A Note on Alternative Regressions
Author(s): Paul A. Samuelson
Source: *Econometrica*, Vol. 10, No. 1 (Jan., 1942), pp. 80-83
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/1907024
Accessed: 11/06/2013 11:33

A NOTE ON ALTERNATIVE REGRESSIONS

By Paul A. Samuelson

In the January issue of this journal Mr. Elliott B. Woolley presented a method of determining a straight-line regression by minimizing the summed *absolute* values of the areas of rectangles formed by the projections of each observation upon the regression line.[1] The resulting line possesses the usual property of passing through the point of means, and its slope is a simple average of the elementary regression slopes derived by minimizing in each direction; it is the geometric mean of the elementary regression coefficients, each referred to the same axis, and has their algebraic sign. It should be pointed out that this is nothing other than Frisch's "diagonal" regression (cf. *Statistical Confluence Analysis* . . .), and a statistical parameter which has long appeared in the literature. In terms of a correlation surface it represents the major axis of the concentric ellipses of equal frequency.

While Mr. Woolley has made an interesting contribution in proving this "minimizing" property of the diagonal regression,[2] his further argument that it is to be preferred in any sense as a method of determining regression lines seems to require brief comment.

(a) The lack of consistency between the elementary regressions is a necessary property of a linear multivariate frequency surface. It is expressed in the purely formal statistical law of *regression towards the average*. The elementary regressions are not thereby "illogical."

(b) If the aim of the investigation is not simply a characterization of the properties of the multivariate distribution, but rather the search for a hypothetical "true" (in some sense) linear relationship, upon which has been superimposed a distribution of errors, then no definite method of determining the regression equation can be specified until some assumptions have been made concerning the nature of the disturbing causes. *These assumptions must be in the nature of postulates; by no possible method can they be determined inductively from an examination of the data, even in an infinitely large sample.* This last statement must be emphasized since some of the recent literature seems at first sight to suggest otherwise. This is because seemingly innocent, but in fact highly restrictive and often arbitrary, assumptions of "noncorrela-

[1] "The Method of Minimized Areas as a Basis for Correlation Analysis," Econometrica, Vol. 9, January, 1941, pp. 38–62; see also "Editorial Note," Econometrica, Vol. 9, July–October, 1941, p. 312.

[2] This holds only in the two variable case. The regression hyperplane which minimizes the sum of the absolute volumes of the parallelepipeds formed by the projections of an observation point upon the hyperplane is not equivalent to the "diagonal" regression plane when three or more variables are involved.

tion," etc., do succeed in determining allegedly "true" regressions or limits upon these. (Cf. the highly interesting treatment in T. Koopmans' *Linear Regression Analysis of Economic Time Series.*)

(c) Why then has Mr. Woolley argued in favor of the diagonal regression? Clearly because it is independent of the designation of one variable as independent and the other as dependent. It is consistent in his sense; namely, invariant under an interchange of variables. However, it is easy to show that an infinity of methods can be devised which have this property. Thus, minimizing the sum of the squares of the shortest distances between the points and the regression locus yields one such. Actually, minimizing any symmetric function of the residuals in all directions will do the same thing.

(d) There are various properties which a method of determining regression equations *might* desirably possess.

Property I. *For perfectly correlated variables the fitted curve should reduce to the correct equation.* All proposed methods possess this property.

Property II. *The fitted equation should be invariant under an interchange of variables.* An interchange of variables is a very special orthogonal transformation. (It represents a 90° rotation of axes followed by a reversal.) This suggests the following stronger condition:

Property III. *The fitted equation should be invariant under any orthogonal transformation of variables.* Another desirable property is the following:

Property IV. *The regression equation should be invariant under a simple dimensional or scale change in any of the variables.* A still stronger condition is implied in the following:

Property V. *The fitted equation should be invariant under any linear transformation of co-ordinates.* Of a different nature is the following condition:

Property VI. *The regression slope must depend only upon the correlation coefficient and ratios of standard deviations; i.e., it must be some function of the elementary regression coefficients alone.*

Still other properties could be enumerated, such as the obvious one that all coefficients should be symmetric functions of the observations, etc. But, are those listed consistent? Clearly, it is in principle impossible to satisfy Property V (unless the correlation is perfect). This means that *all* methods which yield a unique answer depend upon some privileged ("natural") choice of variables.

Properties I and VI together imply an alternative postulate: Property VII. *The fitted regression coefficient must be a (generalized) mean of the elementary regression coefficients, each related to the same axis.* For, when these are equal, Postulate I requires that the fitted coefficient be equal to the common true value.

The method of minimum perpendicular-squared-distances satisfies Properties I, II, III, and VI, but not IV. The proofs are obvious and are omitted. In the case of zero correlation with equal variance, it very properly breaks down, the regression coefficient becoming indeterminate. Near to this position the solution of the determining equation becomes increasingly sensitive, which should serve as a warning to the statistical computer.

The diagonal regression (in two variables) satisfies Properties I, II, IV, and VI, but not III. That it does not satisfy III is clear geometrically from the fact that a rotation of axes through any angle not a multiple of 90° will leave the rectangles whose summed areas are to be minimized in skewed positions relative to the new axes. Not only does the diagonal regression satisfy I, II, IV, and VI, but it is the *only* one which does so. This is proved simply as follows: Let B_{yx} be the fitted regression of y on x, b_{yx} and b_{xy} being the respective elementary regressions. Then by VI

$$B_{yx} = f(b_{yx}, b_{xy}). \tag{1}$$

II requires that

$$B_{yx} = \frac{1}{B_{xy}}$$

or

$$f(b_{yx}, b_{xy})f(b_{xy}, b_{yx}) \equiv 1. \tag{2}$$

Finally IV requires that

$$f(b_{yx}, b_{xy}) \equiv \frac{1}{\lambda} f\left(\lambda b_{yx}, \frac{b_{xy}}{\lambda}\right)$$

for all nonvanishing λ's; or

$$f(b_{yx}, b_{xy}) \equiv \frac{f(b_{yx}b_{xy}, 1)}{b_{xy}} = \frac{h(b_{yx}b_{xy})}{b_{xy}}. \tag{3}$$

Substitution of (3) into (2) gives

$$h(v)h(v) \equiv v,$$

or

$$h(b_{yx}b_{xy}) \equiv \pm\sqrt{b_{yx}b_{xy}} \tag{4}$$

and

$$B_{yx} \equiv \pm\sqrt{\frac{b_{yx}}{b_{xy}}} \equiv \pm\frac{\sigma_y}{\sigma_x}, \tag{5}$$

which is the diagonal regression. Properties I and VII are both corollaries.

(e) This and the following point are independent of the previous discussion, but relate to Mr. Woolley's paper. The reader of his paper will perhaps be interested to know that previous methods of curve fitting can be given a geometrical interpretation as minimizing areas. Thus the regression which minimizes the summed-perpendicular-squared-distances has the property of minimizing the sum of the areas of circles with radii equal to the distances from the line. (This illustrates its invariance under orthogonal transformation since a circle, unlike a quadrilateral, looks the same from all angles.) The elementary regressions are derived by minimizing the areas of circles whose centers lie at the projection of the observation upon the line and with radius equal to its distance from the line. Similarly, any weighted regression can be derived by averages of these respective circles or by the formation of ellipses whose principal axes are related to the weights involved.

(f) The minimizing property of the diagonal regression throws some light upon the ambiguity of sign necessarily involved when a square root is to be extracted. To each possible sign corresponds a different relative minimum of the function to be minimized. Choosing the sign of the correlation coefficient gives the lowest of these ($r \neq 0$). When r vanishes identically, both minimum positions take on equal values, and there is no way of choosing between them. (Taking their average by using the zero sign certainly has no sense since at this value the function to be minimized is infinite.) Unfortunately for *almost* zero correlation the solution of the equations, unlike the case mentioned in (d), gives no indication of proximity to indeterminacy.

Massachusetts Institute of Technology