Annals of Operations Research 124, 69–79, 2003

Multiple Neutral Data Fitting

CHRIS TOFALLIS c.tofallis@herts.ac.uk
Deptament of Stats, Econ, Accounting and Management Systems, Business School, University of Hertfordshire, UK

Abstract. A method is proposed for estimating the *relationship* between a number of variables; this differs from regression where the emphasis is on predicting *one* of the variables. Regression assumes that only one of the variables has error or natural variability, whereas our technique does not make this assumption; instead, it treats all variables in the same way and produces models which are units invariant – this is important for ensuring physically meaningful relationships. It is thus superior to orthogonal regression in that it does not suffer from being scale-dependent. We show that the solution to the estimation problem is a unique and global optimum. For two variables the method has appeared under different names in various disciplines, with two Nobel laureates having published work on it.

Keywords: functional relation, data fitting, regression, model fitting

Introduction

The method of regression is undoubtedly one of the most widely used quantitative methods in both the natural and social sciences. For most users of the technique the term is taken to refer to the fitting of a function to data by means of the least squares criterion. Researchers who probe further into the field will discover that other criteria exist, such as minimising the sum of the absolute deviations, or minimising the maximum absolute deviation. All of these fitting criteria have one thing in common: they measure goodness of fit by referring to deviations in the dependent variable alone. If one is intending to use the regression model for the purpose of predicting the dependent variable then this is a natural track to follow. If, however, one is seeking the underlying relationship between a number of variables then there may not be an obvious choice for the dependent variable. One would then be interested in a method which treated all variables on the same basis. This is the underlying motivation for this paper. We allow that there may be both natural variation and/or measurement error in *all* of the variables (another departure from conventional regression). We do not assume that we have information on error variances (these are rarely known) or the error distributions associated with the data. (If such information is available then there are established statistical approaches, see (Cheng and Van Ness, 1999).) In fact we are simply presenting an estimation procedure without any distributional assumptions. We justify our choice of method by virtue of certain useful properties which other methods do not possess.

1. Total least squares or orthogonal regression

Suppose we wish to fit a line to data, but without basing it solely on the vertical deviations from the line. A possible alternative fitting criterion that one might consider is to minimise the sum of the perpendicular distances from the data points to the regression line, or the squares of such distances. This involves applying Pythagoras' theorem to calculate such distances and so involves summing the squared deviations in each dimension. This is therefore sometimes referred to as "total least squares", as well as "orthogonal regression". The trouble with this approach is that we shall be summing quantities which are in general measured in different units; this is not a meaningful thing to do. One can try to get around this objection by normalising in some way: e.g., divide each variable by its range, or its standard deviation, or express it as a percentage of some base value, etc. Unfortunately each of these normalisations results in a different (or non-equivalent) fitted model for a given set of data.

It is also worth noting that even if a variable is dimensionless, multiplying it by some constant factor will also affect the resulting orthogonal regression model. This is because total least squares will tend to concentrate on reducing the deviations of that variable whose values happen to be greater in absolute magnitude, effectively attaching greater weight to it. Ideally we would like a method that was invariant to a scale change in any of the variables. In general the need for invariance properties to ensure meaningfulness is covered in the subject of measurement theory (Roberts, 1979). For more on total least squares see (Van Huffel and Vandewalle, 1991; Van Huffel, 1997).

2. The multiplicative approach

Instead of adding deviations for each variable, let us multiply them together. In other words, for each data point multiply together the deviations from the fitted line for each variable. Thus in the two variable case we multiply the vertical (y) deviation with the horizontal (x) deviation. This gives twice the area of the triangle described by the data point and its projections (parallel to the axes) onto the regression line. Our fitting criterion is then to minimise the sum of these triangular areas, i.e., the sum of the products of the deviations. Woolley (1941) refers to these triangles as "area deviations". See figure 1.

This idea is not new; in fact, it has surfaced in many fields over the twentieth century. (See the historical notes below.) Unfortunately it has usually appeared under different names in each discipline and this may have played a part in it not being more widely known – there have been small pockets of scattered usage and a critical mass of unified literature has yet to be achieved.

One of the attractions of this method is that it produces the same model irrespective of which variable is labelled x and which is labelled y; this symmetry is apparent from the fact that if we exchange the axes on which each variable is plotted, the triangular areas remain the same – triangles that were above the line now appear below, and vice versa – we have reflected everything through the line $y = x$. For an algebraic proof see

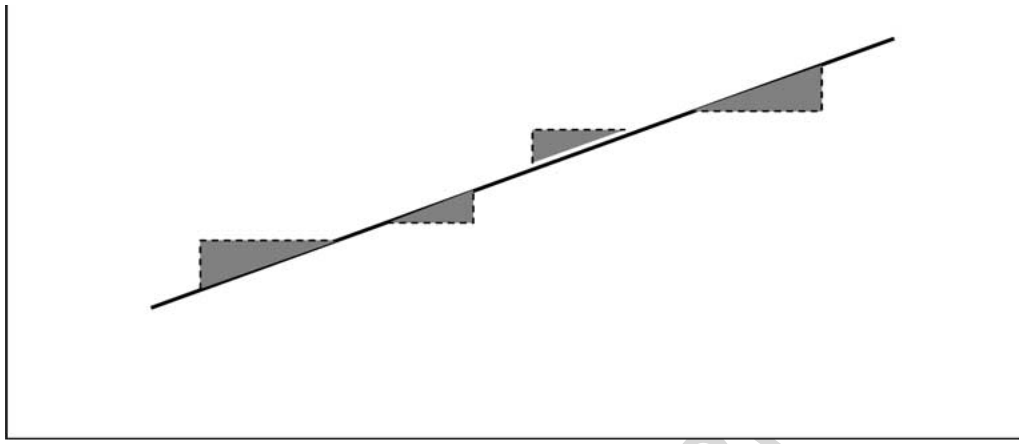

Figure 1. There are four data points: one at each right angle of the shaded triangles. The aim is to fit a line such that the sum of the triangular areas is minimised.

(Woolley, 1941). Because no variable is given special treatment, we choose to call this approach "neutral data fitting".

Note also that this method is invariant to scale change for any variable, since if we scale up one variable by some factor then we are merely stretching the length of the triangles in that dimension – the total triangular area is scaled up by the same factor. In fact, the entire graph is simply stretched in one of the dimensions with everything remaining in the same relative position.

3. Relation to least squares regression: the geometric mean property

Consider the least squares regression lines of y on x, and x on y plotted using the same axes. They each pass through the centroid of the data (the point whose co ordinates are the means of the variables), but in general they will have different slopes and intercepts. The neutral data fitting model will also pass through the centroid, but possesses the attractive property of having a slope (and intercept) lying somewhere between those of the two least squares lines: in fact it will have a slope that is the geometric mean of these two. It is for this reason that the resulting model is sometimes referred to as the *geometric mean functional relation*. This connection has been made at different times in different disciplines (e.g., see (Barker, Soh, and Evans, 1988) for a proof).

4. Properties of neutral data fitting

Samuelson (1942) lists seven properties, which a fitting method might desirably possess. He states (without providing any justification) that one of these is in principle impossible except when the correlation is perfect. The desirability of another of his properties is open to question: namely that the fitted equation should be invariant under any orthogonal transformation of variables (i.e., transformations that leave angles and lengths unchanged). Such transformations include rotation of the axes: thus a real relationship represented by a line that passed through the origin (such as $y = x$) could be transformed into a non-existent relationship by rotating in such a way that the fitted line coincided with the x- or y-axis. This does not therefore seem to be a useful property for a fitting procedure to possess. However, what is particularly remarkable is that Samuelson proves that what we here call neutral data fitting is the *only* method, which satisfies all of his desirable properties, bar the two we have mentioned. His list includes invariance under scale change, invariance under interchange of variables, and reduction to the correct equation for perfectly correlated variables.

Kruskal (1953) also proved a uniqueness theorem: of all the procedures which depend only on first and second moments (i.e., on the means, covariance and standard deviations), neutral data fitting is the only line-fitting procedure which behaves correctly under translation and change of scale. This implies that any linear transformation on any variable is acceptable. The translation invariance is easy to visualise since adding a constant to say, the y-variable, merely shifts the data points *en masse* in the vertical direction, hence the triangular areas are unaffected. As for the scale invariance: imagine multiplying all the values of one of the variables, say y, by some constant, this would stretch the plane containing the data points in the y-direction; the height and hence the area of each triangle would be multiplied by the same factor whilst the x-deviation (the width of each triangle) would be unaffected. Hence the slope of the "least areas line" would scale by the same factor, and hence is scale invariant. The simple geometry of our fitting criterion allows us to see that these arguments extend to higher dimensions and so will apply for the case of multiple variables.

5. Comparison with other line-fitting methods

Babu and Feigelson (1992) noted that "For many applications . . . only a set of correlated points is available to fit a line. The underlying joint distribution is unknown, and it is not clear which variable is 'dependent' and which is 'independent' ". They compared a number of line-fitting techniques and proved the following theorem: whenever the fitting techniques are applied to the same set of data, the estimated slopes will generally differ, with the two extremes provided by the ordinary least squares (OLS) regressions (y regressed on x, and x on y). In between these will lie the slopes obtained from orthogonal regression and the OLS bisector (the line which bisects the angle formed by the two least squares lines). And nested in between the latter will be the slope value obtained by neutral data fitting (which they refer to as the 'reduced major axis' or RMA); i.e., the neutral data fit slope provides the median value from these five fitting techniques. Their paper also includes a useful table providing expressions for the variance of the slope estimate for each technique.

Babu and Feigelson applied the techniques to simulated data using a Monte Carlo approach. They ran 500 simulations for each method using a large data sample (500 points), and a further 500 simulations using a smaller data sample (50 points). These were repeated under a variety of conditions of correlation and standard deviation. They were interested in how accurately the theoretical slope could be reproduced. They found that orthogonal regression consistently gave the poorest accuracy, and that taking the mean of the two OLS slopes gave intermediate accuracy. The RMA (neutral data fitting) and OLS bisector gave the highest accuracy. In their conclusion, however, they reject RMA on the grounds that the expression for the slope does not depend on the correlation between x and y. This is rather a strange objection because the correlation is a measure of the strength of the linear relationship and should be independent of the slope. The slope provides an estimate of the rate of change of y with x; why should this value be determined by the correlation r? After all, two regression lines can have the same slope but the data sets on which they are based can differ in the correlation; conversely, two sets of data can have the same correlation but have different regression slopes. It is conceivable that their objection may be grounded in the "OLS conditioning" that most researchers are imbued with (since in OLS the slope is related to the correlation r, according to: slope $= rs_y/s_x$). The eminent statistician John Tukey has indeed described least squares regression as a statistical idol and feels "it is high time that it be used as a springboard to more useful techniques" (Tukey, 1975).

Another connection between the two OLS slopes and correlation is that $r^2 =$ [slope of OLS(y on x)]/[Slope of OLS(x on y)] where y is plotted on the vertical axis for both cases so that the slope represents the rate of change of y with x. This implies that for a data set with a correlation of 0.7 the usual OLS regression line will have a slope which is less than half the magnitude of the OLS regression of x on y. In general, the lower the correlation the more these two lines will diverge, irrespective of what the "true" underlying relationship is. This dependence on correlation may partly explain why the two OLS lines provide the extreme values when the slopes are placed in order for the five methods mentioned above. If the "true" slope lies somewhere between the two least squares estimates then the usual OLS slope provides an under-estimate for the absolute value and the OLS(x on y) an over-estimate. Riggs, Guarnieri, and Addelman (1978) make the point forcefully: "OLS continues to be by far the most frequently used method even when it is obviously inappropriate. As a result, hundreds if not thousands of regression lines with too-small slopes are being published annually".

Riggs, Guarnieri, and Addelman (1978) managed to come up with a staggering number 34 of different line-fitting methods – admittedly many were variations, which involved different weighting schemes. They rejected 16 of them due to deficiencies such as lack of scale-invariance, and carried out a Monte Carlo study of the remaining 18 methods to see how well they could reproduce the true underlying model that was used to generate the data. Their conclusion is couched in terms of λ, the ratio of the error variance in the y variable to the error variance in the x variable: "In overall performance, judged both by root mean square error and percent bias, the geometric mean (i.e., neutral data fitting) was almost always the best method when $\lambda = 1 \ldots$ For $\lambda < 1$ error-weighted methods, and for $\lambda > 1$ variance-weighted methods tended to be superior". Unfortunately λ is rarely known because it "requires special observations on y when x is accurately known, and on x when y is accurately known" (Ricker, 1973). After some discussion about what to do when λ is not known, Ricker comes to the "conclusion that the geometric mean regression is the best available estimate of the functional relationship for the situation where all the variability of both variates is due to measurement error and there is no supplementary information concerning the relative point errors in x and y". Ricker also argues for the use of this method when the variability in each variable is largely or entirely natural rather than due to measurement error. Thus he recommends its use in biological work.

6. Multiple neutral data fitting

So far we have seen that neutral data fitting possesses a number of desirable theoretical properties, and that comparative simulations with other fitting methods show it to perform well in many situations. This provides us with the motivation to extend the method to the case of multiple variables. We shall do this by generalising the concept of area deviations as shown in figure 1. From each point we extend a line to the fitted plane in a direction parallel to the co-ordinate axes; in three dimensions this will describe a tetrahedron with a right angle at the data point. The volume of this tetrahedron corresponds to the "volume deviation" for that data point. We then fit a plane to the data such that the sum of volume deviations V is a minimum. The same idea carries over into higher dimensions when fitting a hyperplane.

Draper and Yang (1997) have also presented work which treats each variable on the same footing. Their approach differs from ours in that they minimised the sum of the geometric means of the squared deviations in each dimension. This quantity is also related to the volume deviations, and corresponds, in n dimensions, to minimising the sum of quantities of the form $V^{2/n}$. Minimising the sum of geometric means of absolute deviations is covered by Tofallis (2002a). Indeed we can propose a whole family of fitting methods in which one minimises $\sum V^P$ where higher values of p are chosen to emphasise the larger deviations. This is akin to the family of fitting procedures based on L_p norms where $p = 1$ corresponds to minimising the sum of the absolute y-deviations (known as LAV or least absolute value regression), $p = 2$ corresponds to OLS, and $p = \infty$ is the Chebyshev or minimax norm where one minimises the largest residual.

7. Estimation procedure

Given that we are aiming to minimise a certain quantity $\sum V$, it is natural to formulate the problem as one of optimisation. To make matters easier to visualise we first describe the three dimensional case. We shall fit a plane of the form:

$$ax + by + cz = k. \tag{1}$$

The coefficients or parameters are best understood by noting that this plane intersects the axes at $x = k/a$, $y = k/b$, and $z = k/c$, respectively. Let us consider the extreme values of the parameters: if, say, $b = 0$ then the plane is parallel to the y-axis and so there is no dependence on y. An infinite value for b corresponds to a plane perpendicular to the

y-axis, and so once again there is no dependence on the y variable. We shall assume that we do not have such cases, or that the relevant variable is removed from the data set if they do arise. Note that $k = 0$ corresponds to the plane passing through the origin, which is perfectly acceptable.

We can of course multiply through equation (1) by any non-zero factor to obtain an equivalent form. So we can impose one constraint on the parameters to specify a single solution; for instance we might choose the constraint to be

$$a + b + c = 1 \tag{2}$$

or one could instead set $k = 1$ provided one did not expect the plane to pass through the origin.

Consider any data point with coordinates (x_i, y_i, z_i), the associated volume deviation is the volume of the tetrahedron whose sides are the deviations in the x, y, and z directions. From geometry the volume of such a right-angled tetrahedron is given by one sixth the product of the base, height, and width.

From (1) the deviation in the x-direction of the data point from the plane is given by

$$\left| x_i - \frac{k - by_i - cz_i}{a} \right| \quad \text{or} \quad \left| \frac{ax_i + by_i + cz_i - k}{a} \right|.$$

Similarly the deviations in the y and z directions are:

$$\left| y_i - \frac{k - ax_i - cz_i}{b} \right| \quad \text{or} \quad \left| \frac{ax_i + by_i + cz_i - k}{b} \right| \quad \text{and}$$

$$\left| z_i - \frac{k - ax_i - by_i}{c} \right| \quad \text{or} \quad \left| \frac{ax_i + by_i + cz_i - k}{c} \right|.$$

Multiplying these three deviations together, we find the volume deviation associated with this data point is proportional to $|(ax_i + by_i + cz_i - k)^3/(abc)|$.

The optimisation problem is thus to choose values of the parameters a, b, c, k so as to minimise

$$\sum \left| \frac{(ax_i + by_i + cz_i - k)^3}{abc} \right|. \tag{3}$$

If instead of three variables, we have n variables, then this objective function generalises in the obvious way.

8. Properties of the solution: optimality and uniqueness

By formulating the above optimisation problem as a certain type of non-linear programme we can show that the solution possesses some useful properties.

In order to deal with the absolute value operator in (3) we can define non-negative variables u and v:

$$u_i - v_i = ax_i + by_i + cz_i - k.$$

The pairs of non-negative quantities u_i, v_i are a type of residual: those points with a positive u value lie on the opposite side of the fitted plane to those with a positive v value. Note that the minimisation will ensure that at least one of each pair (u_i, v_i) will be zero.

Next, without loss of generality we shall assume that each of the fitted coefficients (a, b, c) is positive; this can always be arranged by multiplying the values of any variable by -1 wherever necessary, and corresponds to axis-reversal.

The optimisation problem (3) then becomes

$$\begin{aligned} &\text{minimise } \sum \frac{u_i^3 + v_i^3}{abc} \\ &\text{subject to} \quad u_i - v_i = ax_i + by_i + cz_i - k, \\ &\qquad\qquad\quad u_i, v_i \geqslant 0. \end{aligned} \tag{4}$$

We have now succeeded in formulating the problem as a particular type of fractional programming problem. The objective function is a ratio of functions with the numerator being non-negative and strictly convex and the denominator being concave and positive. This implies the objective function is strictly explicit quasi-convex (see (Stancu-Minasian, 1997, p. 56)). The constraints in (4) define a convex set. It then follows that any local minimum will be a global minimum, and, furthermore, that it will be unique (by theorems 2.3.5 and 2.3.6, respectively, in Stancu-Minasian (1997)). Clearly both of these properties are retained in higher dimensions, i.e., when we have more than three variables.

These are very valuable properties as it means that we can employ general-purpose optimisation software to seek the solution. In particular this includes the solvers, which are built in as standard in spreadsheet packages, hence the wider community can employ our method.

9. Historical notes

Before concluding we here bring together from various disciplines some of the appearances of the technique for the case of two variables. An early reference is that of Teissier (1948); however, the presence of a paper written in French in an English-language journal could not have helped its dissemination. A paper in English appeared soon after (Kermack and Haldane, 1950) in which the fitted line was referred to as the “reduced major axis”. Both of these deal with allometry, which is that branch of biology which studies the relative size and growth of one part of an organism relative to another part. In such investigations there will be variability in both measurements and there is usually no clear candidate for selecting one of the variables as being “dependent” and the other "independent". As a result the method was later strongly advocated for use in fishery research and biology, in general, by Ricker (1973). In biology the resulting line is also called the "line of organic correlation" (e.g., Kruskal (1953)).

Astronomy has always been a fruitful area for new quantitative techniques. Indeed the least squares method was introduced to deal with the orbits of heavenly bodies by Laplace, Legendre and Gauss in the early 1800's. (See the book by Farebrother (1999) for a history of fitting procedures prior to 1900.) The method which we have been discussing appeared in the astronomical literature in 1940 (Stromberg, 1940) hidden in a paper entitled "Accidental Systematic Errors in Spectroscopic Absolute Magnitudes for Dwarf G_oK_2 Stars". It was subsequently referred to as "Stromberg's impartial line". In cosmology Feigelson and Babu (1992) note that "accurate regression coefficients are crucial to measuring the expansion rate of the universe, estimating the age of the universe, and uncovering large-scale phenomena such as superclustering". The more distant a galaxy, the faster it moves away from us, but one cannot say that that the speed depends on distance or vice versa, hence the need for a method that treats variables symmetrically. They also mention calibration problems, where "one instrument or measurement technique is calibrated against another, with neither one being inherently a standard"; both measurement error and intrinsic scatter about the line are present.

Another field where our technique has been discussed is in economics. Woolley (1941) introduces it as "the method of minimised areas". Samuelson (1942) then commented that "this is nothing other than Frisch's 'diagonal regression' " and gives 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 going back to 1916 (Sverdrup, 1916).

10. Conclusion

We have considered the situation where we wish to estimate a relationship between variables where the data for every variable is treated on the same basis. We may wish to do this rather than use conventional least squares regression for a number of reasons. For instance, there may not be an obvious dependent variable (possibly because all the variables are the effects of a common cause which has not or cannot be measured), or all the variables have uncertainty associated with them but no information is available regarding the error variance so that the measurement error models approach cannot be used.

When only two variables are involved we have seen that work has been done to compare various techniques for estimating an underlying relationship, and that the technique which we here call neutral data fitting has shown itself to be desirable for a number of reasons. Analytically, it has been shown that it is units-invariant (or scale invariant), as well as being invariant to linear transformation of individual variables. It has also been demonstrated that it will always provide a slope value that lies between those given by the two OLS lines. The two OLS slopes will increasingly diverge as the correlation in the data falls, whereas with neutral data fitting the slope value is independent of the correlation. Numerical simulations also show that our technique is very good at unearthing the true underlying model.

In the light of these useful properties and what Draper and Smith (1998) call its "appealing natural symmetry", there is clearly a case for extending the method to the case of multiple variables. We did this by generalising the notion of area deviation to that of volume and hyper-volume deviations. We chose this avenue rather than that of the geometric mean property to ensure that the invariance properties are retained. Generalising the geometric mean property would involve taking all possible least squares regressions (taking each variable in turn to be dependent), and then estimating each coefficient as the geometric mean of the relevant coefficients from all the regressions. A potential obstacle is that it is not clear what sign to attach to a particular coefficient when one finds opposite signs arising from different regressions. Moreover, it is not clear if this approach would also possess the invariance properties discussed. Note that this is not the approach Draper and Yang (1997), rather they took the geometric mean of squared deviations in each dimension; they also showed that their solution for the coefficients is a convex combination of the separate least squares solutions.

Much work needs to be done to explore the method we have proposed. We intend to carry out Monte Carlo simulations to see how well it reproduces the underlying model in comparison with other fitting procedures. Initial results reported in Tofallis (2002b) show it to be superior to least squares in identifying the model. Furthermore, the coefficients were found to be much more stable to small changes in the data; such perturbations tend to have a large effect on a regression model if there is multicollinearity in the data.

Another challenge is to obtain a closed form expression for the coefficients; whilst we have this in the case of two dimensions, it may be too much to expect for the general case.

The appearance of the method for two variables (under the name "geometric mean functional relationship") in the latest edition of the widely known monograph on regression by Draper and Smith (1998) will hopefully do much to raise interest in a wider audience, particularly statisticians. The fact that most of the literature on the subject has appeared in the natural and social sciences is testament to its utility. It is, however, surprising that the statistical community has largely been unaware of this elegant and valuable technique. It is also noteworthy that it has attracted the attention of eminent researchers such as Kruskal and Haldane, as well as Nobel prize-winners Samuelson and Frisch, this too must commend its further investigation in higher dimensions.

References

Babu, G.J. and E.D. Feigelson. (1992). "Analytical and Monte Carlo Comparisons of Six Different Linear Least Squares Fits." *Communications in Statistics: Simulation and Computation* 21(2), 533–549.

Barker, F., Y.C. Soh, and R.J. Evans. (1988). "Properties of the Geometric Mean Functional Relationship." *Biometrics* 44, 279–281.

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

Cheng, C.-L. and J.W. Van Ness. (1999). *Statistical Regression with Measurement Error*. London: Arnold.

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

Draper, N.R. and Y. Yang. (1997). "Generalization of the Geometric Mean Functional Relationship." *Computational Statistics and Data Analysis* 23, 355–372.

Farebrother, R.W. (1999). *Fitting Linear Relationships: A History of the Calculus of Observations* 1750–1900. New York: Springer.

Feigelson, E.D. and G.J. Babu. (1992). "Linear Regression in Astronomy II." *Astrophysical J*. 397, 55–67.

Frisch, R. (1934). "Statistical Confluence Analysis by Means of Complete Regression Systems." University Institute of Economics, Oslo.

Kermack, K.A. and J.B.S. Haldane. (1950). "Organic Correlation and Allometry." *Biometrika* 37, 30–41.

Kruskal, W.H. (1953). "On the Uniqueness of the Line of Organic Correlation." *Biometrics* 9, 47–58.

Ricker, W.E. (1973). "Linear Regressions in Fishery Research." *J. Fisheries Research Board of Canada* 30, 409–434.

Ricker, W.E. (1975). "A Note Concerning Professor Jolicoeur's Comments." *J. Fisheries Research Board of Canada* 32, 1494–1498.

Riggs, D.S., J.A. Guarnieri, and S. Addelman. (1978). "Fitting Straight Lines when Both Variables Are Subject to Error." *Life Sciences* 22, 1305–1360.

Roberts, F.S. (1979). *Measurement Theory: With Applications to Decision Making, Utility and the Social Sciences*. Reading, MA: Addison-Wesley.

Samuelson, P.A. (1942). "A Note on Alternative Regressions." *Econometrica* 10(1), 80–83.

Stancu-Minasian, I.M. (1997). *Fractional Programming: Theory, Methods and Applications*. Dordrecht: Kluwer Academic.

Stromberg, G. (1940). "Accidental Systematic Errors in Spectroscopic Absolute Magnitudes for Dwarf G_0K_2 Stars." *Astrophysical J*. 92, 156ff.

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

Teissier, G. (1948). "La relation d'Allometrie: sa Signification Statistique et Biologique." *Biometrics* 4(1), 14–48.

Tofallis, C. (2002a). "Model Fitting for Multiple Variables by Minimising the Geometric Mean Deviation." In S. Van Huffel and P. Lemmerling (eds.), *Total Least Squares and Errors-in-Variables Modeling: Algorithms, Analysis and Application*. Dordrecht: Kluwer Academic.

Tofallis, C. (2002b). "Model Fitting Using the Least Volume Criterion." In J.C. Mason and J. Levesley (eds.), *Algorithms for Approximation IV*. University of Huddersfield Press.

Tukey, J.W. (1975). "Instead of Gauss–Markov Least Squares, What?" In R.P. Gupta (ed.), *Applied Statistics*. Amsterdam: North-Holland.

Van Huffel, S. (1997). *Recent Advances in Total Least Squares*. Philadelphia, PA: SIAM.

Van Huffel, S. and J. Vandewalle. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM.

Woolley, E.B. (1941). "The Method of Minimized Areas as a Basis for Correlation Analysis." *Econometrica* 9(1), 38–62.

VTEX(Viktorija) PIPS No:5146594 artty:res (Kluwer BO v.2002/10/03)
a5146594.tex; 10/10/2003; 12:28; p. 11