A TEST FOR NONLINEARITY IN THE GENERAL MULTIPLE REGRESSION MODEL BASED UPON SQUARED DIFFERENCES BETWEEN PAIRS OF RESIDUALS

by

Nancy Irene Lyons

Institute of Statistics
Mimeograph Series No. 1039
Raleigh, N.C. 1975

TABLE OF CONTENTS

1	INTRODUCTION	1
2	REVIEW OF LITERATURE	8
3	TEST PROCEDURE	13
	3.1 Distance Function	13
	3.2 Construction of D_L	15
	3.3 Test Statistic	16
	3.4 Moments of r under H_0	20
	3.5 Distribution of r	26
	3.6 Bounds for the Critical Values of the Test Statistic	32
4	ALTERNATIVE HYPOTHESIS	37
	4.1 The Alternative Hypothesis	37
	4.2 Properties of r under H_A	38
	4.3 Method of Approximating the Power of the Test	42
5	LARGE SAMPLE PROPERTIES	46
	5.1 Large Sample Comparisons	46
	5.2 Asymptotic Normality under Pitman's Alternative	48
	5.3 Asymptotic Relative Efficiency of r and F	55
	5.4 Selection of Pairs for D_L	58
6	RESULTS OF SIMULATION STUDIES AND APPLICATIONS	60
	6.1 Accuracy of Approximations to the Critical Region Under H_0	60
	6.2 Power Studies	66
	6.3 Application of the Test Procedure	71
7	SUMMARY AND CONCLUSIONS	74
	7.1 Summary and Conclusions	74
	7.2 Suggestions for Future Research	77
8	REFERENCES	80
9	APPENDIX	83

CHAPTER 1
INTRODUCTION

The motivation for the following research arises from a problem in the field of sociology. Suppose a general multiple linear regression model is proposed to represent a relationship in which each observation of the dependent variable and k independent variables represent a set of measurements describing a particular collection of sociological characteristics of an individual. The general linear model is expressed as

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \tag{1.1.1}$$

where $\underline{y}' = (y_1, y_2, \cdots, y_n)$ is the vector of observations of the dependent variable. The matrix X is an $n \times k$ matrix of independent variables assumed to be fixed, $\underline{\beta}$ is a vector of k unknown parameters, and $\underline{\varepsilon}$ is a vector of n unobservable independent error terms each distributed normally with mean zero and variance σ^2, or, $\varepsilon \sim N(\underline{0}, I\sigma^2)$. It is assumed that the matrix X has full column rank and that a term representing a general mean is included in the model. This assumption implies that one column of the matrix X is an $n \times 1$ vector with each element equal to 1. Model (1.1.1) with its corresponding assumptions will be referred to as H_0.

The matrix X may be written as

$$X = \begin{bmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_n' \end{bmatrix} \tag{1.1.2}$$

where $\underline{x}_i' = (x_{i1}, x_{i2}, \cdots, x_{ik})$ is a $1 \times k$ vector for $i=1,2,\cdots,n$,[1] where the x_{ij}'s are expressed as deviations from the mean for each variable. Then the model (1.1.1) can be written alternatively as

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i, \quad i=1,2,\cdots,n, \tag{1.1.3}$$

where $\underline{\beta}' = (\beta_1, \beta_2, \cdots, \beta_k)$.

According to the methods of least squares, $\underline{\beta}$ is estimated under H_0 by $\underline{b} = (X'X)^{-1}X'\underline{y}$, and the vector $\underline{z} = \underline{y} - \hat{\underline{y}}$, where $\hat{\underline{y}} = X\underline{b}$, is called the vector of residuals. The vector $\underline{z}$ can be written as

$$\begin{aligned} \underline{z} &= \underline{y} - X\underline{b}, \\ &= (I - X(X'X)^{-1}X')\underline{y}, \\ &= M\underline{y}, \end{aligned} \tag{1.1.4}$$

where $M = (I - X(X'X)^{-1}X')$. Furthermore, since $MX = 0$,

$$\begin{aligned} \underline{z} &= M(X\underline{\beta} + \underline{\varepsilon}), \\ &= M\underline{\varepsilon} \end{aligned} \tag{1.1.5}$$

and $\underline{z}$ is therefore a random variable distributed as $N(\underline{0}, M\sigma^2)$.

In this situation a test is often needed for lack of fit of the proposed model which is sensitive to an alternative specification that the relationship between the dependent and the independent variables is expressed by the addition to the model of one or more terms each of which is a nonlinear function of one or more of the independent variables. The total number of observations is often small and there may be no replications at any combination of values of the independent variables.

---

[1] A typical vector from the set $\underline{x}_i$, $i=1,2,\cdots,n$, will often be referred to as $\underline{x}$.

If it is assumed that the alternative specification of the relationship is given by the addition into the model equation of a single term which is a nonlinear function in one of the k independent variables, then the usual procedure for testing the significance of a regression coefficient can be used. This procedure involves including the suspected term in the model equation and performing the partial F or t test that its coefficient is zero. If there is no indication from sources independent of the fitted model as to the functional form of the possible alternatives, then a common procedure is to test the significance of numerous probable alternative models. However, the interpretation of the significance levels of such a sequence of tests is difficult (see Ashar and Wallace (1972)).

The test commonly used against a general nonlinear alternative is the classical lack of fit test due to R. A. Fisher (1922). Under the above assumptions this procedure requires that there be repeated observations at one or more values of the independent variable. Within the framework of model (1.1.1) suppose there are $p < n$ distinct values of $\underline{x}_i$, which are denoted by $\underline{x}_1^*, \underline{x}_2^*, \cdots, \underline{x}_p^*$. Let each $\underline{x}_i^*$ occur n_i times with associated observations y_{ij}, $j=1,\cdots,n_i$, where $n \geq \sum_{i=1}^{p} n_i$. The error sum of squares (SSE) from fitting model (1.1.1) by the method of least squares has n-k degrees of freedom and can be partitioned into two component sums of squares. The first, called the pure error sum of squares (SSPE), is defined by

$$\mathrm{SSPE} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \,, \qquad (1.1.6)$$

where $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$, $i=1,\cdots,p$, and is associated with n-p degrees of freedom. The second sum of squares, called the lack of fit sum of squares (SSLF), is obtained as SSLF = SSE - SSPE and is associated with p-k degrees of freedom. The quantities SSPE/(n-p) and SSLF/(p-k) are independent, the former being an unbiased estimate of random error under both H_0 and H_A, and the latter being an unbiased estimate of σ^2 under H_0 but sensitive to lack of fit of the proposed model. The test statistic

$$F = \frac{SSLF/(p-k)}{SSPE/(n-p)} \tag{1.1.7}$$

is compared to tabulated percentage points of the $F_{(p-k,n-p)}$ distribution to determine the adequacy of the model. When the above assumptions are met, this is a UMP invariant test against a p parameter nonlinear alternative model (see Kendall and Stuart (1963)).

In practice if there are no replications but the number of observations is large enough, groups of observations at "nearly" identical values of the independent variable vector are formed and the F-type test of (1.1.7) is carried out assuming that observations within these groups are replicates. An example of the use of this modification is given in Daniel and Wood (1971). Corresponding to the notation in (1.1.6), suppose there are p disjoint sets G_i, $i=1,2,\cdots,p$, each of n_i near-replicates, with $n_i > 1$. For each $\underline{x}_{i_j} \varepsilon G_i$ let y_{i_j}, $j=1,2,\cdots,n_i$ be the associated observations. In this situation SSPE is now computed as

$$\text{SSPE} = \sum_{i=1}^{p} \sum_{i_j \varepsilon G_i} (y_{i_j} - \bar{y}_i)^2 \ ,$$

where $\bar{y}_i = \sum_{i_j \varepsilon G_i} y_{i_j} / n_i$, and is treated as though it had a chi-square distribution with n-p degrees of freedom under H_0. Otherwise the test is carried out in the same manner.

If replication does not exist, the test is no longer a likelihood ratio test since the actual distribution of the test statistic is altered. This alteration may result in a disruption of the significance levels of the test. Attempting to determine appropriate critical values by approximating the distribution of the test statistic is difficult due to the form of the function. In addition rules for forming the groups G_i can be very arbitrary. When the number of observations is small, it may not be possible to group together values of $\underline{x}$ in G_i which are close together relative to the overall dispersion of the set of data points, and it is reasonable to suspect that the optimal property of the F-type test might be affected when the distances between values within the groups G_i are large. For these reasons it becomes meaningful to consider alternatives to the F-type lack of fit test.

Although the responses associated with near-neighbor pairs of the independent variable can not be considered as being from the same population, the means of the two populations do not differ appreciably compared to means associated with an arbitrary pair of values of $\underline{x}$. The purpose of this investigation is to develop an alternative test procedure based upon a test statistic which is a function of pairs of residuals associated with pairs of near-neighbor values of $\underline{x}$. The basic functional unit is

$$b_{ij} = \frac{(z_i - z_j)^2}{2},$$

where z_i and z_j are the residuals associated with the pair $(\underline{x}_i, \underline{x}_j)$. Under H_0 the quantity b_{ij} is a biased estimate of σ^2, however the bias may be small for near-neighbor pairs relative to that of a b_{ij} associated with an arbitrary pair $(\underline{x}_i, \underline{x}_j)$. In the case of small samples, rather than ignore the distances between these pairs, the approach taken will be to quantify them and allow them to exert their relative influences on the test procedure. This is done by approximating the distribution of the test statistic using its moments under H_0. Analytic expressions for these moments can be found for the particular test statistic selected.

The test procedure is then compared with the modification of the F test in a two variable case as to its power against certain nonlinear alternatives under several patterns of dispersion of the independent variable values. A review of the relevant literature is given in Chapter 2. Chapters 3 and 4 deal with the test statistic and its basic properties under the null and general alternative hypotheses, respectively. In Chapter 3 an approximate test is derived for the case in which the set of near-neighbors consists of disjoint groups of equal size. The asymptotic relative efficiency of the test with respect to the F test is considered in Chapter 5. Data are simulated under several alternatives and discussion and presentation of results concerning the power as well as a demonstration of the application of the test procedure to some sociological data are included in Chapter 6. A summary of the results of the thesis and some suggested problems for future research are discussed in Chapter 7.

CHAPTER 2
REVIEW OF LITERATURE

One of the oldest tests for general lack of fit in the linear regression model is the one discussed in Chapter 1 due to R. A. Fisher. As do many of the classical methods, this test requires an independent estimate of random error about the model. This estimate is usually obtained from the presence of several observations of y at one or more values of $\underline{x}$. Alternatively if the sample is large enough, the $\underline{x}$-values are grouped so that several y-values are associated with each group, and an estimate of error is obtained within groups. Often neither of these two conditions can be met in practice especially for survey data.

When this is the case several approaches may be taken. One approach is to choose statistics which express the overall shape of the data points and are therefore sensitive to departures from the conditions in question. Included in these are graphical techniques such as the plotting of residuals versus fitted values of y (see Anscombe and Tukey (1963)). A second approach, which was introduced in Chapter 1, involves the formation of groups of near-replicates in order to obtain an estimate of the variance of random error. Still another approach is through the use of transformations. For example in the case of the BLUS (Best Linear Unbiased Scalar covariance matrix) residuals of Theil (1965), the residuals from the fitted regression are transformed into a set of n-k BLUS residuals with variance-covariance matrix equal to a scalar times the identity matrix. In a paper by Ramsey (1969) four tests based upon statistics which are functions of the BLUS residuals are developed to test against four basic types of departures from the assumed model. The approach taken in this investigation is to consider only the untransformed residuals. The effects of using transformed residuals could be considered in future research. In this chapter a review is given of a few selected techniques which depend upon different types of functions of the residuals from the fitted regression. It is assumed throughout that there are no replications of observations at any values of $\underline{x}$.

For obvious reasons, more work seems to have been done concerning the problem of lack of fit in the simple rather than the multiple regression model. Many problems concerning the handling of a multiple regression model are best attacked by considering first a simple regression model involving one dependent variable, possibly the residuals from a fitted regression, and one independent variable. Although his emphasis is in applications to design problems, Anscombe (1961) proposed the use of two statistics in this connection. These statistics, g_1 and g_2, are based upon the third and fourth power sums of the residuals and are estimates of the skewness and kurtosis, respectively, of the error distribution of the proposed model. Examples of approximate tests of significance are given. Tukey (1949) proposes a test which he calls "one degree of freedom for additivity", designed to detect a nonlinear relationship between the independent and dependent variables. The technique essentially involves examination of a regression of the residuals on the squares of the predicted y values, $\hat{\underline{y}}$. The test is based upon the statistic

$$f = \frac{(\underline{z}'\hat{\underline{y}}^2)^2}{\hat{\underline{y}}^{2\prime}M\hat{\underline{y}}^2} ,$$

where $\hat{\underline{y}}^{2\prime} = (\hat{y}_1^{\,2}, \hat{y}_2^{\,2}, \cdots, \hat{y}_n^{\,2})$. This quantity is subtracted from the residual sum of squares and compared with the remainder against the values of an F distribution with 1 and n-k-1 degrees of freedom.

In the simple regression case, Thornby (1972) develops a non-parametric test based upon the signs of the residuals for testing linearity of regression against convex alternatives. The procedure involves division of the independent variable axis into at least three disjoint intervals and recording the sign of the residuals in each group for comparison. Using the distribution theory of U statistics, he develops the test procedure and examines its behavior under various distributions of the error term.

The mean-square-successive-difference statistic based on residuals from a simple regression model is used for testing against an alternative of serial correlation between successive observations (see Durbin and Watson (1950), (1951)). The statistic is employed by Prais and Houthakker (1955) to test against nonlinearity. Although it is not mentioned in their discussion, the procedure would seem to give the best results in the case of nearly equally spaced values of the independent variable. In the case of equally spaced observations the statistic proposed in Chapter 3 is equivalent to that of Prais and Houthakker in the simple regression case. However the method of determining significant values of the statistic proposed here is different.

As mentioned earlier the second approach, which is basically the one taken here, is the determination of groups of near replicates through the introduction of an appropriate distance function defined on the k-dimensional space of the independent variable. In a method suggested by Daniel and Wood (1971), an estimate of pure error is obtained from a set of near-neighbors which is determined by the following interpoint distance function,

$$WSSD_{jj'} = \sum_{i=1}^{k} \left[\frac{b_i(x_{ij} - x_{ij}')}{S_y} \right]^2 ,$$

where b_i, $i=1,\cdots,k$, are the estimates of the regression coefficients, $j, j'=1,2,\cdots,n$, $j \neq j'$, and S_y is the error mean square from the fitted regression equation. The first step in the procedure is to order the observations according to the values of $\hat{y}$. Then $WSSD_{jj'}$ is computed for the 4n-10 pairs of points separated by 0, 1, 2, 3, and 4, $\hat{y}$ values. A running average estimate of σ, based on the assumption that the residuals could be treated as if they were independent normal observations, is obtained as

$$S_n(a) = 0.886 \sum_{D(a)} \frac{|z_j - z_{j'}|}{n(a)}$$

where $\sum_{D(a)}$ indicates summation over the first a of the 4n-10 pairs of residuals in increasing order of $WSSD_{jj'}$. The point a at which the value of $S_n(a)$ "stabilizes" is taken as the estimate of σ. This value is then compared with the estimate of σ obtained from the error sum of squares from fitting the original regression model. No explicit decision rule is given either for selecting the value of a or for determining significant values of the ratio. Furthermore Daniel and Wood give no consideration to the possibility that a high positive correlation may exist between the two estimates of σ. The presence of such correlation would tend to bias the results of the test toward acceptance of the model.

In the following investigation relatively few assumptions are made in developing the test procedure. Subsequently the set of alternatives is very large, and the conclusions that can be drawn as to the true relationship when the null hypothesis is rejected are necessarily limited.

CHAPTER 3
TEST PROCEDURE

3.1 Distance Function

In order to determine the set of near-neighbors in the k-dimensional space of the independent variable, a distance function must be chosen which is defined between every possible pair of points $(\underline{x}_i, \underline{x}_j)$. The Mahalanobis generalized interpoint distance, D, is chosen as the criterion. For two sample points $\underline{x}_i$ and $\underline{x}_j$, this function is defined by

$$D(\underline{x}_i, \underline{x}_j) = [(\underline{x}_i - \underline{x}_j)' S^{-1} (\underline{x}_i - \underline{x}_j)]^{1/2} \qquad (3.1.1)$$

where $S^{-1} = (X'X/n)^{-1} = nC_n^{-1}$, with $C_n = X'X$. When X is of full column rank, X'X is nonsingular and positive definite and therefore D is a metric on R^k.[2]

The interest in this investigation lies in detecting non-linearity in one or more of the independent variables. Therefore the distance function for determining the near-neighbor vector pairs should weight the independent variables according to the relative spread of the data points for each component of the vector $\underline{x}$. To illustrate this property of the Mahalanobis distance function it is helpful to begin by making the assumption that

$$\lim_{n\to\infty} (X'X/n) = \Sigma \qquad (3.1.2)$$

where Σ is a $k \times k$ matrix of constants. If the vectors $\underline{x}_i$ are orthogonal, the matrix C_n is diagonal for each n. As n increases

---

[2]See Lancaster (1969).

C_n/n approaches a matrix which must itself be diagonal. The set of all points $\underline{u}$ in R^k which are a constant distance d_0 from a vector $\underline{x}_0$ forms a k-dimensional ellipsoid around $\underline{x}_0$. For fixed n the length of the i-th axis measured on the Euclidean scale E_{0i} is $d_0/\sqrt{n(C_n^{-1})_{ii}}$, where $(C_n^{-1})_{ii}$ is the i-th diagonal element of C_n^{-1}, $(i=1,\cdots,k)$. As n increases the length of the corresponding axis on the Mahalanobis scale tends to the value

$$\frac{E_{0i}}{\sqrt{(\Sigma)_{ii}}},$$

where $(\Sigma)_{ii}$ is the corresponding element of Σ. Therefore in the limit the Mahalanobis Distance weights the relative influence of each component of a vector as if each were measured on the same scale. If the vectors $\underline{x}$ are not orthogonal, the crossproducts of the components are also taken into account in the weighting imposed by D.

The function D is invariant to changes in the scale of the data points. If $\underline{x}_i^* = \lambda\underline{x}_i$, $i=1,\cdots,n$, and $\lambda > 0$, then

$$D(\underline{x}_i^*,\underline{x}_j^*) = [\lambda^2(\underline{x}_i-\underline{x}_j)'(\lambda^2 C_n/n)^{-1}(\underline{x}_i-\underline{x}_j)]^{1/2}$$

$$= D(\underline{x}_i,\underline{x}_j).$$

If the set of all possible unordered pairs of the vectors $\underline{x}_i$ $(i=1,2,\cdots,n)$ is denoted by Π, then the elements of the set Π may be ordered by D. The ordering imposed by D is unchanged by any transformation of the form

$$aD^b, \qquad a,b > 0.$$

Therefore to simplify matters when the actual calculations are performed (*e.g.*, in computer program routines), the function

$$d(\underline{x}_i,\underline{x}_j) = (\underline{x}_i - \underline{x}_j)'C_n^{-1}(\underline{x}_i - \underline{x}_j)$$

$$= (1/n)D(\underline{x}_i,\underline{x}_j)^2$$

is often used when the concern is only with the ordering of the pairs, $(\underline{x}_i,\underline{x}_j)$.

3.2 Construction of D_L.

Let the set of near-neighbor pairs of the vectors $\underline{x}_i$ be denoted by D_L, and let the number of distinct unordered pairs in D_L be denoted by n_L. With respect to the construction of D_L it is useful to introduce the notion of the index of a point. Let the index I_i of a point $\underline{x}_i$, $(i=1,2,\cdots,n)$ be defined as the number of distinct unordered pairs in D_L for which $\underline{x}_i$ is one of the coordinates. It is clear that the index of any point $\underline{x}$ can take values $0,1,2,\cdots,n-1$. In forming D_L it may be necessary to impose a uniform bound on the indices of $\underline{x}$. In other words the restriction

$$I_i \leq I_L, \qquad i=1,2,\cdots,n$$

may be imposed. Alternatively an upper bound (N_L) may be placed upon the size of D_L or a maximum value d_L may be placed upon the distance allowed between a pair of elements in D_L. Then the pairs in Π are ordered by their value of D. In order, beginning with the pair with smallest interpoint distance, all the pairs are included into D_L until one or more of the three restrictions causes the inclusion process to stop, *i.e.*, until either $n_L > N_L$ or

$D > d_L$, or the index of every point is equal to I_L. Specifically, let D_L be written as $D_L = D(N_L, d_L, I_L)$, where N_L, d_L and I_L are respectively, the maximum values allowed for the number of pairs in D_L, the distance between any pair in D_L, and the index of any point $\underline{x}$. Then $D_L = D(N_L,-,-)$ implies that D_L is restricted only by size, $D_L = D(-,d_L,-)$ implies that D_L is restricted only by point indices. Any combination of these three restrictions may also be imposed, yielding a total of seven possibilities.

The set D_L will be called the set of near-neighbor pairs of the vectors $\underline{x}_i$, $i=1,2,\cdots,n$. For convenience in notation D_L considered interchangeably with the set of subscripts of its elements, *i.e.*, alternatively,

$$D_L = \{(i,j) \mid (\underline{x}_i, \underline{x}_j) \text{ is a near-neighbor pair of the set } \underline{x}_i,\ i=1,2,\cdots,n\}$$

3.3 Test statistic

The test statistic proposed in this investigation is

$$r = \frac{n-k}{2n_L} \frac{\sum\limits_{(i,j)\epsilon D_L} (z_i - z_j)^2}{\sum\limits_{i=1}^{n} z_i^2}, \tag{3.3.1}$$

where z_i and z_j are the residuals from the fitted regression model which correspond to the pair $(\underline{x}_i, \underline{x}_j)$. In the case that I_L of (3.2.1) is 2 (*i.e.*, $D_L = D(-,-,2)$) while $k=1$ and the observations are equally spaced, this statistic reduces to the mean-square-successive-difference statistic discussed in Chapter 2.

Suppose the argument $\underline{z}$ of the statistic (3.3.1) is replaced by a vector $\underline{y}$ of independently distributed random normal variables whose observations are taken simultaneously with an equally spaced variable $\underline{t}$ (e.g., time). In this situation Kamat (1963) considers several alternatives to (3.3.1) including

$$\frac{\sum_{i=1}^{n-1} |y_i - y_{i+1}|}{\sqrt{n-1}\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{3.3.2}$$

and

$$\frac{(n-1)}{(n-2)} \frac{\sum_{i=1}^{n-2} (y_i - 2y_{i+1} + y_{i+2})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{3.3.3}$$

in testing independence versus pure quadratic and sinusoidal trend. Using the criterion of asymptotic relative efficiency, Kamat concludes that the form (3.3.1) is the best among the alternatives.

Both the numerator and demoninator of (3.3.1) are quadratic forms in $\underline{z}$, so that r may be represented in matrix notation as the ratio

$$r = \frac{(n-k)}{2n_L} \frac{\underline{z}' P_L \underline{z}}{\underline{z}'\underline{z}}, \tag{3.3.4}$$

where the matrix P_L is defined elementwise by

$$(P_L)_{ij} = \begin{cases} -1, & \text{if } (\underline{x}_i, \underline{x}_j) \varepsilon D_L,\ i \neq j \\ 0, & \text{if } (\underline{x}_i, \underline{x}_j) \notin D_L,\ i \neq j \end{cases}$$

$$(P_L)_{ii} = \left\{ -\sum_{\substack{j=1 \\ j\neq i}}^{n} (P_L)_{ij},\ i=1,2,\cdots,n. \right.$$

Alternatively, using (1.1.5),

$$r = \frac{(n-k)}{2n_L} \frac{\underline{\varepsilon}'MP_LM\underline{\varepsilon}}{\underline{\varepsilon}'M\underline{\varepsilon}} . \tag{3.3.5}$$

Since M is indempotent, M and MP_LM commute with each other. Therefore there exists an orthogonal transformation which simultaneously reduces the matrices of the numerator and demoninator of r to a diagonal matrix and the identity matrix respectively. If $\underline{\varepsilon} = H\underline{\varepsilon}^*$ represents this orthogonal transformation, r becomes

$$r = \frac{(n-k)}{2n_L} \frac{\underline{\varepsilon}^{*\prime}V\underline{\varepsilon}^*}{\underline{\varepsilon}^{*\prime}\tilde{I}\underline{\varepsilon}^*} , \tag{3.3.6}$$

where V is an $n \times n$ diagonal matrix whose elements are the eigenvalues of MP_LM, and $\tilde{I} = \begin{pmatrix} I_{n-k} & 0 \\ 0 & 0 \end{pmatrix}$. Let the rank of MP_LM be denoted by m. Then the rank of V is also m, where m is at most equal to n-k, implying that at least k of the eigenvalues of V are zero. Denote the remaining n-k eigenvalues by $\nu_1,\nu_2,\cdots,\nu_{n-k}$, where n-k-m of these are also zero. Since H is an orthogonal matrix, the vector $\underline{\varepsilon}^*$ is distributed as $\underline{\varepsilon}^* \sim N(\underline{0}, I_n\sigma^2)$, and without loss of generality the statistic r can be considered to have the form

$$r = \frac{(n-k)}{2n_L} \frac{\sum_{i=1}^{n-k} \nu_i\varepsilon_i^{*2}}{\sum_{i=1}^{n-k} \varepsilon_i^{*2}} \tag{3.3.7}$$

under H_0.

Any ratio of the form (3.3.6) is known as a Rayleigh quotient and has the following useful properties (see Lancaster (1969)).

Theorem 3.3.1. The quotient $r = \frac{\underline{\varepsilon}^{*\prime}V\underline{\varepsilon}^*}{\underline{\varepsilon}^*\tilde{I}\underline{\varepsilon}^*}$ is invariant to changes in scale of the vector $\underline{\varepsilon}^*$, i.e., if $\underline{w}^* = \lambda\underline{\varepsilon}^*$, $\lambda \neq 0$, then

$$\frac{\underline{w}^{*\prime} V \underline{w}^*}{\underline{w}^{*\prime} \tilde{I} \underline{w}^*} = \frac{\lambda^2 \underline{\varepsilon}^{*\prime} V \underline{\varepsilon}^*}{\lambda^2 \underline{\varepsilon}^{*\prime} \tilde{I} \underline{\varepsilon}^*} = \frac{\underline{\varepsilon}^{*\prime} V \underline{\varepsilon}^*}{\underline{\varepsilon}^{*\prime} \tilde{I} \underline{\varepsilon}^*} \tag{3.3.8}$$

Theorem 3.3.2. If ν_1 and ν_{n-k} are the smallest and largest eigenvalues, respectively, of V, then for every vector $\underline{\varepsilon}^* \neq 0$,

$$\nu_1 \leq \frac{\underline{\varepsilon}^{*\prime} V \underline{\varepsilon}^*}{\underline{\varepsilon}^{*\prime} \tilde{I} \underline{\varepsilon}^*} \leq \nu_{n-k} \tag{3.3.9}$$

In terms of the distribution theory of r, equation (3.3.8) implies that the value of r is independent of σ^2 under H_0. The inequality (3.3.9) guarantees that for fixed n all finite moments of r exist since r is bounded by two constant functions.

3.4 Moments of r under H_0.

Analytic expressions for the moments of r can be obtained by using the following theorem due to Pitman (1937).

Theorem 3.4.1 (Pitman). Let the random variable w have continuous distribution with density function

$$f_m(w) = \begin{cases} \frac{1}{\Gamma(m)} e^{-w} w^{m-1} & 0 \leq w < \infty \\ 0 & \text{o.w.} \end{cases}$$

where $m > 0$ and $\Gamma(m)$ denotes the gamma function with parameter m. Let $w_1, w_2, \cdots, w_n$ be n independent random variables each with $w_i \sim f_{m_i}(w_i)$, $i=1,2,\cdots,n$, with the m_i's not necessarily equal. Let $H(w_1, w_2, \cdots, w_n)$ be any function of the w_i's which is independent of scale, i.e.,

$$H(\lambda w_1, \lambda w_2, \cdots, \lambda w_n) = H(w_1, w_2, \cdots, w_n)$$

for fixed λ. Then the variables $h(w_1, w_2, \cdots, w_n) = \sum_{i=1}^{n} w_i$ and $H(w_1, w_2, \cdots, w_n)$ are independent. The proof can be found in Pitman (1937).

Let $r = v/h$ where $v = \underline{\varepsilon}^{*}{}'V\underline{\varepsilon}^{*}$ and $h = \underline{\varepsilon}^{*}{}'\tilde{I}\underline{\varepsilon}^{*}$. Under the null hypothesis, each $\underline{\varepsilon}_i^{*2}$ is distributed as $\sigma^2\chi_1^2$, for each i. By property 3.3.7 the function r is independent of scale. Therefore by Pitman's theorem h and r are independently distributed under H_0. In this case for any positive integer s,

$$E_{H_0}(r^s) = \frac{E_{H_0}(v^s)}{E_{H_0}(h^s)} \qquad (3.4.1)$$

Obtaining the moments of r simplifies to that of obtaining the moments of two quadratic forms in independent standard normal variables. In order to evaluate these moments it will be useful at this point to state three well-known results concerning quadratic forms and matrices. The proofs of the first two theorems can be found in Searle (1970), the proof of the last can be found in Lancaster (1969).

Theorem 3.4.2 If $\underline{w} \sim N(\underline{\mu},Q)$, then the s'th cumulant of $\underline{w}'A\underline{w}$ is

$$K_s = 2^{s-1}(s-1)![tr(AQ)^s + s\underline{\mu}'A(QA)^{s-1}\underline{\mu}].$$

In the case that s=1 it is not necessary that $\underline{w}$ be normally distributed.

Theorem 3.4.3 If $\underline{w} \sim N(\underline{\mu},Q)$, then

$$Cov(\underline{w}'A\underline{w},\underline{w}'B\underline{w}) = 2tr(AQBQ) + 4\underline{\mu}'AQB\underline{\mu}.$$

Theorem 3.4.4 If $\lambda_1,\lambda_2,\cdots,\lambda_n$ are the eigenvalues of the matrix A, then for any positive integer s,

$$\sum_{i=1}^{n} \lambda_i^s = tr(A^s).$$

Using Theorems 3.4.1 - 3.4.4, the expected value of r under H_0, $E_{H_0}(r)$ (or μ_{r0}), is equal to

$$\mu_{r0} = \frac{n-k}{2n_L}\frac{tr(V)\sigma^2}{tr(M)\sigma^2} = \frac{tr(V)}{2n_L} \tag{3.4.2}$$

$$= \frac{(n-k)}{2n_L}\frac{\sum_{i=1}^{n-k}\nu_i}{n-k} = \frac{(n-k)}{2n_L}\bar{\nu},$$

where $\bar{\nu} = \frac{\sum_{i=1}^{n-k}\nu_i}{n-k}$. Note that μ_{r0} does not involve the parameter σ^2.

The moments of r about μ_{r0} can be derived using a technique described in Durbin and Watson (1950).

$$\frac{2n_L}{n-k}(r - \mu_{r0}) = \frac{\sum_{i=1}^{n-k} \nu_i \varepsilon_i^{*2}}{\sum_{i=1}^{n-k} \varepsilon_i^{*2}} - \frac{\sum_{i=1}^{n-k} \nu_i}{(n-k)}$$

$$= \frac{\sum_{i=1}^{n-k} \nu_i \varepsilon_i^{*2}}{\sum_{i=1}^{n-k} \varepsilon_i^{*2}} - \frac{\sum_{i=1}^{n-k} \nu_i \cdot \sum_{i=1}^{n-k} \varepsilon_i^{*2}}{(n-k) \sum_{i=1}^{n-k} \varepsilon_i^{*2}}$$

$$= \frac{\sum_{i=1}^{n-k} (\nu_i - \bar{\nu}) \varepsilon_i^{*2}}{\sum_{i=1}^{N-k} \varepsilon_i^{*2}}$$

$$\frac{\sum_{i=1}^{n-k} \xi_i \varepsilon_i^{*2}}{\sum_{i=1}^{n-k} \varepsilon_i^{*2}}, \text{ say.} \tag{3.4.3}$$

As before since $\underline{\varepsilon}^* \sim N(\underline{0}, I\sigma^2)$, $r - \mu_{r0}$ is of the required form for the application of Pitman's Theorem and the central moments of r can be written as

$$E_{H_0}(r - \mu_{r0})^s = \left(\frac{n-k}{2n_L}\right) \frac{E_{H_0}\left(\sum_{i=1}^{n-k} \xi_i \varepsilon_i^{*2}\right)^s}{E_{H_0}\left(\sum_{i=1}^{n-k} \varepsilon_i^{*2}\right)^s}, \tag{3.4.4}$$

which is the s'th central moment of the numerator of 3.4.3 over the s'th noncentral moment of its denominator. Thus the next three higher moments of r are as follows (Durbin and Watson (1950)):

$$\mathrm{Var}_{H_0}(r) = \left(\frac{n-k}{2n_L}\right)^2 \frac{2\sum_{i=1}^{n-k}(\nu_i - \bar{\nu})^2}{(n-k)(n-k+2)} ,$$

$$E_{H_0}(r - \mu_{r0})^3 = \left(\frac{n-k}{2n_L}\right)^3 \frac{8\sum_{i=1}^{n-k}(\nu_i - \bar{\nu})^3}{(n-k)(n-k+2)(n-k+4)} ,$$

$$E_{H_0}(r - \mu_{r0})^4 = \left(\frac{n-k}{2n_L}\right)^4 \frac{48\sum_{i=1}^{n-k}(\nu_i - \bar{\nu})^4 + 12\left[\sum_{i=1}^{n-k}(\nu_i - \bar{\nu})^2\right]^2}{(n-k)(n-k+2)(n-k+4)(n-k+6)} .$$

The first moment of r under H_0 is particularly interesting since it can be expressed as an explicit function of the distances between the pairs in D_L. Let the matrix $C_n = X'X$ be written as $C_n = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i'$, and let the matrix M be rewritten as a matrix of row vectors, i.e.,

$$M = \begin{bmatrix} \underline{m}_1' \\ \underline{m}_2' \\ \vdots \\ \underline{m}_n' \end{bmatrix} \tag{3.4.5}$$

where $\underline{m}_i$, $i=1,2,\cdots,n$ are $n \times 1$ vectors. Elementwise M is given by

$$\begin{bmatrix} 1-\underline{x}_1'C_n^{-1}\underline{x}_1 & -\underline{x}_1'C_n^{-1}\underline{x}_2 & \cdots & -\underline{x}_1'C_n^{-1}\underline{x}_n \\ -\underline{x}_2'C_n^{-1}\underline{x}_1 & 1-\underline{x}_2'C_n^{-1}\underline{x}_2 & \cdots & -\underline{x}_2'C_n^{-1}\underline{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ -\underline{x}_n'C_n^{-1}\underline{x}_1 & -\underline{x}_n'C_n^{-1}\underline{x}_2 & \cdots & 1-\underline{x}_n'C_n^{-1}\underline{x}_n \end{bmatrix} \tag{3.4.6}$$

or $$m_i = \begin{cases} 1-\underline{x}_i'C_n^{-1}\underline{x}_i, & \text{if } i = j \\ -\underline{x}_i'C_n^{-1}\underline{x}_j, & \text{if } i \neq j \end{cases} .$$

Since MP_LM is orthogonally similar to V, $\sum_{\alpha=1}^{n-k} \quad = tr(V)$.

Therefore,

$$MP_LM = \begin{bmatrix} \underline{m}_1' \\ \underline{m}_n' \\ \cdot \\ \cdot \\ \cdot \\ \underline{m}_n' \end{bmatrix} P_L(\underline{m}_1, m_2, \cdots, \underline{m}_n)$$

and

$$tr(MP_LM) = \sum_{\alpha=1}^{n} \underline{m}_\alpha' P_L \underline{m}_\alpha .$$

By the construction of P_L, $\underline{m}_\alpha' P_L \underline{m}_\alpha = \sum_{(i,j)\varepsilon D_L} (m_{\alpha i} - m_{\alpha j})^2$

Therefore, $$tr(MP_LM) = \sum_{\alpha=1}^{n} \sum_{(i,j)\varepsilon D_L} (m_{\alpha i} - m_{\alpha j})^2$$

$$= \sum_{(i,j)\varepsilon D_L} \sum_{\alpha=1}^{n} (m_{\alpha i} - m_{\alpha j})^2 .$$

Consider only one of the terms $\sum_{\alpha=1}^{n} (m_{\alpha i} - m_{\alpha j})^2$. This equals

$$\sum_{\alpha\neq i,j} (m_{\alpha i} - m_{\alpha j})^2 + (m_{ii} - m_{ij})^2 + (m_{ij} - m_{jj})^2 ,$$

$$= \sum_{\alpha\neq i,j} [-(\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_\alpha]^2 + [1 - (\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_i]^2$$

$$+ [1 + (\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_j]^2 ,$$

$$= \sum_{\alpha\neq i,j} (\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_\alpha\underline{x}_\alpha'C_n^{-1}(\underline{x}_i - \underline{x}_j) + 1$$

$$+ (\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_i\underline{x}_i'C_n^{-1}(\underline{x}_i - \underline{x}_j) - 2(\underline{x}_i-\underline{x}_j)'C_n^{-1}\underline{x}_i$$

$$+ 1 + (\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_j\underline{x}_j'C_n^{-1}(\underline{x}_i - \underline{x}_j) + 2(\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_j ,$$

$$= \sum_{\alpha=1}^{n} (\underline{x}_i - \underline{x}_j)'C_n^{-1}(\underline{x}_i - \underline{x}_j) + 2$$

$$+ 2(\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_j - 2(\underline{x}_i - \underline{x}_j)'C_n^{-1}\underline{x}_i .$$

By the distributive property of matrices it is also equal to

$$(\underline{x}_i - \underline{x}_j)'C_n^{-1}(\sum_{\alpha=1}^{n} \underline{x}_\alpha \underline{x}_\alpha')C_n^{-1}(\underline{x}_i-\underline{x}_j) + 2$$

$$- 2(\underline{x}_i - \underline{x}_j)'C_n^{-1}(\underline{x}_i - \underline{x}_j) ,$$

$$= 2 - (\underline{x}_i - \underline{x}_j)'C_n^{-1}(\underline{x}_i - \underline{x}_j) ,$$

$$= 2 - D^2(\underline{x}_i,\underline{x}_j)/n .$$

Therefore

$$tr(MP_LM) = \sum_{(i,j)\epsilon D_L} (2 - D^2(\underline{x}_i,\underline{x}_j)/n)$$

$$= 2n_L - \sum_{(i,j)\epsilon D_L} D^2(\underline{x}_i,\underline{x}_j)/n .$$

From (3.4.2)

$$E_{H_0}(r) = 1/2n_L[2n_L - \sum_{(i,j)\epsilon D_L} (D^2(\underline{x}_i,\underline{x}_j)/n)]$$

$$= 1 - \overline{D^2}/2n , \qquad (3.4.7)$$

where

$$\overline{D^2} = \sum_{(i,j)\epsilon D_L} D^2(\underline{x}_i,\underline{x}_j)/n_L .$$

Since $(z_i - z_j)^2$ is non-negative for all pairs (i,j),

$$E_{H_0}(z_i - z_j)^2 = 2 - D^2(\underline{x}_i, \underline{x}_j)/n \qquad (3.4.8)$$

implies

$$0 \leq D^2(\underline{x}_i, \underline{x}_j)/n \leq 2 \quad \text{for all pairs} \quad (\underline{x}_i, \underline{x}_j) .$$

Since D is always greater than zero, it follows that

$$0 \leq D(\underline{x}_i, \underline{x}_j) \leq \sqrt{2n} \quad \text{for all pairs} \quad (\underline{x}_i, \underline{x}_j) .$$

It can be seen from (3.4.7) that when the specified model is the correct one the expected value of r is in general larger over a set D_L, where $\overline{D^2}$ is small, than over an arbitrary set of n_L pairs. This fact is partly explained by the fact that the least squares technique for fitting a regression model tends to be more sensitive to the remote points of the independent variables. This results in somewhat of an "overfit"[3] of the data at these points.

3.5 Distribution of r.

In the special case that there are equal numbers of replications at several values of the independent variable, the exact distribution of r can be derived. Suppose there are exactly $N \geq 1$ replications at each of $p \geq 1$ different values. Without loss of generality it can be assumed that the observations are arranged so that the vectors, $\underline{x}$, associated with each set of replications occur as adjacent rows of X and, taken together, occur as the first pN rows of X. Then the matrix $\frac{1}{N}P_L$ can be written as

[3]This term is used in Daniel and Wood (1971).

$$\begin{bmatrix} P_1 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & P_2 & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & P_p & & \cdot \\ \cdot & & & & & \tilde{0} \end{bmatrix} . \qquad (3.5.1)$$

Each P_i of (3.5.1) is an $N \times N$ matrix of the form

$$P_i = I - \frac{1}{N} J \text{ ,} \qquad (3.5.2)$$

where J is the matrix with all elements equal to 1, and $\tilde{0}$ is an $n\text{-}pN \times n\text{-}pN$ zero matrix.

It is well-known (see Searle (1971)) that a matrix of the form (3.5.2) is idempotent, and that the rank of each P_i is equal to N-1. Therefore the matrix P_L is also idempotent and has rank p(N-1). Accordingly the eigenvalues of $\frac{1}{N} P_L$ are

$$\underbrace{1, 1, 1, \cdots, 1}_{p(N-1)}, \underbrace{0, 0, \cdots, 0}_{n - p(N-1)} .$$

Since $X'P_L = 0$ in the case of replications,

$$\begin{aligned} MP_L &= (I - X(X'X)^{-1}X')P_L \\ &= P_L . \end{aligned}$$

Therefore the eigenvalues of MP_LM and P_L are identical and from expression (3.3.7), r is distributed like the quantity

$$\frac{N(n-k)}{2n_L} \frac{\sum_{i=1}^{p(N-1)} \varepsilon_i^{*2}}{\sum_{i=1}^{n-k} \varepsilon_i^{*2}} , \qquad \underline{\varepsilon}^* \sim N(\underline{0}, I\sigma^2),$$

$$= \frac{N(n-k)}{2n_L} \frac{\sum_{i=1}^{p(N-1)} \varepsilon_i^{*2}}{\sum_{i=1}^{p(N-1)} \varepsilon_i^{*2} + \sum_{i=p(N-1)+1}^{n-k} \varepsilon_i^{*2}} .$$

This in turn is distributed like the quantity

$$\frac{N(n-k)}{2n_1} \beta \left(\frac{p(N-1)}{2}, \frac{n-k-p(N-1)}{2}\right),$$

where $\beta(a,b)$ denotes a random variable having a beta distribution with parameters a and b.

Furthermore, in the special case that N=2, the statistic r can be written explicity as a function of the lack of fit statistic F. For any pair $(\underline{x}_i, \underline{x}_j)$,

$$\begin{aligned}(z_i - z_j)^2 &= (y_i - \hat{y}_i - (y_j - \hat{y}_j))^2 \\ &= (y_i - y_j)^2\end{aligned}$$

because the predicted values of y associated with identical $\underline{x}$'s are the same. Since there are only two observations per group

$$\sum_{i=1}^{2} (y_i - \bar{y})^2 = \frac{(y_i - y_j)^2}{2} ,$$

where $\bar{y} = (y_1 + y_2)/2$. Therefore r becomes

$$r = \frac{\sum_{\substack{(i,j)\varepsilon D_L \\ i<j}} \sum_{i=1}^{2} (y_i - \bar{y})^2/n_L}{\sum_{i=1}^{n} z_i^2/(n-k)}$$

$$= \frac{SSPE/n_L}{SSE/(n-k)} ,$$

$$= \left(\frac{n-k}{n_L}\right)\left[\frac{1}{1 + \frac{n-k-n_L}{n_L} F}\right] .$$

where F is the usual lack of fit test statistic and is distributed as $F(n-k-n_L, n_L)$. In this case r is distributed as

$$\frac{(n-k)}{n_L} \beta\left(\frac{n_L}{2}, \frac{n-k-n_L}{2}\right), \qquad \text{for } n-k-n_L > 0 .$$

The statistic r can be viewed as a special case of the more general statistic

$$U = \frac{\underline{w}'A\underline{w}}{\underline{w}'\underline{w}} \qquad \text{where } \underline{w} \sim N(\underline{0}, I\sigma^2) . \tag{3.5.3}$$

This type of statistic is known as a serial correlation coefficient (see, e.g., Anderson (1948)). For a completely general $n \times n$ matrix A, the distribution of U under H_0 is not expressible in closed form. However an exact expression for the density is given by R. L. Anderson (1942) in two cases. These are: (1) n-k is even and the eigenvalues of A are equal in pairs, and (2) n-k is odd and the eigenvalues of A are equal in pairs with the remaining eigenvalue being either the smallest or largest value. J. von Neumann (1941) derives an exact expression for the $[\frac{1}{2}(n-k)-1]$ order derivative of the density function of U for the case that n-k is even and all the eigenvalues are different. However these cases do not occur in general for the statistic r. Anderson's expression becomes very complicated for n-k moderately large and similarly, von Neumann's result can only be used for very small n-k. In all cases it is necessary to know the individual eigenvalues. For practical applications a simple method of approximation is desirable.

T. W. Anderson (1948) shows that $\sqrt{n}U$ is asymptotically normally distributed for a sequence of matrices, A(n), for which the eigenvalues satisfy fairly general regularity conditions. However there are indications that the convergence to normality is slow enough that use of a normal approximation does not always give satisfactory results, even for moderately large sample sizes (see, e.g., Anderson (1942), and Koopmans (1942)).

From (3.3.7) the statistic r can be written as

$$r = \left(\frac{n-k}{2n_L}\right) \sum_{i=1}^{n-k} \nu_i \frac{w_i^2}{\left(w_i^2 + \sum_{j\neq i}^{n-k} w_j^2\right)}$$

$$= \left(\frac{n-k}{2n_L}\right) \sum_{i=1}^{n-k} \nu_i W_i, \text{ say,} \tag{3.5.4}$$

where the random variables $W_i (i=1,2,\cdots,n)$ are correlated beta variables, each with parameters 1/2 and (n-k-1)/2. This expression suggests that a Pearson type I (beta) distribution be used to approximate the distribution of r. To this result a series of beta approximations arise in the literature for special forms of A. In all cases the parameters of the approximating distributions are estimated by the method of moments.

In the case that (3.5.3) is the mean-square-successive-difference statistic, Durbin and Watson (1951) suggest a two parameter beta distribution on the interval (0,4). Bloch and Watson (1967) use a constant times a beta distribution.

Press (1969) suggests that for small n the distribution of any statistic of the form

$$\frac{\sum_{i=1}^{n} \lambda_i w_i^2}{\sum_{i=1}^{n} w_i^2}$$

can be approximated by

$$\lambda_n \beta(a,b) \ ,$$

where λ_n is the largest value of the λ_i, $i=1,2,\cdots,n$. This method essentially fits a two parameter beta distribution with the proper end points. A four parameter beta distribution is used by Henshaw (1966) to approximate the distribution of the serial correlation coefficient of Durbin and Watson, yielding better results than the two parameter beta approximation for small n-k but involving substantially more computation. For larger n the two parameter approximation of Durbin and Watson gives relatively good results considering the amount of computation involved (see Durbin and Watson (1971)). For purposes of approximating the entire distribution function of r, a technique similar to one or more of the above can be used. However in practical application of the test all that may be needed are several cirtical values of the test statistic.

The Pearson Curves are a family of absolutely continuous probability distributions whose density function, f(x), and distribution function, F(x), satisfy

$$\frac{1}{f(x)} \ \frac{df(x)}{dx} = \frac{x+a}{F(x)} \ ,$$

and whose first four moments exist. Assuming different relationships between a and these moments yields three main, and nine transitional types of Pearson Curves. Included among these are the beta and normal distributions. Methods are available to estimate the parameters of the appropriate approximating density function, if this is desired. By Theorem 3.3.2 the moments of r exist and, as was seen in the previous section, exact expressions for these moments are available. For this reason the Pearson Curves are recommended to obtain approximate critical values. This method was also suggested by von Neumann, Kent, Bellinson, and Hart (1941), for the mean-square-successive-difference statistic.

3.6 Bounds for the Critical Values of the Test Statistic

The critical values for the test statistic which are obtained according to the methods of the previous section depend upon the matrix X. Similar to the situation for replications described in Section 3.5, suppose the data points, $\underline{x}_i$, $i=1,2,\cdots,n$, are arranged in k-space in such a manner that p disjoint groups each containing N points can be formed. Suppose further that the function D generates a set of near-neighbors, D_L, such that a pair of points is in D_L if and only if both points belong to one and the same group. In this case bounds can be obtained for the critical values of the statistic r under H_0 which do not depend upon the values in the matrix X. These bounding values are actually critical values of two bounding random variables and can be used in an approximate test procedure to test against alternative hypotheses of the type described in Chapter 4. In this section the derivation of the bounding random variables and their distributions are presented. The general method for conducting the approximate test for nonlinearity is indicated. The derivations draw heavily upon the work of Durbin and Watson (1950) and Anderson (1958).

First some bounds upon the eigenvalues of MP_LM are derived using the following theorem adapted from Anderson (1958).

Theorem 3.6.1. Let $M=I-X(X'X)^{-1}X'$, where X is an $n \times k$ matrix of rank k. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of the symmetric matrix P_L, and let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{n-k}$ be the n-k eigenvalues of MP_LM other than k zeros, arranged in descending order. Then

$$\lambda_{i+k} \leq \mu_i \geq \lambda_i \qquad i=1,2,\cdots,n-k$$

The proof of the theorem is found in Durbin and Watson (1950) and is omitted here.

Letting $\underline{\varepsilon}^* = H\underline{\varepsilon}$ be the orthogonal transformation diagonalizing MP_LM, it follows immediately that

$$\sum_{i=1}^{n-k} \lambda_{i+k}(\varepsilon_i^*)^2 \leq \sum_{i=1}^{n-k} \mu_i(\varepsilon_i^*)^2 \leq \sum_{i=1}^{n-k} \lambda_i(\varepsilon_i^*)^2 .$$

Furthermore,

$$\frac{\sum_{i=1}^{n-k} \lambda_{i+k}(\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2} \leq \frac{\sum_{i=1}^{n-k} \mu_i(\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2} \leq \frac{\sum_{i=1}^{n-k} \mu_i(\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2} .$$

Therefore for every real number r_0,

$$P\{d_U < r_0\} \leq Pr\{d < r_0\} \leq Pr\{d_L < r_0\} ,$$

where

$$d_U = \frac{\sum_{i=1}^{n-k} \lambda_i(\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i)^2} , \quad d_L = \frac{\sum_{i=1}^{n-k} \lambda_{i+k}(\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2} ,$$

and $$d = \frac{\sum_{i=1}^{n-k} \mu_i (\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2}$$

In certain circumstances in which the columns of X are eigenvectors of P_L these bounds can be improved. Consider the following theorem adapted from Anderson (1958).

Theorem 3.6.2. Under the hypothesis of Theorem 3.6.1., suppose that $s(\leq k)$ linearly independent linear combinations of the columns of X are s eigenvectors of the symmetric matrix P_L. Let $\lambda_1' \geq \lambda_2' \geq \cdots \geq \lambda_{n-s}'$ be the eigenvalues of P_L remaining after deleting the s eigenvalues corresponding to these s vectors. Then

$$\Pr\{d_U' < r_0\} \leq \Pr\{d < r_0\} \leq \Pr\{d_L' < r_0\} ,$$

where
$$d_U' = \frac{\sum_{i=1}^{n-k} \lambda_i' (\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2}, \quad d_L' = \frac{\sum_{i=1}^{n-k} \lambda_{i+k-s} (\varepsilon_i^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i^*)^2}$$

The proof is in Durbin and Watson (1950).

From the previous section the eigenvalues of P_L are

$$\underbrace{N, N, \cdots, N}_{p(N-1)}, \underbrace{0, 0, \cdots, 0}_{n-p(N-1)}$$

In general the value of s will be at least 1 since the vector $\underline{j}$, where $\underline{j}'=(1,1,\cdots,1)$, appears as a column of X and is an eigenvector of P_L associated with the eigenvalue 0. In the case that a subset of the independent variables is replicated at particular points, s can be greater than 1. For example, let $m \leq k$ and let each vector $\underline{x}_i'$ be written as $(\underline{x}_{i1}' | \underline{x}_{i2}')$, where $\underline{x}_{i1}'$ is a $1 \times m$ vector and $\underline{x}_{i2}'$ is a $1 \times (k-m)$ vector. If the vectors $\underline{x}_i$, $i=1,2,\cdots,n$, occur so that for every $(i,j)\varepsilon D_L$, $\underline{x}_{i1} = \underline{x}_{i2}$, then s=m and each of these s vectors is associated with a 0 eigenvalue, provided the s vectors are linearly independent. In this case the n-s eigenvalues, $\lambda_i{}'$, are

$$\underbrace{N, N, \cdots, N}_{p(N-1)}, \underbrace{0, 0, \cdots, 0}_{n-p(N-1)-s}$$

Letting j=i+k-s, $d_L{}'$ simplifies to

$$d_L = \frac{\sum_{j=k-s+1}^{n-s} \lambda_j{}'(\varepsilon^*_{j-k+s})^2}{\sum_{j=1}^{n-k} (\varepsilon_j{}^*)^2} .$$

Therefore substituting in the values for $\lambda_i{}'$,

$$d_L = \frac{\sum_{j=k-s+1}^{p(N-1)} N(\varepsilon^*_{j-k+s})^2}{\sum_{i=1}^{n-k} (\varepsilon_j{}^*)^2}, \quad d_U{}' = \frac{\sum_{i=1}^{p(N-1)} N(\varepsilon_i{}^*)^2}{\sum_{i=1}^{n-k} (\varepsilon_i{}^*)^2}$$

Under H_0 the random variables $\varepsilon_i{}^{*2}$ are independent and each is distributed as

$$\sigma^2\chi_1^2$$

Therefore $d_L{}'$ is distributed as

$$N \frac{\chi^2_{m_1}}{\chi^2_{m_1} + \chi^2_{m_2}}$$

where $m_1 = p(N-1) - k + s$, $m_2 = n-k-m_1 = n - p(N-1) - s$, and $\chi^2_{m_1}$ and $\chi^2_{m_2}$ are independent chi-square random variables with m_1 and m_2 degrees of freedom, respectively. Thus it follows that $d_L{}'$ is distributed as

$$N\cdot\beta(m_1/2, m_2/2).$$

Similarly, $d_U' \sim N\beta(n_1/2,\ n_2/2)$, where $n_1 = p(N-1)$ and $n_2 = n-k-n_1 = n-k-p(N-1)$. It follows, since $r = \frac{(n-k)}{n_L}d$, that r satisfies

$$r_L \leq r \leq r_U$$

where $r_L \sim N\frac{(n-k)}{n_L} \cdot \beta(m_1/2,\ m_2/2)$ and $r_U \sim N\frac{(n-k)}{n_L}\ \beta(n_1/2,\ n_2/2)$, for every r_0

$$\Pr\{r_U \leq r_0\} \leq \Pr\{r \leq r_0\} \leq \Pr\{r_L \leq r_0\}\ .$$

Therefore if $r'(\alpha)$ satisfies $\Pr\{r \leq r'(\alpha)\} = \alpha$, then

$$r_L'(\alpha) \leq r'(\alpha) \leq r_U'(\alpha),$$

where $r_L'(\alpha)$ is a lower level α critical value for the distribution of r_L, and $r_U'(\alpha)$ is the lower level α critical value of the distribution of r_U. The values $r_L'(\alpha)$ and $r_U'(\alpha)$ depend only upon n, n_1, k, and s.

CHAPTER 4
ALTERNATIVE HYPOTHESIS

4.1 The Alternative hypothesis.

In general the alternative hypothesis for a test for lack of fit of H_0 would result when any one or more of the assumptions of H_0 are relaxed or are differently specified. However the possible departures from the assumptions which are considered in this investigation are restricted to those which can be expressed by the addition to the model equation (1.1.1) of a term which is itself a nonlinear function of one or more of the independent variables. Two basic types of terms are considered here, 1) a second degree polynomial function of the x_{ij}'s, and 2) a general nonlinear function of the x_{ij}'s. The model for the alternative relationship in each of these two cases can be represented as

$$\underline{y} = X\underline{\omega} + \gamma\underline{g}(X) + \underline{u} \tag{4.1.1}$$

where $\underline{\omega}$ is a $k \times 1$ vector of unknown parameters, $\underline{u}$ is an $n \times 1$ vector of unobserved random error terms with $\underline{u} \sim N(\underline{0}, I\sigma_u^2)$, and $\underline{g}(X)$ is an $n \times 1$ vector of unknown nonlinear functions of X with unknown coefficient γ. The vector $\underline{g}$ can be written as

$$\underline{g}(X) = \begin{bmatrix} g_1(X) \\ g_2(X) \\ \cdot \\ \cdot \\ \cdot \\ g_n(X) \end{bmatrix} \tag{4.1.2}$$

In case 1), $g_i(X)$ may be defined as

$$g_1(X) = \underline{x}_i{}'A\underline{x}_i \qquad i=1,2,\cdots,n \tag{4.1.3}$$

for some $k \times k$ matrix A of fixed coefficient parameters. For case 2) $g_i(X) = f(\underline{x}_i)$, $i=1,2,\cdots,n$, where $f(\underline{x}_i)$ is some nonlinear function of the x_{ij}'s. In the case that there are replications at p values of the independent variable it was noted in Chapter 1 that the alternative hypothesis for which the F test is uniformly most powerful is that $g_i(X)$ is a p parameter ($1 \leq p \leq n-k$) nonlinear function of the k independent variables. The same alternative will be assumed for the r test as well as for the F test when there are no replications.

4.2 Properties of r under H_A.

Under the alternative hypothesis, H_A, the residual vector $\underline{z}$ is equal to

$$\underline{z} = M\underline{y} = M(\gamma\underline{g} + \underline{u}) .$$

Therefore

$$r = \left(\frac{n-k}{2n_L}\right) \frac{(\underline{u} + \gamma\underline{g})'MP_LM(\underline{u} + \gamma\underline{g})}{(\underline{u} + \gamma\underline{g})'M(\underline{u} + \gamma\underline{g})} .$$

If $H = (\underline{h}_1, \underline{h}_2, \cdots, \underline{h}_n)$ is the matrix of the orthogonal transformation used in Section 3.2 and $\underline{w}^* = H(\underline{u} + \gamma\underline{g})$, then $\underline{w}^* \sim N(\gamma H\underline{g},\ I\sigma_u^2)$. Subsequently r becomes

$$\left(\frac{n-k}{2n_L}\right) \frac{\underline{w}^{*\prime} V \underline{w}^*}{\underline{w}^{*\prime} \tilde{I} \underline{w}^*}$$

$$= \left(\frac{n-k}{2n_L}\right) \frac{\sum_{i=1}^{n-k} \nu_i (w_i^*)^2}{\sum_{i=1}^{n-k} (w_i^*)^2}$$

Under this alternative each of the variables $(w_i^*)^2$ is distributed as $\sigma_u^2 \chi_1^{2\prime}(\lambda_i)$, where $\chi_1^{2\prime}(\lambda_i)$ is a random variable having a noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter $\lambda_i = \gamma^2 \underline{g}'\underline{h}_i\underline{h}_i'\underline{g}/\sigma_u^2$, $(i=1,2,\cdots,n)$. Using the methods of Section 3.5 r can be expressed as

$$r = \frac{(n-k)}{2n_L} \sum_{i=1}^{n-k} \nu_i W_i^*,$$

although now the (W_i^*)'s, $i=1,2,\cdots,n-k$, are n-k correlated doubly-noncentral beta random variables. The fact that the distribution of the doubly-noncentral beta random variable can be approximated by a constant times a beta distribution (see Scheffe (1959)) suggests that if necessary the distribution of r under H_A could be approximated by one of the variations of a beta distribution suggested in section 3.5.

Under H_A the moments of r are now not only functions of the omitted vector $\underline{g}$, but also are dependent upon σ_u^2, the variance of the error term in model (4.1.1). The distribution of the $(w_i^*)^2$'s is not in the form of the Γ distribution discussed in Section 3.4, therefore (Pitman's) Theorem (3.4.1) is not applicable and exact analytic expressions for the moments of r under H_A are not obtainable. However the moments of r can be approximated by using a Taylor series expansion of r about its mean. These approximations are used to study the behavior of the statistic r under H_A as compared to its behavior under H_0.

Let $r = r(v,h) = v/h$, where $v = \underline{w}^*{}'V\underline{w}^*/2n_L$ and $h = \underline{w}^*{}'\tilde{I}\underline{w}/(n-k)$, and denote the means of v and h under H_A by $E_A(v)$ and $E_A(h)$, respectively. Since the variances of v and h and the covariance of v and h are all of order 1/n, a first and second order Taylor series approximation to the moments of r have error terms of order 1/n and

$1/n^2$, respectively. Consider the first order Taylor series approximation to $E_A(r)$.[4]

$$E_A(r(v,h)) = \frac{E_A(v)}{E_A(h)} + 0(1/n) \tag{4.2.1}$$

Using Theorems 3.4.2 - 3.4.4 the moments of v and h under H_A can be expressed as functions of the moments of v and h under H_0 as follows:

$$\begin{aligned} E_A(v) &= E_0(v)\rho + G_p(n) \\ E_A(h) &= E_0(h)\rho + G(n) \\ \mathrm{Var}_A(h) &= \frac{2\sigma_u^2}{n-k}\,[E_0(h)\rho + 2G(n)] \\ \mathrm{Cov}_A(v,h) &= \frac{2\sigma_u^2}{n-k}\,[E_0(v)\rho + 2G_p(n)] \ , \end{aligned} \tag{4.2.2}$$

where $G_p(n) = \gamma^2\underline{g}'MP_LM\underline{g}/2n_L$, $G(n) = \gamma^2\underline{g}'M\underline{g}/(n-k)$, and $\rho = \sigma_u^2/\sigma^2$.

Therefore

$$E_A(r) = \frac{E_0(v)\rho + G_p(n)}{E_0(h)\rho + G(n)} + 0(1/n) \tag{4.2.3}$$

and for large n, $E_A(r) < E_0(r) = \dfrac{E_0(v)}{E_0(h)}$ if

$$\frac{G_p(n)}{G(n)} < E_0(r) \tag{4.2.4}$$

It is useful to note that (4.2.4) is equivalent to

$$(1 - \overline{D^2}/2n) > \frac{G_p(n)}{G(n)} \tag{4.2.5}$$

In the limit (4.2.4) becomes equivalent to

$$G > G_p$$

where $\quad G = \lim\limits_{n\to\infty} G(n)$, and $\quad G_p = \lim\limits_{n\to\infty} G_p(n)$

---

[4]The second order approximation leads to the same result.

For large n, condition (4.2.4) implies that $E_A(r) \leq E_0(r)$, therefore the rejection region of the test for nonlinearity using r is in the lower tail of its distribution.

Noting that $\underline{z} = M\underline{u} + \gamma M\underline{g}$, the sum of a fixed and a random component, suppose the function r is computed on the elements of the vector representing the fixed portion of $\underline{z}$. If the value of this function is less than the expected value of the same statistic, r, calculated on the random component of $\underline{z}$, then (4.2.4) holds and for large n $E_A(r) \leq E_0(r)$. It is also noteworthy that the approximate condition (4.2.4) is independent of the values of γ and σ_u^2.

In the case that there are replications at values of $\underline{x}$, D_L may be defined by $D_L = D(-,0,-)$. Since the function D is a metric it follows that $\underline{x}_i \equiv \underline{x}_j$ for every $(\underline{x}_i, \underline{x}_j) \varepsilon D_L$. By the definition of P_L, $X'P_L = 0$ in this special case. Therefore $MP_L =$ $(I - X(X'X)^{-1}X')P_L = P_L$ and the numerator of r becomes

$$(\underline{u} + \gamma\underline{g})'P_L(\underline{u} + \gamma\underline{g})$$

Furthermore since $g_i(X) = f(\underline{x}_i)$, $\underline{g}'P_L = \underline{0}$. Therefore r simplifies to

$$\left(\frac{n-k}{2n_L}\right) \frac{\underline{u}'P_L\underline{u}}{(\underline{u} + \gamma\underline{g})'M(\underline{u} + \gamma\underline{g})}$$

$$= \left(\frac{n-k}{2n_L}\right) \frac{\sum\limits_{(i,j)\varepsilon D_L} (u_i - u_j)^2}{\sum\limits_{i=1}^{n-k} z_i^2}$$

with $\underline{g}'MP_LM\underline{g} = 0$. Condition (4.2.4) always holds in this case since

$$(1 - \overline{D^2}/2n) = 1 > 0$$

4.3 Method of Approximating the Power of the Test

For purposes of determining the power of the test procedure the interest lies in tail probabilities of the form

$$P_{H_A}\{r \leq r_0\} ,$$

where r_0 is defined by $P_{H_0}\{r \leq r_0\} = \alpha$, α being the size of the test. However since the denominator is never negative, this quantity can be rewritten as

$$P_{H_A}\{(\underline{u} + \gamma\underline{g})'M\Omega M(\underline{u} + \gamma\underline{g}) \leq 0\} ,$$

where $\Omega = (1/2n_L)P_L - (r_0/(n-k))I$. The quantity $\ell = (\underline{u} + \gamma\underline{g})'M\Omega M(\underline{u} + \gamma\underline{g})$ is an indefinite quadratic form in the variable $\underline{w} = \underline{u} + \gamma\underline{g}$. The distribution of ℓ can be approximated by a Pearson three-moment chi-square distribution. This method, as described by Imhoff (1961), assumes that the distribution of ℓ is approximately that of

$$\frac{(\chi_q^2 - q)}{\sqrt{2q}} \sqrt{\mathrm{Var}_{H_A}(\ell)} + E_{H_A}(\ell) \tag{4.3.1}$$

where χ_q^2 is a central chi-square random variable with q degrees of freedom. The parameter q is estimated using the first three moments of ℓ. This estimate, $\hat{q}$, is [5]

$$\hat{q} = \frac{2[\mathrm{Var}_{H_A}(\ell)]^3}{[\mu_{3H_A}(\ell)]^2} \tag{4.3.2}$$

Although Imhoff cautions that this method of approximation gives

---

[5] It is assumed that $\mu_3(\ell)>0$, otherwise the distribution of $-\ell$ should be approximated with the power of the test being approximated by $P_{H_A}(-\ell>0)$, (Imhoff (1961)).

more accurate results in the case of definite quadratic forms than for indefinite forms it will be considered sufficiently accurate here to indicate trends in the power of the test against specific types of alternatives.

The first three central moments of ℓ are

$$E_{H_A}(\ell) = \mathrm{tr}(M\Omega M)\,\sigma_u^2 + \gamma^2 \underline{g}' M\Omega M\underline{g} ,$$

$$\mathrm{Var}_{H_A}(\ell) = 2\mathrm{tr}(M\Omega M)^2\sigma_u^4 + 4\gamma^2\underline{g}'(M\Omega M)^2\underline{g}\sigma_u^2 , \text{ and}$$

$$\mu_{3H_A}(\ell) = 8\mathrm{tr}(M\Omega M)^3\sigma_u^6 + 24\gamma^2\underline{g}'(M\Omega M)^3\underline{g}\sigma_u^4 .$$

The terms $\mathrm{tr}(M\Omega M)^k$, $k=1,2,3$, can be expanded into a sum of terms each of the form $\mathrm{tr}M(P_LM)^j$, $j=0,\cdots,k-1$. In case the individual eigenvalues of the matrix MP_LM are known the evaluation of these terms can be simplified.

Let the matrix H be the orthogonal matrix defined in Section 3.3. Then

$$\mathrm{tr}(M\Omega M)^k = \mathrm{tr}(M\Omega MHH')^k = \mathrm{tr}(H'M\Omega MH)^k ,$$

since $HH' = I$. But from the choice of the matrix H this becomes $\mathrm{tr}(D(\lambda_i))^k$, where $D(\lambda_i)$ is an $n \times n$ diagonal matrix with elements λ_i given by

$$\lambda_i = \nu_i/2n_L - r_0/(n-k), \qquad i=1,2,\cdots,n,$$

and ν_i, $i=1,2,\cdots,n$, are the eigenvalues of MP_LM (including the k zero eigenvalues). In this case

$$\mathrm{tr}(M\Omega M)^k = \sum_{i=1}^{n} \lambda_i^k, \quad k=1,2,3.$$

Similarly the terms $\underline{g}'(M\Omega M)^k\underline{g}$ can be expressed as a sum of terms of the

$$\frac{\underline{g}'M(P_LM)^j\underline{g}}{(2n_L)^j(n-k)^{k-j}}\ , \quad j = 1,\cdots,k, \tag{4.3.3}$$

for k=1,2,3, and the term $\underline{g}'M\underline{g}/(n-k)^k$. For small samples when there is no replication these terms influence the power of the test and reflect the degree of departure from the ideal condition of replication.

An ordering exists among the terms of the form (4.3.3). The matrix P_L is a real symmetric definite matrix. Therefore there exists a definite matrix U of the same rank as P_L which is real and symmetric such that $UU' = P_L$.

Letting $\underline{g}^{*\prime} = \underline{g}'MU$, then

$$\begin{aligned} \underline{g}'MP_LM\underline{g} &= \underline{g}'MUU'M\underline{g} \\ &= \underline{g}^{*\prime}\underline{g}^*\ , \end{aligned}$$

and

$$\begin{aligned} \underline{g}'(MP_LM)^2\underline{g} &= \underline{g}'MUU'MUU'M\underline{g} \\ &= \underline{g}^{*\prime}U'MU\underline{g}^*\ . \end{aligned} \tag{4.3.4}$$

By Theorem 3.3.2 the two quadratic forms satisfy the inequality

$$\lambda_1 \leqq \frac{\underline{g}^{*\prime}U'MU\underline{g}^*}{\underline{g}^{*\prime}\underline{g}} \leqq \lambda_n\ , \tag{4.3.5}$$

where λ_1 and λ_n are the smallest and largest eigenvalues, respectively, of U'MU. But the eigenvalues of a product of matrices are not affected by a rotation of the product, so that U'MU and $MUU' = MP_L$ have the same eigenvalues.

Therefore

$$0 \leqq \frac{\underline{g}'(MP_LM)^2\underline{g}}{\underline{g}'MP_LM\underline{g}} \leqq \nu_n\ . \tag{4.3.6}$$

Similarly, defining $\underline{g}^{*\prime} = \underline{g}'MP_LM$, it can be shown that

$$0 \leq \frac{\underline{g}'(MP_LM)^3\underline{g}}{\underline{g}'MP_LM\underline{g}} \leq \nu_n \quad . \tag{4.3.7}$$

Inequalities (4.3.5) - (4.3.7) imply that

$$\underline{g}'(MP_LM)^3\underline{g} \leq \nu_n\underline{g}'(MP_LM)^2\underline{g} \leq \nu_n^{\,2}\underline{g}'(MP_LM)\underline{g} \ . \tag{4.3.8}$$

Since MP_LM is positive semidefinite, $\nu_i \geq 0$ for every i, so that

$$\nu_n \leq \sum_{i=1}^{n-k} \nu_i = 2n_L(1-\overline{D^2}/2n).$$

Therefore

$$0 \leq \frac{\underline{g}'(MP_LM)^3\underline{g}}{(2n_L)^3} \leq (1-\overline{D^2}/2n)\frac{\underline{g}'(MP_LM)^2\underline{g}}{(2n_L)^2} \leq (1-\overline{D^2}/2n)^2\ \frac{\underline{g}'MP_LM\underline{g}}{2n_L} \quad . \tag{4.3.9}$$

If the set D_L consists of pairs of replicates then $\underline{g}'MP_L = \underline{g}'P_L = \underline{0}$, implying that all the terms of inequality (4.3.9) are identically zero.

The process leading to (4.3.9) can be extended by induction to yield the inequality

$$\frac{\underline{g}'(MP_LM)^m\underline{g}}{(2n_L)^m} \leq \frac{\underline{g}'(MP_LM)\underline{g}}{2n_L}$$

for any positive integer m. If the eigenvalues of MP_LM are uniformly bounded by some value λ_N, say, for all n, then from (4.3.8) ,

$$\underline{g}'(MP_LM)^m\underline{g} \leq \lambda_N^{\,m}\underline{g}'(MP_LM)\underline{g} \quad . \tag{4.3.10}$$

Therefore the term $\underline{g}'(MP_LM)^m\underline{g}$ has the same order of magnitude for any fixed positive integer m. This result is used in Chapter 5 when the large sample properties of the test are considered.

CHAPTER 5
LARGE SAMPLE PROPERTIES

5.1 Large Sample Comparisons

Any comparison of the F test and the r test depends upon the number of observations available. When there is a sufficient number of replications in the data, of course the F test would be preferable to use at any sample size. As discussed in Chapter 1, the F test is also preferred when the number of observations is large, regardless of the dispersion of the data points. An additional argument in favor of the F test for large samples is that as the number of observations increases, the exact moments of r under H_0 become increasingly expensive to compute.

When the sample size is small, problems arise in the determination of the groups of near-neighbors. Although computer algorithms for selecting such groups are available, the criterion for determining the best grouping often depends upon some arbitrary starting point - especially in a procedure which allows the groups to form to variable sizes. With the r test this problem is not encountered since the formation of groups is replaced by the selection of pairs of nearest-neighbors which do not necessarily have to be disjoint pairs. Since it is a straightforward task to order all pairs, $(\underline{x}_i, \underline{x}_j)$, by a well-defined distance function, determining a usable set of pairs is a simpler task. Although some arbitrariness is introduced in deciding upon how many pairs to include in D_L, this same type of decision is made in forming groups for the F test.

A meaningful comparison of the two tests should involve a consideration of the relative power of the tests. Therefore for small sample sizes this comparison is done by simulation of the behavior of the test statistic under several alternative model specifications. A description of the simulation experiment and discussion of these results are given in Chapter 6. In the present chapter the two tests are compared for large samples by using the criterion as asymptotic relative efficiency (ARE) as defined by Pitman (see Noether (1955)).

For large sample considerations it is necessary to make the following assumptions:

$$\lim_{n\to\infty} \frac{X'X}{n} = \Sigma \ , \tag{5.1.1}$$

where Σ is a fixed, finite-valued, $k \times k$ matrix, and

$$\lim_{n\to\infty} \gamma^2 \frac{\underline{g}'\underline{g}}{n} = \sigma_g^2 < \infty \ , \tag{5.1.2}$$

where $\underline{g}$ is the vector of nonlinear terms of model (4.1.1). Assumption (5.1.1) implies that the vectors $\underline{x}_i$, $i=1,2,\cdots$, are increased so that their matrix of cross-products converges to a limit. It follows that

$$\lim_{n\to\infty} n(X'X)^{-1} = \Sigma^{-1} \ .$$

Subsequently for a fixed pair, $(\underline{x}_i,\underline{x}_j)$,

$$\begin{aligned}\lim_{n\to\infty} D^2(\underline{x}_i,\underline{x}_j) &= \lim_{n\to\infty} [n(\underline{x}_i-\underline{x}_j)'C_n^{-1}(\underline{x}_i-\underline{x}_j)] \\ &= (\underline{x}_i-\underline{x}_j)'\Sigma^{-1}(\underline{x}_i-\underline{x}_j) \\ &= \delta^2(\underline{x}_i,\underline{x}_j) \ , \quad \text{say.}\end{aligned}$$

This guarantees that

$$\lim_{n\to\infty} \frac{D^2(\underline{x}_i, \underline{x}_j)}{n} = 0.$$

Throughout the remainder of the chapter it will be implicitly assumed that (5.1.1) and (5.1.2) are satisfied.

5.2 Asymptotic Normality of r under Pitman's Alternative

Given the two tests F and r of identical size and of the same hypothesis, the relative efficiency of r to F is generally defined as the ratio n_F/n_r, where n_r is the sample size required using test r to give the same power as test F with sample size n_F. In general the ratio n_F/n_r is a complex function of the parameters of the particular alternative chosen. However in the limiting case often this dependence can be simplified. The asymptotic relative efficiency of r relative to F (E_{rF}) is defined by

$$E_{rF} = \lim_{n\to\infty} n_F/n_r \, . \tag{5.2.1}$$

Suppose that the set of hypotheses to be tested is given by H_0: $\gamma = 0$ versus H_A: $\gamma \neq 0$, where γ is defined as in (4.1.1). Consider an alternative parameter γ_n which satisfies $\lim_{n\to\infty} \gamma_n = 0$.

For particular forms of γ_n $\lim_{n\to\infty} n_F/n_r$ can be evaluated under H_1: $\gamma = \gamma_n$ under certain conditions given by Pitman (see Noether (1955)). The value of E_{rF} obtained as such is considered to be a reasonably good indication of the relative efficiency of the two tests against local alternatives, i.e., alternatives in which the value of γ is in the neighborhood of 0. At this stage in order to evaluate (5.2.1) using Pitman's theory it must be assumed that limit $\mathrm{tr}(MP_LM)^2/2n_L$ exists. This condition is satisfied for the r test under certain regularity conditions on the eigenvalues of the matrix MP_LM.

One of the requirements of Pitman's theorem is that $\sqrt{n}\cdot r$ have limiting normal distribution under both H_0 and H_1. The demonstration of this property is given here in considerable detail because it illustrates the development of the regularity conditions imposed upon the eigenvalues of MP_LM. The general method of approach is that used by Kamat (1962) to show the asymptotic normality of the mean-square-successive-difference statistic under alternatives of serial dependence. The satisfaction of the other requirements for Pitman efficiency can be demonstrated using first order Taylor series approximations to the first two moments of r under H_1. These are straightforward and are omitted here. In the following derivations it is necessary to assume that n_L, the number of pairs in D_L, remains proportional to n and n increases, i.e., $n_L/n = 0(1)$. The following two normal convergence theorems are used. The proof of both theorems are found in Hoeffding and Robbins (1948).[8]

Theorem 5.2.1. Let $\{p(n)\}$ be a non-decreasing sequence of positive integers approaching infinity. Let (X_{nj}, Y_{nj}), $j=1,2,\cdots,p(n)$, be an infinite sequence of finite sequences of p(n) random variables in R^2 which are independent for each fixed n.

[8]See Orey (1955) for an extension of theorem 5.2.1 to sequences of m-dependent random variables.

Suppose $E(X_{nj}) = E(Y_{nj}) = 0$, and define

$$\mu_{i2-i}(nj) = E(X_{nj}^{i} \cdot Y_{nj}^{2-i}) \ , \quad i=0,1,2$$

$$\rho_{nj}^{3} = \max(E|X_{nj}|^{3}, E|Y_{nj}|^{3}), \quad \text{and}$$

$$\rho_{n}^{3} = \sum_{j=1}^{p(n)} \rho_{nj}^{3} \ .$$

If $\quad \lim_{n\to\infty} \sum_{j=1}^{p(n)} \mu_{i2-i}(nj))/p(n) = \mu_{i2-i}$,

and $\quad \lim_{n\to\infty} \rho_n/\sqrt{p(n)} = 0,$

then the random vector

$$\frac{1}{\sqrt{p(n)}} \left(\sum_{j=1}^{p(n)} X_{nj}, \sum_{j=1}^{p(n)} Y_{nj} \right)$$

has a limiting normal distribution with mean (0,0) and variance-covariance matrix

$$\begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix} .$$

Theorem 5.2.2. If S and T are random variables with $E(S) = s_0$ and $E(T) = t_0$ and if $\sqrt{n}(S - s_0)$ and $\sqrt{n}(T - t_0)$ have a limiting bivariate normal distribution with nonsingular variance-covariance matrix $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, then $\sqrt{n}(S/T - s_0/t_0)$ has limiting normal distribution with mean 0 and variance

$$\frac{a}{t_0^{2}} - \frac{2bs_0}{t_0^{3}} + \frac{cs_0^{2}}{t_0^{4}} \ . \qquad (5.2.2)$$

Necessary for the results of this section is the following lemma. The proof appears in the appendix.

Lemma 5.2.1. Let the eigenvalues of MP_LM be uniformly bounded for all n. Then for every finite fixed positive integer m, and for every $\alpha > 0$

$$\mathrm{tr}(MP_LM)^m = o(n^{1+\alpha})$$

and
$$\underline{g}'(MP_LM)^m\underline{g} = o(n^{1+\alpha}). \tag{5.2.3}$$

The Pitman alternative takes the specific form H_1: $\gamma^2 = \theta/\sqrt{n}$ for some $\theta \neq 0$.

In the notation of theorem 5.2.1, let p(n) = n-k and define

$$X_{nj} = \frac{\nu_j(n)(n-k)}{2n_L}[w_j{*}2 - E(w_j{*}2)]$$

$$Y_{nj} = [w_j{*}2 - E(w_j{*}2)].$$

Then under H_1,

$$E(X_{nj}) = E(Y_{nj}) = 0, \quad \text{and } \mu_{i2-i}(nj) \text{ are}$$

$$\mu_{20}(nj) = \frac{\nu_j^2(n)(n-k)^2}{(2n_L)^2}\mathrm{Var}(\chi_1'^2(\lambda_j))$$

$$\mu_{02}(nj) = \mathrm{Var}(\chi_1'^2(\lambda_j))$$

$$\mu_{11}(nj) = \nu_j(n)\ \mathrm{Var}(\chi_1'^2(\lambda_j)), \quad \text{where}$$

$$\lambda_j = (\theta/\sqrt{n})\ \underline{g}'\underline{h}_j\underline{h}_j'\underline{g}/\sigma_u^2. \quad \text{Furthermore}$$

$$\frac{1}{(n-k)}\sum_{j=1}^{n-k}\mathrm{Var}(X_{nj}) = \frac{n-k}{(2n_L)^2}\mathrm{Var}(\underline{w}^{*}{}'V\underline{w}^{*})$$

$$= \frac{n-k}{(2n_L)^2}\,[2tr(MP_LM)^2\sigma_u^4 + \frac{4\theta}{\sqrt{n}}\,\underline{g}'MP_LM\underline{g}\sigma_u^2]$$

Therefore

$$\mu_{20} = \underset{n\to\infty}{\text{limit}}\ \frac{n-k}{(2n_L)^2}\sum_{j=1}^{n-k}\mu_{20}(nj)\ ,$$

$$= S^2(2\sigma_u^4)\quad ,$$

where $S^2 = \underset{n\to\infty}{\text{limit}}\ (tr(MP_LM)^2/(n-k))$.

Similarly $\frac{1}{(n-k)}\sum_{j=1}^{n-k}\mu_{11}(nj)$

$$= \frac{1}{2n_L}\,[2tr(MP_LM)\sigma_u^4 + (\theta/\sqrt{n})\underline{g}'MP_LM\underline{g}\sigma_u^2]\ ,$$

$$= \left[\frac{4n_L(1-\overline{D^2}/2n)\sigma_u^4}{2n_L} + O(1/\sqrt{n})\right]\ .$$

Therefore $\mu_{11} = 2\sigma_u^4$. Finally μ_{02}

$$= \underset{n\to\infty}{\text{limit}}\ \frac{1}{(n-k)}\sum_{j=1}^{n-k}\mu_{02}(nj)$$

$$= \underset{n\to\infty}{\text{limit}}\ \frac{1}{(n-k)}[2(n-k)\sigma_u^4 + (4\theta/\sqrt{n})\underline{g}'M\underline{g}\sigma_u^2]$$

$$= 2\sigma_u^4$$

It is well-known (see Cramer (1946), p176) that

$$(E|X|^a)^{1/a} \leqq (E|X|^{a+1})^{1/(a+1)},\quad a=1,2,3,\cdots,$$

for any random variable whose absolute moments exist. Therefore for a = 3,

$$\frac{1}{\sqrt{n-k}}\sqrt[3]{\sum_{j=1}^{n-k} E|X_{nj}|^3} \leqq \frac{1}{\sqrt{n-k}}\sqrt[3]{\sum_{j=1}^{n-k} (E(X_{nj}^{\;4}))^{3/4}}$$

But $E(x_{nj}^{\;4})^{3/4} \leqq E(X_{nj}^{\;4}) + 1$ for all j, n. Therefore

$$\sum_{j=1}^{n-k} E(X_{nj}^{\;4})^{3/4} \leqq \sum_{j=1}^{n-k} E(X_{nj}^{\;4}) + (n-k).$$

Using Kolmogorov's Inequality for a = 4,

$$\sum_{j=1}^{n-k} E(X_{nj}^{\;4}) \leqq \sum_{j=1}^{n} \nu_j^4(n)[60 + 240\theta\underline{g}'\underline{h}_j\underline{h}_j'\underline{g}/\sqrt{n}\sigma_u^2 +$$

$$+ 48(\theta\underline{g}'\underline{h}_j\underline{h}_j'\underline{g}/\sqrt{n}\sigma_u^2)^2]$$

$$\leqq \sum_{j=1}^{n} \nu_j^{\;4}(n)[60 + 240(\theta\underline{g}'\underline{g}/\sqrt{n}\sigma_u^2) + 48(\theta\underline{g}'\underline{g}/\sqrt{n}\sigma_u^2)^2].$$

(5.2.4)

The values $\nu_j^{\;4}(n)$ are uniformly bounded by some number λ_N, say. Therefore (5.2.4) is bounded above by

$$n\lambda_N^4[60 + 240\ 0(\sqrt{n}) + 48\ 0(n) + (n-k)] \;, \qquad (5.2.5)$$

which is a term of order n^2. Therefore

$$\rho_n/\sqrt{n-k} = \frac{1}{\sqrt{n-k}}\sqrt[3]{\sum_{j=1}^{n-k} E|X_{nj}|^3}$$

is less than a term of order $n^{-1/6}$, and $\lim_{n\to\infty} \rho_n/\sqrt{n-k} = 0$. By letting $\lambda_N = 1$ in (5.2.5) the same procedure yields the corresponding result for Y_{nj}. Replacing λ_N by $\max(\lambda_N,1)$ in (5.2.5), it follows that $\lim_{n\to\infty} \rho_n/\sqrt{n-k} = 0$. Therefore, under H_1,

$$\left(\frac{1}{\sqrt{n-k}}\sum_{j=1}^{n-k} X_{nj},\ \frac{1}{\sqrt{n-k}}\sum_{j=1}^{n-k} Y_{nj}\right) \quad \text{is}$$

asymptotically normally distributed with limiting mean (0,0) and variance-covariance matrix

$$\begin{pmatrix} S^2 & 1 \\ 1 & 1 \end{pmatrix} 2\sigma_u^4 .$$

The main result of this section is stated the the following two theorems.

Theorem 5.2.3. Under H_1, $\sqrt{n}\cdot r$ is asymptotically normally distributed with mean 1 and variance $2(S^2 - 1)$ if the eigenvalues of MP_LM are uniformly bounded for all n.

Define $S = \frac{1}{2n_L} \underline{w}^{*\prime}V\underline{w}^*$ and $T = \underline{w}^{*\prime}\tilde{I}\underline{w}^*/(n-k)$. Then $r = S/T$. By theorem 5.2.1 $\sqrt{n}(S-s_0)$ and $\sqrt{n}(T - t_0)$ have an asymptotically bivariate normal distribution where $t_0 = s_0 = 2\sigma_u^2$. Therefore s and T satisfy the conditions of Theorem 5.2.2. and $\sqrt{n}\cdot r$ has limiting normal distribution with mean 1 and variance $2(S^2 - 1)$.

Corollary 5.2.1. $\sqrt{n}\cdot r$ is asymptotically normally distributed under H_1 if the indices of the vectors $\underline{x}_i (i=1,2,\cdots,n)$ are uniformly bounded for all n.

Proof. There exists a value N_r such that $I_i \leq N_r$ for all i. It follows immediately that any diagonal element of P_L is less than N_r, i.e.,

$$(P_L)_{jj} \leq N_r \qquad \text{for every } j=1,2,\cdots,n$$

Define

$$\rho_j = \sum_{\substack{k=1 \\ k\neq j}}^{n} |(P_L)_{kj}|$$

Then $$\max_j \rho_j \leq \sum_{\substack{k=1 \\ k\neq j}}^{n} \delta_{kj} \leq N_r ,$$

where $\quad \delta_{kj} = \begin{cases} 1 & \text{if } (i,j)\varepsilon D_L \ . \\ 0 & \text{otherwise} \ . \end{cases}$

By Gersgorin's Theorem (see Lancaster (1969)) every eigenvalue of P_L lies in one of the intervals

$$0 \leq z \leq (P_L)_{jj} + \rho_j , \qquad j=1,2,\cdots,n,$$

or
$$0 \leq z \leq N_r + N_r = 2N_r \ .$$

Let $\lambda_1(n) \geq \lambda_2(n) \geq \cdots \geq \lambda_n(n)$ be the eigenvalues of P_L. By Theorem 3.6.1,

$$\lambda_{i+k}(n) \leq \nu_i'(n) \leq \lambda_i(n) \qquad i=1,2,\cdots,n-k,$$

where $\nu_1'(n) \geq \nu_2'(n) \geq \cdots \geq \nu_n'(n)$ are the eigenvalues of MP_LM arranged in decending order. Since the $\lambda_i(n)$'s are uniformly bounded by $2N_r$, there exists a value, λ_N, such that $\nu_i'(n) \leq \lambda_N$ for all i and for all n. Since the eigenvalues of MP_LM are bounded, the result follows from Theorem 5.2.3.

That all the conditions of this section hold for H_0 as well as H_1 follows immediately by letting $\theta = 0$ in the preceding results.

5.3 Asymptotic Relative Efficiency of r and F.

The classical lack of fit statistic developed in Chapter 1 takes the form

$$F = \frac{\underline{y}'(M-A)\underline{y}/(n-k-a)}{\underline{y}'A\underline{y}/a} \tag{5.3.1}$$

where A is the idempotent matrix of rank a of the quadratic form which represents the pure error sum of squares, $\underline{y}'A\underline{y}$. In the usual case of true replications $A(M-A) = 0$, implying that $AM = A$. However this is only an approximation when near-replicates are used. To evaluate the Pitman efficiency of r relative to F some restrictions must be placed upon the manner in which n tends to infinity so that the conditions are compatible with both tests.

For the r test suppose that for each fixed n the $\underline{x}_i$'s are grouped into m(n) intervals or clusters so that $(\underline{x}_i, \underline{x}_j) \varepsilon D_L$ if and only if $\underline{x}_i$ and $\underline{x}_j$ belong to one and the same cluster. Suppose for each n there exists a $d_L(n)$ which, with D, generates this set D_L. Furthermore, it is assumed that the index of any point $\underline{x}$ is not greater than N_r and that the proportion of points whose indices are equal to N_r tends to one as n tends to infinity.

For the F test it is assumed that the $\underline{x}_i$'s can be grouped into p(n) groups, G_i, $i=1,2,\cdots,p(n)$, with each G_i containing n_i points on which pure error sum of squares is computed. Suppose that $n_i \leq N_F$ for some N_F and for all i. As was the case for the r test, some regularity assumptions must be made concerning the noncentrality parameters of the numerator and denominator of F for the case that there are no replications. These are

$$\begin{aligned} \underline{\beta}'XAX\underline{\beta} &= o(a) \\ \underline{g}'A\underline{g} &= o(a) \end{aligned} \tag{5.3.2}$$

Then using second order Taylor series approximations to the moments of the F ratio, it is straightforward to show that the Pitman alternative hypothesis for the F test is also H_1: $\gamma^2 = \theta/\sqrt{n}$. The numerator and denominator of F both have central chi-square distributions under H_0 and noncentral chi-square distributions under H_1 with noncentrality parameter of order $1/\sqrt{n}$. Using the methods of Section 5.2, it can be shown that F is asymptotically normally distributed under both H_0 and H_1.

The ARE of r relative to F (E_{rF}) in the general case is given by (Noether (1955)):

$$E_{rF} = \underset{n\to\infty}{\text{limit}} \frac{\left[\frac{\underline{g}'M\underline{g}}{n-k} - \frac{\underline{g}'MP_LM\underline{g}}{2n_L(1-\overline{D^2}/2n)}\right]^2 \left[\frac{2n_L}{n-k} + \frac{2n_L(1+2B_n)}{a(1+B_n)} - \frac{4n_L tr(MA)}{a(n-k)(1+B_n)}\right]}{\left[\frac{\underline{g}'M\underline{g}}{n-k} - \frac{\underline{g}'A\underline{g}}{a(1+B_n)}\right]^2 \left[\frac{tr(MP_LM)^2}{2n_L(1-\overline{D^2}/2n)} - \frac{2n_L}{n-k}\right]},$$

(5.3.3)

where $B_n = \underline{\beta}'\underline{X}'A\underline{X}\underline{\beta}/a\sigma_u^2$. The term B_n appears in the expression because the pure error sum of squares is calculated on near-replications. Using assumptions (5.1.1), (5.1.4), and (5.3.2), the limit of expression (5.3.3) simplifies to

$$E_{rF} = \underset{n\to\infty}{\text{limit}} \frac{\left[\frac{2n_L}{a} + \frac{2n_L}{n-k} - \frac{2n_L 2tr(MA)}{(n-k)\cdot a}\right]}{\left[\frac{tr(MP_LM)^2}{2n_L} - \frac{2n_L}{n-k}\right]}, \tag{5.3.4}$$

provided the limit exists.[9] Furthermore it can be shown that

$$\underset{n\to\infty}{\text{limit}} \frac{tr(MA)}{a} = 1 \qquad \text{and}$$

$$\underset{n\to\infty}{\text{limit}} \frac{tr(MP_LM)^2}{2n_L} = N_r + 1$$

(see appendix for proof). In the limit $2n_L/a \to N_rN_F/(N_F-1)$ and $2n_L/n \to N_r$, therefore (5.3.4) reduces to

$$N_r/(N_F-1) \ , \qquad N_F > 1 \ , \tag{5.3.5}$$

[9] If the data points are increased in the manner described, the term $\underline{g}'MP_LM\underline{g}/2n_L$ tends to zero by the continuity of the vector function $\underline{g}'M$.

where N_r is the maximum index of a point $\underline{x}_i$ and N_F is the maximum number of elements in each set G_i, $i=1,2,\cdots,p(n)$.

In the simple regression model, suppose that the $\underline{x}_i$'s are bounded and equally spaced for each n, so that the G_i's are formed by partitioning the range of values of $\underline{x}$ into m equal intervals. The G_i's consist of points for which the maximum distance between any pair does not exceed the width of a subinterval G_i. In order to make the r test equivalent in terms of distances between pairs of near-neighbors, the maximum value that N_r is considered to attain is $2(N_F-1)$. For example if each G_i contains N_F points, then a reasonable restriction is that N_r satisfies

$$1 \leq N_r \leq 2(N_F-1) \quad . \tag{5.3.6}$$

For the case $N_r = N_F - 1$, E_{rF} is 1 regardless of the value of N_F. In general E_{rF} satisfies

$$1/(N_F-1) \leq E_{rF} \leq N_F \quad . \tag{5.3.7}$$

5.4 Selection of Pairs for D_L.

The results of the previous section indicated that the ARE of an r-type test with smaller index per point relative to the same test with a larger index per point is less than 1. It is assumed that the distances between all pairs in the two sets tend to zero in the limit. This would suggest that the preferred scheme for selecting pairs for the set D_L would be to include as many pairs as possible and specifically all pairs whose distance is less than the critical value d_L. At the same time the index of any point should not be allowed to become considerably larger than the others.

Suppose r_1 is an r-type test statistic based upon a set of near-neighbors, D_1, and let r_2 be a test statistic of the same form using the same set of data points but based upon a different set of near-neighbors D_2. Then using the results of the previous section, a comparison can be made between the two tests involving r_1 and r_2. Let N_1 be the maximum index of any point using set D_1 and N_2 be the maximum index using D_2. Suppose that for each test the proportion of points whose index is less than its maximum value tends to zero as n tends to infinity. Then the ARE of r_1 relative to r_2 reduces to

$$E_{12} = \lim_{n\to\infty} \frac{\sigma_{20}{}^2(n)}{\sigma_{10}{}^2(n)}, \qquad (5.4.1)$$

where $\sigma_{10}{}^2(n)$ and $\sigma_{20}{}^2(n)$ are the respective variances of the statistic r_1 and r_2 evaluated at $\gamma = 0$ for sample size n. Since the tests are of the same form (see Noether (1955)), this simplifies further to

$$E_{12} = N_1/N_2, \qquad N_2 > 0. \qquad (5.4.2)$$

CHAPTER 6
RESULTS OF SIMULATION STUDIES AND APPLICATIONS

6.1 Accuracy of Approximations to the Critical Regions Under H_0.

A relatively small-scale Monte Carlo experiment was performed to compare the behavior of the statistics r and F under H_0 for small sample sizes. Variations of the following factors were taken into account in the study.

1. sample size
2. arrangement of the data points of the independent variable
3. magnitude of the variance of random error about the model
4. significance level

A sequence of twenty-one sets of X values were chosen for a sequence of regression models in two variables. The data generated for this experiment were used to evaluate the Pearson Curve approximation to the critical values of r, and to determine the amount of error in using critical values of F assuming replications.

To generate the $\underline{x}$ vectors, initially a set of twelve equally spaced two-dimensional lattice points was chosen. These are:

(-6, 4)	(-2, 4)	(2, 4)	(6, 4)
(-6, 0)	(-2, 0)	(2, 0)	(6, 0)
(-6,-4)	(-2,-4)	(2,-4)	(6,-4)

At each of the twelve lattice points, sets of 2, 3, and 4 data points were generated in turn, to give the three different sample sizes of 24, 36, and 48, respectively. The values were generated by independently selecting each of the two coordinates of each vector $\underline{x}$ from a normal distribution with mean equal to the corresponding coordinate value of the associated lattice point. The dispersion of the x_{ij}'s, which was the same for both coordinates, varied from 0.0 to 1.8 in steps of 0.09 to yield the twenty-one sets of $\underline{x}$ values.

The model used to generate the values of y under H_0 was:

$$y_i = 10.0 + 2.0\, x_{1i} + 2.0\, x_{2i} + \varepsilon_i \qquad (6.1.1)$$

$i=1,2,\cdots,n$, where $\underline{x}_i' = (x_{1i}, x_{2i})$ represents the displ ced lattice point values, and $\varepsilon_i \sim \text{IIDN}(0, \sigma^2)$. For each sample, values of the normal deviates ε_i were selected by a computer subroutine which generates random normal deviates according to a method developed by Chen (1971). Based upon the results of several trial runs two values of σ were chosen (8.0 and 9.0).[10]

The pairs of points in D_L were selected as follows. First the modified squared interpoint distance, d, between two adjacent lattice points was determined. One-half of this distance was taken as the value of d_L and the set D_L^* was formed as

$$D_L^* = D^*(-, d_L, -).$$

The pairs in D_L^* were then ordered according to the values of their interpoint distances. The smallest n_L pairs in D_L^* were then chosen for the set D_L. The number n_L was determined by

---

[10] Reasons for selecting these particular values are discussed in Section 6.2.

$$n_L = \min(m_r,\ n_L^*)\ , \quad \text{where}$$

n_L^* = number of elements in D_L^*, and

$$m_r = \frac{12n_r(n_r-1)}{2}\ ,$$

with n_r equal to the number of replications at each of the twelve lattice points. The result was equivalent to choosing $D_L = D(n_L,d_L,-)$, but the two stage procedure resulted in more efficient use of computing time due to the substantial reduction in the number of elements to be sorted. The exact moments of the statistic r were determined in a separate computer run along with the eigenvalues of MP_LM, and the parameters necessary for using the Pearson Curve approximation for the r test. Critical values of the r statistic were then determined for five levels of significance (0.10, 0.05, 0.025, 0.01, 0.005) from tables of the Pearson Curves (Johnson, Nixon and Amos (1963)).

The sets, G_i, $i=1,2,\cdots,p$, were formed using the pairs in D_L. The size of each G_i was set equal to n_r for each i and for each set of X values. The value of p was allowed to vary over the sets of X's. A criterion distance variable, d_c, was used. Beginning at zero, the value of d_c was increased until it exceeded the smallest interpoint distance available. These two points were then included in G_1 and the points associated with this pair were removed from further consideration. The value of d_c was increased further until n_r such points were included in G_1 or until the set D_L was exausted. Then d_c was returned to zero and G_2 was formed in a similar manner with the remaining points. This process yielded p groups. Critical values for the F test were obtained for identical levels of significance (α) from tables in Pearson and Hartley (1966).

Results of this study are based upon 1500 samples of observations at each sample size (24, 36, and 48) and at each value of σ_X. These results for $\alpha = 0.05$ appear in Table 6.1.1. The entries in the table represent the number of rejections of H_0 out of the 1500 samples of data generated under model (6.1.1).[11]

The number of rejections of H_0 for each model follows a binomial distribution with mean $n\alpha$ and variance $n\alpha(1-\alpha)$. A test for a significant difference between the true and the estimated critical values for both tests was made using the normal approximation. Subsequently under H_0 the number of rejections at significance level $\alpha = 0.05$ should fall within the interval [57.6, 92.4] for approximately 95% of the samples.

For the r test only 3 out of $126 = 21 \times 3 \times 2$ (or 2.4%) of the counts fell outside the interval. All three of these values occured for samples of size 24. For the F test a total of 19 out of 126 (or 15.1%) of the counts fell outside the interval [57.6, 92.4]. A breakdown of these values by sample size and by value of σ_u are given in Table 6.1.2.

---

[11]Results for the other values of α are available from the author.

Table 6.1.1 Number of rejections of H_0 out of 1500 samples generated under H_0 using the r test and the F test with $\alpha = 0.05$.

$\sigma_u = 8.0$

Test	F	r	F	r	F	r
Sample Size	24	24	36	36	48	48
Sample Number						
1	81	81	70	70	75	75
2	71	73	77	72	82	79
3	64	65	58	64	71	65
4	65	69	76	82	65	76
5	71	76	59	70	62	82
6	70	69	65	75	58	63
7	65	55	52	64	58	69
8	74	68	74	85	65	67
9	66	77	76	74	64	67
10	63	82	62	77	61	70
11	86	74	58	89	56	63
12	81	77	80	78	55	69
13	66	69	58	86	31	75
14	90	91	64	83	46	75
15	60	74	65	73	49	71
16	54	73	56	69	60	77
17	74	95	57	67	36	63
18	65	73	49	90	42	76
19	79	76	49	78	44	70
20	67	81	60	74	64	78
21	79	81	79	81	73	78

Table 6.1.1 Continued

$\sigma_u = 9.0$

Test Sample Size	F 24	r 24	F 36	r 36	F 48	r 48
Sample Number						
1	77	78	79	79	70	70
2	63	62	77	78	73	78
3	70	71	67	64	81	83
4	82	82	78	83	71	65
5	66	68	71	72	65	70
6	68	72	74	70	57	72
7	70	63	63	69	69	72
8	60	63	70	89	64	68
9	79	83	59	75	54	66
10	63	81	82	89	67	88
11	77	64	58	84	65	83
12	87	93	72	77	47	73
13	69	69	66	74	34	78
14	74	91	69	75	54	84
15	76	92	74	85	61	64
16	62	77	69	73	68	68
17	81	68	62	77	42	77
18	71	82	58	84	65	76
19	65	69	57	72	62	83
20	66	75	59	74	45	77
21	63	75	70	63	69	73

Table 6.1.2: Number of Rejection counts out of 126 Samples falling outside the 95% Confidence Interval [57.6, 92.4] for the F test With $\alpha = 0.05$.

	σ_u 8.0	σ_u 9.0	Total
n = 24	1	0	1
n = 36	4	0	4
n = 48	8	6	14
Total	13	6	19

A simple chi-square test of independence indicates no change in the distribution of rejections over the three sample sizes for the two values of σ_u. However a similar test procedure used on the three marginal totals corresponding to the three sample sizes indicates a significant difference. For smaller sample sizes (and also smaller size groups of near-neighbors) the critical values of the F test are adequate. However for n = 48 (using groups of near-replicates of size 4) the approximation to the true critical value is inadequate in terms of numbers of rejections falling outside the 95% confidence interval.

6.2 Power Studies

A second simulation experiment was simultaneously conducted to compare the power of the r test and the F test for small samples. In terms of power, it was hypothesized that the r test would do no better than the F test if there were replications of the independent variable. As the data points changed from a clumped distribution gradually to a random distribution it was assumed that the F test would begin to look worse either in terms of power or in terms of the deterioration of the significance levels under H_0, or both. It was desirable to determine if a point existed in such a sequence of dispersions at which the r test would give better results than the F test. For this reason relatively more consideration was given to the arrangement of the values of the independent variable than to variations of other parameters. Subsequently the sequence of twenty-one sets of values of the independent variable described in the previous section were used to form a corresponding sequence of twenty-one regression models in two variables. The data was the same as that used for the experiment described in Section 6.1. The method of generating the data is described there as well as the method for choosing D_L and forming the disjoint groups for the F test. The model proposed for fitting the data was

$$y_i = \mu + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

$i=1,\cdots,n$ where $\varepsilon_i \sim IIDN(0,\sigma^2)$.

Three alternative models were selected for generating values of the dependent variable. In each case the true relationship was expressed by the addition to the basic model of Section 6.1, of one term which was non-linear in one or more of the independent variables. Then a simulation of random error about the model was conducted. Two quadratic terms and one interaction term were considered, yielding the three models. Specifically, they are

$$\text{I:} \quad y_i = 10.0 + 2.0x_1 + 2.0x_2 + 0.5x_1^2 + \mu_i$$

$$\text{II:} \quad y_i = 10.0 + 2.0x_1 + 2.0x_2 + 0.5x_2^2 + \mu_i$$

$$\text{III:} \quad y_i = 10.0 + 2.0x_1 + 2.0x_2 + 0.5x_1 x_2 + \mu_i$$

For each sample the variables $u_i \sim IIDN(0,\sigma_u^2)$ were selected by computer as described in Section 6.1.[12] Simulations were done for two values of σ_u (8.0 and 9.0). Based upon several preliminary trial runs of the simulation program it was found that these values of σ_u resulted in estimated power for both tests that was sufficiently far from both 0 and 1 for all 63 = 21 × 3 regression models. This allowed for meaningful comparisons of power for the two tests. Estimates of the power of the tests were obtained from the number of rejections of the null hypothesis out of 500 trials at each of three sample sizes. The results of the power considerations are presented in Table 6.2.1 for $\sigma_u = 9.0$. Results for $\sigma_u = 8.0$ are similar in all aspects.

The results indicate that both tests are sensitive to the non-linear alternatives. Except in only a very few cases, the estimated power of the r test is at least as great as the estimated power of the F test. The magnitude of the difference increases both with the sample size and with the value of σ_X, the parameter representing the dispersion of the x values, for each alternative. As discussed in Section 6.1, this is in part due to the apparent reduction in the actual significance level of the F test when it is based upon near-replicates. As the values of the independent variable become more randomly dispersed (σ_X increases), the power of the r test seems to increase relative to that of the F test. However it steadily increases up to a point about half way through the sequence, and

[12] The same random normal deviates were used in this experiment as were used in the experiment of the previous section.

Table 6.2.1 Number of rejections of H_0 out of 500 samples of data generated under models I, II, and III with σ_u = 9.0, using the r test and the F test with α = 0.05.

	Model	I			II			III		
	N	24	36	48	24	36	48	24	36	48
Sample Number	Test									
1	r	285	460	492	73	132	191	255	465	485
	F	285	484	492	72	132	191	254	465	485
2	r	274	460	489	63	139	198	233	422	490
	F	278	488	490	62	144	196	228	421	489
3	r	287	451	494	77	104	196	281	383	481
	F	286	472	492	82	112	189	274	380	480
4	r	299	446	494	82	140	189	249	441	479
	F	299	476	493	74	137	175	246	439	475
5	r	275	443	483	69	137	179	221	431	483
	F	276	484	483	66	137	159	210	427	484
6	r	281	464	487	55	113	170	221	405	472
	F	276	474	478	57	103	131	213	371	462
7	r	165	420	495	62	110	185	211	385	488
	F	201	422	489	69	78	169	212	326	482
8	r	200	452	472	69	126	181	168	377	475
	F	167	462	466	49	80	151	138	322	459
9	r	267	362	491	72	131	176	262	376	470
	F	267	329	469	52	85	132	238	286	413
10	r	257	445	489	83	148	160	211	392	478
	F	195	439	478	47	107	113	158	340	443

Table 6.2.1 Continued

Model		I			II			III		
N		24	36	48	24	36	48	24	36	48
Sample Number	Test									
11	r	167	450	493	56	117	236	164	398	437
	F	145	450	479	47	87	124	119	328	351
12	r	249	437	497	90	87	226	247	276	475
	F	210	413	465	78	59	139	192	178	402
13	r	242	430	497	143	169	312	191	400	486
	F	206	313	480	100	92	157	151	283	448
14	r	169	457	494	73	135	321	191	403	451
	F	119	394	458	60	69	208	139	244	266
15	r	354	236	468	74	247	240	240	308	391
	F	274	220	349	77	124	182	182	247	235
16	r	305	469	452	119	313	303	196	289	448
	F	217	479	360	78	222	177	110	206	339
17	r	195	495	491	61	162	276	175	439	478
	F	109	484	429	36	103	173	97	328	427
18	r	159	454	500	46	185	376	112	446	488
	F	113	450	485	37	131	247	57	334	431
19	r	255	433	499	121	249	338	262	455	465
	F	143	350	469	91	162	214	186	334	411
20	r	342	494	481	92	286	312	201	419	445
	F	205	465	451	74	168	197	97	277	372
21	r	482	491	498	100	400	399	339	314	470
	F	414	448	483	76	298	305	249	154	395

then the power begins to vary considerably. This may be due to the chance variation possible in the method of generation of the data. This trend is also evident for the F test.

The r test as well as the F test seem to be most powerful against model I and least powerful against model II-both quadratic models - while the interaction alternative falls between these two in terms of power. This seems to indicate that the power for small samples may be influenced more by the relative number of points appearing in a line in the direction of the nonlinearity than by the functional form of the alternative term. The quadratic term of model I involves the variable with the lattice points having the largest number of points in the direction of the nonlinearity, while those of model II have the fewest number. Both tests are more powerful for $\sigma_u = 8.0$ than for $\sigma_u = 9.0$.

The test was also compared with the usual t test by including the nonlinear term in the model and testing that its coefficient is zero. In nearly all cases considered the value of the power for this test was estimated to be nearly equal to one. Subsequently the results are not presented here. This might have been expected since the test makes use of prior indications as to the nature of the functional form of the nonlinear alternative.

6.3 Application of the Test Procedure

The test procedure was illustrated on a set of sociological data. The data represent observations of five sociological characteristics on 51 individual school children. The model proposed to represent the relationship is

$$y_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_4 X_{4i} + \varepsilon_i$$

$i=1,2,\cdots,51$.

A graph was made of the smallest 5% of the pairwise distances. There were four noticeable jump points in the graph near the values 0.012, 0.014, 0.017, and 0.020. The value of d_L was set equal to the smallest of these points 0.012. There were 15 pairs of points with interpoint distance less than d_L. The maximum value for the index of each point was one so that the beta bounds test procedure could be illustrated. Thus D_L was defined by $D_L = D(0.012, 15, 1)$. This yielded a total of 9 disjoint pairs in D_L. The same nine pairs were used as nine groups of two for the comparison with the F test. The first four power sums of the eigenvalues of MP_LM are computed as $\sum_{i=1}^{n} \nu_i^s = tr(MP_LM)^s$[13] for s=1,2,3,4. The first four moments of r $(\mu_1, \mu_2, \mu_3, \mu_4)$ are then computed using the formulae of Section 3.4. The Pearsonian parameters β_1 and β_2 are then computed as

$$\beta_1 = \mu_3^2/\mu_2^3 \quad \text{and}$$

$$\beta_2 = \mu_4/\mu_2^2 .$$

This yielded the values $\sqrt{\beta_1} = 0.60$ and $\beta_2 = 3.29$ indicating a type I, bell-shaped, (beta) approximation to the distribution of standardized r. The lower 5% critical value from the tables of the Pearson Curves is -1.44. The actual value of the statistic r and its first two moments under H_0 are: $r = 0.895$, $\mu_1 = 0.996$, and $\sqrt{\mu_2} = 0.412$.

The smallest and largest eigenvalues of MP_LM are 0.0000 and 1.9999. Therefore the value of r can fall in the interval [0, 5.11].

---

[13] A computer program for application of the test procedure is available from the author.

The critical value for the distribution of the non-standardized r statistic is $(-1.44)(0.412) + 0.996 = 0.403$. Therefore it was concluded that there was no indication of significant nonlinearity in the model.

Using the beta bounds approximate test the estimated lower critical value is the lower critical value of the random variable r_L where

$$r_L \sim \frac{46}{9}\, \beta(5/2,\ 41/2) \ .$$

The lower 5% critical value for r_U is 0.120.

Using the F test with the assumption of replications corresponds to using as an approximation for r, the distribution of r_U, where

$$r_U \sim \frac{46}{9}\, \beta(9/2,\ 37/2) \ .$$

The lower 5% critical value of r_U is 0.400. In all three tests the assumed model cannot be rejected against any general p(9) parameter nonlinear alternative model.

CHAPTER 7
SUMMARY AND CONCLUSIONS

7.1 Summary and Conclusions.

The statistic r was developed for use in a test against non-linearity of the general linear regression model. The classical lack of fit tests usually involve the estimation of the variance of random error about the model from observations associated with repeated values of the independent variable. An alternative to this procedure was developed for small to moderate sample sizes when there are no replications at any values of the independent variable. The alternative test is based upon the statistic r which is a ratio of an appropriate estimate of the error variance, to the error mean square. This estimate is the average squared difference between residuals associated with a set of near-neighbor pairs of the independent variable. The Mahalanobis inter-point distance was used to determine the set of near-neighbors.

The function was expressed as the ratio of two quadratic forms in standard normal variables and the relevant theory concerning the exact distribution of a general statistic of this type was reviewed. It was noted that the statistic has an asymptotically normal distribution under the null hypothesis, but that the convergence to normality in general may be very slow. The exact distribution of r was derived for several special cases. All of these involve linear combinations of beta random variables. It was shown that when the independent variable consists of pairs of replicates, the r statistic can be expressed as a one-to-one function of the classical F statistic used to test for lack of fit.

Considerable attention was given to a review of the existing methods for approximating the distribution of the test statistic, since the exact expressions available are not practical for routine application. All of these approximations involve variations of the beta distribution. For purposes of conducting the test, it was recommended that critical values be approximated using Pearson Curves. The recommendation was based upon the existence of the moments or r, and upon the fact that the availability of extensive tables allows for the elimination of the intermediate step of determining the specific form of the distribution function used in the approximation. Finally it was noted that the family of Pearson Curves includes the beta and the normal distributions.

A small scale Monte Carlo experiment was conducted to check the accuracy of the Pearson Curve approximation at six critical values. At the same time the significance levels of the corresponding F test were examined. It was concluded that the Pearson Curves gave relatively good approximations to the critical regions examined. However the F approximation tended to have smaller tails than the true distribution of the modified F statistic when the number of points involved in the groups of replicates was increased.

When the proposed regression model expresses the actual relationship between the variables, the expected value of the test statistic was expressed as a simple function of the average, over the pairs of near-neighbors, of the squared interpoint distances. Using a technique described in Durbin and Watson (1950), random variables which bound the statistic r were derived. For certain arrangements of the data points, it was shown that the distribution of these random variables could be expressed as a constant times a simple beta distribution whose parameters do not depend upon the specific values of the independent variables. A possible test procedure using these results was described.

A general expression for the nonlinear alternative model was given. Large sample conditions were derived which indicated the behavior of the statistic r under H_A, and a method was described to approximate the power of the test. A large sample comparison of the test with the F test was made using the criterion of asymptotic relative efficiency as defined by Pitman. Considerable attention was given to determining conditions under which the statistic is asymptotically normal under the Pitman alternative hypothesis. Sufficient conditions are essentially that the eigenvalues of the matrix of the quadratic form of the numerator of r be bounded for all n. This can be accomplished by requiring that the index of any point be uniformly bounded for all points as n increases. In these cases the Pitman efficiency of r relative to F is in the neighborhood of one. When the index pairs which overlap the boundaries of the groups of points formed for the F test, the efficiency becomes greater than one.

As a small sample comparison, a simulation experiment was conducted for a two-variable regression model to compare the power of the r and F tests against alternative models containing an additional term which is nonlinear in form. Two quadratic terms and one interaction term were considered, yielding three alternatives. Results indicated that the r test is at least as powerful as the F test in nearly all of the cases considered. However the simulation results also indicated that the actual significance levels of the F test are decreased when near-replications are used. Apparently the sensitivity of the r test, as well as that of the F test, depended more upon the number of points available in the direction of the nonlinear variable than on the functional form of the alternative for the cases considered in this experiment.

7.2 Suggestions for Future Research

Numerous problems related to the work done in this thesis seem to deserve further consideration. This is especially true if the r test is ever to be considered for routine use. Many of these problems are purely computational and some are theoretical in nature. Of major concern is the creation of the set D_L. Subsequently the choice of the parameter d_L (or N_L) should be more rigorously investigated than is done here. The computation of moments, pairwise distances, and possibly eigenvalues prohibits the use of the r test in large samples routinely. A study should be made of computer methods and computational shortcuts to minimize computer time needed for larger sample sizes. If pairwise distances are computed only for pairs of points within subsets formed from one or two initial partitionings of the x-space considerable computations can be saved, while sacrificing relatively few near-neighbor pairs. Other methods similar to this may be discovered.

Although the r test is suggested for testing against alternatives of nonlinearity, most of the theoretical development was done in the framework of testing independent observations against correlated error structures (see Anderson (1948)). The use of the r test against such alternatives should be considered. The sensitivity of the test to other types of misspecifications should be examined theoretically.

A more extensive study could be made to determine the relationship between the patterns of indices of the points and the patterns of eigenvalues of MP_LM for small samples. A study of the type should be made to determine if a test procedure could be developed which, like the beta approximate bounds test, is independent of the actual values in the matrix X. The relative merits of pre-choosing the indices of the points in patterns could be considered for use in the case of data in which exact replicates cannot be selected.

As mentioned in Chapter 1, the behavior of the r test when used on transformed residuals should be studied. For example, consider the BLUS residuals of Theil (1965) which are independently distributed. Many of the problems of the r test resulting from the dependence of the usual residuals might be avoided if the r statistic were based upon residuals, such as the BLUS residuals referred to in Chapter 1, which are independently distributed. This is a particularly important consideration since the r test is recommended for small samples.

Since the r test is suitable for small samples, research concerning the properties of the test leads itself to almost endless possibilities for Monte Carlo studies. Studies of the type described in the previous sections could be extended to include different emphases than are given here. Particularly important is the case in which the x values have a more normal-shaped dispersion pattern, since observations would be clumped in the center. By the same methods, the r test could be compared with respect to sensitivity and power to other tests of nonlinearity such as the Prais-Houthakker test (1955) and Thornby's nonparametric test (1972). The work done in this investigation is of a rather general nature, even though numerous assumptions are made throughout. It is believed that many useful analytical results can be obtained by considering several more specific cases. For example interesting cases would include alternatives of polynomial trand using equally-spaced variables or indicator variables in the simple regression scheme.

REFERENCES

Anderson, R. L. 1942. Distribution of the serial correlation coefficient. The Annals of Mathematical Statistics 13: 1-13.

Anderson, T. W. 1948. On the theory of testing serial correlation. Skandinavisk Aktuarietidskrift 31: 81-116.

Anderson, T. W. 1958. The Statistical Analysis of Time Series. John Wiley and Sons, Inc., New York.

Anscombe, F. J. 1961. Examination of residuals. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1: 1-36.

Anscombe, F. J. and Tukey, J. W. 1963. The examination and analysis of residuals. Technometrics 5: 141-160.

Ashar, V. G. and Wallace, T. D. 1972. Sequential methods in model construction. Review of Economics and Statistics 54: 172-178.

Chen, E. H. 1971. A random normal number generator for 32-bit-word computers. Journal of the American Statistical Association 66: 400-403.

Cramér, H. 1946. Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J.

Bloch, D. A. and Watson, G. S. 1967. A Bayesian study of the multinomial distribution. The Annals of Mathematical Statistics 38: 1423-1435.

Daniel, C. and Wood, F. S. 1971. Fitting Equations to Data. John Wiley and Sons Inc., New York.

Durbin, J. and Watson, G. S. 1950. Testing for serial correlation in least squares regression I. Biometrika 37: 409-428.

Durbin, J. and Watson, G. S. 1951. Testing for serial correlation in least squares regression II. Biometrika 38: 159-178.

Durbin, J. and Watson, G. S. 1971. Testing for serial correlation in least squares regression III. Biometrika 58: 1-19.

Fisher, R. A. 1922. The goodness of fit of regression formulae and the distribution of regression coefficients. Journal of the Royal Statistical Society 85: 597-612.

Henshaw, R. C. 1966. Testing single-equation least squares regression models for autocorrelated disturbances. Econometrica 34: 646-660.

Hoeffding, W. and Robbins, H. E. 1948. The central limit theorem for dependent variables. Duke Mathematical Journal 15: 773-780.

Imhoff, J. P. 1961. Computing the distribution of quadratic forms in normal variables. Biometrika 48: 419-426.

Johnson, N. L., Nixon, E. and Amos, D. E. 1963. Tables of percentage points of Pearson Cruves for given $\sqrt{\beta_1}$ and β_2 expressed in standard measure. Biometrika 50: 459-498.

Kamat, A. R. 1963. Asymptotic relative efficiencies of certain test criteria based on first and second differences. Journal of the Indian Society of Agricultural Statistics 15: 213-222.

Kendall, M. G. and Stuart, A. 1963. The Advanced Theory of Statistics, volume 2, Charles Griffin and Company, Inc., London.

Koopmans, T. 1942. Serial correlation and quadratic forms in normal variables. The Annals of Mathematical Statistics 13: 14-33.

Lancaster, P. 1969. Theory of Matrices. Academic Press, New York.

Loeve, M. 1955. Probability Theory. van Nostrand, New York.

Noether, G. 1955. On a theorem of Pitman. The Annals of Mathematical Statistics 26: 64-69.

Orey, S. 1957. A Central limit theorem for m-dependent random variables. Duke Mathematical Journal 25: 543-546.

Pearson, E. S. and Hartley, H. O. 1966. Biometrika Tables for Statisticians. Cambridge University Press.

Pitman, E. J. G. 1937. The 'closest' estimates of statistical parameters. Proceedings of the Cambridge Philosophical Society 33: 212-222.

Prais, S. J. and Houthakker, H. S. 1955. The Analysis of Family Budgets. Cambridge University Press.

Press, S. J. 1969. On serial correlation. The Annals of Mathematical Statistics 40: 188-190.

Ramsey, J. B. 1969. Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society, series B, 31: 350-371.

Scheffe, H. 1959. The Analysis of Variance. John Wiley and Sons, Inc., New York.

Searle, S. R. 1970. Linear Models. John Wiley and Sons, Inc., New York.

Thornby, J. I. 1972. A robust test for linear regression. Biometrics 28: 533-543.

Tukey, J. W. 1959. One degree of freedom for nonadditivity. Biometrics 5: 232-242.

Theil, H. 1965. Analysis of distrubances in regression analysis. Journal of the American Statistics Association 60: 1067-1079.

von Neumann, J. 1941. Distribution of the mean square successive difference to the variance. The Annals of Mathematical Statistics 12: 367-395.

von Neumann, J., Kent, R. H. Bellinson, H. R., and Hart, B. I. 1941. The mean square successive difference. The Annals of Mathematical Statistics 12: 153-162.

APPENDIX

Lemma A1: Let I_i be the index of the point $\underline{X}_i$ $(i=1,\cdots,n)$. Suppose $I_i \leq N_r$ for all i and let $\ell(n)$ be the number of points for which $I_i < N_r$. If $\lim_{n\to\infty} \ell(n)/n = 0$, then

$$\lim_{n\to\infty} \frac{\mathrm{tr}(MP_LM)^2}{2n_L} = N_r + 1.$$

Proof. Let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ be the eigenvalues of P_L and let $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_n$ be the eigenvalues of MP_LM. Since X is of full column rank

$$0 \leq \mu_{i+k} \leq \nu_i \leq \mu_i, \qquad i=1,2,\cdots,n-k,$$

by Theorem 3.6.1. Therefore

$$\sum_{i=1}^{n} \mu_i^2 - \sum_{i=1}^{k} \mu_i^2 \leq \sum_{i=1}^{n-k} \nu_i^2 \leq \sum_{i=1}^{n-k} \mu_i^2$$

but $\sum_{i=1}^{n} \mu_i^2 = \mathrm{tr}(P_L^2) = \sum_{i=1}^{n} I_i(I_i+1)$.

Since the indices of $\underline{x}_i$ are uniformly bounded, the eigenvalues of P_L are also uniformly bounded by some value, λ_N, by the definition of P_L. Therefore

$$\sum_{i=1}^{n} I_i(I_i+1) - k\lambda_N^2 \leq \sum_{i=1}^{n-k} \nu_i^2 \leq \sum_{i=1}^{n} I_i(I_i+1)$$

or,

$$(n-\ell(n))N_r(N_r+1) - k\lambda_N^2 \leq \sum_{i=1}^{n-k} \nu_i^2 \leq n\cdot N_r(N_r+1)$$

Since $\qquad (n-\ell(n))N_r \leq 2n_L \leq nN_r$,

$$\frac{(n-\ell(n))N_r(N_r+1)}{nN_r} - \frac{k\lambda_N^2}{2n_L} \leq \frac{\sum_{i=1}^{n-k} \nu_i^2}{2n_L} \leq \frac{nN_r(N_r+1)}{(n-\ell(n))N_r}$$

Or, $$(1-\frac{\ell(n)}{n})(N_r+1) - O(1/n) \leq \frac{tr(MP_LM)^2}{2n_L} \leq \frac{N_r+1}{(1-\ell(n)/n)}$$

But $$\lim_{n\to\infty} (1-\frac{\ell(n)}{n})(N_r+1) = \lim_{n\to\infty} \frac{(N_r+1)}{(1-\frac{\ell(n)}{n})} = N_r + 1$$

Therefore $$\lim_{n\to\infty} \frac{tr(MP_LM)^2}{2n_L} = N_r + 1.$$

Lemma A2. For fixed n let a(n) be the rank of the matrix A. If a(n) is of order n, then

$$\lim_{n\to\infty} \frac{tr(MA)}{a} = 1$$

Proof. Let $\mu_1(n) \geq \mu_2(n) \geq \cdots \geq \mu_n(n)$ be the eigenvalues of MA, and let $\alpha_1(n) \geq \cdots \geq \alpha_n(n)$ be the eigenvalues of A. Then since X is of full column rank

$$\alpha_{i+k}(n) \leq \mu_i(n) \leq \alpha_i(n), \qquad i=1,2,\cdots,n-k,$$

by Theorem 3.6.1. Therefore

$$\sum_{i=k+1}^{n} \alpha_i(n) \leq \sum_{i=1}^{n-k} \mu_i(n) \leq \sum_{i=1}^{n} \alpha_i(n) ,$$

$\alpha_i(n) \geq 0$ for every i and for all n. Since A is idempotent, the eigenvalues of A are

$$\underbrace{1, 1, \cdots, 1}_{a}, \underbrace{0, 0, \cdots, 0}_{n-a}$$

Therefore,

$$\sum_{i=k+1}^{a(n)} 1 \leq \sum_{i=1}^{n-k} \mu_i(n) \leq \sum_{i=1}^{a(n)} 1$$

or, $$a(n)-(k+1) \leq \sum_{i=1}^{n-k} \mu_i(n) \leq a(n) ,$$ implying

$$1 - \frac{(k+1)}{a(n)} \leq \frac{tr(MA)}{a(n)} \leq 1$$

Since $a(n) = 0(n)$, $\lim_{n\to\infty} \frac{k+1}{a(n)} = 0$, and

therefore, $\lim_{n\to\infty} \frac{tr(MA)}{a(n)} = 1.$