$S_{\hat{x}} \approx \frac{S}{b} \left[\frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right]^{1/2} = \frac{S_y}{b}. \quad (19)$$

Letting $\hat{x} = x_o$ gives

$$S_{x_o} \approx \frac{S}{b} \left[\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right]^{1/2} \quad (20)$$

where, from Eq.(18) with $y = 0$,

$$x_o = -a/b. \quad (21)$$

In summary, to determine the precision indices for derivatives or intercepts estimated with a least squares straight line, first, calculate the precision indices S_x and S_y of the x - and y -variables using standard error analysis techniques. Then, calculate the combined index, S , with Eq.(4). For the linear model of Eq.(1), the precision index, S_b , of the derivative, b , is given by Eq.(16). The precision index, S_a , of the y -intercept, a , is given by Eq.(15). Finally, the precision index, S_{x_o} , of the x -intercept, x_o , is given by Eq.(20) with x_o defined by Eq.(21).

UNCERTAINTY ESTIMATES

A bias error, for example, flow angularity, does not add uncertainty to the derivative. Therefore, the total uncertainty in the derivative, for a confidence level of 95%, is given by [1],

$$U_b = t_{95} S_b \quad (22)$$

where t_{95} is the 95th percentile point for the two-tailed Student's t -distribution with ν degrees of freedom. The degrees of freedom associated with the uncertainties in the derivative and intercepts can be estimated by the Welch-Satterthwaite equation [1,2],

$$\nu = \frac{(S_y^2 + b^2 S_x^2)^2}{\frac{S_y^4}{\nu_y} + \frac{b^4 S_x^4}{\nu_x}} \quad (23)$$

where ν_x and ν_y are the degrees of freedom associated with S_x and S_y , respectively.

Bias errors will affect the uncertainty in the x - and y -intercepts and must be included in the analysis. Assuming that there are bias limits, B_x and B_y , in the x - and y -measurements, the bias limit, $B_{\hat{x}}$, in a predicted value of x is given by

$$B_{\hat{x}} = (B_y^2/b^2 + B_x^2)^{1/2}. \quad (24)$$

The bias limit, $B_{\hat{y}}$, in a predicted value of y is given by

$$B_{\hat{y}} = (B_y^2 + b^2 B_x^2)^{1/2}. \quad (25)$$

The total uncertainty of a result, for a confidence level of 95%, is defined in Ref. 1 as the root-sum-square of the bias limit and the product of the precision index and t_{95} . Then, the uncertainty in predicted values, $\hat{x}$ or $\hat{y}$, is given by,

$$U_{\hat{x}} = [B_{\hat{x}}^2 + (t_{95} S_{\hat{x}})^2]^{1/2} \quad (26)$$

and

$$U_{\hat{y}} = [B_{\hat{y}}^2 + (t_{95} S_{\hat{y}})^2]^{1/2}. \quad (27)$$

For the x - and y -intercepts, x_o and a , the uncertainty is given by

$$U_{x_o} = [B_{\hat{x}}^2 + (t_{95} S_{x_o})^2]^{1/2} \quad (28)$$

and

$$U_a = [B_{\hat{y}}^2 + (t_{95} S_a)^2]^{1/2}. \quad (29)$$

It should be reiterated that, if there are errors in the independent variable, this technique is valid only for controlled experiments, where the independent variable values are preselected.

MONTE CARLO SIMULATION

To verify the technique described above, a Monte Carlo simulation was used. The model parameters, α and β , and the n values of x_i were defined. The dependent variable values, y_i , were then calculated with Eq.(3) using normally distributed pseudo-random errors with standard deviations of S_x and S_y for the errors ϵ_{x_i} and ϵ_{y_i} , respectively. Data generated in this way are consistent with Berkson's definition of a controlled experiment. A straight line least squares fit with the model of Eq.(1) was obtained using the generated data. The errors in a , b , and x_0 were calculated. This process was repeated 1000 times for each set of base values and sample standard deviations were calculated for a , b , and x_0 . As would be expected, the errors in the derivatives and intercepts were also normally distributed. This procedure only evaluates the effect of precision errors, since the effect of bias errors is obvious. Many cases were evaluated and the Monte Carlo results agreed closely with the estimates. Input values for four typical cases are listed in Table 1. The results for these cases are presented in Table 2, where the estimates, indicated by the appropriate equation numbers, are compared to the Monte Carlo results, indicated by "MC". For these cases, the y -intercept, α , was set equal to $-\beta$ which gave an x -intercept of 1.0.

Case	β	x range	n	S_x	S_y	S , Eq.(4)
1	0.01	-4 to +4	9	0.0396	0.000400	0.000562
2	1.0	-4 to +4	9	0.0396	0.0400	0.0563
3	100.0	-4 to +4	9	0.0398	4.05	5.68
4	200.0	0 to +5	6	0.0101	9.86	10.1

Table 1: Input

Case	S_a		S_b		S_{x_0}	
	Eq.(15)	MC	Eq.(16)	MC	Eq.(20)	MC
1	0.000188	0.000190	0.0000726	0.0000712	0.0201	0.0200
2	0.0188	0.0190	0.00726	0.00734	0.0201	0.02015
3	1.89	1.91	0.733	0.741	0.0203	0.0204
4	7.29	7.32	2.41	2.45	0.0274	0.0274

Table 2: Results

CONCLUSIONS

The uncertainty in a straight line fit can be estimated with conventional least squares techniques, even though there are errors in the independent variable data, if the data were acquired in a controlled experiment. A controlled experiment is one in which the independent variable is preset at selected values prior to making the dependent variable measurements. Data of this type is common in instrument calibrations and aerospace test measurements. Equations are derived for the precision indices, bias limits and combined uncertainties of straight line intercepts and derivatives. The equations are verified with Monte Carlo calculations for typical data sets.

REFERENCES

1. American Society of Mechanical Engineers, "Instruments and Apparatus, Part 1, Measurement Uncertainty," ANSI/ASME PTC 19.1-1985, ASME, 1985.
2. Abernathy, R. B., et al, "Measurement Uncertainty Handbook," (Reprint of AEDC-TR-73-5), ISBN: 087664-483-3, Instrument Society of America, 1980.
3. Berkson, J., "Are There Two Regressions?," *American Statistical Association Journal*, Vol. 45, June 1950, pp. 164-180.
4. Mandel, J., *The Statistical Analysis of Experimental Data*, Sect. 12.6, Dover Publications, New York, 1964.
5. Brownlee, K. A., *Statistical Theory and Methodology in Science and Engineering*, Second Edition, Chap. 11, John Wiley & Sons, New York, 1965.
6. Beers, Y., *Introduction to the Theory of Error*, Second Edition, Sect. VI-B, Addison-Wesley Publishing Co., Reading, Mass., 1957.
7. Draper, N. R. and Smith, H., *Applied Regression Analysis*, Second Edition, Sect. 1.7, John Wiley & Sons, New York, 1981.

APPENDIX

LEAST SQUARES STRAIGHT LINE FIT THROUGH THE ORIGIN

In some cases it may be known that the intercept, a , should be zero. The least squares estimator is then,

$$\hat{y} = bx.$$

The dependent variable residual, R_i , is defined as,

$$R_i = y_i - \hat{y}_i = y_i - bx_i.$$

The method of least squares provides a value for the parameter, b , which minimizes the sum of the residuals squared. The parameter is estimated by differentiating $\sum R_i^2$ with respect to b and equating the result to zero. Then,

$$b = \frac{\sum x_i y_i}{\sum x_i^2}, \quad i = 1, n.$$

Since b is a linear function of the y_i , and $V[x_i] = 0$, the precision index, S_b , of the derivative, b , is given by

$$S_b = \sqrt{V[b]} = \sqrt{\frac{\sum x_i^2 V[y_i]}{(\sum x_i^2)^2}} = \frac{S}{\sqrt{\sum x_i^2}}$$

where S is defined in Eq.(4). The uncertainty in the derivative, for a confidence level of 95%, is given by

$$U_b = t_{95} S_b.$$

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.

- END -

DATE FILMED

04 / 03 / 91