EARTH AND PLANETARY SCIENCE LETTERS 5 (1969) 320-324. NORTH-HOLLAND PUBLISHING COMP., AMSTERDAM

LEAST SQUARES FITTING OF A STRAIGHT LINE WITH CORRELATED ERRORS

Derek YORK

Geophysics Division, Department of Physics, University of Toronto, Toronto, Canada

Received 23 December 1968

Earlier least squares treatments of the fitting of a straight line when both variables are subject to errors are generalized to allow for correlation of the x and y errors. The method is illustrated by reference to lead isochron fitting.

It has been shown that when both x and y coordinates are subject to errors the slope of the best straight line on a plot of y versus x is given by a root of the "least squares cubic" equation, assuming there is no correlation between the x and y errors (York [1]; McIntyre, Brooks, Compston and Turek [2]). The least squares cubic may be reduced algebraically to a quadratic equation (the "least squares quadratic") which may be solved in the usual fashion for quadratic equations. However, it frequently happens in practice that the error in the y-coordinate of a point is correlated with the error in the x-coordinate. In such circumstances the least squares cubic and quadratic equations should be used in generalized forms. It is the purpose of this communication to give such generalizations.

Two equivalent approaches may be adopted, corresponding to the two different starting points adopted in the uncorrelated-errors case by York [1] and McIntyre et al. [2]. Firstly one may begin by minimizing the expression

$$S = \sum_i \{\omega(X_i)(x_i - X_i)^2 - 2r_i\sqrt{\omega(X_i)\,\omega(Y_i)}\,(x_i - X_i)(y_i - Y_i) + \omega(Y_i)(y_i - Y_i)^2\}\,\frac{1}{(1 - r_i^2)} \qquad (1)$$

subject to the requirement

$$y_i = a + bx_i\,, \quad i = 1, ..., n\,.$$

X_i, Y_i are the observations, x_i, y_i are the adjusted values of these, $\omega(X_i)$, $\omega(Y_i)$ are the weights of the various observations, and the r_i are the correlations between the x and y errors. Alternatively one may start by minimizing the expression

$$S = \sum_i Z_i(Y_i - bX_i - a)^2\,, \qquad (2)$$

where

$$Z_i = \frac{\omega(X_i)\omega(Y_i)}{b^2\omega(Y_i) + \omega(X_i) - 2br_i\sqrt{\omega(X_i)\omega(Y_i)}}\,.$$

Pursuing the analysis we find the following generalized versions, for the case of correlated x and y errors, of the least squares cubic and quadratic equations, either of which may be solved for b to yield the best slope:

$$b^3 \sum_i \frac{Z_i^2 U_i^2}{\omega(X_i)} - b^2 [\, 2 \sum_i \frac{Z_i^2 U_i V_i}{\omega(X_i)} + \sum_i \frac{Z_i^2 r_i U_i^2}{\alpha_i} \,]$$

$$-b \left[\sum_i Z_i U_i^2 - 2 \sum_i \frac{Z_i^2 r_i U_i V_i}{\alpha_i} - \sum_i \frac{Z_i^2 V_i^2}{\omega(X_i)} \right] + \sum_i Z_i U_i V_i - \sum_i \frac{Z_i^2 r_i V_i^2}{\alpha_i} = 0 \ , \qquad (3)$$

and

$$b^2 \sum_i Z_i^2 \left[\frac{U_i V_i}{\omega(X_i)} - \frac{r_i U_i^2}{\alpha_i} \right] + b \sum_i Z_i^2 \left[\frac{U_i^2}{\omega(Y_i)} - \frac{V_i^2}{\omega(X_i)} \right] - \sum_i Z_i^2 \left[\frac{U_i V_i}{\omega(Y_i)} - \frac{r_i V_i^2}{\alpha_i} \right] = 0 \ , \qquad (4)$$

where

$$U_i = X_i - \bar{X} \ , \quad V_i = Y_i - \bar{Y} \ , \quad \bar{X} = \sum_i Z_i X_i / \sum_i Z_i \ , \quad \bar{Y} = \sum_i Z_i Y_i / \sum_i Z_i \ , \quad \alpha_i^2 = \omega(X_i)\omega(Y_i) \ .$$

The similarity between expression (3) and the uncorrelated least squares cubic equation is striking and the uncorrelated equations may be immediately obtained from eqs. (3) and (4) by setting $r_i = 0$. It might be noted that the best straight line goes through $(\bar{X}, \bar{Y})$ when $\bar{X}$ and $\bar{Y}$ are defined as above. The slope of the best straight line may now be written as a root of eq. (4) as

$$b = -\sum_i Z_i^2 \left[\frac{U_i^2}{\omega(Y_i)} - \frac{V_i^2}{\omega(X_i)} \right] + \left\{ \left(\sum_i Z_i^2 \left[\frac{U_i^2}{\omega(Y_i)} - \frac{V_i^2}{\omega(X_i)} \right] \right)^2 + 4 \sum_i Z_i^2 \left[\frac{U_i V_i}{\omega(X_i)} - \frac{r_i U_i^2}{\alpha_i} \right] \right.$$

$$\left. \times \sum_i Z_i^2 \left[\frac{U_i V_i}{\omega(Y_i)} - \frac{r_i V_i^2}{\alpha_i} \right] \right\}^{\frac{1}{2}} \Bigg/ \left\{ 2 \sum_i Z_i^2 \left[\frac{U_i V_i}{\omega(X_i)} - \frac{r_i U_i^2}{\alpha_i} \right] \right\}^{-1} \ , \qquad (5)$$

U_i, V_i and Z_i contain b, of course, so the solution is obtained by inserting an approximate value for b into these terms and calculating a new b from eq. (5). This new b is then inserted into the U_i, V_i and Z_i and a better value for b is recalculated from eq. (5), reiterating until the calculated best b no longer changes, to the degree of accuracy desired.

The correlated least squares quadratic may alternatively be solved by analogy with the method used by Williamson [3] for the uncorrelated least squares quadratic. In this case the best slope may be found from the equation

$$b = \frac{\sum_i Z_i^2 V_i \left[\frac{U_i}{\omega(Y_i)} + \frac{b V_i}{\omega(X_i)} - \frac{r_i V_i}{\alpha_i} \right]}{\sum_i Z_i^2 U_i \left[\frac{U_i}{\omega(Y_i)} + \frac{b V_i}{\omega(X_i)} - \frac{b r_i U_i}{\alpha_i} \right]} \ . \qquad (6)$$

Again an approximate slope is inserted where necessary on the right hand side of the equation and b is calculated. This new b is re-inserted on the right hand side and a better b found, and so on. Computer calculations show that eqs. (5) and (6) yield identical solutions for b, as they should. The best intercept, as usual, is found from the equation

$$a = \bar{Y} - b\bar{X} \ . \qquad (7)$$

The x and y residuals in the present case are given by

$$x_i - X_i = \frac{Z_i(a + bX_i - Y_i)(c_i - b\omega(Y_i))}{\omega(X_i)\omega(Y_i)} , \tag{8}$$

and .

$$y_i - Y_i = \frac{Z_i(a + bX_i - Y_i)(\omega(X_i) - bc_i)}{\omega(X_i)\omega(Y_i)} \tag{9}$$

where

$$c_i = r_i\alpha_i .$$

These expressions reduce to those given in York [1] when r_i is set equal to zero.

A simple pictorial illustration of how the above methods work is now given. We will take as an example the classic method of calculating the age of the meteorites, and by inference the Earth. The $^{207}Pb/^{204}Pb$ and $^{206}Pb/^{204}Pb$ isotope ratios are measured in stone and iron meteorites and these data are plotted as y versus x. The data define a linear array and the "age of meteorite formation" is calculated (under several assumptions) from the slope of the best straight line through the data. Consider one such ($^{207}Pb/^{204}Pb$, $^{206}Pb/^{204}Pb$) point, as shown in fig. 1. The actual line on which the meteorite points would fall if there were no errors of measurement is also drawn in the figure. The actual observational point we are considering does not of course fall on this line because of experimental error. Now in the mass spectrometric analysis in which the $^{207}Pb/^{204}Pb$ and $^{206}Pb/^{204}Pb$ ratios were determined, the ^{204}Pb ion beam would be an order of magnitude smaller than the ^{207}Pb and ^{206}Pb beams. The ^{204}Pb beam would accordingly be measured with far less precision than either the ^{207}Pb or ^{206}Pb beams. The errors in the $^{207}Pb/^{204}Pb$ and $^{206}Pb/^{204}Pb$ ratios are therefore due almost entirely to the ^{204}Pb error. If the same ^{204}Pb measurement is used in calculating both the $^{207}Pb/^{204}Pb$ and $^{206}Pb/^{204}Pb$ ratios, then the x and y errors of the point will be very highly correlated. If the ^{207}Pb and ^{206}Pb beams are essentially free from error of measurement and all the error in the ratios is due to ^{204}Pb error then the x and y errors will be perfectly correlated and we would have $r_i = 1$, and we will assume this to be the case in our illustration. Under these assumptions, let us suppose that the $^{206}Pb/^{204}Pb$ value of the point shown in fig. 1 is in error by an amount α due to ^{204}Pb error. Then the $^{207}Pb/^{204}Pb$ value will be in error by the amount $(^{207}Pb/^{206}Pb) \times \alpha$ because of the same ^{204}Pb error. From fig. 1 it is therefore immediately apparent that the observation point is shifted from the true value along a line whose slope is equal to the value of the ratio $^{207}Pb/^{206}Pb$. This error line is known to lead isotope workers as the "204 error line". Our method of least squares fitting should therefore adopt a line of adjustment whose slope equals that of the "204 error line": i.e., each data point should be adjusted according to our least squares method along a line of slope $(^{207}Pb/^{206}Pb)_i$, the value of this ratio for the ith point. That the method described in the earlier part of this communication does in fact achieve this may now be readily shown. In fig. 2 "ℓ" represents the best straight line as found by the method of this paper; (X_i, Y_i) is an observed point and (x_i, y_i) is the adjusted value: "L" is the line of adjustment joining (X_i, Y_i) and (x_i, y_i) whose slope we wish to show has the value $(^{207}Pb/^{206}Pb)_i$. By simple geometry

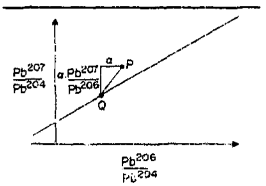

Fig. 1. P is the observed point, Q is the true point.

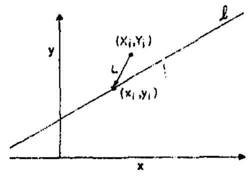

Fig. 2. (X_i, Y_i) is the observed point. (x_i, y_i) is the adjusted point on the best straight line "ℓ".

$$\text{slope of } L = \frac{Y_i - y_i}{X_i - x_i} = \frac{y\text{-residual}}{x\text{-residual}} = \frac{\omega(X_i) - bc_i}{c_i - b\omega(Y_i)} , \tag{10}$$

from eqs. (8) and (9). In our particular example the x and y errors of any one point are perfectly correlated, so $r_i = 1$ and $c_i = \sqrt{\omega(X_i)\,\omega(Y_i)}$ and

$$\text{slope } L = \frac{\omega(X_i) - b\sqrt{\omega(X_i)\omega(Y_i)}}{\sqrt{\omega(X_i)\omega(Y_i)} - b\omega(Y_i)} = \sqrt{\frac{\omega(X_i)}{\omega(Y_i)}} .$$

But, as we have already seen,

$$\frac{\text{error in } ^{207}\text{Pb}/^{204}\text{Pb}}{\text{error in } ^{206}\text{Pb}/^{204}\text{Pb}} = {}^{207}\text{Pb}/^{206}\text{Pb} ,$$

and since the weights are inversely proportional to the squares of the errors we have

$$\sqrt{\frac{\omega(X_i)}{\omega(Y_i)}} = {}^{207}\text{Pb}/^{206}\text{Pb} .$$

$\therefore$ the slope of $L = {}^{207}\text{Pb}/^{206}\text{Pb}$.

Thus the method of fitting given in this paper causes the points to be adjusted along lines of slope $({}^{207}\text{Pb}/^{206}\text{Pb})_i$, which is what our earlier considerations led us to require.

Eq. (10) is the generalized version of eq. (2) in York [4] which was given by Deming [5].

The standard errors in the slope (σ_b) and intercept (σ_a) are found by the usual method of partial differentiation. Reasonable approximate values for these quantities are given by the expressions

$$\sigma_b^2 = 1/\sum_i Z_i U_i^2 , \tag{11}$$

$$\sigma_a^2 = \sigma_b^2/\sum_i Z_i . \tag{12}$$

The exact expressions are long and are given in the appendix.

The minimized quantity S has a χ^2 distribution and the goodness of fit of the points to the line is found by reference to χ^2 tables. On average S should be about $n-2$, where n is the number of points plotted, if the points do fit the line. If it is desired to incorporate in the error estimates some measure of the degree of scatter of the points about the best line, the values for σ_a and σ_b calculated from expressions (11) and (12) may be multiplied by $\{S/(n-2)\}^{\frac{1}{2}}$. More elaborate analyses of variance along the lines of McIntyre et al. [2] may be adopted if desired.

Recently Brooks, Wendt and Harre [6] have given a method for least squares fitting of a straight line and have applied it to the fitting of Rb-Sr isochrons and suggested it is a suitable approach to fitting lead isochrons. Examination of the quantity minimized by these investigators shows that they have minimized the quantity we have called S in eqs. (1) and (2), if r_i is set equal to -1. Their method, therefore, is the correct one to adopt when r_i is in fact equal to -1, i.e., when the x and y errors at each point are in perfect negative correlation. This is, however, not the case in Rb-Sr and Pb isochron plotting.

Numerical considerations will be given elsewhere. The computer programme which carries out fitting according to the above treatment is available on request.

Appendix

The exact standard error in the slope, σ_b, is calculated in the usual way from the expression

$$\sigma_b^2 = \left\{ \sum_i \left[\left(\frac{\partial \phi}{\partial X_i} \right)^2 \frac{1}{\omega(X_i)} + \left(\frac{\partial \phi}{\partial Y_i} \right)^2 \frac{1}{\omega(Y_i)} + \frac{2r_i}{\sqrt{\omega(X_i)\omega(Y_i)}} \left(\frac{\partial \phi}{\partial X_i} \right) \left(\frac{\partial \phi}{\partial Y_i} \right) \right] \right\} \Big/ \left(\frac{\partial \phi}{\partial b} \right)^2 ,$$

where ϕ represents the left hand side of eq. (4).

$$\frac{\partial \phi}{\partial X_i} = \sum_j Z_j^2 \left[\delta_{ij} - \frac{Z_i}{\Sigma Z} \right] \left[\frac{b^2 V_j}{\omega(X_j)} - \frac{2b^2 r_j U_j}{\alpha_j} + \frac{2bU_j}{\omega(Y_j)} - \frac{V_j}{\omega(Y_j)} \right] ,$$

$$\frac{\partial \phi}{\partial Y_i} = \sum_j Z_j^2 \left[\delta_{ij} - \frac{Z_i}{\Sigma Z} \right] \left[\frac{b^2 U_j}{\omega(X_j)} - \frac{2bV_j}{\omega(X_j)} - \frac{U_j}{\omega(Y_j)} + \frac{2r_j V_j}{\alpha_j} \right] ,$$

$$\frac{\partial \phi}{\partial b} = \sum_i Z_i^2 \left[U_i^2 \left(\frac{1}{\omega(Y_i)} - \frac{2br_i}{\alpha_i} \right) + \frac{V_i}{\omega(X_i)} (2bU_i - V_i) \right] + b^2 \left(\frac{\partial A}{\partial b} \right) + b \left(\frac{\partial B}{\partial b} \right) - \left(\frac{\partial C}{\partial b} \right) ,$$

where A, B and C are the coefficients of the correlated least squares quadratic (i.e. eq. (4)) read from left to right. δ_{ij} is the Kronecker delta.

The exact standard error in the intercept, σ_a, is calculated from the expression

$$\sigma_a^2 = \sum_i \left\{ \left(\frac{\partial a}{\partial X_i} \right)^2 \frac{1}{\omega(X_i)} + \left(\frac{\partial a}{\partial Y_i} \right)^2 \frac{1}{\omega(Y_i)} + \frac{2r_i}{\sqrt{\omega(X_i)\omega(Y_i)}} \left(\frac{\partial a}{\partial X_i} \right) \left(\frac{\partial a}{\partial Y_i} \right) \right\} ,$$

where

$$\frac{\partial a}{\partial X_i} = \left\{ - \frac{bZ_i}{\sum_i Z_i} + \left(\frac{2}{\sum_i Z_i} \sum_i \frac{Z_i^2}{\alpha_i^2} \left[c_i - b\omega(Y_i) \right] \left[V_i - bU_i \right] - \bar{X} \right) \left(- \frac{(\partial \phi / \partial X_i)}{(\partial \phi / \partial b)} \right) \right\} ,$$

and

$$\frac{\partial a}{\partial Y_i} = \left\{ \frac{Z_i}{\sum_i Z_i} + \left(\frac{2}{\sum_i Z_i} \sum_i \frac{Z_i^2}{\alpha_i^2} \left[c_i - b\omega(Y_i) \right] \left[V_i - bU_i \right] - \bar{X} \right) \left(- \frac{(\partial \phi / \partial Y_i)}{(\partial \phi / \partial b)} \right) \right\} .$$

Acknowledgements

The author is indebted to Mr. R.J. Doyle for providing considerable assistance with the numerical computations involved in checking the various expressions. Useful discussions were enjoyed with Drs. W.Compston, G.A. McIntyre, J.Richards, P.Arriens and P.H.Reynolds in the Geophysics Department at the Australian National University, Canberra. The work reported in this paper was carried out in Canberra while the author was a guest on Sabbatical Leave from the University of Toronto as a National Research Council of Canada Senior Research Fellow. This research was supported by a grant from the National Research Council of Canada.

References

[1] D.York, Least-squares fitting of a straight line, Can. J. Phys. 44 (1966) 1079.
[2] G.A.McIntyre, C.Brooks, W.Compston and A.Turek, The statistical assessment of Rb-Sr isochrons, J. Geophys. Res. 71 (1966) 5459.
[3] J.H.Williamson, Least-squares fitting of a straight line, Can. J. Phys. 46 (1968) 1845.
[4] D.York, The best isochron, Earth Planet. Sci. Letters 2 (1967) 479.
[5] W.E.Deming, Statistical adjustment of data (Wiley, New York, 1943).
[6] C.Brooks, I.Wendt and W.Harre, A two-error regression treatment and its application to Rb-Sr and initial Sr^{87}/Sr^{86} ratios of younger Variscan granitic rocks from the Schwarzwald Massif, Southwest Germany, J. Geophys. Res. 73 (1968) 6071.