LETTERS TO THE EDITOR

Parameter estimation with error in the observables

Richard H. Luecke
Department of Chemical Engineering
University of Missouri
Columbia, Missouri 65201
Herbert I. Britt
Union Carbide Corporation
South Charleston, West Virginia 25300
(Received 6 June 1974; revised 24 July 1974)

In a recent paper,[1] Barker and Diana discuss the method of effective variances for fitting data when both dependent and independent variables have uncertainties. They point out the importance of the problem and the fact that real and significant errors in estimation can result from neglect of uncertainties in the independent variables.

The method of effective variances has received broad usage in estimation of thermodynamic parameters after being described by Hust and McCarty[2] in 1967. Another widely used method, for vector models, was presented by Deming[3] in 1943. Barker and Diana[1] show results, using a method that is a special case of Hust and McCarty's, that agree closely with those given by Deming[3] for quadratic data. If Deming had carried the iterative process to completion, his results would have been identical to those obtained by Barker and Diana, since Luecke, Britt, and Hall[4] have shown that the method of effective variances yields results identical to those obtained with the Deming method.

As pointed out by Barker and Diana, these methods are approximate. However, three publications in the last two years have described algorithms leading to the exact maximum likelihood solution to this problem.[5-7] The method of Britt and Luecke,[5] which is the most general and is an extension of Deming's method, applies to problems where observed variables are related to unknown parameters through implicit functions of the form:

$$F(z,p)=0 \tag{1}$$

where z is a q vector of observables and p is an n vector of parameters. For zero mean, normally distributed error, the maximum likelihood estimate for p is obtained by minimizing

$$Q(z)=(z_m-z)^T R^{-1}(z_m-z) \tag{2}$$

subject to the constraint of Eq. (1), where z_m are the measured values and R is the variance–covariance matrix of the measurement error. An iterative solution was presented for estimates for both the parameters and the measurements:

$$p_{i+1}=p_i-[F_p{}^T(F_zRF_z{}^T)^{-1}F_p]^{-1}F_p{}^T(F_zRF_z{}^T)F, \tag{3}$$

$$z_{i+1}=z_m-RF_z{}^T(F_zRF_z{}^T)^{-1} \times[F+F_p(p_{i+1}-p_i)+F_z(z_m-z_i)] \tag{4}$$

where F_p is the Jacobian matrix of F with respect to the parameters and where F_z is the Jacobian matrix of F with respect to the observables. F, F_p, and F_z are evaluated using the most recent estimates for z and p.

[1]D. R. Barker and L. M. Diana, Am. J. Phys. **42**, 224 (1974).
[2]J. G. Hust and R. D. McCarty, Cryogenics **7**, 200 (1967).
[3]W. E. Deming, *Statistical Adjustment of Data* (Wiley, New York, 1943).
[4]R. H. Luecke, H. I. Britt, and K. R. Hall, Cryogenics **14**, 284 (1973).
[5]H. I. Britt and R. H. Luecke, Technometrics **15**, 233 (1973).
[6]D. R. Powell and R. J. Macdonald, Comput. J. **15**, 148 (1972).
[7]A. Celmins, U.S.A. Ballistic Research Laboratory Report No. 1658, Aberdeen Proving Ground, MD, 1973.

Comment on "Simple method for fitting data when both variables have uncertainties"

J. Ross Macdonald*
Texas Instruments
Dallas, Texas 75222
(Received 4 April 1974; revised 16 August 1974)

Barker and Diana[1] (BD) have recently illustrated the approximate *effective variance* method of Clutton–Brock[2] and Hust and McCarty[3] for handling least-squares data fitting when all the variables involve measurement uncertainties. This general "errors-in-variables" situation is the common experimental one, yet it is usually anaylzed using ordinary least squares (OLS) procedures where uncertainties in independent-variable measurements are implicitly or explicitly taken to be negligible. Since such OLS analysis of a situation where significant random errors occur in measured values of all variables is known to lead to unnecessary errors of estimation, more publicity on improved alternatives, such as the effective variance method, is indeed welcome.[1] The aim of the present note is, therefore, to add additional information related to BD's results and discussion and to mention two other errors-in-variables analysis methods which converge to yield exact solutions of the least squares minimization equations (unlike the effective variance method) and are likely to be just as easy to use as that method.

Although BD cite the slowly convergent exact solution of the errors-in-variables problem for polynomial fitting given by O'Neill, Sinclair, and Smith,[4] two subsequent papers[5,6] were not cited which provide rapidly convergent exact least-squares solutions for arbitrary functions, linear or nonlinear in variables and parameters. As is usual in nonlinear least-squares problems, exactness assumes iterative convergence to the absolute minimum sum of squares. The method of Britt and Luecke[6] requires the user to provide both the function to be fitted and partial first derivatives, sometimes a messy task for complicated functions. By contrast, the Powell–Macdonald method[5] numerically approximates needed derivatives automatically and requires the user to provide only the function to be fitted and a derivative step size parameter or approximation interval. It is thus often easier to use, but, in its present form applies only to functions which can be written $y = f(x)$ or $x = g(y)$. Both methods converge very rapidly, much faster than that of O'Neill *et al.*, usually reaching converged values after only four or five cycles for most functions.

As BD mention, the effective variance method does not yield an exact least-squares solution. Its estimated parameter values thus do not lead to the *least* sum of squares. BD find in specific cases studied that the effective variance results are very nearly the same as those following from an earlier solution of Deming.[7] The small differences in parameter estimates obtained with the two methods for the example presented in Table I arise entirely from differences in round-off (L. M. Diana, private communication). It has, in fact, been shown by Luecke, Britt, and Hall[8] that the two different methods must yield identical results on convergence. But it has also been shown[5,6,8] that the Deming method does not lead to satisfaction of the least-squares minimization conditions. The Deming solution is, in fact, used to provide starting parameter estimates in the Powell–Macdonald method, which only ceases iteration when the least-squares conditions are very well satisfied.

Some of the results of various methods are illustrated in Table I, where data of Pearson[9] using the x and y weights of York[10] are fitted to the straight line function $y = \alpha_1 + \alpha_2 x$. The estimated values of α_1 and α_2 are given by a_1 and a_2 in the table. The first two rows of the table show the results of ordinary weighted least squares (OWLS) regressions. The values given in the first four rows come from Ref. 1 and those in the last two from Ref. 5. More significant figures are given in the a_1, a_2 values of Table I than are justified by the uncertainties of the estimates in order to allow these results to be used in the well-defined, purely mathematical problem of showing that these values yield the minimum sum of squares of residuals for the present data and weighting. The table also shows the weighted sum of squares,[5] S, for runs with weighting of both x and y residuals. Since the value of S is not given by either York or BD for York's method and specific example, it has been calculated here directly from York's results. No S value is given for the effective variance method, both because none was given by BD and because, as we have seen, the value must be the same, on convergence, as that obtained from Deming's method. The Deming results shown here are fully converged.

Table I. Comparison of various parameter estimates for Pearson–York data.

Method of calculation	Parameter estimates a_1	Parameter estimates $-a_2$	Sum of squares, S
OWLS: y on x (x data taken exact)	6.100	0.611	
OWLS: x on y (y data taken exact)	5.945	0.630	
York[10] "exact"	5.463 [a]	0.477	11.8702
Effective variance	5.396	0.463	
Deming[7]	5.3961	0.46345	11.9564
Powell–Macdonald[5] exact	5.4799	0.48053	11.86635

[a] York gives no figure here; that listed above follows from his results.

Although the BD analysis is, as stated,[1] an approximate one, BD compare their results with those of two treatments[4,10] which they term "exact." Since the published results[1,4,10] of these two methods differ when applied to the present data and weighting, the term "exact" evidently needs some clarification. Although the discrepancy was recognized by BD, they left it largely unresolved and termed only York's results exact in their tabular comparison of results. The explanation is that while York's[10] *method* is indeed exact for linear fits, his *calculations* were not carried to full convergence, leading to errors in his parameter estimates. The O'Neill *et al.*[4] algorithm is also exact for polynomials in that it leads to satisfaction of the least-squares conditions upon convergence. After full convergence, the O'Neill solution yields[4] values identical to those of Powell and Macdonald. These parameter estimates have been shown[5] to lead to the smallest sum of squares yet found for the present data and to excellent satisfaction of the least-squares minimization conditions. They are thus almost certainly *least*-squares estimates.

Had York, O'Neill *et al.*, and BD presented values of S, it would have been immediately clear that $S(\text{York}) > S(\text{O'Neill})$, and thus that York's numerical results were not associated with a *least*-squares solution. It thus appears that it is always safest to examine the value of S and to require an explicit numerical test of satisfaction of the least-squares minimum conditions[5] before concluding that a given set of numerical results represents the exact *least* squares solution. Although the Deming, effective-variance approaches are relatively simple, they are somewhat inaccurate. York's accurate method is only applicable for straight-line situations. The accurate Powell–Macdonald method is so simple to apply that it should usually be preferable whenever weighting of both variables is appropriate and the fitting function can be solved for one of the two variables. When such a solution is impossible, or there are more than two variables, the Britt–Luecke[6] method should be used.

A listing of the Powell–Macdonald program is available on request.

In conclusion, it is amusing and perhaps useful to speculate on the features a new, generalized least-squares program might usefully incorporate. The following is what the author would like to have available and see widely used. The program would be based on the solution of Britt and Luecke,[6] which allows functions nonlinear in both parameters and variables and of the form $f(x_i) = 0$ to be handled directly. Here the number of variables, $i = 1,2, \ldots, n$, which can be involved should be at least 10.

Simultaneous weighting of all variables would be possible. The partial derivatives needed in the program would be calculated exactly internally, using symbolic formula manipulation, as recently described by Wolberg and Isenberg.[11] To facilitate rapid convergence to the correct estimates, the program might incorporate a variable projection algorithm,[12] appropriate when some of the parameters in the model enter linearly, some nonlinearly. Finally, it might incorporate as an option the iterative procedure of Schlossmacher[13] which converts a least-*squares* solution into one with least sum of absolute deviations, more appropriate for certain error distributions. Any takers?

*Present address: Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC 27514.

[1]D. R. Barker and L. M. Diana, Am. J. Phys. **42,** 224 (1974).
[2]M. Clutton-Brock, Technometrics **9,** 261 (1967).
[3]J. G. Hust and R. G. McCarty, Cryogenics **6,** 200 (1967).
[4]M. O'Neill, I. G. Sinclair, and F. J. Smith, Comput. J. **12,** 52 (1969).
[5]D. R. Powell and J. R. Macdonald, Comput. J. **15,** 148 (1972); **16,** 51 (1973). In this work the expression for estimates of parameter variances is too small by a factor of two. See also J. R. Macdonald, J. Comp. Phys. **11,** 620 (1973); Nucl. Instrum. Methods **121,** 203 (1974).
[6]H. I. Britt and R. H. Luecke, Technometrics **15,** 233 (1973).
[7]W. E. Deming, *Statistical Adjustment of Data* (Wiley, New York, 1943; Dover, New York, 1964).
[8]R. H. Luecke, H. I. Britt, and K. R. Hall, Cryogenics **14,** 284 (1974).
[9]K. Pearson, Phil. Mag. **2,** 559 (1901).
[10]D. York, Can. J. Phys. **44,** 1079 (1966).
[11]J. R. Wolberg and J. Isenberg, Nucl. Instrum. Methods **112,** 533 (1973). See also, "An Efficient Application of FORMAC to the Nonlinear Least Squares Problem," J. R. Wolberg, J. Isenberg, M. Rafal, and S. Malvadker, Technion Computer-Applications Group Report TME-204, presented at Math Software Conference II, Purdue University, 29–31 May 1974.
[12]F. T. Krogh, Comm. ACM **17,** 167 (1974).
[13]E. J. Schlossmacher, J. Am. Stat. Assn. **68,** 857 (1973).

Comment on "Teaching and learning"

R. F. Cooper
Department of Mathematics and Science
Hong Kong Polytechnic
Hung Hom, Kowloon, Hong Kong
(Received 21 March 1974)

Although I agree with practically everything in Frank Oppenheimer's interesting article,[1] I am not sure that I agree that a lazy teacher is necessarily a bad teacher.

I was taught physics by a master who came into the class every Monday morning, when he set work for that week, collected the work from the previous week and returned the work from the week prior to that. The work he handed back had rarely been corrected in any way, but by collecting it he ensured we had carried out the past week's assignment. The only time we saw him again for the remainder of that week was if the class made any noise, in which case he would come from his adjoining room, frighten the wits out of us, and then return to continue reading his newspaper and drinking his tea, secure in the knowledge that we would work for the remainder of that week at least. His justification for "teaching" us in this manner was that he was training us to work on our own when we left school. For him the method succeeded in that we applied ourselves diligently to our studies and very few of his students failed their examinations.

Teaching is an art and each artist has his own individual way of presenting his material. A style which suits one artist will not necessarily suit another, so the qualities which might be present in one teacher to make him a good teacher might also be present in another to make him a bad one. I would be tempted to say that the only characteristic which definitely goes to make a bad teacher is lack of control of the class, although having said that, no doubt someone will be able to give an example which invalidates it.

[1]F. Oppenheimer, Am. J. Phys. **41,** 1310 (1973).

The world should love lovers, but not theoreticians. Never theoreticians! Show them the door. Ladies, throw out these gloomy bastards!
—Saul Bellow