The Annals of Statistics
1989, Vol. 17, No. 2, 673–683

LACK-OF-FIT TESTS BASED ON NEAR OR EXACT REPLICATES

BY RONALD CHRISTENSEN

University of New Mexico

This article shows how to modify an approximate lack-of-fit test proposed by Neill and Johnson (1985) to obtain an exact F test. The modified test is shown to be consistent and to yield uniformly most powerful invariant tests for two specific types of lack of fit. Pure types of lack of fit are (1) lack of fit that exists between clusters of near replicates and (2) lack of fit that is contained within clusters of near replicates. Lack of fit that is a mixture of these two types can be difficult or impossible to find depending on the nature of the mixture.

1. Introduction. Neill and Johnson (1985) studied a lack-of-fit test based on near replications. They established the asymptotic distribution of the test and obtained results on consistency. Their test has also been discussed in Neill (1982) and in a review article, Neill and Johnson (1984a).

The Neill–Johnson (NJ) test is based on modifying a regression model in which near replicates have been identified so that the classical lack-of-fit test based on exact replication can be used. To do this, Neill and Johnson estimate the parameters of the regression model prior to performing the classical lack-of-fit test. This prior estimation invalidates the usual distribution theory.

In this article it will be shown that the prior estimation is not necessary and that an exact F test is available. This exact F test is related to work by Shillington (1979) and Tsiatis (1980).

Section 2 sets the notation to be used, discusses the NJ test, gives the exact test procedure and discusses its relation to other work. Section 3 examines the types of lack of fit that can be detected by the exact test, establishes the consistency of the test and discusses power.

2. Lack-of-fit tests based on near replicates. A regression model with near replicates can be written as

$$(1) \qquad y_{ij} = x'_{ij}\beta + e_{ij}, \qquad i = 1, \dots, c, \; j = 1, \dots, n_i.$$

The number of groups of near replicates is c, there are n_i near replicates in the ith group. The dependent variable is y_{ij}, the vector x_{ij} is a vector of independent variables, β is a vector of regression coefficients and the e_{ij}'s are independent $N(0, \sigma^2)$. In matrix notation model (1) can be rewritten as

$$(1') \qquad Y = X\beta + e, \qquad e \sim N(0, \sigma^2 I).$$

Received March 1987; revised March 1988.
AMS 1980 *subject classifications*. Primary 62J05; secondary 62F03.
Key words and phrases. Regression, linear model, model adequacy, pure error.

Let

$$\bar{x}_{i.} = \sum_{j=1}^{n_i} x_{ij}/n_i.$$

A model similar to (1) but which contains exact replicates is

$$(2) \qquad y_{ij} = (\bar{x}_{i.})'\beta + e_{ij}.$$

Model (2) will be written in matrix notation as

$$(2') \qquad Y = X_0\beta + e.$$

If $x_{ij} = \bar{x}_{i.}$ for all i and j, then models (1) and (2) are identical and contain exact replicates. The classical lack-of-fit test based on exact replications is to test model (2) against a larger model, namely the one-way analysis of variance (ANOVA) model

$$(3) \qquad y_{ij} = \mu_i + e_{ij}.$$

In matrix notation model (3) is rewritten as

$$(3') \qquad Y = Z\gamma + e.$$

It can be shown [cf. Christensen (1987), Section VI. 6] that model (3) is in some sense the largest model that subsumes model (2).

A naive lack-of-fit test based on near replicates [cf. Draper and Smith (1981), Section 1.5] is to test model (1) against model (3). Models (1) and (3) are not hierarchical models so an exact F test is not achieved. Moreover, Neill (1982) has shown that this test does not approximate its nominal size.

Neill and Johnson have improved on the naive test in the following way. Let $\Delta = X - X_0$. Model (1) can be rewritten as

$$Y = X_0\beta + \Delta\beta + e$$

or

$$(4) \qquad Y - \Delta\beta = X_0\beta + e.$$

The right-hand side of (4) involves exact replicates. The classical lack-of-fit test is to test model (4) against

$$(5) \qquad Y - \Delta\beta = Z\gamma + e.$$

The only problem with this test is that since β is unknown, the left-hand sides of both equations (4) and (5) are also unknown. Neill and Johnson avoid this problem by substituting $\hat{\beta} = (X'X)^{-1}X'Y$ for β on the left-hand sides and then doing the test. They establish that the test based on this substitution has nice asymptotic properties and in Neill and Johnson (1984b), they examine the test's small sample properties.

In fact, to test model (4) against model (5), it is not necessary to estimate β. Models (4) and (5) can both be rewritten in forms that allow an exact F test. Let $M_Z = Z(Z'Z)^{-}Z'$ be the perpendicular projection operator onto the column space of Z [denoted $C(Z)$]. Note that since Z is the design matrix for the one way

ANOVA model (3), $X_0 = M_Z X$. Thus, model (5) is equivalent to

$$\begin{aligned} Y &= Z\gamma + \Delta\beta + e \\ &= Z\gamma + (X - X_0)\beta + e \\ &= Z\gamma + (X - M_Z X)\beta + e \\ &= Z\gamma + (I - M_Z)X\beta + e. \end{aligned}$$

Because $C(Z, (I - M_Z)X) = C(Z, X)$, model 5 is also equivalent to

$$Y = X\beta + Z\gamma + e, \tag{6}$$

where the γ's in (5) and (6) will typically be different. Recall that model (4) is equivalent to model (1) so the test of model (4) versus model (5) is identical to the test of model (1) versus model (6). In other words, an exact F test for lack of fit is available by testing model (1) against model (6). Note that if all near replicates are exact replicates, then this is precisely the classical test.

In testing for lack of fit, the sum of squares for error in model (6) is analogous to the sum of squares for pure error. The difference in the sum of squares for error between models (1) and (6) is analogous to the sum of squares for lack of fit.

Tsiatis (1980) and Fienberg and Gong (1984) have proposed similar tests for logistic regression. Landwehr, Pregibon and Shoemaker (1984) used the linear structure $X\beta + Z\gamma$ to estimate pure error in logistic regression. This linear structure is also implicit in the exact F test for lack of fit proposed by Shillington (1979).

Shillington's test procedure is considerably simplified if presented in terms of models. Shillington's test uses the mean squared error from model (6) as a mean square for pure error. However, Shillington's test uses the difference in the sums of squares for error between models (2) and (3) as a sum of squares for lack of fit. Because $C(X_0) \subset C(Z) \subset C(X, Z)$, it is a simple matter to check that Shillington's procedure gives an exact F test. On the other hand, a direct test of model (1) versus model (6) is more appealing.

3. Models for the alternative and consistency. This section begins with a discussion of the types of lack of fit that can be detected by tests based on near or exact replicates. Lack of fit can be described as a continuum from existing between clusters to existing within clusters. In the middle of the continuum there is little hope of detecting lack of fit. However, for either of the pure types of lack of fit it is shown that the test presented here is consistent. We begin by discussing the continuum of lack of fit for the classical test based on exact replicates.

The classical lack-of-fit test is closely tied to one-way analysis of variance in which clusters of replicates are identified with different treatment groups. The classical test is designed to identify lack of fit that exists between clusters. Somewhat surprisingly, the classical test can also be used to identify lack of fit that exists within clusters. What the test has problems with is lack of fit that is a combination of between-cluster and within-cluster lack of fit.

Between-cluster lack of fit is the kind typically associated with the classical lack-of-fit test. It is assumed that all observations in a cluster have exactly the same mean and some model is postulated (e.g., a line) for the relationship between the means. If the postulated model does not adequately describe the relationship between the cluster means we have a between-cluster lack of fit.

Within-cluster lack of fit is a more unfamiliar idea. Consider a simple linear regression with replication

$$y_{ij} = \beta_0 + \beta_1 x_i + e_{ij},$$

where there are two observations for each value x_i (i.e., $j = 1, 2$). Now suppose that the lack of fit is contained within the clusters. For example, suppose that the mean of the first of the two observations is always ν units greater than the mean for the second observation [i.e., $E(y_{i1}) = E(y_{i2}) + \nu$]. Although the classical lack-of-fit test is not designed to detect such a phenomenon, it is capable of doing just that. As will be seen below, such a phenomenon will cause the F statistic for lack of fit to be unusually *small*.

The test presented in this article and other tests based on clusters of near replicates share these characteristics of the classical test. They can detect lack of fit either between or within clusters. This will be shown explicitly for the test presented here. Since the classical test based on exact replicates is a special case of the test presented here, the arguments given for the new test also apply to the classical test.

In the test based on exact replicates, it is traditionally assumed that $E(y_{ij})$ is the same for all j. In other words, it is assumed that there is no lack of fit within clusters. The only lack of fit considered is that the current model does an inadequate job of explaining the relationship between the different clusters $(i = 1, \dots, c)$ of $E(y_{ij})$'s.

Lack-of-fit tests using near replicates behave similarly. These tests are based on the hope that, within each cluster, $E(y_{ij})$ will be nearly the same. If this is true, the F statistic will be large when there is a lack of fit between clusters. However, the F statistic will be small if the lack of fit exists within clusters. If both kinds of lack of fit are present, the F statistic will lean toward the predominant type of lack of fit. If the two types of lack of fit cancel each other the F statistic will be unable to detect lack of fit.

Lack of fit within clusters can be illustrated for near replicates with the following example. Consider a simple linear regression on the four (x, y) pairs $(-2, 1)$, $(-1, -1)$, $(1, -1)$, $(2, 1)$. These form a perfect V shape with a vertex at $(0, -3)$. If we cluster points based on their x values, it is natural to form two clusters. One cluster consists of the points with negative x values: $(-2, 1)$ and $(-1, -1)$. The other cluster contains $(1, -1)$ and $(2, 1)$. The downward trend on the left-hand side of the V is contained entirely within the first cluster and the upward trend on the right-hand side of the V is entirely within the second cluster. The lack of fit is contained entirely within clusters. For this example, the proposed lack-of-fit test yields a mean square lack of fit which is zero. As indicated, the test picks up lack of fit that exists only within the clusters by having an F statistic of zero.

One characteristic of the exact replicate lack-of-fit test is that it is hard to see its deficiencies. In a simple (x, y) plot it is *impossible* to see lack of fit within clusters. The lack of fit must depend on some variable that distinguishes observations within clusters. With exact replicates this variable must be something other than x. Thus the lack of fit cannot show up in an (x, y) plot.

On the other hand, with near neighbors, the x variable still distinguishes observations within a cluster. Thus the lack of fit can depend on x and may be apparent in an (x, y) plot. The example given above with the V-shaped data illustrates this point.

To investigate consistency properties of the test presented here, we need to specify a true model that contains lack of fit, say

$$Y = W\delta + e, \quad \text{where } C(X) \subset C(W). \tag{7}$$

With the condition that $C(X) \subset C(W)$, model (7) may not be the most succinct version of the true model but any true model can always be enlarged to another true model that has the property that $C(X) \subset C(W)$. Now take a matrix W_2 with the properties that $C(W) = C(X, W_2)$ and $C(X) \perp C(W_2)$. We can rewrite model (7) as

$$Y = X\delta_1 + W_2\delta_2 + e. \tag{8}$$

Thus the lack of fit of model (1) consists of not accounting for the term $W_2\delta_2$.

If model (8) is true, the test statistic proposed in this article typically has a doubly noncentral F distribution. The expected value of the mean square lack of fit is

$$\begin{aligned} E(\text{MSLF}) &= E\left[Y'(M_{XZ} - M)Y/(c + p' - p)\right] \\ &= \sigma^2 + \delta'W'(M_{XZ} - M)W\delta/(c + p' - p) \\ &= \sigma^2 + \delta_2'W_2'(M_{XZ} - M)W_2\delta_2/(c + p' - p) \\ &= \sigma^2 + \delta_2'W_2'M_{XZ}W_2\delta_2/(c + p' - p), \end{aligned}$$

where M and M_{XZ} are the perpendicular projection operators onto $C(X)$ and $C(X, Z)$, respectively, $c + p' = r(X, Z)$ and $p = r(X)$. For tests based on exact replicates, typically $p' < p$. For tests based exclusively on near replicates, typically $p' = p$.

The expected value of the mean square pure error is

$$\begin{aligned} E(\text{MSPE}) &= E\left[Y'(I - M_{XZ})Y/(n - c - p')\right] \\ &= \sigma^2 + \delta'W'(I - M_{XZ})W\delta/(n - c - p') \\ &= \sigma^2 + \delta_2'W_2'(I - M_{XZ})W_2\delta_2/(n - c - p'), \end{aligned}$$

where $n = n_1 + \cdots + n_c$.

By analogy with one-way analysis of variance, effects that are between clusters exist in $C(Z)$ and effects that are within clusters are orthogonal to $C(Z)$. If the lack of fit is between clusters, then $W_2\delta_2 \in C(Z)$. If it is within clusters, then $W_2\delta_2 \perp C(Z)$. If the lack of fit is between clusters, then $(I - M_{XZ})W_2\delta_2 = 0$, $E(\text{MSPE}) = \sigma^2$ and the F statistic has the usual noncentral F distribution (under normality). If the lack of fit is within clusters then $W_2\delta_2 \perp C(Z)$, $M_{XZ}W_2\delta_2 = 0$, $E(\text{MSLF}) = \sigma^2$ and the F statistic has an inverted noncentral F distribution. (An inverted noncentral F is compressed toward zero so the test should be rejected for small F values.) If the lack of fit is some mixture of these archetypes, then the F statistic has a doubly noncentral F distribution. In general, rejection of the F test depends on the relative sizes of the two noncentrality parameters. It is also of interest to note that (a) based on the expectations of the numerator and denominator, the tendencies for the F statistic to become either large or small under the alternatives hold even for nonnormal data and (b) the NJ test and Shillington's test behave similarly.

For asymptotic results we assume that (1) the total number of observations $n \to \infty$, (2) that as n becomes large, the number of clusters is $c_n \to \infty$ and (3) that the relative rate of convergence is $c_n/n \to \eta$, where $0 \le \eta < 1$. This allows the special case where all clusters are of the same size, say $n_i = N > 1$. It also allows some or all of the n_i's to approach infinity. Model (8) can be rewritten as

$$y_{ij} = x'_{ij}\delta_1 + w'_{2ij}\delta z_2 = e_{ij}. \tag{9}$$

While the parameter vectors δ_1 and δ_2 and the ranks of X and (X, W_2) are assumed fixed, the structure of Z and the rows of X and W_2 depend on c_n and n. This dependence will be supressed in the notation but should not be forgotten. Finally, we assume that the e_{ij}'s are independent and that for all i and j $E(e_{ij}) = 0$, $\text{Var}(e_{ij}) = \sigma^2$, $E(e_{ij}^3) = \mu_3$ and $E(e_{ij}^4) = \mu_4$.

3.1. *Between-cluster lack of fit.* First we consider the case in which, asymptotically, the entire lack of fit lies between clusters. For this to happen we need the within-cluster lack of fit to approach zero. The within-cluster lack of fit is $(I - M_Z)W_2\delta_2$. For fixed δ_2, this vector gets progressively larger as the sample size increases because the dimension of the space increases. However, on the average this vector can approach zero. Technically, we take the statement that "asymptotically the lack of fit lies between clusters" to mean that

$$\|(I - M_Z)W_2\delta_2\|^2/n \to 0. \tag{10}$$

PROPOSITION 1. *If condition* (10), *holds, then MSPE* $\to_p \sigma^2$.

PROOF. As seen earlier

$$E(\text{MSPE}) = \sigma^2 + \|(I - M_{XZ})W_2\delta_2\|^2/(n - c_n - p'). \tag{11}$$

Note that, since $C(I - M_{XZ}) \subset C(I - M_Z)$, we have

$$\|(I - M_{XZ})W_2\delta_2\|^2 \le \|(I - M_Z)W_2\delta_2\|^2$$

and

$$\|(I - M_{XZ})W_2\delta_2\|^2/(n - c_n - p') \to 0.$$

Therefore $E(\text{MSPE}) \to \sigma^2$.

We now show that $\text{Var(MSPE)} \to 0$. Let m_0 be a vector consisting of the diagonal elements of $(I - M_{XZ})$. The variance depends on m_0 so we begin by establishing some inequalities relating to m_0. Recall that the diagonal elements of projection operators lie between 0 and 1. If J is a column of ones, we have

$$m_0'm_0 \leq J'm_0 = \operatorname{tr}(I - M_{XZ}) = n - c_n - p'.$$

Also note that by the Cauchy–Schwarz inequality

$$\begin{aligned}|\delta_2'W_2'(I - M_{XZ})m_0| &\leq \|(I - M_{XZ})W_2\delta_2\| \, \|m_0\| \\ &\leq \|(I - M_{XZ})W_2\delta_2\|\sqrt{n - c_n - p'}.\end{aligned}$$

From these results, the idempotence of projection operators and taking absolute values of any negative terms in Theorem 1.8 of Seber (1977) we get

$$\begin{aligned}\operatorname{Var}(\text{MSPE}) \leq (n - c_n - p')^{-2}\Big[&|\mu_4 - 3\sigma^4|(n - c_n - p') \\ &+ 2\sigma^4(n - c_n - p') + 4\sigma^2\|(I - M_{XZ})W_2\delta_2\|^2 \\ &+ 4|\mu_3| \, \|(I - M_{XZ})W_2\delta_2\|\sqrt{n - c_n - p'}\Big].\end{aligned} \tag{12}$$

The upper bound converges to zero so Var(MSPE) does also. By Chebyshev's inequality

$$\text{MSPE} \to_p \sigma^2. \qquad \square$$

Similar arguments lead to a result for MSLF.

PROPOSITION 2. *If* $\|(M_{XZ} - M)W_2\delta_2\|^2/(c_n + p' - p) \to \gamma$ *where* $0 \leq \gamma$, *then* $MSLF \to_p \sigma^2 + \gamma$.

PROOF.

$$\begin{aligned}E(\text{MSLF}) &= E[Y'(M_{XZ} - M)Y/(c_n + p' - p)] \\ &= \sigma^2 + \delta_2'W_2'(M_{XZ} - M)W_2\delta_2/(c_n + p' - p),\end{aligned} \tag{13}$$

$$\begin{aligned}\operatorname{Var}(\text{MSLF}) \leq\; &(c_n + p' - p)^{-2} \\ &\times\Big[|\mu_4 - 3\sigma^4|(c_n + p' - p) + 2\sigma^4(c_n + p' - p) \\ &\quad + 4\sigma^2\|(M_{XZ} - M)W_2\delta_2\|^2 \\ &\quad + 4|\mu_3| \, \|(M_{XZ} - M)W_2\delta_2\|\sqrt{c_n + p' - p}\Big].\end{aligned} \tag{14}$$

Since $E(MSLF) \to \sigma^2 + \gamma$ and $\operatorname{Var}(MSLF) \to 0$ the proposition holds. $\square$

Note that even if $\|(M_{XZ} - M)W_2\delta_2\|^2/(c_n + p' - p) \to \infty$, the rate of convergence of E(MSLF) is greater than that of Var(MSLF) so MSLF will converge in probability to infinity.

The condition $\|(M_{XZ} - M)W_2\delta_2\|^2/(c_n + p' - p) \to \gamma$ is rather artificial. We now relate this to a more intuitive idea of lack of fit. If lack of fit exists, the vector $W_2\delta_2$ is nonzero. For asymptotic results, we need this vector to remain substantial as the models change. A reasonable assumption is that

$$\|W_2\delta_2\|^2/n \to \gamma > 0. \tag{15}$$

A possibly weaker assumption is that

$$\|W_2\delta_2\|^2/c_n \to \gamma > 0. \tag{16}$$

If $c_n/n \to \eta > 0$, these assumptions are equivalent. If $c_n/n \to 0$, then the assumption (16) implies comparatively less lack of fit.

Assumption (16) taken with a (possibly) stronger version of (10), the assumption that the lack of fit exists between clusters, ensures the consistency of the test. The (possibly) stronger version of (10) is that

$$\|(I - M_Z)W_2\delta_2\|^2/c_n \to 0. \tag{17}$$

Again, if $c_n/n \to 0$, this is a stronger assumption about the lack of fit. If $c_n/n \to \eta > 0$, then this condition is equivalent to (10).

THEOREM 1. *If* (16) *and* (17) *hold, then MSLF/MSPE* $\to_p 1 + \gamma/\sigma^2$ *and the lack-of-fit test is consistent.*

PROOF. If (17) holds, then (10) holds and Proposition 1 applies. If Proposition 2 also holds, the result is proven.

To see that Proposition 2 holds, note that $c_n/(c_n + p' - p) \to 1$ so it is enough to show that $\|M_{XZ} - M)W_2\delta_2\|^2/c_n \to \gamma$. Using the multidimensional version of the Pythagorean theorem,

$$\begin{aligned}\|W_2\delta_2\|^2 &= \|MW_2\delta_2\|^2 + \|(M_{XZ} - M)W_2\delta_2\|^2 + \|(I - M_{XZ})W_2\delta_2\|^2 \\ &= \|(M_{XZ} - M)W_2\delta_2\|^2 + \|(I - M_{XZ})W_2\delta_2\|^2\end{aligned}$$

and clearly

$$\begin{aligned}\|(M_{XZ} - M)W_2\delta_2\|^2/c_n &\le \|W_2\delta_2\|^2/c_n \\ &= \|(M_{XZ} - M)W_2\delta_2\|^2/c_n + \|(I - M_{XZ})W_2\delta_2\|^2/c_n.\end{aligned}$$

Note that if $\|(I - M_{XZ})W_2\delta_2\|^2/c_n \to 0$, then (16) implies that

$$\|(M_{XZ} - M)W_2\delta_2\|^2/c_n \to \gamma.$$

By (17), $\|(I - M_{XZ})W_2\delta_2\|^2/c_n \le \|(I - M_Z)W_2\delta_2\|^2/c_n \to 0$, and we are done. □

Allowing for division of positive numbers by zero, we also get

THEOREM 2. *If* (10) *and* (15) *hold, then MSLF/MSPE* $\to_p 1 + \gamma/\sigma^2\eta$ *and the lack-of-fit test is consistent.*

PROOF. The proof is similar to that of Theorem 1. Proposition 1 holds because (10) holds. MSLF $\to_p \sigma^2 + \gamma/\eta$ because (15) implies that $\|W_2\delta_2\|^2/c_n \to \gamma/\eta$. □

Finally, we take a closer look at condition (10). Since M_Z is the projection operator for a one-way ANOVA, we can use model (9) and the Cauchy–Schwarz inequality to get

$$\|(I - M_Z)W_2\delta_2\|^2 = \sum_i \sum_j \left[(w_{2ij} - \bar{w}_{2i.})'\delta_2\right]^2 \le \sum_{ij} \left[\|w_{2ij} - \bar{w}_{2i.}\|^2 \|\delta_2\|^2\right].$$

Since δ_2 is assumed fixed, the condition (10) will be satisfied if

$$\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} \|w_{2ij} - \bar{w}_{2i.}\|^2/n \to 0.$$

In particular, think of each cluster as being contained in a hypersphere centered at $\bar{w}_{2i.}$ with radius r_i. Then

$$\sum_{i=1}^{c_n} \sum_{j=1}^{n_i} \|w_{2ij} - \bar{w}_{2i.}\|^2/n \le \sum_{i=1}^{c_n} n_i r_i^2/n.$$

Clearly, the r_i's depend on n and if $\max_{1 \le i \le c_n}\{r_i\} \to 0$, then condition (10) holds. Moreover, for any finite number of clusters such that $n_i/n \to 0$ we can allow the r_i's to remain bounded away from zero and condition (10) still holds. Of course, the r_i's for these clusters must remain bounded above.

It is important to note that the size of clusters is defined in the space $C(W_2)$. Since W_2 is unknown, we cannot ensure that the cluster sizes are getting small. Typically, we choose clusters of small volume in the space $C(X)$ and hope that they are small in $C(W_2)$. It is possible to choose clusters with zero volume in $C(X)$ (i.e., exact replications) and still have clusters of substantial volume in $C(W_2)$. This is precisely what occurs when the lack of fit exists within clusters.

3.2. *Within-cluster lack of fit.* Restricting the lack of fit to be within clusters, is the restriction that

$$Z'W_2\delta_2 = 0.$$

If this is true, the lack-of-fit test will be consistent. A more appropriate asymptotic result might be based on assuming that $Z'W_2\delta_2$ is converging to zero, for example, $\|Z'W_2\delta_2\|^2/c_n \to 0$. However, because of possible collinearity in $C(X, Z)$, this condition is not enough to ensure consistency. If $W_2\delta_2 \perp C(X)$ and $W_2\delta_2 \perp C(Z)$, then $W_2\delta_2 \perp C(X, Z)$. However, if $W_2\delta_2 \perp C(X)$ and $W_2\delta_2$ is almost orthogonal to $C(Z)$ we have no assurance that $W_2\delta_2$ is nearly orthogonal to $C(X, Z)$. The condition that $W_2\delta_2$ is nearly orthogonal to $C(X, Z)$ is the condition required for consistency.

THEOREM 3. *If* $\|W_2\delta_2\|^2/n \to \gamma$ *and* $\|M_{XZ}W_2\delta_2\|^2/(c_n + p') \to 0$, *then* $MSLF \to_p \sigma^2$, $MSPE \to_p \sigma^2 + \gamma/(1 - \eta)$ *and the test is consistent.*

PROOF.

$$\|M_{XZ}W_2\delta_2\| = \|(M_{XZ} - M)W_2\delta_2\|$$

so

$$\|(M_{XZ} - M)W_2\delta_2\|^2/(c_n + p' - p) \to 0.$$

From (13) and (14) $E(\text{MSLF}) \to \sigma^2$ and $\text{Var(MSLF)} \to 0$. The Pythagorean theorem and arguments in the previous subsection lead to $\|(I - M_{XZ})W_2\delta_2\|^2/(n - c_n - p') \to \gamma/(1 - \eta)$. Thus from (11) and (12) $E(\text{MSPE}) \to \sigma^2 + \gamma/(1 - \eta)$ and $\text{Var(MSPE)} \to 0$. □

Although mathematically the results on within-cluster lack of fit are as general as the results on between-cluster lack of fit, they do not apply in the usual paradigm of taking more and more clusters in which the volume of each cluster is getting progressively smaller. This is an appealing paradigm but it is not the only one of interest. Another possibility is to consider a sequence of clusters with each cluster retaining some kind of intrinisic structure. The lack-of-fit test can be viewed as testing for treatment effects in a one-way analysis of covariance. To do this, one thinks of the clusters as being treatments. If the "treatments" seem very different, there is a between-cluster lack of fit. If the "treatments" seem too similar, it may be because there is structure within the clusters causing the "pure error" to be overestimated. Suppose that the clusters have some sort of intrinsic structure, for example, they are observations taken on the same day. Then it makes sense to think of a sequence of "treatments" (clusters) in which the volume of the clusters is not getting arbitrarily small. The concept of an asymptotic within-cluster lack of fit only makes sense when the sequence of clusters retains some sort of physical meaning. If the clusters become arbitrarily small, it is unreasonable to imagine that they will retain the lack of fit within them.

3.3. *Power*. Assuming normal errors, the power of tests based on comparing models (1) and (6) can be computed exactly because the distribution under the alternative is a doubly noncentral F. While programs for computing the necessary probabilities are not readily available, the formulas in Bulgren (1971) are easily programmed.

From standard linear model theory, it is known that no uniformly most powerful or uniformly most powerful unbiased test exists for testing lack of fit. However, the test proposed here (if rejected only for large F values) gives a uniformly most powerful invariant (UMPI) test against the alternative that the lack of fit vector can be written as the sum of a vector in the between clusters space $C(Z)$ and a vector in the space determined by the projection of X onto the within clusters space. Moreover, the test (if rejected only for small F values) gives a UMPI test against the alternative that the orthogonal lack of fit $W_2\delta_2$ lies within clusters. To see this note that the alternative model for lack of fit within clusters is

$$Y = X\beta + (I - M_{XZ})\gamma + e. \tag{18}$$

It is easily seen that rejecting for small values of F as described above gives the likelihood ratio test for testing model (1) against model (18) and the likelihood ratio test is known to be UMPI.

Acknowledgments. The author would like to thank the Associate Editor for making a number of very valuable suggestions and a referee for his careful reading of the first revision.

REFERENCES

BULGREN, W. G. (1971). On representations of the doubly noncentral F distribution. *J. Amer. Statist. Assoc.* **66** 184–186.

CHRISTENSEN, R. (1987). *Plane Answers to Complex Questions: The Theory of Linear Models.* Springer, New York.

DRAPER, N. and SMITH, H. (1981). *Applied Regression Analysis*, 2nd ed. Wiley, New York.

FIENBERG, S. E. and GONG, G. D. (1984). Comment on "Graphical methods for assessing logistic regression models" by J. M. Landwehr, D. Pregibon and A. C. Shoemaker. *J. Amer. Statist. Assoc.* **79** 72–77.

LANDWEHR, J. M., PREGIBON, D. and SHOEMAKER, A. C. (1984). Graphical methods for assessing logistic regression models. *J. Amer. Statist. Assoc.* **79** 61–71.

NEILL, J. W. (1982). Testing regression function adequacy without replication. Ph.D dissertation, Dept. Statistics, Kansas State Univ.

NEILL, J. W. and JOHNSON, D. E. (1984a). Testing for lack of fit in regression—A review. *Comm. Statist. A—Theory Methods* **13** 485–511.

NEILL, J. W. and JOHNSON, D. E. (1984b). Small sample behavior of a test for regression function adequacy without replication. Unpublished manuscript.

NEILL, J. W and JOHNSON, D. E. (1985). Testing linear regression function adequacy without replication. *Ann. Statist.* **13** 1482–1489.

SEBER, G. A. F. (1977). *Linear Regression Analysis.* Wiley, New York.

SHILLINGTON, E. R. (1979). Testing lack of fit in regression without replication. *Canad. J. Statist.* **7** 137–146.

TSIATIS, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model. *Biometrika* **67** 250–251.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
HUMANITIES BUILDING 419
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 87131