USE OF AN EXTRANEOUS INFORMATION IN ESTIMATION OF FEED-MILK RELATIONSHIP SUBJECT TO COLLINEARITY

By

T. Jacob

Institute of Agricultural Research Statistics, New Delhi.

(Received : October, 1976)

Introduction

In estimation of regression relationships the researcher sometimes is faced with the problem of multi-collinearity, arising out of high inter-correlation between two or more of the explanatory variables. The topic has been discussed at length by several authors like Theil (1971), Goldberger (1964), Klein (1969), Johnston (1972), Huang (1970) and Rao (1975). Use of supplementary information has been often suggested as a way out for breaking the deadlock. This procedure has been dealt with in the paper with reference to estimation of parameters of a linear production function relating nutrients intake to milk production.

Procedure of Analysis

Consider the model

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+e$$

where y represents in kg. the milk yield per day of a cow, x_1 and x_2 respectively represent in kg. the daily intake of digestible crude protein (D.C.P.) and total digestible nutrients (T.D.N.) which form the major nutrients recommended by nutritionists (Morrison, 1950).

Consider the situation where x_1 and x_2 are highly correlated. It may be mentioned here that some workers use x_1 and $x_2^*=x_2-x_1$ as the two explanatory variables to reduce the collinearity. x_2^* is termed as digestible non-nitrogenous nutrients or D.N.N. (Morrison). The implicit estimates of β_1 and β_2 will still remain the same. Even with the new variables, the collinearity persists in the type of data dealt with here. While, if x_1 and x_2 are perfectly correlated, the regression parameters are not uniquely estimable, a high inter-correlation may produce large standard errors for the

estimates. How 'high' the correlation should be is a question which cannot be easily answered by any single criterion. If γ_{12}, the total correlation between x_1 and x_2 is 0.80 or greater (Huang, 1970) one may suspect collinearity. Klein has given a thumb rule that collinearity is tolerable if

$$\gamma_{12} < R$$

where R is the multiple correlation coefficient based on the fitted regression. A general procedure for testing the presence of multi-collinearity has been given by Farrar and Glauber (1967). The procedure gives indication as to which of the explanatory variables are most affected by multi-collinearity. For the case of two explanatory variables, as in the present study, the test of simple correlation coefficient will suffice. In the presence of collinearity, the individual coefficients may not be very precisely estimated (large standard error) even though the multiple correlation would be high. It may sometimes result in difficulties in the interpretation of the estimated coefficients as the value may be quite contrary to the expectations.

One procedure followed is to impose zero restriction on the basis of the observed data, *i.e.*, to drop the variable with which the other independent variable is highly correlated. This, however, may introduce specification bias. Another procedure similar to zero restriction is to bring in information extraneous to the sample and use the same to impose linear restriction on the parameters. For instance, if β_1 can be taken to be equal to $k\beta_2$ where k is determined on the basis of supplementary information, then the model becomes

$$\begin{aligned} y &= \beta_0 + k\beta_2 x_1 + \beta_2 x_2 + e \\ &= \beta_0 + \beta_2 X + e \end{aligned}$$

where $X = x_2 + kx_1$

Now regress y on X to estimate β_2 and finally by multiplying the estimate of β_2 by k, β_1 can be estimated (Goldberger, 1964). It is, however, not an easy task to obtain appropriate supplementary information for the purpose.

In the present context of the problem, let us examine the standards prescribed for feeding dairy cattle. Morrison, a pioneer in the field of animal nutrition undertook extensive studies by way of conducting feeding experiments with different levels of protein and energy and its effect on milk production. On the basis of the results of such studies supplemented by similar results from other investigations, feeding standards were formulated by him. The work was spread over a period of 40 years and the standards also underwent modifications before leading to the ones currently prescribed.

A perusal of the standards (Morrison, 1950) indicates the relative feed value of D.C.P. as 7.5 times (0.33/0.044) that of T.D.N. This is true for milk of 4.5% fat which is agreeable with the fat per cent of Indian cow milk. Even if the fat per cent is between 3·5 and 4.5, the ratio remains at 7·5. If the observed value is outside this range the ratio is only slightly altered. In such cases the ratio can be easily worked out by using the appropriate figures given in Morrison's tables. Feeding trials were conducted in India too at different research stations. The standards generally used by many workers in India are those given by Sen and Ray (1964) where also the relative value remains at 7·5.

Consider the foregoing aspects *vis-a-vis* the production function. β_1 gives the marginal product with respect to D.C.P. and β_2 the marginal product with respect to T.D.N. The ratio of the marginal productivities provides a basis for relative evaluation of feed values (Heady and Dillon, 1961) and this would lead to the formulation, namely, $\beta_1/\beta_2 = 7{\cdot}5$. Using this information, the parameters can be estimated as indicated earlier. Instead of x_1 and x_2, if x_1 and x_2^* (*see* page 85) are used as explanatory variables the ratio (k) works out to be 8·5.

ILLUSTRATION

Using data on the intakes of D.C P. (x_1) and D.N.N. (x_2^*) and milk output (y) of 16 cows collected under an I.A.R.S. survey, the following production function was fitted by the least-squares procedure:

$$y = -2{\cdot}945 + \underset{(5{\cdot}424)}{11.536}\ x_1 + \underset{(0{\cdot}607)}{0{\cdot}415}\ x_2^*$$

The multiple correlation coefficient worked out to be 0.915. The coefficient of x_1 just falls short of significance while that of x_2^* was found to be far from significance. A look for collinearity shows a high correlation of the order of 0.932 between x_i and x_2^*. Using the restriction $k = 8{\cdot}5$ the equation was refitted and is given by

$$y = -2{\cdot}760 + \underset{(0{\cdot}884)}{7{\cdot}432}\ x_1 + \underset{(0{\cdot}104)}{0{\cdot}874}\ x_2^*$$

The multiple correlation worked out to be 0.913 and the coefficients were found to be highly significant. It is desirable to convert x_2^* back to x_2 in order to conform to the conventional units of D.C.P. and T.D.N. while interpreting the results. The marginal productivities of D.C.P. and T.D.N. worked out to be 6.558 and 0.874 respectively.

REMARKS

It may be mentioned that if the supplementary information is correct, the least-squares estimates of β_1 and β_2 are unbiased and efficient. If not, the estimates will be biased, the extent of bias depending upon the divergence from the true value. But as pointed out by Huang (1970), in case of multicollinearity, one should be willing to tolerate some bias for a smaller variance (in the mean square sense) of the estimate with linear restrictions, if the restrictions are approximately true.

ACKNOWLEDGEMENT

The author is thankful to Dr. D. Singh, Director, Institute of Agricultural Research Statistics for the encouragement given in preparing the paper.

REFERENCES

[1] Farrar, D.E. and Glauber, R.R. (1967) : Multicollinearity in Regression Analysis : The problem revisited. Rev. of Economics and Statistics, Vol. 49.

[2] Goldberger, A.S. (1964) : Econometric Theory. Wiley, New York.

[3] Heady, E.O. and Dillon, J.L. (1961) : Agricultural Production Functions. Iowa State University Press, Iowa.

[4] Huang, D.S. (1970) : Regression and Econometric Methods. Wiley, New York.

[5] Johnston, J. (1972) : Econometric Methods. McGraw-Hill Co., Tokyo.

[6] Klein, L.R. (1969) : Introduction to Econometrics, Prentice Hall, New Delhi.

[7] Morrison, F.B. (1950) : Feeds and Feeding. The Morrison Publishing Co., New York.

[8] Rao, C.R. (1975) : Some thoughts on Regression and Prediction. Sankhya, Series C, Vol. 37, Part 2.

[9] Sen, K.C. and Ray, S.N. (1964) : Nutritive Values of Indian Cattle Feeds and Feeding of Animals. Res. Ser. No. 25, 5th Edn., Indian Council Agri. Res., New Delhi.

[10] Theil, H. (1971) : Principles of Econometrics. Wiley, New York.