[9] Champetier, J. L., "Interprétation Théorique de l'Évolution du Plasma Créé par Focalisation d'un Faisceau Laser dans l'Air," *Compte Rendus à l'Academie des Sciences de Paris*, Vol. 261, Nov. 1965, pp. 3954–3957.

[10] Raizer, Y., "Heating of a Gas by a Powerful Light Pulse," *Soviet Physics JETP*, Vol. 21, No. 5, Nov. 1965, p. 1009.

[11] Colin, C. et al., "Laser Produced Plasma from Solid Deuterium Targets," *Journal of Applied Physics*, Vol. 39, No. 7, June 1968, pp. 2991–2993.

[12] Champetier, J. L. et al., "Utilisation de Lentilles Asphériques pour l'Obtention d'Éclairements Élevés," *Compte Rendus à l'Academie des Sciences de Paris*, Vol. 266, March 1968, pp. 838–841.

[13] de Metz, J., Terneaud, A., and Veyrié, P., "Etude Optique du Faisceau Émis par un Laser de Grande Intensité," *Applied Optics*, Vol. 5, No. 5, May 1966, pp. 819–822.

[14] Bobin, H. L. et al., "X Rays from a Laser Created Plasma," *Physics Letters*, Vol. 28 A, No. 6, 1968, pp. 398–399.

[15] Veyrie, P., "Contribution à l'Etude de l'Ionisation et du Chauffage des Gaz par le Rayonnement d'un Laser Déclenché—Résultats Experimentaux," *Le Journal de Physique*, Vol. 29, Jan. 1968, pp. 33–41.

[16] Caruso, A., Bertotti, B., and Guiponni, P., "Ionization and Heating of Solid Materials by Means of a Laser Pulse," Il Nuovo Cimento, Vol. XLV, No. 2, Oct. 1966, pp. 176–189.

[17] Zel'dovich, Y. and Raizer, Y., *Physics of Shock Waves and High Temperature Hydrodynamic Phenomena*, Vol. II, Academic Press, New York, 1967.

[18] Marshak, R. E., "Effect of Radiation on Shock Wave Behavior," *The Physics of Fluids*, Vol. 1, No. 1, 1958, pp. 24–28.

AUGUST 1970 AIAA JOURNAL VOL. 8, NO. 8

Criteria for Selecting Curves for Fitting to Data

T. J. DYLEWSKI*

Lockheed Missiles & Space Company, Sunnyvale, Calif.

The purpose of curve-fitting is the prediction of future data; thus, criteria should discriminate strongly in favor of curves having high predictive merit for a given phenomenon. If the laws of nature pertaining to a phenomenon are known, they should form the basis for choosing the curve to be fitted. Otherwise, bias is almost sure to exist in the representation of data by an arbitrary curve. Bias is a systematic discrepancy between the fitted curve and the true equation governing the data. Bias can arise from either oversmoothing or undersmoothing. A criterion has been developed which selects a curve of minimum bias for improving the predictability of future data even if the true equation is unknown. The minimum-bias criterion accepts that type of curve for which the ratio of over-all sum of squares of deviations to subset sum of squares of deviations is a minimum.

I. Purpose and Nature of Curve-Fitting

THE purpose of curve-fitting is assumed to be the summarizing of experimental evidence for making predictions with regard to future data. Although this purpose has been clearly stated in the past,[1,2] the literature of the subject reveals little evidence of the explicit use of predictive ability as a criterion of curve-fitting.

The two kinds of prediction possible through curve-fitting, in order of rapidly decreasing accuracy, are 1) interpolation, or prediction of the dependent variable within the range of observed values of the independent variable and 2) extrapolation, or prediction of the dependent variable beyond the range of observed values of the independent variable.

Of course, all predictions at best have only a formal, *ceteris-paribus* validity; that is, all conditions, deterministic as well as probabilistic, must prevail in the future.

Two distinct but interdependent steps in the curve-fitting process determine how well its predictive purpose is realized. These steps are 1) selection of the type of curve to be fitted and 2) evaluation of the constants of the curve.

The two steps are mutually complementary since each satisfies requirements that the other is unable to fulfill. Both must always be done with care. If one is slighted, no amount of meticulous attention to the other can compensate for the neglect. Moreover, execution of these steps should be preceded by a critical review of the data for the purpose of determining whether a curve fitted to them can serve the intended use.

Regardless of the functional relation that it represents, a curve fitted to data can be thought of as simple or complicated, depending on the number of constants that must be evaluated for defining it. The greater the number of constants, the more closely will the curve follow the experimental data. The suitability of a particular curve, however, is always dependent upon the amount of real random fluctuation present in the data.

If the selected curve is too simple, there will be large deviations between the curve and the data; consequently, excellent data may look bad. Conversely, if the curve is too complicated, it will follow closely not only the true variation of the data, but the random fluctuation as well, producing thereby a false impression of high accuracy.

Figures 1 and 2 illustrate how an arbitrary choice of the type of curve in place of the proper one of Fig. 3 can misrepresent the data through either oversmoothing or undersmoothing. Oversmoothing means that some deterministic variation has been regarded as random variation and discarded. Undersmoothing means that some random variation has been regarded as deterministic and has been retained.

It is unfortunate that the phrase "best fit" is commonly used to describe various functional approximations without adequate attention to the criteria upon which the so-called best fit is based. Perhaps the most insidious is the Tschebycheff method of fitting polynomials, which is superficially

Received September 4, 1969; presented as Paper 69-950 at the AIAA Aerospace Computer System Conference, Los Angeles, Calif., September 8–10, 1969; revision received February 2, 1970. The author is grateful to D. Colvin for programing the bias criterion test cases for the Univac 1108 computer and SC 4020 plotter.

* Senior Staff Scientist, Scientific Computing Division. Member AIAA.

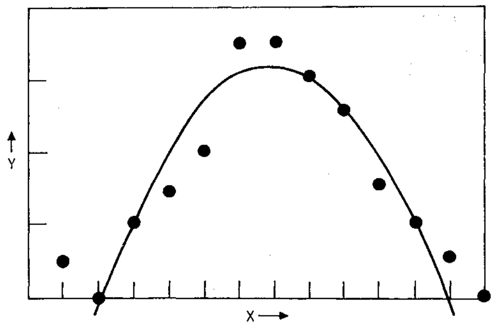

Fig. 1 Bias due to oversmoothing.

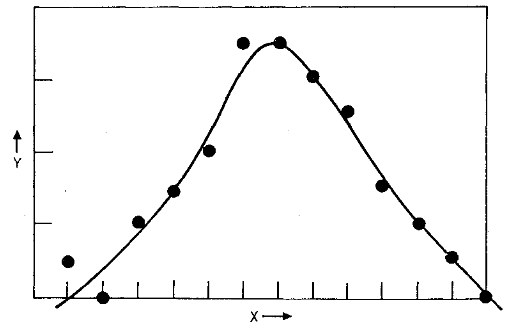

Fig. 3 Curve fitted with minimum bias.

appealing but often disappointing or even misleading. In this method the criterion of curve-fit is stated as the achievement of a minimum value for the maximum deviation.[3] It is often not realized that the criterion is met by tacitly assuming that the deviations of the data about the chosen curve type are sinusoidal. As a consequence of this assumption, successive deviations acquire opposite signs and all have the maximum magnitude.

Since deviations of empirical data from an appropriate approximating curve are random and not systematically sinusoidal, Tschebycheff polynomials are inherently not well suited for expressing experimental results. The hopefully minimized maximum deviations are associated with all observations and may be very large indeed. At the same time, the approximating function usually has such large fluctuation that numerical differentiation is ruled out. Tschebycheff polynomials have utility in the generation of approximations to theoretically error-free functions when it is desired to keep the approximation error within definite bounds. This, however, is not the same as the problem being considered here, namely: the fitting of a curve to experimental data subject to random error. The least-squares criterion of curve-fit is more fundamentally based on statistical principles and, therefore, much better suited for the present purpose.

II. Shortcomings of the *F*-Test of Significance as Criterion for Curve Selection

It is possible to consider conventional statistical tests of significance as criteria for the selection of a curve for representing data. Use of the F, or variance-ratio, test is the one commonly encountered in the literature, and usually in connection with the fitting of orthogonal polynomials.

It is noteworthy that although Sir Ronald A. Fisher,[4] the leader in the development of methods of statistical estimation, originated both the variance-ratio test and convenient methods of evaluating orthogonal polynomials, he refrained from using the former for selecting the degree of the latter.

Guest[2] and Snedecor[5] illustrate the sequential use of the F-test as orthogonal polynomials of successively higher degree are fitted to the same data. In particular, they consider the ratio of reduction in variance to residual variance for each increase of degree of polynomial, but do not call attention to the shortcomings of this procedure. Snedecor,[5] however, reveals his distrust of the method when in an illustrative example he discards the criterion's indication of a cubic in favor of a subjective preference for a fourth-degree polynomial.

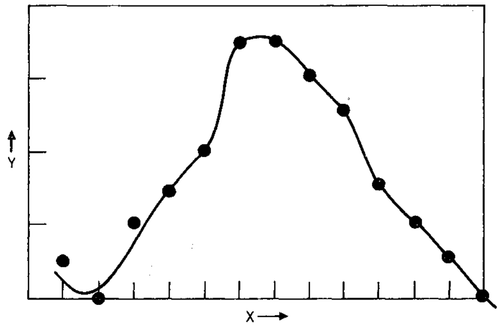

Fig. 2 Bias due to undersmoothing.

First of all, it is a necessary condition of tests of significance that the rejection-acceptance boundary be fixed in advance of testing. Ex post facto setting of significance levels for decision is a form of Monday-morning-quarterbacking which is not readily defensible.

Another difficulty arises if there is a gap in the series of polynomial terms expressing the mathematical law underlying the data. If, for instance, the term of power n is missing, it may be erroneously decided that $n - 1$ is the highest degree of importance whereas $n + 1, n + 2, \ldots n + i$, may also be important.

Furthermore, the distribution of the F-statistic was derived for normally-distributed random deviations. If a polynomial of degree n is fitted to data really depending on a polynomial of a higher degree m, the disparity between the two is far from being either normal or random.

Finally, even when the deviations are random and normal, available tabulations of the distribution of the F-statistic apply to one test only, not to a sequence of tests for the purpose of locating an extremal probability. Only very limited tables are available for the latter case[6] and these are inadequate for the present purpose.

III. Almost Indispensible Criterion

Although there are borderline cases, two major categories of curve-fitting problems can be named: cases in which the law of nature is known and those in which it is not.

As an assurance of predictability of future data, there is no substitute for knowledge of the laws of nature underlying an observed phenomenon.[7] Working from fundamentals, it is often possible to derive the general form of the equation to be taken as a mathematical model of exprimental observations.

All known effects involved in a phenomenon need not be explicably represented if only the over-all result is the one of interest. In approximate models, complex second-order effects can be merged advantageously with simpler major effects. For example, Newton's law of cooling is a good linear approximation of the effects of thermal conduction, radiation, and convection for moderate temperatures and ranges, although radiative heat transfer follows a fourth-power law and no simple general relation is available for convection.

A check of relevant literature can aid in establishing reasonable lower and upper limits on the functional complexity needed for representing the data. The minimum requirements for even a rough model is that it shall be able to provide the correct sign and order of magnitude of the dependent variable and its derivatives of interest over the range of interest of the independent variable and often, when the latter becomes zero and ± infinity.

Many mechanical systems operate under conditions of approximately constant acceleration; thus a polynomial of second degree in time may be adequate representation of displacement. A system with aperiodic acceleration may not need representation by a polynomial of much higher degree, since friction and deformation can severely limit response in a real system.

Data dependent on areas and volumes can obviously be represented by quadratic and cubic functions of linear dimensions. Recent experiences in the use of cubic spline curves for data representation indicates the frequent utility and adequacy of this class of simple functions.[8]

Because of their simplicity and versatility, polynomials are the representations of choice for mathematical models whenever possible.[9,10] Nevertheless, there are many instances where polynomial representations are so inappropriate as to be grossly inaccurate and misleading. If the data exhibit regular periodicities that are to be represented, a Fourier series may be more suitable than a polynomial. A polynomial having alternate positive and negative coefficients for successive terms of odd or even power is open to suspicion on two points. The data may be periodic and should be so represented, and the alternating signs can give unreliable evaluations if there are small differences of large quantities. If the data pertain to a growth process or a biological cause-effect relationship, an exponential or logarithmic representation is usually preferable to polynomial.[11-13] The same holds true for data arising from counting processes of any sort.[14]

The ultimate use that will be made of the fitted curve needs to be considered in the selection of the curve type. If the curve is to be integrated, reliability of the initial value may be an overriding requirement. Differentiation always demands that slopes be validly represented throughout the range of data. Frequent evaluation of the fitted curve, as for interpolation, may emphasize computational ease. For extrapolation, observed data should span an abscissa range at least as great, or preferably several times greater, than the extrapolated range and should have essentially uniform variance about the fitted curve throughout its range. The uniformity of variance is an indication that the ceteris paribus assumption may be realized in the extrapolated range provided that no discontinuities are encountered.

IV. Minimum Bias as a Criterion of Curve-Fit

Even if we are unwilling or unable to make use of the laws of nature governing the data, the need to assure predictive ability must still guide the choice of the curve.

It is not sufficient to minimize the effect of noise, or random fluctuations, in the data. An important requirement is to minimize the effect of bias, or systematic discrepancy between the true and the assumed curve types. This requirement exists even if the true curve type is unknown.

It is interesting to note that for the more difficult multivariable case, Box[15] and Highleyman[16] give criteria which essentially depend on minimum bias for enhancing predictability. Gadd and Wold[17] have a similar measure for the accuracy of forecasts.

The general approach taken here is to accept that type of curve for data representation for which residual deviations show least bias. The minimum-bias criterion, therefore, demands that deviations about the curve be as random as possible. Accordingly, as much as possible of the systematic variation of the data is accounted for by the deterministic relationship expressed by the equation of the curve.

Application of the minimum-bias criterion requires the testing of a number of candidate curve types so that the most suitable one may be selected. Although a closed-form solution is not available, various algorithms can be devised for carrying out the search for a curve of minimum bias. Among

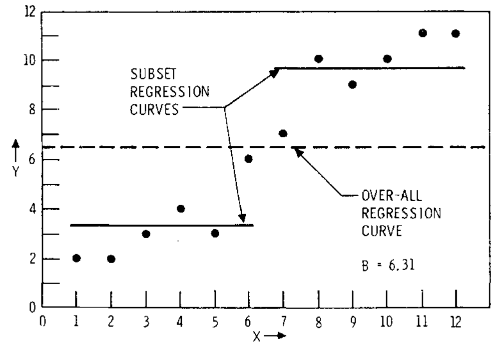

Fig. 4 Test of zero-degree polynomial.

equally valid algorithms, preference should rest on a computational strategy that can lead efficiently to a curve of an acceptably small bias in a particular data-processing situation.

At this stage of development of the minimum-bias criterion, it is intended only as a decision-making aid for identifying the most suitable curve without attempting to evaluate relative probabilities of suitability of the various curves considered.

When large volumes of data are handled in a highly automated manner, the sum-of-squares ratio form of the minimum-bias criterion is suitable.

As is customary in discussions of curve-fitting, only a unique sampling of a single-valued function of one uniformly and finitely changing variable is taken up explicitly. The comments and criteria presented can be extended to several variables, and continuous, irregular, or multiple sampling.

V. Sum-of-Squares Ratio Form of Minimum-Bias Criterion

To use the sum-of-squares ratio minimum-bias criterion, the data are divided into two subsets, each containing as close to one-half of the values of the independent variable as possible. The same type of curve is fitted three times: once to all of the data and then individually to each of the two subsets. The sums of squares of deviations resulting from these operations reveal the relative bias inherent in the use of the type of curve under test.

The rationale of this criterion is as follows. Residual deviation after curve-fitting are due to two components: systematic discrepancy, or bias, between the assumed and the real functional relations, and random fluctuation of data values about the real function. The effect of bias increases with the amount of data; thus, it is greater when a curve is fitted to all of the data than when a curve of the same form is fitted separately to the two subsets of data. Conversely, the effect of random fluctuations is independent of the amount of data.

Let $2N$ or $2N + 1$ be the total number of data points in the data space. Divide the data space into two non-overlapping subsets of N or $N + 1$ consecutive points. Let C be the number of constants to be evaluated for an equation under test; for example if the equation is a polynomial of degree R, $C = R + 1$. During the testing process, equations with progressively larger numbers of constants are investigated, but not more than N-1.

Fit a curve of given type to the data space. Let S' be the sum of squares of data deviations about the curve. Next fit the same given type of curve to each of the data subsets individually. Let S'' be the sum of squares of deviations of data in both subsets about their respective curves. Finally, take the ratio B of the over-all sum of squares of deviations to the within-subset sum of squares of deviations,

$$B = S'/S''$$

as the measure of bias in the representation of the data by the curve under test.

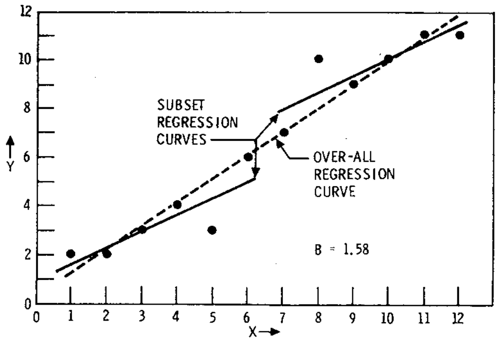

Fig. 5 Test of first-degree polynomial.

Even when random fluctuation is small, S' will be large compared to S'' for an unsuitable type of curve. Regardless of the amount of random fluctuation present in the data, S' will approach S'' for a curve of suitable type and the ratio S'/S'' will tend to the value of 1.

In the evaluation of B, two circumstances potentially capable of producing ambiguities must be provided for. In the first place, all data have finite precision. It is essential, therefore, to limit the number of significant figures in S', S'', and B in accordance with the precision of the original data.

Secondly, in the unlikely case of the perfect curve fit, $S' = S'' = 0$ and the value of S'/S'', is indeterminate. In order to circumvent this difficulty, add the same small quantity d to both S' and S''. Chose d to be of the order of $2Nq^2$, where q is the value of the least significant digit in the data.

The B parameter can be used in several ways for selecting a suitable type, for example: 1) of several curves, choose the one for which B is a minimum, 2) choose the first encountered curve for which B is less than k, where k is some preassigned number greater than 1, and 3) choose the simplest curve (one with fewest constants) for which B is less than k.

VI. Illustrative Example of the Use of the Sum-of-Squares Ratio Form of the Minimum-Bias Criterion

Figures 4–8 present graphically the operations constituting the application of the criterion of minimum bias to curve fitting. In each of the figures, the same data are plotted as dots. It is desired to test polynomials of successively higher degree in order to identify that degree for which the fit is most nearly unbiased over the range of the independent variable. Here $2N = 12$, therefore, polynomials of degree 0–4 can be tested.

In testing each degree, a polynomial of that degree is fitted to all twelve points; then, individual polynomials are fitted to the points in each of the two subsets. The polynomial fitted to all the data is plotted as a broken line and is called the over-all regression curve. The polynomials fitted separately to data within the subsets are plotted as solid lines and are called subset-regression curves.

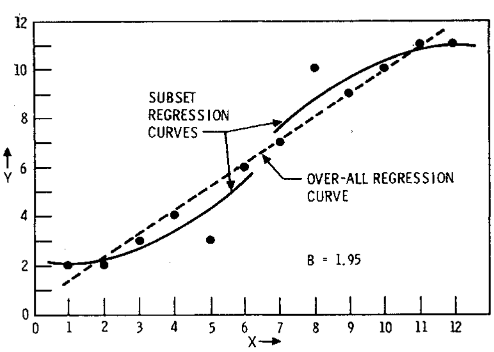

Fig. 6 Test of second-degree polynomial.

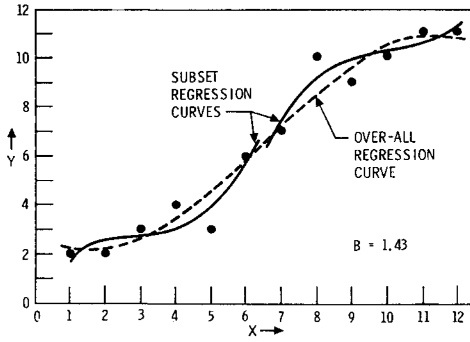

Fig. 7 Test of third-degree polynomial.

The data space chosen for the example exhibits perfect odd-order symmetry, that is, $f(\bar{x}) - f(\bar{x} - k) = f(\bar{x} + k) - f(\bar{x})$. Thus, polynomial over-all regression curves have no even-order terms above the zero order. This choice of data was made since it appears to clarify the example without restricting its validity.

The values of the B parameter obtained for the five polynomials tested are listed in Table 1.

The cubic equation exhibits a minimum sum-of-squares ratio, but one not much different from that of its nearest rival. Accordingly, the cubic is to be preferred for representing the data in this example, but the linear equation might be given consideration if simplicity is important.

VII. Conclusion

The minimum-bias criterion is free from the shortcomings of conventional tests of significance used as curve-selection criteria.

It has at least these six additional advantages.

1) The user is not required to select a level of significance. The arbitrary nature of this selection is an unsatisfactory aspect of most conventional statistical tests, and especially those associated with curve-fitting.

2) Comparisons can be made of the relative quality of fit of entirely different types of curves, for example, logarithmic versus polynomial.

3) Comparisons can be made of the relative suitability of different methods of curve-fitting, such as least-squares, Tschebycheff, and average.

4) Curve-fitting theory calls attention to the possibility of improving the quality of fit by assigning weights to data on the basis of local variance.[2] Nevertheless, data are rarely weighted for curve-fitting because of the added complication. Since variance is more nearly uniform for a curve identified by a minimum value of the B-ratio of the order of unity, the need for weighting of data is much decreased.

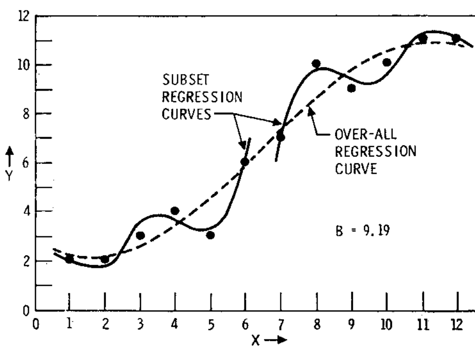

Fig. 8 Test of fourth-degree polynomial.

Table 1 Bias parameter for test cases

Degree of polynomial	B
0	6.31
1	1.58
2	1.95
3	1.43
4	9.19

5) If, as the number of constants in a given form of equation is increased, the B-ratio approaches a minimum of the order of unity slowly, investigation of a different form of curve may be indicated; for instance, exponential rather than polynomial.

6) If fine detail in a theoretically correct representation of data is masked by large random errors, the minimum-bias criterion properly accepts as preferable a simpler curve which is adequate for representing the low-grade data.

References

[1] Deming, W. E., *Statistical Adjustment of Data*, Dover, New York, 1964, pp. 128–147.

[2] Guest, P. G., *Numerical Methods of Curve Fitting*, Cambridge University Press, Cambridge, England, 1961, pp. 89, 257, 88.

[3] Aleksandrov, A. D., Kolmogorov, A. N., and Lavrent'ev, M. A., eds., *Mathematics—Its Content, Method, and Meaning*, Vol. 1, Pt. 4, American Mathematical Society, Providence, R.I., 1963, pp. 90–96.

[4] Fisher, R. A., *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh, Scotland, 1958, pp. 147–173.

[5] Snedecor, G. W., *Statistical Methods*, Iowa State College Press, Ames, Iowa, 1956, pp. 447–472.

[6] Pearson, E. S. and Hartley, H. O., eds., *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, Cambridge, England, 1962, p. 164.

[7] Whittle, P., *Prediction and Regulation*, Van Nostrand, Princeton, N. J., 1963, pp. 83–97.

[8] Ahlberg, J. H., Nilson, E. N., and Walsh, J. L., *Theory of Splines and Their Applications*, Academic Press, New York, 1967, pp. 50–52.

[9] Hildebrand, F. B., *Introduction to Numerical Analysis*, McGraw-Hill, New York, 1956, pp. 258–311.

[10] Householder, A. S., *Principles of Numerical Analysis*, McGraw-Hill, New York, 1953, pp. 185–225.

[11] Hexter, A., "Selective Advantages of the Sickel-Cell Trait," *Science*, Vol. 160, No. 3826, April 26, 1968, p. 436.

[12] Finney, D. J., *Probit Analysis*, Cambridge University Press, London, 1952, pp. 8–12.

[13] Finney, D. J., *Statistical Methods in Biological Assay*, Hafner, New York, 1964, pp. 39, 66.

[14] Acton, F. S., *Analysis of Straight-Line Data*, Dover, New York, 1966, p. 223.

[15] Box, G. E. P., "Fitting Empirical Data," *Annals of the New York Academy of Sciences*, Vol. 86, Art. 3, May 25, 1960, pp. 792–816.

[16] Highleyman, R. W., "The Design and Analysis of Pattern Recognition Experiments," *Bell System Technical Journal*, Vol. XLI, No. 2, March 1962, pp. 723–744.

[17] Gadd, A. and Wold, H., "The Janus Coefficient: A Measure for the Accuracy in Prediction" *Economic Model Building: The Causal Chain Approach*, edited by H. Wold, North-Holland, Amsterdam, 1964, pp. 229–235.