THE UNIVERSITY OF HERTFORDSHIRE

BUSINESS SCHOOL

WORKING PAPER SERIES

The Working Paper series is intended for rapid dissemination of research results, work-in-progress, innovative teaching methods, etc., at the pre-publication stage. Comments are welcomed and should be addressed to the individual author(s). It should be remembered that papers in this series are often provisional and comments and/or citation should take account of this.

For further information about this, and other papers in the series, please contact:

University of Hertfordshire Business School

College Lane

Hatfield

Hertfordshire

AL10 9AB

United Kingdom

NEUTRAL DATA FITTING IN TWO AND THREE DIMENSIONS

Abstract

We consider fitting a line (or plane) to data on two (or three) variables for the purpose of representing scientific or law-like relationships.
If these variables are related in a manner in which no variable plays a special role (i.e. no distinction between dependent and independent variables), then it is appropriate that the fitting procedure should treat all variables in the same way. Starting from obviously desirable properties for the 'error' measure we deduce the form of this error measure and show it to be unique.
We derive fitting procedures that aggregate the proposed measure over all points using the sum of squares. For the case of three variables our procedure is new and we analyse completely the solution to the resulting minimisation problem.

Keywords: data fitting, functional relation, line fitting, plane fitting, functional regression.

1. Introduction

Given data on two or three variables we are interested in fitting a line or plane for the purposes of modelling the relationship between these variables. Most of the literature on this subject approaches this problem by selecting one of these variables and treating it differently from the others in the fitting procedure. This is acceptable if the purpose is to make predictions of that variable, but if we are seeking an underlying scientific law or a law-like relationship then it would seem more reasonable to treat all variables in the same way, unless there are particular reasons for not doing so. A scientific law is a unique relation between variables: consider the relation between energy and mass in $E = mc^2$, the regression of E on m will give a different relationship from the regression of m on E: one will give a slope that is too low and the other gives a slope that is too high. Hence we have a different law according to which variable is selected as being dependent. Thus in this paper we consider procedures for fitting lines or planes which treat all the variables equally.

With a plethora of fitting procedures in the literature, some means is required of deciding which to choose. One fitting procedure that might be considered is orthogonal regression (also known as total least squares): here the sum of squared perpendicular distances to the line or plane is minimised. If the variables are not measured in the same units then this perpendicular distance is not a meaningful quantity since it involves adding together quantities measured in different units (incommensurability). One can transform the data in various ways to make it dimensionless, but the resulting relationship will depend on which particular transformation is chosen.

Our approach to choosing a method is to stipulate certain desirable properties which one would expect a fitting procedure to possess, and then prove that there is a unique procedure which satisfies these properties.

An early attempt at laying down desirable properties for fitting a line to data on two variables is due to Nobel laureate Paul Samuelson (1942). He suggested the following:

I. For perfectly correlated data the fit should reduce to the correct equation.
II. The fitted equation should be invariant under an interchange of variables. By this it is meant that the co-ordinate axes can be reversed.
III. The fitted equation should be invariant under any orthogonal transformation of variables i.e. those transformations that preserve length and angle.
IV. The equation should be invariant under a scale change in any of the variables.
V. The equation should be invariant under any linear transformation of individual co-ordinates.
VI. The slope must depend only upon the correlation coefficient and ratio of standard deviations.

We note that property II is precisely the condition we require, as discussed in the first paragraph. We also agree that properties I, IV and V are desirable. However we feel that property III is too restrictive. To see why, we note in particular, that a rotation by 45 degrees would change a perfect correlation situation $y = x$ to a situation where the data lie on the x-axis, so that there is no relation whatever between the variables. While condition VI is satisfied by traditional least squares regression, and also by the procedure which satisfies the properties we shall stipulate, we do not find it clear why it should be assumed a priori.

A different set of criteria was suggested by the noted statistician William Kruskal (1953). These apply to fitting a line (not a plane) to data in any number of dimensions and are as follows:

Criterion 1. Dependence on first and second moments only (i.e. on the mean, variance and covariance).

Criterion 2. Proper behaviour (i.e. invariance) under translation.

Criterion 3. Proper behaviour under change of scale.

Criterion 4. Proper direction of association. By this it is meant that the sign of the correlation between any two variables should be reflected in the signs of the estimated coefficients in the equation.

Criterion 5. Consistency under omission of variables i.e. if some variables are removed from the analysis then the coefficients of the remaining variables are unchanged. In other words we have a projection onto the lower dimensional space.

We note that property III, which we felt was too restrictive, has now been dropped. However property II, which we particularly desire has also been dropped. Criterion 4 is new and criterion 5 is particularly related to fitting a line (not a plane) in more than two dimensions.

To consider a line which `best fits' given data suggests that the line minimises some measure of error between the data and the line. Indeed in all procedures for fitting a model to data, some objective function is optimised.
For this reason we differ from the approaches cited above by considering desirable properties for the error measure itself. In Section 2 we stipulate and motivate seven such properties and show that they define the measure of error uniquely, up to a scaling factor. In fact the error is given by the square root of the area of the triangle bounded by the line and the horizontal and vertical lines through the data point. Our procedure for fitting a line to data is then to choose the line which minimises the sum of the squares of these errors i.e. the line minimises the sum of the areas of these triangles.

The resulting method is the same as that considered by Samuelson (1942) and Kruskal (1953) in connection with their conditions discussed above. Earlier Stromberg (1940) had employed it in a paper in the astronomical literature, and it was subsequently referred to as `Stromberg's impartial line'. In 1941 Woolley introduced it as `the method of minimised areas'. Samuelson (1942) then commented that ‘this is nothing other than Frisch's diagonal regression’ and gave a reference dating back to 1934. However the earliest of all references appears to have been found by Ricker (1975) who cites a German paper on meteorology by Sverdrup (1916). Because of the variation of names and accreditations for this method, and because it gives neither variable special treatment, we shall refer to it as ‘neutral data fitting’.

In Section 3 we extend neutral data fitting to the case of fitting a plane to data in three dimensions. Analogous properties for the error between a data point and a plane lead to a corresponding definition for the error. Then our choice of plane to fit the data is that which minimises the sum of the squares of the errors of the data. For neutral data fitting in two dimensions there are certain exceptional cases where there are two lines which best fit the data, and similarly for three dimensions there are exceptional cases where there are two, three or four planes of best fit. We show that in all other cases the plane of best fit is unique and is determined by the unique solution in a given interval of a quartic equation. In general this root will need to be determined numerically but we show that for certain special classes of data there are closed form solutions.

To finish this section, we discuss some situations where we suggest that neutral regression could be applied.

Some applications

Strictly speaking, the term 'regression' applies when we want to predict one variable given values of one or more explanatory variables. There is therefore a clear distinction between the predicted variable and the predictors. In estimating a regression equation these variables are treated differently. By contrast, our interest is in treating the variables symmetrically in deriving some intrinsic relationship; such relationships are sometimes called functional or structural equations with the latter term often being reserved for cases where we assume knowledge of the probability distributions involved. The term 'law-like relationship' has been suggested to embrace all of these cases.

Let us take a look at some disciplines where this work is appropriate. Deeming (1968) states that 'the astronomer is always interested in the physical relationship between, say x and y in $y = ax + b$. His interest in this relationship is usually of a physical rather than a statistical nature - he wants the "best" values of a and b. Only rarely will he need the special statistical properties of the regression lines'. He points out that the scatter about such a line will arise from two sources: observational or measurement error, and intrinsic random variation (as in differences between individuals) where a number of cases of the same x-value can possess different y-values, or vice-versa. One illustration is the relationship between the apparent magnitude (brightness) of galaxies and their radial velocities away from us arising from the gradual expansion of the universe. There may be a cluster of galaxies of the same type and brightness, and yet because of local motions their individual radial velocities will differ. The need for symmetric treatment of data in the physical sciences is very apparent because here we are dealing with scientific laws. Conventional least squares regression would provide two different laws depending on which of the two variables is selected as being the explanatory variable; this is clearly not satisfactory.

Biometrics seems to be the area where the symmetrical treatment of variables is most well known. Here one is dealing with measurements of different parts of a plant or animal; these may be lengths, weights, or body proportions for representing shape etc. The textbook by Sokal and Rohlf (1995) provides a thorough discussion of what they call 'model II regression lines'. (It is a notable feature of this topic that different disciplines give different names to the same approach.) They provide a set of rules for choosing a fitting method depending on what information is known about the errors, together with a worked example dealing with fish weight and number of eggs produced. In fact these methods have been particularly studied by fisheries researchers (see Ricker (1984) and references therein). For Ricker the need for symmetric treatment of variables is apparent as 'it is just as true to say that a fish is long in relation to the fin as that the fin is short relative to the fish – operationally there is no difference. In the same way, if a fish is "heavy for its length", it makes no sense to argue whether it is obese or whether it is dwarfed'.

Psychometric tests (e.g. intelligence quotient or personality tests) are widely used in the selection of staff by organisations. School and college examinations are similarly used. Given two tests that are meant to measure the same thing (possibly at different times), we wish to match equivalent scores on the two tests. Otis saw the need for this as long ago as 1922 and discusses generating a symmetric correspondence for the construction of a conversion table between the two test scores. In recent years it has become an annual feature of the British summer that school results will be announced indicating improvements over previous years. There then follows heated debate as to whether this is due to harder working pupils and teachers, or whether it arises from easier examinations. One would have thought it a simple enough task to pilot test two exams prior to their publication by getting the same set of people to take both and then fit a symmetric correspondence relationship to the results. The same could be done to ensure uniformity of standards in examination papers set by different examination boards aimed at the same level.

2. Neutral Data Fitting in Two Dimensions

Suppose two real variables x and y are connected through a relationship which is symmetric in the sense that y does not depend on x any more or less than x depends on y. We are given a sequence of data (x, y) = $\{\alpha_i, \beta_i\}$ in ∇^2 , i = l, … n, n ≥ 2, and wish to find a straight line which 'best fits' the data. The usual procedure is to define some measure of the 'error' between a point (α, β) and a line L, and then choose the line L which minimises some 'aggregate' of these errors over the points (α_i, β_i,), i = 1, …, n. The measure of the error between (α, β) and L is some non-negative number which we denote by F(α,β, L). The usual choice of aggregate is the sum of the squares of the errors, i.e. we minimise $\sum_{i=1}^{n} F(\alpha_i, \beta_i, L)^2$ over all lines L. Of course we could choose other aggregates, e.g. $\sum_{i=1}^{n} F(\alpha_i, \beta_i, L)^p$ for some p, 1 ≤ p < ∞, or $\min \max_{i=1\ldots,n} F(\alpha_i, \beta_i, L)$, but our first concern here is not with this but with the choice of error function F(α, β, L).

Any line L has an equation of the form ax + by + c = 0, for real numbers a, b, c. Since we are assuming that there is some relationship between x and y, we do not consider lines which are parallel to the x- or y- axes. So we may denote the error between a point (α, β) and a line with equation ax + by + c = 0 by F(α, β, a, b, c), for α, β, a, b, c in ∇ with a, b ≠ 0. We shall consider various properties which we would reasonably expect such a function F to satisfy and we shall prove that these properties determine

F up to a positive constant multiple. (The formulae below hold for all α, β, a, b, c in ∇ with a, b $\neq$ 0.)

Property 1 For any number $\lambda \neq 0$, the equation $\lambda ax + \lambda by + \lambda c = 0$ gives the same line as the equation $ax + by + c = 0$. So we must have

(1) $F(\alpha, \beta, \lambda a, \lambda b, \lambda c) = F(\alpha, \beta, a, b, c), \lambda \neq 0$.

Property 2 Clearly the error should be zero if and only if (α,β) lies on L, i.e.

(2) $F(\alpha, \beta, a, b, c) = 0$ if and only if $a\alpha + b\beta + c = 0$.

Property 3 We would not expect the error to depend on the choice of origin of co-ordinates, i.e. if we shift both the point and line by the same vector, then the error should be unchanged. If the vector is (u, v), then the point (α, β) is shifted to $(\alpha + u, \beta + v)$ and the line $ax + by + c = 0$ is shifted to $a(x - u) + b(y - v) + c = 0$. Thus we have

(3) $F(\alpha + u, \beta + v, a, b, c - au - bv) = F(\alpha, \beta, a, b, c), \quad u, v \in \nabla$.

Property 4 A crucial assumption is that we treat x and y equally. Thus the error should be unchanged if we interchange x and y, i.e.

(4) $F(\beta, \alpha, b, a, c) = F(\alpha, \beta, a, b, c)$.

Property 5 We would expect the error to be unchanged under a reflection of the x-variable, i.e. a reflection in the y-axis. Thus we have

(5) $F(-\alpha, \beta, -a, b, c) = F(\alpha, \beta, a, b, c)$.

Of course, (4) and (5) imply that the same holds for a reflection of the y-variable.

Property 6 If we scale x and y by a factor $\lambda > 0$, it is reasonable to have a corresponding change for the error, i.e.

(6) $F(\lambda\alpha, \lambda\beta, a, b, \lambda c) = \lambda F(\alpha, \beta, a, b, c), \ \lambda > 0$.

Property 7 In Property 6 we considered scaling both x and y. We now consider a change of scale in an individual variable, say x. It would seem reasonable that the scaling of the error is independent of the choice of point and line, which is equivalent to the optimal line being always preserved by a change of scale in the x-variable.

Thus we have

(7) $F(\lambda\alpha, \beta, a, \lambda b, \lambda c) = f(\lambda)\, F(\alpha, \beta, a, b, c), \quad \lambda > 0$

for some function $f : (0, \infty) \to (0, \infty)$. Of course a similar result follows if we consider a scaling of y.

<u>Theorem 1</u> If F is a function from $\{(\alpha, \beta, a, b, c) \in \nabla^5, ab \neq 0\}$ to $[0, \infty)$ satisfying (1) – (7), then for some k > 0 this error function will take the form

$$F(\alpha,\beta,a,b,c) = k\frac{|a\alpha + b\beta + c|}{|ab|^{1/2}}, \qquad \alpha,\beta,a,b,c \in \nabla.$$

<u>Proof</u> By (7), (4) and (1) we have for α, β, a, b, c in ∇, $ab \neq 0$

$$\begin{aligned} F(\lambda\alpha,\lambda\beta,a,b,\lambda c) &= f(\lambda)F\left(\alpha,\lambda\beta,a,\frac{b}{\lambda},c\right) \\ &= f(\lambda)F\left(\lambda\beta,\alpha,\frac{b}{\lambda},a,c\right) \\ &= f(\lambda)^2 F\left(\beta,\alpha,\frac{b}{\lambda},\frac{a}{\lambda},\frac{c}{\lambda}\right) \\ &= f(\lambda)^2 F(\alpha,\beta,a,b,c) \end{aligned}$$

and so by (6),

(8) $f(\lambda) = \sqrt{\lambda}, \lambda > 0$

By (1) and (3), for α, β, a, b, c in ∇, $ab \neq 0$ we have

(9) $F(\alpha,\beta,a,b,c) = F\left(\alpha,\beta + \frac{c}{b},\frac{a}{b},1,0\right)$.

By (2) let us define G using:

(10) $F(\alpha,\beta,a,1,0) = \frac{|a\alpha + \beta|}{|a|^{1/2}}G(\alpha,\beta,a),$

for a function G defined for a, α, β in ∇, $a \neq 0$.

By (6) we have

(11) $G(\lambda\alpha, \lambda\beta, a) = G(\alpha, \beta, a),$

and by (1), (7) and (8),

(12) $G\left(\lambda\alpha,\beta,\frac{a}{\lambda}\right)=G(\alpha,\beta,a)$,

where (11) and (12) hold for $\lambda > 0$, α, β, $a \in \nabla$, $a \neq 0$. Applying (11) and (12) gives

(13) $G(\alpha,\beta,a)=G\left(\frac{\alpha}{\beta},1,a\right)=G\left(1,1,\frac{a\alpha}{\beta}\right)$

for α, β, > 0, $a \neq 0$.

Now by (5),

(14) $G(-\alpha, \beta, -a) = G(\alpha, \beta, a)$,

and by (4)

(15) $G(\alpha, -\beta, -a) = G(\alpha, \beta, a)$,

where (14) and (15) hold for α, β, a in ∇, $a \neq 0$. Thus (13) holds for any $\alpha, \beta \neq 0$.

For convenience, we write

$$g(t)=G(1,1,t),\ t\neq 0\ ,$$

so that by (9), (10) and (13),

(16) $F(\alpha,\beta,a,b,c)=\frac{|a\alpha+b\beta+c|}{|ab|^{1/2}}\ g\left(\frac{a\alpha}{b\beta+c}\right)$

for α, β, a, b, c in ∇, $ab \neq 0$, $\alpha \neq 0$, $b\beta + c \neq 0$.

Applying (3) also gives for $u \in \nabla$,

(17) $F(\alpha,\beta,a,b,c)=\frac{|a\alpha+b\beta+c|}{|ab|^{1/2}}\ g\left(\frac{a\alpha+au}{b\beta+c-au}\right)$

provided $\alpha + u \neq 0$, $b\beta + c \neq au$. Putting $\alpha = t$, $a = b = \beta = 1$, $c = 0$ in (16) and (17) gives

$$g(t)=g\left(\frac{t+u}{1-u}\right)$$

provided $t \neq 0$, $t + u \neq 0$, $u \neq 1$. In particular, $g(1)=g\left(\frac{1+u}{1-u}\right)$,

$u \neq \pm 1$ and so $g(t) = g(1)$ for all $t \neq 0, -1$.

Putting $g(1) = k$, we see from (2) that $k > 0$. Now for any α, β, a, b, c in ∇, $ab \neq 0$, choosing u with $u + \alpha \neq 0$, $au \neq b\beta + c$ and substituting into (17) gives the result. □

We now return to the problem of finding a line with equation ax + by + c = 0 which best fits the data (x, y) = $\{\alpha_i, \beta_i,\}$, i = 1, … n, $n \geq 2$. We shall take the aggregate of the errors to be the sum of squares, and hence must find a, b, c in ∇, $ab \neq 0$, which minimises

$$f(a,b,c) = \sum_{j=1}^{n} \frac{(a\alpha_j + b\beta_j + c)^2}{|ab|} .$$

(Clearly the constant k in Theorem 1 is irrelevant.) We shall be generalising this approach in Section 3.

First suppose ab > 0. Then there is no loss of generality in supposing a > 0, b > 0, ab = 1 (since we can always transform one of the variables by multiplying by −1 to ensure the coefficients are positive, and we can divide through the equation by a constant such as $(ab)^{1/2}$ to ensure ab = 1.

The problem becomes to minimise

$$f(a,b,c) = \sum_{j=1}^{n} (a\alpha_j + b\beta_j + c)^2$$

over a, b, c in ∇, a, b > 0, ab = 1. Since f(a, b, c) $\to \infty$ as we approach the boundary of this region, the minimum occurs when the Lagrangian

$$g(a,b,c) = \sum_{j=1}^{n} (a\alpha_j + b\beta_j + c)^2 + \lambda(ab - 1)$$

satisfies $\frac{\partial g}{\partial a} = \frac{\partial g}{\partial b} = \frac{\partial g}{\partial c} = 0$, i.e.

$$2\sum_{j=1}^{n} \alpha_j(a\alpha_j + b\beta_j + c) + \lambda b = 0 ,$$

$$2\sum_{j=1}^{n} \beta_j(a\alpha_j + b\beta_j + c) + \lambda a = 0 ,$$

$$2\sum_{j=1}^{n} (a\alpha_j + b\beta_j + c) = 0 .$$

Putting

$$\bar{\alpha} = \frac{1}{n}\sum_{j=1}^{n}\alpha_j, \quad \bar{\beta} = \frac{1}{n}\sum_{j=1}^{n}\beta_j, \quad \sigma^2 = \frac{1}{n}\sum_{j=1}^{n}\alpha_j^{\,2}, \quad \tau^2 = \frac{1}{n}\sum_{j=1}^{n}\beta_j^{\,2},$$

$$\nu = \frac{1}{n}\sum_{j=1}^{n}\alpha_j\beta_j, \quad \bar{\lambda} = \frac{1}{2n}\lambda$$

where σ, $\tau > 0$, this becomes

(18) $a\sigma^2 + b\nu + c\bar{\alpha} + \bar{\lambda}b = 0$

(19) $a\nu + b\tau^2 + c\bar{\beta} + \bar{\lambda}a = 0$

(20) $a\bar{\alpha} + b\bar{\beta} + c = 0$.

Substituting for c from (20) into (18) and (19) gives

$$a\left(\sigma^2 - \bar{\alpha}^2\right) + b\left(\nu - \bar{\alpha}\,\bar{\beta}\right) + \bar{\lambda}b = 0$$
$$a\left(\nu - \bar{\alpha}\,\bar{\beta}\right) + b\left(\tau^2 - \bar{\beta}^2\right) + \bar{\lambda}a = 0$$

and eliminating $\bar{\lambda}$ gives

(21) $a^2\left(\sigma^2 - \bar{\alpha}^2\right) = b^2\left(\tau^2 - \bar{\beta}^2\right)$.

We are not considering the trivial case $\alpha_j = \bar{\alpha}, j = 1,\dots,n$ when all the data lie on the line $x = \bar{\alpha}$. Thus

$$\sigma^2 - \bar{\alpha}^2 = \frac{1}{n}\left(\sum_{j=1}^{n} \alpha_j^{\,2} - n\bar{\alpha}^2\right) = \frac{1}{n}\sum_{j=1}^{n}\left(\alpha_j - \bar{\alpha}\right)^2 > 0.$$

Similarly,

$$\tau^2 - \bar{\beta}^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\beta_j - \bar{\beta}\right)^2 > 0.$$

Now from (20) and ab = 1, we have

$$\begin{aligned}\frac{1}{n}f(a,b,c) &= a^2\sigma^2 + b^2\tau^2 + c^2 + 2\nu + 2ac\bar{\alpha} + 2bc\bar{\beta} \\ &= a^2\sigma^2 + b^2\tau^2 + 2\nu - \left(a\bar{\alpha} + b\bar{\beta}\right)^2 \\ &= a^2\left(\sigma^2 - \bar{\alpha}^2\right) + b^2\left(\tau^2 - \bar{\beta}^2\right) + 2\left(\nu - \bar{\alpha}\,\bar{\beta}\right)\end{aligned}$$

and by (21) and ab = 1

(22) $\frac{1}{2n}f(a,b,c) = \left(\sigma^2 - \bar{\alpha}^2\right)^{1/2}\left(\tau^2 - \bar{\beta}^2\right)^{1/2} + \nu - \bar{\alpha}\,\bar{\beta}$.

If ab < 0, we similarly assume ab = –1 and again derive (18)- (20). In this case

(23) $\frac{1}{2n}f(a,b,c) = \left(\sigma^2 - \alpha^2\right)^{1/2}\left(\tau^2 - \bar{\beta}^2\right)^{1/2} - \nu + \bar{\alpha}\,\bar{\beta}$.

So, if $\nu - \bar{\alpha}\bar{\beta} = \sum_{j=1}^{n}(\alpha_j - \bar{\alpha})(\beta_j - \bar{\beta}) < 0$, then f attains its minimum when ab > 0, while if $\nu - \bar{\alpha}\bar{\beta} > 0$, the minimum occurs when ab < 0.

To summarise, the optimal line passes through the mean of the data $(\bar{\alpha}, \bar{\beta})$, by (20), and has slope m, where by (21),

$$m^2 = \frac{\sum_{j=1}^{n}(\beta_j - \bar{\beta})^2}{\sum_{j=1}^{n}(\alpha_j - \bar{\alpha})^2}.$$

If $\sum_{j=1}^{n}(\alpha_j - \bar{\alpha})(\beta_j - \bar{\beta}) \neq 0$, then by (22) and (23), m has the same sign as $\sum_{j=1}^{n}(\alpha_j - \bar{\alpha})(\beta_j - \bar{\beta})$, the covariance. If $\sum_{j=1}^{n}(\alpha_j - \bar{\alpha})(\beta_j - \bar{\beta}) = 0$, then there are two optimal lines with slopes ±m. In statistical terms, our line has a slope of magnitude given by the ratios of the standard deviations of the variables. This line turns out to be what is variously known as the reduced major axis or geometric mean functional relationship. The latter name arising from the fact that the slope of the line is the geometric mean of the slopes from the two possible regression lines. So our theoretical development from first principles has in fact provided a foundation for a fitting technique that is little known but has been recommended elsewhere (e.g. Draper and Smith, 1998).

3. **Three Dimensions**

Suppose that (α,β,γ) is a point in ∇^3 and ax + by + cz + d = 0 is the equation of a plane. Then it can be shown, as in Section 1, that if F(α,β,γ,a,b,c,d) represents a measure of the error of the point with respect to the plane which satisfies properties analogous to 1-7 in Section 1, then

$$F(\alpha,\beta,\gamma,a,b,c,d) = \frac{k|a\alpha + b\beta + c\gamma + d|}{(abc)^{1/3}} ,$$

for a constant k > 0. Since the equation of the plane is invariant to multiplication by a non-zero number, we may assume abc > 0.

Now take points $\{\alpha_i, \beta_i, \gamma_i\}$, i = 1, …, n, n ≥ 3. We shall again take the aggregate of the errors of these points from a plane to be the sum of squares. Thus to find a plane with equation ax + by + cz + d = 0 which best fits the above data, we need to find, a,b,c,d in ∇ abc > 0, which minimise

$$(3.1) \quad f(a,b,c,d) = \sum_{j=1}^{n} \frac{(a\alpha_j + b\beta_j + c\gamma_j + d)^2}{(abc)^{2/3}} .$$

At the minimum we shall have $\frac{\partial f}{\partial d} = 0$ and so

$$(3.2) \quad a\overline{\alpha} + b\overline{\beta} + c\overline{\gamma} + d = 0 ,$$

where $\overline{\alpha} = \frac{1}{n}\sum_{j=1}^{n}\alpha, \quad \overline{\beta} = \frac{1}{n}\sum_{j=1}^{n}\beta_j, \quad \overline{\gamma} = \frac{1}{n}\sum_{j=1}^{n}\gamma_j$,

i.e. the required plane passes through the mean of the data $(\overline{\alpha}, \overline{\beta}, \overline{\gamma})$. Substituting from (2) into (1), we need to find a, b, c in ∇, abc > 0, which minimise

$$(3.3) \quad f(a,b,c) = \sum_{j=1}^{n} \frac{\left(a(\alpha_j - \overline{\alpha}) + b(\beta_j - \overline{\beta}) + c(\gamma_j - \overline{\gamma})\right)^2}{(abc)^{2/3}} .$$

We may ignore the trivial case $\alpha_j = \overline{\alpha}, j = 1,...,n$, i.e. when all the data lie in the plane $x = \overline{\alpha}$. Similarly, we ignore the cases $\beta_j = \overline{\beta}, j = 1,...,n$ and $\gamma_j = \overline{\gamma}, j = 1,...,n$. Then we define $s_1, s_2, s_3 > 0$ by

$$s_1^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\alpha_j - \bar{\alpha}\right)^2, \quad s_2^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\beta_j - \bar{\beta}\right)^2, \quad s_3^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\gamma_j - \bar{\gamma}\right)^2.$$

We also define

$$s_{12} = \frac{1}{n}\sum_{j=1}^{n}\left(\alpha_j - \bar{\alpha}\right)\left(\beta_j - \bar{\beta}\right), \quad s_{23} = \frac{1}{n}\sum_{j=1}^{n}\left(\beta_j - \bar{\beta}\right)\left(\gamma_j - \bar{\gamma}\right), \quad s_{13} = \frac{1}{n}\sum_{j=1}^{n}\left(\alpha_j - \bar{\alpha}\right)\left(\gamma_j - \bar{\gamma}\right)^2.$$

Putting

$x = as_1$, $y = bs_2$, $z = cs_3$,

$\lambda = \dfrac{s_{23}}{s_2\, s_3}$ $\mu = \dfrac{s_{13}}{s_1\, s_3}$ $\nu = \dfrac{s_{12}}{s_1\, s_2}$

we have

$$f(a,b,c) = n\left(s_1 s_2\, s_3\right)^{2/3}\, g(x,y,z) \quad ,$$

where

(3.4) $$g(x,y,z) = \frac{x^2 + y^2 + z^2 + 2\lambda yz + 2\mu xz + 2\nu xy}{(xyz)^{2/3}}.$$

Thus the problem is equivalent to minimising g(x,y,z) over x,y,z in ∇, $xyz > 0$.

Now suppose $\lambda = 1$. Thus there is some constant $\alpha \neq 0$ such that

$\beta_j - \bar{\beta} = \alpha\left(\gamma_j - \bar{\gamma}\right)$, $j = 1,\ldots,n$,

and so the data lie on the plane $y - \alpha z = 0$. This plane is not among those considered by this method. Indeed, we assume that there is some relation involving all the variables and since the case $\lambda = 1$ would deem the variable $\{\alpha_j\}$ to be irrelevant, we ignore this case. Similarly, by making a transformation, we may assume $\lambda \neq -1$, i.e. $|\lambda| < 1$. Similarly, we assume $|\mu| < 1$, $|\nu| < 1$.

Note that g(x, y, z) is invariant under making the same permutation of (x, y, z) and of (λ, μ, ν). It is also invariant under the following transformations:

$x \to -x$, $\mu \to -\mu$, $\nu \to -\nu$,
$y \to -y$, $\lambda \to -\lambda$, $\nu \to -\nu$,
$z \to -z$, $\lambda \to -\lambda$, $\mu \to -\mu$

There is therefore no loss of generality in assuming either $0 \le \lambda \le \mu \le \nu$ or $\lambda < 0$, $\mu, \nu \ge 0$.

We now give a result describing the complete solution to the above minimisation problem. We first give several special cases where the solution can be described explicitly, and then in Case 6 we give the generic case where the solution is given in terms of a solution of a quartic equation. Of course, any solution for (x, y, z) can be multiplied by any non-zero constant to give another solution.

Theorem 2 The minimum value of (3.4) over x, y, z in ∇, $xyz > 0$ is given as follows:

Case 1 If $\lambda = \mu = \nu = 0$, then

$$(x,y,z) = \left(-\sqrt{1-\lambda}, 1, -1\right) \text{ or } \left(-\sqrt{1-\lambda}, -1, 1\right) \text{ for } \lambda > 0,$$

$(x, y, z) = (1, 1, 1), (1, -1, -1), (-1, 1, -1)$ or $(-1, -1, 1)$ for $\lambda = 0$.

Case 2 If $\lambda = \mu = \nu > 0$, then

$(x,y,z) = (1 + \lambda, -1, -1), (-1, 1 + \lambda, -1)$ or $(-1, -1, 1 + \lambda)$.

Case 3 If $0 \le \lambda < \mu = \nu$ or $\lambda < 0 < \mu = \nu$, then

$$(x, y, z) = \left(\mu + \sqrt{\mu^2 + 4\lambda + 4}, -2, -2\right) .$$

Case 4 If $0 < \lambda = \mu < \nu$, then

$(x, y, z) = (-\alpha, \beta, -1)$ or $(\beta, -\alpha, -1)$,

where

$$\alpha = \frac{1}{2}\left(\sqrt{\mu^2 + \frac{4(1-\mu^2)}{1-\nu}} - \mu\right), \beta = \frac{1}{2}\left(\sqrt{\mu^2 + \frac{4(1-\mu^2)}{1-\nu}} + \mu\right).$$

Case 5 If $\mu \ge 0$, $\nu > 0$, $\lambda = -\nu$, then

$$(x, y, z) = \left(2, -\nu - \sqrt{\nu^2 - 4\mu + 4}, -2\right).$$

Case 6 Suppose $0 \le \lambda < \mu < \nu$ or $\lambda < 0 \le \mu < \nu$, $\mu, \nu \ne -\lambda$.

Then $(x,y,z) = (1, -\alpha, -\beta)$,

where α satisfies $P(-\alpha) = 0$, where

(3.5) $$P(Y) = (1-\lambda^2)Y^4 + \lambda(\mu - \lambda\nu)Y^3 + 2(\lambda\mu\nu - 1)Y^2 + \mu(\lambda - \mu\nu)Y + 1 - \mu^2,$$

and $$\beta = \frac{1-\alpha^2}{\mu + \lambda\alpha}.$$

If $\lambda > -\mu$, then $-\alpha$ is the unique zero of P in $(-1, 0)$ and $0 < \beta < 1$.

If $-\nu < \lambda < -\mu$, then $-\alpha$ is the unique zero of P in $(-\infty, -1)$ and $0 < \beta < 1$.

If $\lambda < -\nu$, then $-\alpha$ is the unique zero of P in $(-\infty, -1)$ and $\beta > 1$.

Proof Since g is invariant under multiplying w, y, z by a non-zero number, we may assume $xyz = 1$. With this assumption, $g(x, y, z) \to \infty$ as we approach the boundary of the region considered. We introduce a Lagrange multiplier $\bar{\lambda}$ and putting the derivatives of g with respect to x, y, z equal to zero gives

$$2x + 2\mu z + 2\nu y + \bar{\lambda} yz = 0,$$

$$2y + 2\lambda z + 2\nu x + \bar{\lambda} xz = 0,$$

$$2z + 2\lambda y + 2\mu x + \bar{\lambda} xy = 0.$$

Eliminating $\bar{\lambda}$ gives

(3.6) $$x^2 + \mu xz + \nu xy = y^2 + \lambda yz + \nu xy = z^2 + \lambda yz + \mu xz.$$

Putting $Y = \frac{y}{x}$, $Z = \frac{z}{x}$, the first equation in (6) becomes

(3.7) $$1 + \mu Z = Y^2 + \lambda YZ,$$

and similarly, equating the first and last terms in (6) gives

(3.8) $$1 + \nu Y = Z^2 + \lambda YZ.$$

Case 1 Suppose $\mu = \nu = 0$. Then (3.7) and (3.8) become

$$1 + Y^2 + \lambda YZ = Z^2 + \lambda YZ.$$

So $Y^2 = Z^2$ and either $Y = Z$, $Y^2 = \frac{1}{1+\lambda}$ or $Y = -Z$, $Y^2 = \frac{1}{1-\lambda}$.

Thus $(x, y, z) = (1+\lambda)^{-1/6}\left(\sqrt{1+\lambda}, 1, 1\right)$,

$(1+\lambda)^{-1/6}\left(\sqrt{1+\lambda}, -1, -1\right)$, $(1-\lambda)^{-1/6}\left(-\sqrt{1-\lambda}, 1, -1\right)$ or

$(1-\lambda)^{-1/6}\left(-\sqrt{1-\lambda}, -1, 1\right)$. The first two cases give $g(x, y, z) = 3(1+\lambda)^{2/3}$ and the last two give $g(x, y, z) = 3(1-\lambda)^{2/3}$. So for $0 < \lambda < 1$ there are two solutions, $\left(-\sqrt{1-\lambda}, 1, -1\right)$ and $\left(-\sqrt{1-\lambda}, -1, 1\right)$, where we have multiplied through by $(1-\lambda)^{1/6}$ for simplicity. Similarly, for $\lambda = 0$ there are four solutions (1, 1, 1), (1, −1, −1), (−1,1,−1) and (−1,−1,1).

<u>Case 2</u> Suppose $\lambda = \mu = \nu > 0$. From (3.7) and (3.8) we have $\lambda(Z-Y) = Y^2 - Z^2$ and so $Y = Z$ or $Y + Z = -\lambda$. This gives the solutions $(Y, Z) = (1,1), \left(\frac{-1}{1+\lambda}, \frac{-1}{1+\lambda}\right), (1, -\lambda-1), (-\lambda-1, 1)$.

The first case gives $g(x, y, z) = 3(1 + 2\lambda) > 3$, while the remaining three cases give $g(x, y, z) = 3(1+\lambda)^{-2/3}\left(1-\lambda^2\right) < 3$. Thus the minimum is attained for $(x, y, z) = (1 + \lambda, -1, -1), (-1, 1 + \lambda, -1)$ and $(-1, -1, 1 + \lambda)$.

For the remaining cases we shall use the following result:

<u>Lemma 1</u> If $\lambda < 0$, $\mu \geq 0$, $\nu > 0$ or if $0 \leq \lambda < \mu \leq \nu$, then (3.4) with, $xyz > 0$, attains its minimum only in the octant $\{x > 0, y < 0, z < 0\}$. If $0 < \lambda = \mu < \nu$, then (3.4) with $xyz > 0$ attains its minimum in both the octants $\{x > 0, y < 0, z < 0\}$ and $\{x < 0, y > 0, z < 0\}$.

<u>Proof</u> Suppose $\lambda < 0$, $\mu \geq 0$, $\nu > 0$. Then for $x, y, z > 0$, $xyz = 1$, we have

$$g(x, y, z) = x^2 + y^2 + z^2 - 2|\lambda|\, yz + 2\mu xz + 2\nu xy,$$
$$g(x, -y, -z) = x^2 + y^2 + z^2 - 2|\lambda|\, yz - 2\mu xz - 2\nu xy,$$
$$g(-x, y, -z) = x^2 + y^2 + z^2 + 2|\lambda|\, yz + 2\mu xz - 2\nu xy,$$
$$g(-x, -y, z) = x^2 + y^2 + z^2 + 2|\lambda|\, yz - 2\mu xz + 2\nu xy.$$

Thus, for all $x, y, z > 0$,

$$g(x, -y, -z) < g(x, y, z),\ g(-x, y, -z),\ g(-x, -y, z),$$

and so the minimum occurs only for $x > 0$, $y < 0$, $z < 0$.

Next suppose $0 \leq \lambda < \mu \leq \nu$. Then for $x, y, z > 0$, $xyz = 1$,

$$g(x, -y, -z) = x^2 + y^2 + z^2 + 2\lambda yz - 2\mu xz - 2\nu xy$$

$$< x^2 + y^2 + z^2 + 2\lambda yz + 2\mu xz + 2\nu xy$$
$$= g(x, y, z),$$

and since $\lambda < \mu$,

$$g(x, -y, -z) < x^2 + y^2 + z^2 + 2\mu yz - 2\lambda xz - 2\nu xy = g(-y, x, -z).$$

Similarly, for $x, y, z > 0$,

$$g(x, -y, -z) < g(-z, -y, x) .$$

So the minimum of $g(x, y, z)$ for $xyz > 0$ occurs only for $x > 0$, $y < 0$, $z < 0$.

Finally suppose $0 < \lambda = \mu < \nu$. As in the previous case, for $x, y, z > 0$,

$$g(x, -y, -z) < g(x, y, z),\ g(-z, -y, x) .$$

By symmetry, $g(x, -y, -z) = g(-y, x, -z)$ for all $x, y, z > 0$. So the minimum of $g(x, y, z)$ for $xyz > 0$ occurs for both $x > 0$, $y < 0$, $z < 0$ and $x < 0$, $y > 0$, $z < 0$.

We now continue the proof of Theorem 2.

Case 3 Suppose $0 \le \lambda < \mu = \nu$ or $\lambda < 0 < \mu = \nu$. From (3.7) and (3.8) we have $Y = Z$ or $Y + Z = -\mu$. By Lemma 1 we know that $Y < 0$, $Z < 0$. If $Y \neq Z$, $Y + Z = -\mu$, then Y and Z are the roots of

(3.9) $$(1 - \lambda) Y^2 + \mu(1 - \lambda) Y + \mu^2 - 1 = 0$$

and so $YZ = \frac{\mu^2 - 1}{1 - \lambda} < 0$, which is not possible. So $Y = Z$.

Then $(1 + \lambda) Y^2 - \mu Y - 1 = 0$ and so $Y = -\alpha^{-1}$, where $\alpha = \frac{1}{2}\left(\mu + \sqrt{\mu^2 + 4\lambda + 4}\right)$.

Thus the minimum is attained for $(x, y, z) = (\alpha, -1, -1)$. Note that if $\lambda = -\mu$, then the solution is $(1, -1, -1)$.

Case 4 Suppose $0 < \lambda = \mu < \nu$. It will be convenient to interchange λ and ν, and therefore x and z, so that $0 < \mu = \nu < \lambda$. Then by Lemma 1, the solution is in the octants $\{x < 0, y < 0, z > 0\}$ and $\{x < 0, y > 0, z < 0\}$ and so $YZ < 0$. As in Case 3, we must have $Y + Z = -\mu$ and Y, Z are the roots of (3.9), i.e.

$$\frac{1}{2}\left\{-\mu \pm \sqrt{\mu^2 + \frac{4\left(1 - \mu^2\right)}{1 - \lambda}}\right\}.$$

Returning to our original case $\left(\frac{y}{z}, \frac{x}{z}\right)$ equals $(\alpha, -\beta)$ or $(-\beta, \alpha)$, for α, β as in Case 4 of Theorem 2. The result follows.

Case 5 Suppose $\mu \geq 0$, $\nu > 0$, $\lambda = -\nu$. For convenience we interchange λ and μ, x and y, so that $\lambda \geq 0$, $\nu > 0$, $\mu = -\nu$. Then from (3.7) and (3.8) we have $Y + Z = 0$ or $Z - Y = \nu$. By Lemma 1, the solution is in the octant $\{x < 0, y > 0, z < 0\}$ and so $Y < 0$, $Z > 0$.

If $Z - Y = \nu$, then Y satisfies $(1 + \lambda) Y^2 + \nu(1 + \lambda) Y + \nu^2 - 1 = 0$. Let the roots of this equation be $Y_1 < 0$ and $Y_2 = -\nu - Y_1 > 0$. Then $Y = Y_1$, $Z = Y_1 + \nu < 0$, which is not possible. So $Y + Z = 0$. Then $(1 - \lambda) Y^2 - \nu Y - 1 = 0$ and so $Y = -\alpha^{-1}$, where $\alpha = \frac{1}{2}\left(\nu + \sqrt{\nu^2 - 4\lambda + 4}\right)$. Then $Z = \alpha^{-1}$ and so the solution is (x, y, z) $= (-\alpha, 1, -1)$. Returning to our original case gives the result.

Case 6 Suppose $0 \leq \lambda < \mu < \nu$ or $\lambda < 0 \leq \mu < \nu$, $\mu, \nu \neq -\lambda$. From (3.7) we have

(3.10) $$Z = \frac{1 - Y^2}{\lambda Y - \mu}.$$

(Note that by (3.7), $Y = \frac{\mu}{\lambda}$ would give $\lambda^2 = \mu^2$ which is not true.)

Substituting from (3.10) into (3.8) gives $P(Y) = 0$, where P is given by (3.5). By Lemma 1 we have $Y < 0$, $Z < 0$. Thus $Y = -\alpha$ for $\alpha > 0$, satisfying $P(-\alpha) = 0$, and $Z = -\beta$ where $\beta > 0$ and by (3.10),

(3.11) $$\beta = \frac{1 - \alpha^2}{\mu + \lambda\alpha}.$$

Now $P(-1) = (\nu - 1)(\lambda + \mu)^2 < 0$, $P(0) = 1 - \mu^2 > 0$, $P(1) = -(\lambda - \mu)^2 (1 + \nu) < 0$, and $\lim_{Y \to \infty} P(Y) = \lim_{Y \to -\infty} P(Y) = \infty$. Thus P has a zero in each of the intervals $(-\infty, -1)$, $(-1, 0)$, $(0, 1)$ and $(1, \infty)$.

First suppose $\lambda \geq 0$. If $\alpha > 1$, then (11) gives $\beta < 0$ which is false. So $0 < \alpha < 1$. Note that from (3.8),

$$\alpha = \frac{1-\beta^2}{\nu + \lambda\beta}. \tag{3.12}$$

From (3.12) we see $0 < \beta < 1$, which gives the required result.

Next suppose $-\mu < \lambda < 0$. Now

$$P\left(\frac{\mu}{\lambda}\right) = \left(\frac{\mu^2}{\lambda^2} - 1\right)^2 > 0. \tag{3.13}$$

If $\alpha > 1$, then (11) gives $\mu + \lambda\alpha < 0$, i.e. $-\alpha < \frac{\mu}{\lambda} < -1$, which contradicts (3.13). So again $0 < \alpha < 1$. Interchanging μ and ν and noting that $-\nu < \lambda < 0$, we see by symmetry that $-\beta$ is the unique solution in

$(-1, 0)$ of $\widetilde{P}(Z) = 0$, where $\widetilde{P}$ is derived from P by interchanging μ and ν. Thus again $0 < \beta < 1$, as required.

Now suppose $-\nu < \lambda < -\mu$. If $0 < \alpha < 1$, then (11) gives $\mu + \lambda\alpha > 0$ and so $-\alpha > \frac{\mu}{\lambda} > -1$, which contradicts (3.13) since $P(Y) < 0$ for $-1 \leq Y < -\alpha$. So $\alpha > 1$. As in the previous case, we have $0 < \beta < 1$.

Finally, we suppose $\lambda < -\nu$. As in the previous case, we have $\alpha > 1$. With $\widetilde{P}$ as before, we see by symmetry that $-\beta$ is the unique solution in $(-\infty, -1)$ of $\widetilde{P}(Z) = 0$. Thus $\beta > 1$ and the proof is complete. □

4. Concluding remarks

In conventional regression where one uses the fitted model to predict a value of one variable, it is clear that the error measure must involve the absolute difference between the predicted value and the actual value. One can then take powers (usually the square) of this error measure – this has the effect of emphasising larger errors. If however prediction is not our aim, so that we no longer have such a 'special variable', then the question arises as to what criterion should be used for fitting purposes. Generally we would wish to treat all variables in the same way, hence switching of axes should not affect our criterion. Likewise a translation of the origin of coordinates should not affect it either, nor should a change of units of measurement. From such axioms we have been able to derive the form of the fitting criterion (Theorem 1). It has a simple geometric interpretation: it corresponds to the geometric mean of the absolute deviations from the data point to the line, measured parallel to each axis. We expect that this important result extends to any number of dimensions.

As before, one may wish to take powers of this fitting criterion. We showed that for two variables, if we minimise the sum of squares then we arrive at what is known as the reduced major axis or geometric mean functional relationship. This has a very simple form for the slope magnitude: it is the ratio of the standard deviations of the two variables. For three variables however we will generally need to solve a quartic equation to fit the plane (Theorem 2).

The general case involving multiple variables has been studied by Draper and Yang (1997). They too used the second power and computed the coefficients iteratively in a nonlinear weighted least squares approach. They prove that the coefficients are a convex combination of those arising from carrying out all possible conventional regressions i.e. by taking each variable in turn as the dependent variable and regressing it on the rest.

Tofallis (2002) has looked at the multiple variable case using the first power of the criterion i.e. the geometric mean of the absolute deviations in each direction. He formulates the problem as a mathematical programme whose objective function is linear, as are all but one of the constraints. The method was then tested on two data sets from Belsley (1991) which involve four variables and suffer from multicollinearity. This data was generated from a known model and so it was possible to see how well the model parameters were reproduced. Conventional regression was poor in this regard: two of the parameters came out with the wrong sign, and another was 15 times too large. By contrast the neutral data fitting method gave all the correct signs and the parameter values were much closer to those of the underlying model. Another interesting effect was in the response to small changes in the data. The two data sets were very similar to each other – differing only in the third digits, yet the conventional regressions produced very different models. Once again the neutral fitting method was superior in that it generated very similar models. This provides some initial evidence for its stability.

References

Belsley, D.A. (1991). *Conditioning Diagnostics*. Wiley, New York.

Deeming, T.J. (1968). The analysis of linear correlation in astronomy. In *Vistas in Astronomy*, volume 10, Pergamon Press, Oxford.

Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*, 3rd edition. Wiley, New York.

Draper, N.R. and Yang, Y. (1997). Generalization of the geometric mean functional relationship. *Computational Statistics and Data Analysis*, 23, 355-372.

Kruskal, W.H. (1953). On the uniqueness of the line of organic correlation. *Biometrics* (9) 47-58.

Otis, A.S. (1922). The method for finding the correspondence between scores in two tests. *J. of Educational Psychology*, XIII, 524-545.

Ricker, W.E. (1973). Linear regressions in fishery research. *J. Fisheries Research Board of Canada*, 30, 409-434.

Samuelson, P.A. (1942). A note on alternative regressions. *Econometrica* 10(1), 80-83.

Stromberg, G (1940). "Accidental systematic errors in spectroscopic absolute magnitudes for dwarf GoK2 stars". Astrophysical J. 92, 156ff.

Sverdrup, H (1916). "Druckgradient, Wind und Reibung an der Erdoberfläche". Ann. Hydrogr. u. Maritimen Meteorol. (Berlin) 44, 413-427.

Sokal, R.R., and Rohlf, F.J. (1995). *Biometry*, 3rd edition. Freeman, New York.

Tofallis, C. (2002). Model fitting for multiple variables by minimising the geometric mean deviation. In *Total Least Squares and Errors-in-Variables Modeling,* (eds. S. van Huffel and P. Lemmerling), Dordrecht: Kluwer Academic.