REVIEW OF MARKETING AND AGRICULTURAL ECONOMICS
Vol. 45, No. 3 (*September*, 1977)

AGRICULTURAL ECONOMISTS, RESPONSE FUNCTIONS AND LACK-OF-FIT TESTS

George E. Battese*

The significance of statistical lack-of-fit tests for hypothesized response functions describing experimental data is traced in some of the recent agricultural economics literature. The statistical test is reviewed in a regression framework and it is shown that the test is related to testing that some of the coefficients in a linear regression model have zero values. Comments are made on the relationship between the lack-of-fit test and the coefficient of determination.

1 INTRODUCTION

Linear regression analysis has been an important tool for empirical research in agricultural economics for over three decades. The advent of large computers in the last two decades has lead to a marked increase in the use of regression methods in research. During this time, however, agricultural economists have not consistently given attention to certain statistical considerations in their empirical research. In particular, some agricultural economists have neglected to consider statistical lack-of-fit tests for their hypothesized models in the analysis of livestock and crop response data. It is not uncommon to find agricultural researchers using experimental data to estimate hypothesized response functions and proceed to somewhat intricate economic analyses provided the coefficient of determination is "reasonably large". Some researchers consider t-tests on individual coefficients, or F-tests on groups of coefficients, without first considering whether the model itself is an adequate representation of the experimental data.

In section 2 of the paper a review of some of the agricultural economics literature is presented in order to support the above claims. A review of the relevant statistical lack-of-fit test is presented in section 3. This test, which has been presented in basic statistical method books for many years is explicitly shown to be related to the well-known F-test for testing that a subset of the parameters of a linear model has zero values. The concluding section of the paper refers to an empirical example in which two response functions were estimated and the function with the larger coefficient of determination was judged not to adequately represent the experimental data, whereas the opposite conclusion was made for the second function. Finally, a few concluding remarks are made on the teaching of statistics to agricultural economists.

* Department of Economic Statistics, University of New England.

G 36493—2

2 AGRICULTURAL ECONOMISTS AND RESPONSE FUNCTIONS

The book, *Agricultural Production Functions*, by Heady and Dillon [11] has made a significant contribution to empirical agricultural research. It documents most of the important literature on production functions up to 1960 and has influenced research in agricultural response analysis since that time. On the matter of statistical lack-of-fit tests, for hypothesized response functions, however, statements that are ambiguous and unclear are sometimes made.

It appears that the first reference to traditional lack-of-fit testing in the book is on page 104 where it is stated—

> "If he (a researcher) feels that the initial hypothesis of functional form should be retained, he may follow this procedure, after applying a simple criterion such as that the lack of fit term is no larger than experimental error for the regression equation."

This statement, however, follows an earlier one dealing with the "Basis for Selection" of forms of production functions, namely—

> "Other related statistics which might be used as empirical criteria include the F ratio and the mean square of deviations from regression (lack of fit). A larger F ratio (the mean square due to regression divided by the mean square of experimental error) or the smaller mean square of deviations from regression (i.e., the lack of fit) is taken to indicate a model most appropriate for the particular set of experimental or sample observations." [11, p. 102.]

It is unfortunate that "lack of fit" was used in these statements because the natural interpretation appears to be that deviations from a regression line are indicative of inappropriate functional forms for response.

Chapter 4 on "Data Analysis for Production Function Analysis" contains no discussion of lack-of-fit tests. Most of the statistical tests that are reviewed assume that the true response model is known to be a linear model with independent and identically distributed normal errors. However, in chapter 5 on "Data Collection for Production Function Estimation" explicit discussion is given on statistical lack-of-fit tests in the context of analysis of experimental data having replicated treatments.

The authors state—

> "Without knowledge of the experimental error, the researcher could not satisfactorily compare the goodness of fit of the various hypothesized functions he might fit. He would have no way of knowing what proportion of the deviations from regression were due to experimental error and what to the intrinsic unsuitability of the fitted function." [11, p. 157.]

In the eight empirical chapters (chapter 8-15) in the book, six deal with livestock response data and the estimation of cumulative gain functions. The two chapters dealing with crop response to fertilizer give explicit attention to lack-of-fit testing (e.g., Heady and Dillon [11, pp. 479, 521, 529, 536, 539]). The difficulty of obtaining true replicates for estimation of gain functions from *ad libitum* livestock feeding trials is presumably the basic reason for the omission of lack-of-fit consideration in chapters 8-13. Throsby [18] states that up to 1961 there had been little application of production function (regression) techniques to Australian experimental data. He estimates a quadratic response function and states—

> "Before being justified in proceeding, however, it is necessary to apply several tests of significance to determine (a) how well it describes the original data, (b) whether the individual regression coefficients are significantly different from zero, and (c) whether in fact there was a significant increase in Y due to the treatments in the experiment." [18, p. 127.]

Throsby incorrectly claims that point (a) is achieved by computing the value of the multiple correlation coefficient (page 127). However, he introduces analysis of variance and traditional lack-of-fit considerations on page 131. The lack-of-fit statistic is not actually calculated in the analysis of variance table on page 132.[1] Throsby concludes his paper by stating that—

> "In the past, analysis of the results of such experiments have often used only the classical analysis of variance model, thereby disregarding the fact that the underlying physical or biological response is continuous rather than discrete. By adopting a functional approach in the analysis of data from many experiments, both the technical understanding and the reliability and extent of recommendation can be increased. This approach does not presume to supersede the analysis of variance, rather it supplements it." [18, p. 146.]

Despite the last sentence, Throsby did not explicitly show the relationship between the so-called "analysis-of-variance approach" and regression analysis.

It appears that since the early 1960's some agricultural economists have *abandoned* any analysis-of-variance (or lack of fit) considerations. Some statements by Dillon [7], [8] may have contributed to this shift in the methods of analysis of experimental data. For example, Dillon states—

> "Additional levels of a factor can substitute for replications of a particular level since additional observations help to locate the response curve more accurately whether they come from replications or extra factor levels." [7, p. 66.]

The statement that observations at different factor levels "substitute" for replications should not be taken to mean that replication is not important for response analysis. Indeed replicated observations and observations at different factor levels are required for quite different purposes. Replications enable experimental error to be estimated and treatment means to be estimated with greater precision, whereas observations at

---

[1] It is noted that if the "replicates" and "error" terms are combined to form the "experimental error" for testing purposes, then the lack-of-fit effect is significant at probability level of 0.10.

more treatment levels permit the estimation of response functions with a larger number of parameters.

The paper by soil scientists Colwell and Esdaile [6] reports analyses for several fertilizer trials involving two replications. The authors state that—

> "An examination of regression fits of the (exponential, square-root and quadratic) models to the data shows that statistically there is little difference in goodness of fit." [6, p. 419].

It is not clear, however, that statistical lack-of-fit tests were conducted to provide a basis for their claim. Colwell and Esdaile [6] estimate regression models with "replications" as a variable along with polynomial terms involving fertilizer inputs. They erroneously claim that the statistical significance of the replication coefficient in twelve of the forty-nine sites considered indicated "appreciable within-site variance". In fact, "within-site variance" is estimated with use of replicated observations at the same treatment levels.

The biometricians Williams and Baker [20], in a paper dealing with efficiency of estimation of response functions from factorial and so-called response-surface designs, summarize the recommended analysis of experimental data as follows—

> "Agricultural scientists using the response surface analysis employ the analysis of variance and its associated lack-of-fit mean square mainly to ensure that the statistical model proposed adequately represents the data, and, having been reassured on that count, make objective predictions using a prediction equation based on the regression coefficients." [20, p. 169].

Anderson [1] in a review paper designed for agricultural scientists makes no mention of the place of, or need for, replication in agricultural experiments for response analysis. Colwell [5], while stating that "R^2 . . . should not be used to assess goodness of fit or the reliability of a regression" [5, p. 191], does not explicitly present lack-of-fit tests to support the claim that the estimated quadratic function (in $\sqrt{N}$ and $\sqrt{P}$) was a "good representation of yield as a function of fertilizer application rates".

The above review is sufficient to indicate that agricultural economists have not consistently drawn attention to lack-of-fit tests in the analysis of experimental data. The present author has had a number of statistical consulting experiences in which agricultural economists dealing with experimental data were ignorant of the existence of such tests. In addition to the possible influence of some of the agricultural economics literature, this may be associated with certain deficiencies in the statistical training of agricultural economists. Comment on this is made in section 4 of the paper.

In order to draw attention to the statistical lack-of-fit test that can be applied in the analysis of experimental data involving replication, we present a review of the test in general notation in section 3.

3 STATISTICAL MODELS AND LACK-OF-FIT TESTS

Lack-of-fit tests for hypothesized response functions are presented in several statistics texts, *e.g.*, Draper and Smith [9, pp. 26–31]; Graybill [10, pp. 183–4]; Ostle and Mensing [14, pp. 208–210]; Rao [15, pp. 284–7]; and Snedecor and Cochran [17, pp. 456–9]. The tests are often presented in the context of considering particular models, such as linear or quadratic response functions with one control variable. The statistical lack-of-fit test is reviewed here with matrix notation familiar in linear statistical models.

We assume that in an experimental situation replicated observations are taken at the same treatment levels. Let Y_{ij} denote the j-th replication at the i-th treatment level, where $j = 1, 2, \ldots, n_i$ and $n_i \geqslant 1$ for all $i = 1, 2, \ldots, t$. The number of replications are not required to be equal, but we assume that for at least one of the t treatment levels n_i exceeds one. We assume that the response observations, Y_{ij}, are independently distributed as normal random variables with the same variance but possibly different means for the different treatment levels. For the i-th treatment level we denote the expected response by μ_i and the variance of responses by σ^2. That is, $E(Y_{ij}) = \mu_i$ for all $j = 1, 2, \ldots, n_i$ and var $(Y_{ij}) = \sigma^2$ for all i and j.

These model specifications are expressed in linear model form by

$$(1) \qquad Y = Z\mu + U$$

where $Y = (Y'_1, Y'_2, \ldots, Y'_t)'$; $Y'_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i\eta_i})$, $i = 1, 2, \ldots, t$; Z is the $(n \times t)$ matrix of the t indicator (dummy) variables for the t different treatment levels, where $n = \sum_{i=1}^{t} n_i$; μ is the $(t \times 1)$ vector of expected responses; and U is a $(n \times 1)$ vector of independently and identically distributed normal random variables with mean zero and variance σ^2.

Suppose that it is hypothesized that the expected responses for the different treatments are a linear function of the levels of several independent variables. This is expressed in matrix notation by

$$(2) \qquad \mu = X\beta$$

where X is a $(t \times k)$ matrix of rank $k < t$ of known constants; and β is a $(k \times 1)$ vector of unknown constants. The elements of the matrix X may be known functions of the experimental control variables that determine the different treatments.

It should be noted that both μ and β in equation (2) are vectors of unknown constants and what is required is a statistical test for the adequacy of the hypothesized relationship of (2).

If the hypothesis of (2) is true, then the model representation of the data is given by

$$(3) \qquad Y = (ZX)\beta + U.$$

It is evident that the first n_1 rows of ZX are the same as the first row of X, the next n_2 rows of ZX are the same as the second row of $X, \ldots,$ and the last n_t rows of ZX are the same as the last row of X.

The computations required for the lack-of-fit test for the linear model of (3) are given in the analysis of variance of table 1.

TABLE 1: *Analysis of Variance for Hypothesized Response Function*

Source of Variation	d.f.	Sums of squares*	$E(S.S.)$†
Between Treatments ..	$t - 1$	$\hat{\mu}'Z'Y - n(\overline{Y}..)^2$	$(t - 1)\sigma^2 + R_1$
Response Model ..	$k - 1$	$\hat{\beta}'W'Y - n(\overline{Y}..)^2$	$(k - 1)\sigma^2 + R_2$
Lack of Fit	$t - k$	$\hat{\mu}'Z'Y - \hat{\beta}'W'Y$	$(t - k)\sigma^2 + R_3$
Experimental Error ..	$n - t$	$Y'Y - \hat{\mu}'Z'Y$	$(n - t)\sigma^2$
Total	$n - 1$	$Y'Y - n(\overline{Y}..)^2$	

* Note that $\hat{\mu} = (Z'Z)^{-1}Z'Y$; $\hat{\beta} = (W'W)^{-1}W'Y$, where $W = ZX$; and $\overline{Y}..$ is the average of the n Y-responses.

† The remainder terms in the expectations of the sums of squares are defined by—

$$R_1 = \sum_{i=1}^{t} n_i\mu_i^2 - \left(\sum_{i=1}^{t} n_i\mu_i\right)^2/n;$$

$$R_2 = \mu'Z'W(W'W)^{-1}W'Z\mu - \left(\sum_{i=1}^{t} n_i\mu_i\right)^2/n;$$

$$R_3 = \mu'Z'MZ\mu, \text{ where } M = I - W(W'W)^{-1}W'.$$

Note that $R_1 = 0$ if and only if $\mu_i = \mu_j$ for all $i, j = 1,2, \ldots, t$;
$R_1 = R_2 + R_3$; and $R_3 = 0$ if $\mu = X\beta$, but $R_3 > 0$ if $\mu \neq X\beta$.

The "Between Treatments Sum of Squares" is decomposed into a sum of squares associated with the hypothesized response function and a residual component that is associated with the inadequacy of the hypothesized response model to describe the treatment effects. The "Experimental Error Sum of Squares" is often called the "Pure-Error Sum of Squares" because it yields estimates for the error variance, σ^2, independently of the hypothesized response model. The "Experimental Error Sum of Squares" is expressed in familiar algebraic form by $\sum_{i=1}^{t}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i.)^2$, where $\overline{Y}_i.$ is the average of the Y-responses for the i-th treatment.

It should be noted that the "Experimental Error Sum of Squares" is obtained as the residual sum of squares from the regression of the Y-responses on the t indicator variables for the treatments. The "Between Treatments Sum of Squares" is the difference between the "Total Sum of Squares" (corrected for the mean) and the "Experimental Error Sum of Squares". Further, the regression sum of squares for the hypothesized response model in Table 1 is obtained by subtracting the correction factor term from the regression sum of squares from regressing the Y-responses on the k independent variables of the model in equation (3). The lack-of-fit sum of squares is obtained as the difference between the "Between Treatments Sum of Squares" and "Response Model Sum of Squares".

The test statistic for testing if the hypothesized response model is an adequate representation of the experimental data is expressed as

$$(4) \qquad F = \frac{(\text{Lack-of-fit Sum of Squares})/(t - k)}{(\text{Experimental Error Sum of Squares})/(n - t)} .$$

The statistic, F, has F-distribution with degrees of freedom $(t - k)$ and $(n - t)$ if the hypothesis $\mu = X\beta$ is true.[2] A test of size α, where $0 < \alpha < 1$, of the hypothesized response function is defined by: "Reject the hypothesis $\mu = X\beta$ if and only if the value of F in equation (4) exceeds the $100(1 - \alpha)$-th percentage point of the $F(t - k, n - t)$ distribution". For example, suppose an experiment contains 16 different treatments, each is replicated twice, and it is hypothesized that the appropriate response function for the data is a linear regression model with 6 independent variables. In this case $t = 16$, $n_i = 2$ for all $i = 1,2, \ldots, 16$ and $k = 6$. The hypothesized model would be rejected (at the 10 per cent level of significance) if the value of F in (4) exceeded 2.03, which is the 90th percentage point for the $F(10, 16)$ distribution.

It is worth noting that if the experimental data were summarized so that only the "pure-error" variance estimate [the denominator of the right-hand side of (4)] and the average treatment responses, $\overline{Y}_i$, $i = 1,2, \ldots, t$, were available, then the lack-of-fit sum of squares for the test statistic, F, is obtained by the residual sum of squares from the weighted regression of the average responses on the independent variables in the hypothesized response model. That is, the average Y-responses and the corresponding rows of the matrix X are multiplied by the square root of the number of replications and an ordinary least-square regression is obtained for the transformed data.

Since agricultural economists having knowledge of basic regression analysis are familiar with the F-test for testing if some of the coefficients of a regression model have zero values, it is of interest to explicitly show the relationship of that F-test to the above lack-of-fit test. Since the t expected responses in μ cannot necessarily be described by linear combinations of a reduced number, k, of parameters as in equation (2), it is possible to express the expected resonses by

$$(5) \qquad \mu = X\beta + X^*\beta^*$$

where X^* is any $tx(t - k)$ matrix such that the augmented square matrix $(X{:}X^*)$ is nonsingular; and β^* is a $(t - k)$ column vector of parameters. The equation (5) states that the expected responses at the t different treatment levels can always be expressed in terms of the k X-variables of equation (2) and additional variables. Since equation (5) expresses the t

[2] This result follows with use of basic distribution theory required for hypothesis tests on a subset of the parameters of a linear statistical model (e.g., *see* Graybill [10], pp. 128–139).

expected responses in terms of t other parameters, it is not particularly informative, except in so far as it explicitly shows that $\mu = X\beta$ is true if and only if the elements of β^* are all zero.

If equation (5) is substituted into the linear model of equation (1), then we obtain the model

$$(6) \qquad Y = [Z(X{:}X^*)] \begin{pmatrix} \beta \\ \beta^* \end{pmatrix} + U.$$

It is readily seen that the residual sum of squares for the regression model of (6) is equivalent to the residual sum of squares for the regression model (1). Thus the test statistic for testing the hypothesis that the elements of β^* are zero is defined by equation (4). This establishes the relationship between the lack-of-fit test for a hypothesized response model and the well-known F-test for testing that a subset of parameters in a linear regression model has zero values.

We conclude this section by noting the relationship between the coefficient of determination, R^2, for a hypothesized response function and the lack-of-fit test. It is clear from Table 1 that the value of R^2 for the hypothesized response function (ratio of the "Response Model Sum of Squares" to the "Total Sum of Squares") is no larger than the ratio of the "Between Treatments Sum of Squares" to the "Total Sum of Squares". Thus the lack-of-fit test above can be considered a test of whether the R^2 for the given response model is *significantly less* than the largest possible value for any response model that seeks to describe the observed treatment effects.

4 CONCLUSIONS

In empirical analyses, one may find that a significant value for the lack-of-fit statistic (4) is obtained for a response function with a "large" value of the coefficient of determination. This is likely to occur in experimental situations in which the experimental error variance is relatively small. Conversely, one may estimate a response function that has a "small" value for the coefficient of determination and yet the lack-of-fit test indicates that the function adequately fits the experimental data.

The statistical analysis of a livestock-feeding experiment reported in Battese et al [3] considers lack-of-fit tests. The experimental data involved were generated by feeding pigs at a *sub ad libitum* level in proportion to the animals' body weights. The rations for a given pig in a given week were quantities of wheat and skim milk determined as constant proportions of the weight of the animal at the beginning of the week. Battese et al [3] estimated quadratic functions in terms of proportions of milk and grain to describe the total quantities of feed consumed during the feeding period. Coefficients of determinations of 0.9393 and 0.9557 were obtained for the quadratic milk and grain functions, respectively. However, the lack-of-fit statistic for the grain function (with the larger R^2 value) was significant at the 5 per cent level of significance, whereas the milk function fitted adequately. The authors did not estimate a new grain function before undertaking their economic analysis. However, the note by Townsley [19] prompted the present author to recognize that the significant lack-of-fit test for the grain function had a simple interpretation. Because of the experimental feeding regime (i.e., the grain fed was approximately a proportion of the milk fed) the total quantity of grain consumed would not be a quadratic function of the experimental diets given that the quadratic function was appropriate for describing the total quantity of milk consumed (*see* Battese [2]).

It is obvious that in many research studies that involve the estimation of linear regression functions it is not possible to perform a statistical lack-of-fit test for hypothesized models. However, agricultural economists seeking to analyze crop and livestock response data from well-designed experiments are advised to give attention to lack-of-fit tests for their hypothesized response functions. It appears that the ignorance of some agricultural economists about lack-of-fit tests may be due to their exposure to statistical methods (and regression in particular) being through standard texts on econometric methods, such as Johnston [13] or Rao and Miller [16]. These books deal with the basic elements of regression methods, but make no mention of methods of particular reference to experimental data. At the University of New England agricultural economists have in the past not been exposed to lack-of-fit testing for hypothesized response functions in courses on applied regression analysis. It is planned, however, to include such subjects in these courses in future years.

It is noted in conclusion that the subject of lack-of-fit testing of hypothesized response functions for experimental data is related to questions of preliminary tests of significance in statistics. At the present time this is a fruitful area of statistical research as is evident by the papers by Bock, Yancey and Judge [4], Johnson, Bancroft and Han [12] and references given in these papers.

REFERENCES

[1] ANDERSON, J. R., "Guidelines for Applied Agricultural Research: Designing, Reporting and Interpreting Experiments", this *Review*, Vol. 39, No. 3, 1971, pp. 3–14.

[2] BATTESE, G. E., Reply to "Optimal Rations for Pigs Fed Separated Milk and Grain: A Comment", *Journal of Agricultural Economics*, Vol. 20, 1969, p. 359.

[3] BATTESE, G. E., DULOY, J. H., HOLDER, J. M. and WILSON, B. R., "The Determination of Optimal Rations for Pigs Fed Separated Milk and Grain", *Journal of Agricultural Economics*, Vol. 19, 1968, pp. 355–364.

[4] BOCK, M. E., YANCEY, T. A. and JUDGE, G. G., "The Statistical Consequences of Preliminary Test Estimators in Regression", *Journal of the American Statistical Association*, Vol. 68, 1973, pp. 109–116.

[5] COLWELL, J. D., "Assessment of the Relative Values of Compound Nitrogen-Phosphorus Fertilizers for Wheat Production", *Australian Journal of Agricultural Economics*, Vol. 17, No. 3, 1973, pp. 189–199.

[6] COLWELL, J. D. and ESDAILE, R. J., "The Application of Production Function Analysis for the Estimation of Fertilizer Requirements of Wheat in Northern New South Wales", *Australian Journal of Experimental Agriculture and Animal Husbandry*, Vol. 6, No. 23, 1966, pp. 418–424.

[7] DILLON, J. L. "Economic Considerations in the Design and Analysis of Agricultural Experiments", This *Review*, Vol. 34, No. 2, 1966, pp. 64–75.

[8] DILLON, J. L., *The Analysis of Response in Crop and Livestock Production*, Pergamon Press, Oxford, 1968.

[9] DRAPER, N. R. and SMITH, H., *Applied Regression Analysis*, Wiley, New York, 1966.

[10] GRAYBILL, F. A., *An Introduction to Linear Statistical Models*, McGraw-Hill, New York, 1961.

[11] HEADY, E. O. and DILLON, J. L., *Agricultural Production Functions*, Iowa State University Press, Ames, 1961.

[12] JOHNSON, J. P., BANCROFT, T. A. and HAN, C. P., "A Pooling Methodology for Regression in Prediction", *Biometrics*, Vol. 33, No. 1, 1977, pp. 57–67.

[13] JOHNSTON, J., *Econometric Methods*, Second Edition, McGraw-Hill, New York, 1972.

[14] OSTLE, B. and R. W. MENSING, *Statistics in Research*, Iowa State University Press, Ames, 1975.

[15] RAO, C. R., *Linear Statistical Inference and Its Applications*, Second Edition, Wiley, New York, 1973.

[16] RAO, P. and R. L. MILLER, *Applied Econometrics*, Wadsworth, Belmont, 1971,

[17] SNEDECOR, G. W. and W. G. COCHRAN, *Statistical Methods*, Sixth Edition. Iowa State University Press, Ames, 1967.

[18] THROSBY, C. D., "Fitting Production Functions to Experimental Data", this *Review*, Vol. 29, No. 3, 1961, pp. 112–147.

[19] TOWNSLEY, R. J., "Optimal Rations for Pigs Fed Separated Milk and Grain: A Comment", *Journal of Agricultural Economics*, Vol. 20, 1969, pp. 357–359.

[20] WILLIAMS, R. J. and J. R. BAKER, "A Comparison of Response Surface and Factorial Designs in Agricultural Research", This *Review*, Vol. 36, No. 4, 1968, pp. 165–177.