TECHNOMETRICS, NOVEMBER 1987, VOL. 29, NO. 4

Diagnostics for Mixed-Model Analysis of Variance

Richard J. Beckman
Statistics Group
Los Alamos National Laboratory
Los Alamos, NM 87545

Christopher J. Nachtsheim
Department of Management Sciences
University of Minnesota
Minneapolis, MN 55455

R. Dennis Cook
Department of Applied Statistics
University of Minnesota
St. Paul, MN 55108

We describe a new method for assessment of model inadequacy in maximum-likelihood mixed-model analysis of variance. In particular, we discuss its use in diagnosing perturbations from the usual assumption of constant error variance and from the assumption that each realization of a given random factor has been drawn from the same normal population. Computer implementation of the procedure is described, and an example is presented, involving the analysis of filter cartridges used with commercial respirators.

KEY WORDS: Curvature; Likelihood displacement; Local influence; Random effects.

1. INTRODUCTION

Interest in regression diagnostics has grown steadily in recent years. The importance of diagnostics as an area of active research can be appreciated by a scanning of the extensive bibliographies in books by Cook and Weisberg (1982) and Belsley, Kuh, and Welsch (1980) and in review papers such as that by Beckman and Cook (1983). The speed with which many of the new methods have been implemented in such packages as SAS, BMDP, SPSS, and MINITAB underscores the statistical community's view as to their worth in practice.

Yet despite this rapid expansion in techniques for diagnosing the adequacy of regression and fixed-effect analysis of variance (ANOVA) models, there has been almost no parallel development in the area of random-effects or mixed-model ANOVA. In fact, it is only very recently that papers dealing with diagnostics for mixed-model ANOVA have begun to appear. For example, Hocking, Green, and Bremer (1986) discussed a new class of unbiased estimators and demonstrated its use in diagnosing the causes of negative variance-component estimates. Fellner (1986) developed outlier-resistant estimates of parameters in mixed models through the use of an influence function and showed how the method could be used to construct diagnostic displays for identification of outliers.

Recently, Cook (1986) gave a completely general method for assessing the influence of local departures from assumptions in statistical models. The method assumes only a well-behaved likelihood and is thus applicable when unrestricted maximum likelihood estimates (MLE's) are of interest. Extensive analyses of data sets based on Jennrich and Sampson's (1976) algorithm for computing both restricted and unrestricted MLE's motivated a need for regression-type diagnostics. This led to an eventual adaptation of Cook's likelihood-based methods.

The purpose of this article is to document methods developed for assessing the effects of perturbations from the usual assumptions in mixed-model ANOVA. Statistical models are generally approximate descriptions of more complicated processes and, because of this inexactness, an assessment of the influence of model perturbations is important. Statistical conclusions that are more or less insensitive to extraneous perturbations are robust and on that account useful, but conclusions that are sensitive to such perturbations must be treated with caution.

We start by giving a brief sketch of the local-influence approach in Section 2, and in Section 3 we develop methodology required for application to mixed-model ANOVA. Fine points of implementation are discussed in Section 4, and an example is considered in Section 5 that involves the analysis of filter cartridges used with commercial respirators.

Concluding remarks are presented in the final section.

2. THE LOCAL-INFLUENCE APPROACH

As indicated previously, a description of methods for assessing local influence was given by Cook (1986). For our purposes, a brief review of the essential aspects of this work is necessary.

Let $L(\boldsymbol{\theta})$ represent the log-likelihood for a postulated model and observed data, where $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters with maximum MLE $\hat{\boldsymbol{\theta}}$. Introducing perturbations into the model via the $m \times 1$ vector $\boldsymbol{\omega} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ represents the set of relevant perturbations, let $L(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ denote the log-likelihood corresponding to the perturbed model, and let $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denote the corresponding MLE. Assume also the existence of a "null" perturbation $\boldsymbol{\omega}_0$ such that $L(\boldsymbol{\theta} \mid \boldsymbol{\omega}_0) = L(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$ and that $\boldsymbol{\theta}$ has the same meaning in both $L(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ and $L(\boldsymbol{\theta})$.

When interest centers on comparing $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$, the influence of the perturbation $\boldsymbol{\omega}$ can be assessed by using the *likelihood displacement* (Cook and Weisberg 1982)

$$LD(\boldsymbol{\omega}) = 2[L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})].$$

Large values of $LD(\boldsymbol{\omega})$ indicate that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ differ considerably relative to the contours of the unperturbed log-likelihood $L(\boldsymbol{\theta})$. Moreover, since $\hat{\boldsymbol{\theta}}$ maximizes $L(\boldsymbol{\theta})$, we must have $LD(\boldsymbol{\omega}) \geq 0$.

To fix ideas, we use the results of a recent Los Alamos study of high efficiency particulate air (HEPA) cartridges (Kerschner, Ettinger, DeField, and Beckman 1984). Such cartridges are used with commercial respirators to provide respiratory protection against dusts, toxic fumes, mists, radionuclides, and other particulate matter. The primary objective of the study was to determine whether the current standard aerosol used to test these filters could be replaced by any of three alternate aerosols for quality-assurance testing of HEPA respirator cartridges. HEPA respirator filters fail a quality-assurance test if the challenge penetration is greater than some percent penetration breakpoint; otherwise, it is considered to have passed.

A secondary objective was to identify those factors that contribute most to the variability in the penetrations of the filters. This led to the model

$$y_{ijkl} = \mu + A_i + M_j + F_{jk} + \varepsilon_{ijkl}, \tag{1}$$

where y_{ijkl} is the percent penetration, A_i is a fixed effect for the ith aerosol type, M_j is a fixed effect for the jth filter manufacturer, F_{jk} is a random effect for the kth filter nested within the jth manufacturer, and ε_{ijkl} is the error associated with the lth replication in the ijkth cell, subject to the usual restrictions: $\sum_i A_i = \sum_j M_j = 0$. A subset of the data that are used here for illustration is listed in Table 1. In this case $i = 1, 2; j = 1, 2; k = 1, 2, 3$; and $l = 1, 2, 3$.

We used the random-effects model (1) to construct the unperturbed log-likelihood $L(\boldsymbol{\theta})$ with $\boldsymbol{\theta}^T = (\mu, A_i, M_j, \sigma_F^2, \sigma^2)$. The log-likelihood for the perturbed model is constructed in the same way except that the variances of ε_{1221} and ε_{1222} are replaced by $\omega_1\sigma^2$ and $\omega_2\sigma^2$, respectively. These two filter-5 cases were chosen for use in this illustration because the corresponding responses listed in Table 1 seem out of line with the rest of the data. A more complete analysis to be described later confirms that these cases are anomalous. Both ω_1 and ω_2 are constrained to lie in the interval (.5, 2.0), which represents a twofold increase or decrease in the assumed variances. In this example $\boldsymbol{\Omega} = (.5 \times 2.0)^2$ and $\boldsymbol{\omega}_0 = \mathbf{1}$. The surface in Figure 1 is an *influence graph* of $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$, as $\boldsymbol{\omega}$ varies in $\boldsymbol{\Omega}$. The likelihood displacement is 0 at

Table 1. Percent Penetration of Two Aerosols Using Three Filters From Two Manufacturers

Aerosol 1 ($i = 1$)		*Aerosol 2 ($i = 2$)*	
Manuf. 1 ($j = 1$)	*Manuf. 2 ($j = 2$)*	*Manuf. 2 ($j = 1$)*	*Manuf. 1 ($j = 2$)*
Filter 1 ($k = 1$)	Filter 4 ($k = 1$)	Filter 1 ($k = 1$)	Filter 4 ($k = 1$)
.750	.600	1.120	.910
.770	.680	1.100	.830
.840	.870	1.120	.950
Filter 2 ($k = 2$)	Filter 5 ($k = 2$)	Filter 2 ($k = 2$)	Filter 5 ($k = 2$)
.082	4.100	.160	.660
.085	5.000	.110	.830
.096	1.800	.260	.610
Filter 3 ($k = 3$)	Filter 6 ($k = 3$)	Filter 3 ($k = 3$)	Filter 6 ($k = 3$)
.082	1.000	.150	2.170
.076	1.800	.120	1.520
.077	2.700	.120	1.580

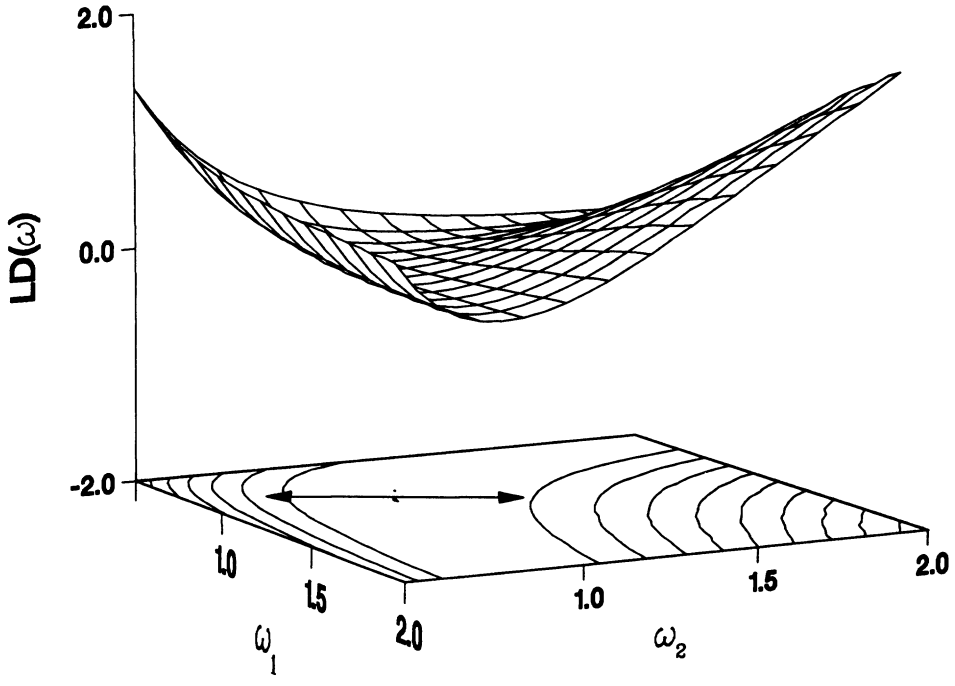

Figure 1. Likelihood Displacement Surface as a Function of Perturbed Error Variances, $\omega_1\sigma^2$ and $\omega_2\sigma^2$, Corresponding to the Two Fillter-5 Aerosol Data Points, y_{1221} and y_{1222}.

$\boldsymbol{\omega}^T = (1, 1)$, and increases as we move away from this point in any direction. For example, at the point $\boldsymbol{\omega}^T = (.5, .5)$, where the variances of the cases in question are set to a half of the common variance of the remaining cases, the likelihood displacement is about 1.7. Clearly, the analysis is very sensitive to the assumption of constant error variances for the two cases in question. If we are convinced that the assumption is accurate, this may not be a problem. Otherwise, this apparent sensitivity and the validity of the cases in question should be cause for concern.

Influence graphs of this form are possible in only the simplest situations ($m \leq 2$). More generally, it will be necessary to investigate other methods for extracting relevant information from an influence graph. One method is based on studying the *local* behavior of an influence graph around $\boldsymbol{\omega}_0$, the perturbation giving the target model. We begin by selecting a direction $\mathbf{h}$ in $\boldsymbol{\Omega}$, $\|\mathbf{h}\| = 1$; one such direction is shown in Figure 1. This direction determines a plane that is normal to $\mathbf{Q}$. The intersection of this plane with the surface is called a *lifted line*. Each direction $\mathbf{h}$ determines a lifted line that may be graphed by plotting $LD(\boldsymbol{\omega}_0 - a\mathbf{h})$ versus a $(a \in R^1)$. The family of lifted lines formed by letting $\mathbf{h}$ range over all unit vectors in $\boldsymbol{\Omega}$ forms the basis for the local characterization of an influence graph. As in Cook (1986), it will be useful to view an influence graph as the surface formed by the values of the $(m + 1) \times 1$ vector

$$\boldsymbol{\alpha}^T(\boldsymbol{\omega}) = (\boldsymbol{\omega}^T, LD(\boldsymbol{\omega}))$$

as $\boldsymbol{\omega}$ varies in $\mathbf{Q}$.

Each lifted line can be characterized by considering its *geometric normal curvature* C_h at $\boldsymbol{\omega}_0$. The curvature C_h is also called the *normal curvature of the surface* $\boldsymbol{\alpha}(\boldsymbol{\omega})$ *in the direction* h. It can be visualized as the inverse radius of the best-fitting circle at $\boldsymbol{\omega}_0$. Large values of C_h indicate sensitivity to the induced perturbations in the direction $\mathbf{h}$. The maximum curvature $C_{\max}$, the corresponding direction $\mathbf{h}_{\max}$, and a plot of the lifted line in the direction $\mathbf{h}_{\max}$ $[LD(\boldsymbol{\omega}_0 + a\mathbf{h}_{\max})$ vs. $a]$ are the main diagnostic quantities in this approach. Larger values of $C_{\max}$ indicate stronger local influence, and $\mathbf{h}_{\max}$ shows how to perturb the postulated model $(\boldsymbol{\omega} = \boldsymbol{\omega}_0)$ to obtain the greatest local changes in the likelihood displacement. A plot of the lifted line can be used to confirm the indications of $C_{\max}$.

In Figure 1, the direction of maximum curvature $\mathbf{h}_{\max} = (-.467, -.884)$ or $\mathbf{h}_{\max} = (.467, .884)$ is shown

by the arrows in the plane. The double arrows indicate that the direction eigenvector $\mathbf{h}_{max}$ may take on either sign. In this case it is obvious from the contours in Figure 1 that the largest changes will occur by moving in the negative direction. Modifications of $(\text{var}(\varepsilon_{1221}), \text{var}(\varepsilon_{1222}))$ in this direction will result in the greatest local changes in the estimates as measured by the likelihood displacement. Further, the analysis will be more sensitive to changes in $\text{var}(\varepsilon_{1222})$ $(y = 5.0)$ than to corresponding changes in $\text{var}(\varepsilon_{1221})$ $(y = 4.1)$, since $|h_{\max,2}| > |h_{\max,1}|$, where $h_{\max,i}$ is the ith component of $\mathbf{h}_{max}$. The vector $\mathbf{h}_{max}$ is invariant under identical one-to-one coordinate transformations of $\boldsymbol{\omega}$. In other words, reformulating the perturbation scheme and Figure 1 in terms of $\log(\omega_i)$ or $1/\omega_i$ would not change $\mathbf{h}_{max}$.

The maximum curvature of Figure 1 is $C_{max} = 7.91$. The ability to assess the importance of curvature values requires some experience because, unlike $\mathbf{h}_{max}$, C_{max} is not invariant under identical coordinate transformations of $\boldsymbol{\omega}$. For a particular perturbation scheme, a few plots of the lifted line in the direction of $\mathbf{h}_{max}$ for various data sets will provide the necessary experience.

In general, if C_{max} is the maximum curvature for perturbations ω_i, then the corresponding curvature C^*_{max} for perturbations ω_i^*, $\omega_i = k(\omega_i^*)$, is $[\partial k/\partial\omega^*]^2 C_{max}$, where the derivative is evaluated at the value of the ω^* that yields the postulated model. In the aerosol data, for example, suppose that we had chosen to describe the variances of ε_{1221} and ε_{1222} as $(\omega_1^*\sigma)^2$ and $(\omega_2^*\sigma)^2$, respectively. The relationship between the original scheme and the new scheme is $\omega_i = (\omega_i^*)^2$, $[\partial k/\partial\omega^*]^2 = 2$, and $C^*_{max} = 2(7.91) = 15.82$. For the original aerosol data setup, our experience suggests that values of C_{max} larger than about 4 indicate notable local sensitivity. Regardless of the value of C_{max}, however, $\mathbf{h}_{max}$ should always be inspected, since it can direct attention to influential cases and other global problems that are not manifest locally.

To calculate C_{max} and $\mathbf{h}_{max}$, define the $m \times m$ matrix

$$\mathbf{F} = \boldsymbol{\Delta}^T\mathbf{Q}^{-1}\boldsymbol{\Delta}, \tag{2}$$

where $\boldsymbol{\Delta}$ is the $p \times m$ $[p = \dim(\boldsymbol{\theta}),\ m = \dim(\boldsymbol{\omega})]$ matrix with elements

$$\Delta_{ij} = \partial^2 L(\boldsymbol{\theta} \mid \boldsymbol{\omega})/\partial\theta_i\,\partial\omega_j\Big|_{\hat{\boldsymbol{\theta}},\,\boldsymbol{\omega}_0}, \tag{3}$$

and $-\mathbf{Q}$ is the observed information for the postulated model

$$\mathbf{Q} = \partial^2 L(\boldsymbol{\theta})/\partial\theta_i\,\partial\theta_j\Big|_{\hat{\boldsymbol{\theta}}}. \tag{4}$$

Then (Cook 1986)

$$C_{max} = \max_{\|\mathbf{h}\| = 1} 2|\mathbf{h}^T\mathbf{F}\mathbf{h}|, \tag{5}$$

and $\mathbf{h}_{max}$ is the eigenvector corresponding to the largest absolute eigenvalue C_{max} of $\mathbf{F}$.

When a subset $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$ is of special interest, the corresponding likelihood displacement is defined as

$$LD(\boldsymbol{\omega}) = 2[L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}), g(\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}})],$$

where $g(\boldsymbol{\theta}_1)$ is the vector that maximizes $L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for each fixed $\boldsymbol{\theta}_1$ and $\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}$ is determined from the partition $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}^T = (\hat{\boldsymbol{\theta}}_{1\boldsymbol{\omega}}^T, \hat{\boldsymbol{\theta}}_{2\boldsymbol{\omega}}^T)$. The methods discussed previously for the full parameter vector carry over to this situation in a straightforward manner. In particular, Cook (1986) showed that with perturbations of only one parameter, say the first, the matrix $\mathbf{F}$ reduces to

$$\mathbf{F} = \boldsymbol{\Delta}^T[\mathbf{Q}^{-1} - \mathbf{B}_{22}]\boldsymbol{\Delta},$$

where

$$\mathbf{B}_{22} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{22}^{-1} \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} q_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix},$$

each matrix being partitioned according to the partition of $\boldsymbol{\theta}$. The matrix $\mathbf{Q}^{-1} - \mathbf{B}_{22}$ can be reduced to a product of a vector $\mathbf{H}$ and its transpose (see App. A), and hence $\mathbf{F} = \boldsymbol{\Delta}^T\mathbf{H}\mathbf{H}^T\boldsymbol{\Delta}$, where $\mathbf{H}^T = (1, -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})/\sqrt{c}$ and $c = q_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}$. For a single parameter then, the direction of maximum curvature is $\mathbf{h}_{max} = \boldsymbol{\Delta}^T\mathbf{H}$ and the corresponding curvature is $C_{max} = \|\boldsymbol{\Delta}^T\mathbf{H}\|^2$.

3. APPLICATION TO MIXED-MODEL ANOVA

3.1 The Model

In this section we consider the construction of diagnostics for unrestricted maximum likelihood estimation in mixed-model ANOVA. Suppose the model is of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}_1\mathbf{b}_1 + \mathbf{U}_2\mathbf{b}_2 + \cdots + \mathbf{U}_c\mathbf{b}_c + \boldsymbol{\varepsilon}, \tag{6}$$

where $\mathbf{X}$ and $\mathbf{U}_i$ are, respectively, $n \times q$ and $n \times q_i$ matrices of known values, $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown constants, $\mathbf{b}_i$ is a $q_i \times 1$ vector of random values with independent component $N(0, \sigma_i^2)$ distributions, and $\boldsymbol{\varepsilon}$ is $n \times 1$ with the ε_i independently and identically distributed. The full parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \sigma_1^2, \ldots, \sigma_c^2)^T$ is $p \times 1$. The random vectors $\{\mathbf{b}_i\}$ and $\boldsymbol{\varepsilon}$ are assumed independent. The response $\mathbf{y}$, then, has a multivariate normal distribution with a mean vector $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix

$$\mathbf{V} = \sum_{i=1}^{c} \mathbf{U}_i\mathbf{U}_i^T\sigma_i^2 + \sigma^2\mathbf{I}.$$

The estimated fixed effects $\hat{\beta}_1, \ldots, \hat{\beta}_q$ and variance components $\hat{\sigma}^2, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_c^2$ are those values that maximize the log-likelihood:

$$\lambda = -(n/2)\log(2\pi) - (1/2)\log|\mathbf{V}| - (1/2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

In each of the subsections that follow, we introduce specific perturbations to the baseline model (6). One or more of these are often of particular concern to the analyst. As indicated in Section 2, to assess the local influence of a particular perturbation scheme, it is necessary to compute the Hessian matrix $\mathbf{F} = \boldsymbol{\Delta}^T\mathbf{Q}^{-1}\boldsymbol{\Delta}$. An important advantage provided by the Jennrich and Sampson (1976) implementation of the Newton–Raphson algorithm is the immediate availability of $\mathbf{Q}$ upon convergence. Computationally, each perturbation scheme requires specification of the $\boldsymbol{\Delta}$ matrix. Since this can involve tedious algebra, details are left to the appendixes.

3.2 An Idealized Example

In the development of any new methodology it is always helpful to consider situations that allow for a relatively straightforward investigation of its operating characteristics. If the answers behave reasonably and can be justified based on past experiences and general knowledge of the situation, then we have gained valuable insight and the reassurance that the methodology will produce sensible results in more complicated settings. Analytic characterizations of the maximum curvature and its direction for simple models would certainly help, but this type of approach has proven difficult.

Instead, we chose to investigate each diagnostic procedure graphically by using a simple example involving a one-way ANOVA. Specifically, for each of the perturbation schemes described in the following sections we start with an idealized data set and then add two standard deviations first to a single data point and later to a single random effect. The associated changes are displayed by index plots of the direction of maximum curvature. The null or idealized data set was constructed from the model

$$y_{ij} = \mu + A_i + e_{ij},$$

where $i = 1, \ldots, 10$, $j = 1, 2, 3$, and $\mu = 0$. Here the A_i's were taken to be the expected values of the 10 order statistics for a sample of size 10 from an $N(0, \sigma_A^2)$, $\sigma_A = 3$, distribution and for each i, e_{ij} was the expected value of the jth order statistic from a sample of size 3 of an $N(0, \sigma^2)$, $\sigma = 1$, distribution. These data are given in Table 2.

Since the model is simple and the data conform exactly to expectation, we should be able to anticipate somewhat the consequences of adding two standard deviations to a single data point or to a single random effect. For example, adding something to A_i should influence $\hat{\sigma}_A^2$ but not $\hat{\sigma}^2$, and the appropriate directions of maximum curvature in fact show this. In general, this simple example will serve to illustrate that relatively small or large components of the direction on maximum curvature correspond to departures from the hypothesized model.

Table 2. Artificial Data Set Constructed From Order Statistics

Case	i	A_i	j	e_{ij}	$y_{ij} = A_i + e_{ij}$
1			1	−.84	−5.46
2	1	−4.62	2	0	−4.62
3			3	.84	−3.78
4			1	−.84	−3.84
5	2	−3.00	2	0	−3.00
6			3	.84	−2.16
7			1	−.84	−2.82
8	3	−1.98	2	0	−1.98
9			3	.84	−1.14
10			1	−.84	−1.97
11	4	−1.13	2	0	−1.13
12			3	.84	−.29
13			1	−.84	−1.20
14	5	−.36	2	0	−.36
15			3	.84	.48
16			1	−.84	−.48
17	6	.36	2	0	.36
18			3	.84	1.20
19			1	−.84	.29
20	7	1.13	2	0	1.13
21			0	.84	1.97
22			1	−.84	1.14
23	8	1.98	2	0	1.98
24			3	.84	2.82
25			1	−.84	2.16
26	9	3.00	2	0	3.00
27			3	.84	3.84
28			1	−.84	3.78
29	10	4.62	2	0	4.62
20			3	.84	5.46

The direction vector of maximum curvature for each of the perturbation schemes to be discussed is illustrated by one of the 24 panels of Figure 2. Rows of this figure correspond to different perturbation schemes. Columns of the figure correspond to likelihood displacement considering the full parameter vector (col. a) or subsets of this vector (cols. b, c, and d).

3.3 Perturbation of Error Variances

As implied by (6), our model assumes homogeneity of error variances: $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. We relax this assumption via the $n \times 1$ perturbation vector $\boldsymbol{\omega}$ such that $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{D}(\boldsymbol{\omega})$, where $\mathbf{D}(\boldsymbol{\omega})$ denotes the $n \times n$

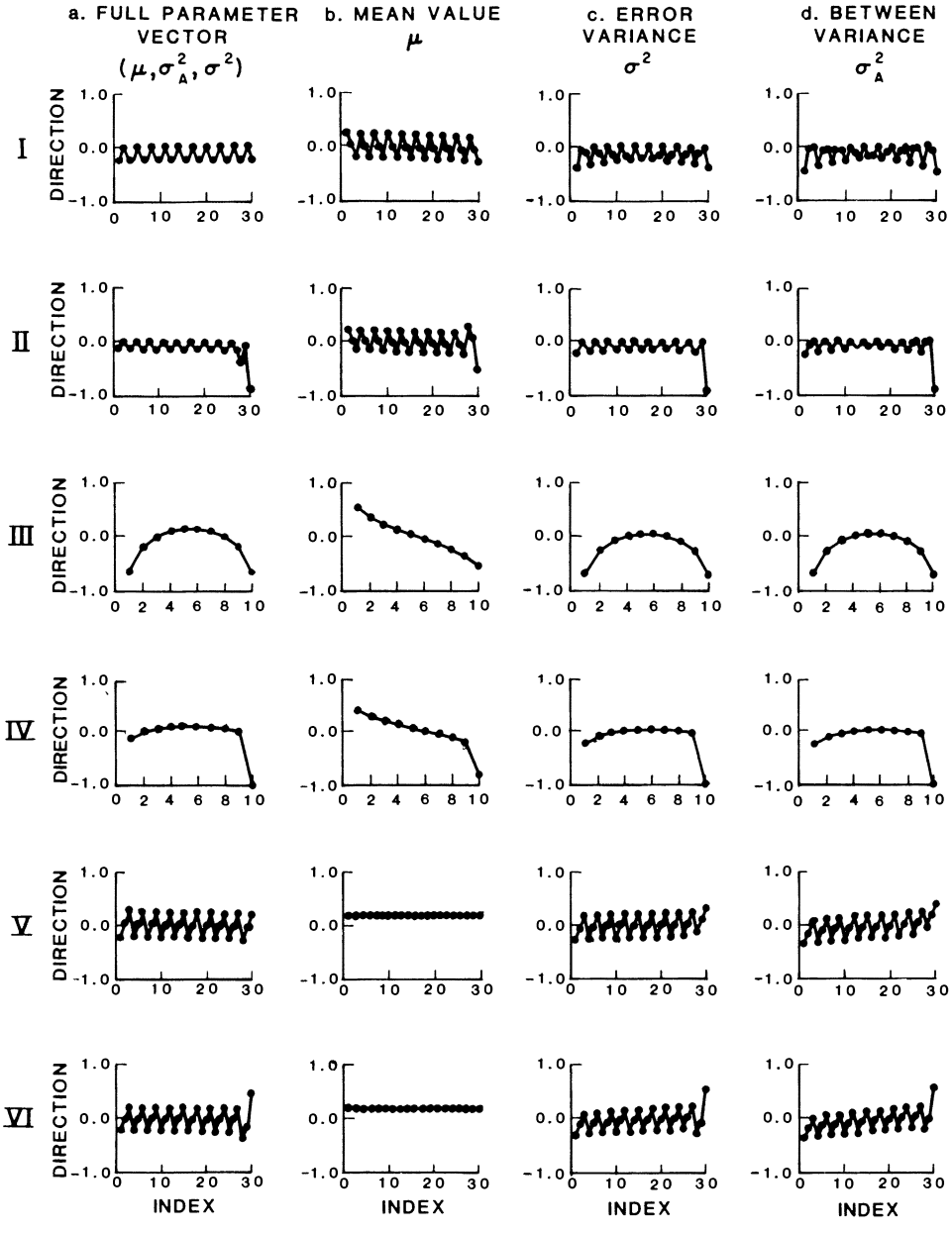

Figure 2. Various Panels Depicting the Direction of Maximum Curvature Vectors for Different Perturbation Schemes and Data Sets for the Idealized Example. Panels in column a represent the direction vector for the maximum curvature of the likelihood displacement considering the entire parameter vector. The direction vector for the maximum curvature considering the mean only, the error variance only, and the between variance only are in the panels of columns b, c, and d, respectively. *Rows I, III, and V, respectively, correspond to perturbation of the error variance, the between variance, and the y values of the artificial data set of Table 2. Row II corresponds to perturbation of the error variance when $2\sigma = 2$ has been added to y_{30}. Row IV represents perturbation of the between variance when $2\sigma_A = 6$ has been added to the last group of artificial data, y_{28}, y_{29}, y_{30}. Row VI corresponds to perturbation of the y values when $2\sigma = 2$ has been added to y_{30}.*

diagonal matrix in which the kth diagonal element is ω_k.

In what follows we shall denote the ith columns of the respective matrices $\mathbf{X}$, $\mathbf{U}_j$, $\mathbf{V}^{-1}$ and $\mathbf{V}^{-1}(\boldsymbol{\omega})$ by $\mathbf{X}_i$, $\mathbf{U}_j^i$, $\mathbf{V}_i^{-1}$, and $\mathbf{V}_i^{-1}(\boldsymbol{\omega})$. Similarly, the (k, h)th elements of the matrices $\mathbf{U}_i$ and $\mathbf{V}^{-1}$ will be written as U_{kh}^i and V_{kh}^{-1}. The kth row of $\boldsymbol{\Delta}$ is

$$\boldsymbol{\Delta}_k^T = \left\{\frac{\partial^2\lambda(\boldsymbol{\omega})}{\partial\omega_k\partial\beta_1}, \ldots, \frac{\partial^2\lambda(\boldsymbol{\omega})}{\partial\omega_k\partial\beta_p}, \frac{\partial^2\lambda(\boldsymbol{\omega})}{\partial\omega_k\partial\sigma^2}, \frac{\partial^2\lambda(\boldsymbol{\omega})}{\partial\omega_k\partial\sigma_1^2}, \ldots, \frac{\partial^2\lambda(\boldsymbol{\omega})}{\partial\omega_k\partial\sigma_c^2}\right\},$$

where each derivative is evaluated at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. We have by differentiation of the explicit expression for $\lambda(\boldsymbol{\omega})$, for $k = 1, \ldots, m$ and with $\mathbf{r} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$,

$$\left.\frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial \omega_k \, \partial \beta_j}\right|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \mathbf{X}_j^T \mathbf{D}_k \mathbf{r}, \qquad j = 1, \ldots, p$$

$$\left.\frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial \omega_k \, \partial \sigma^2}\right|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \frac{1}{2} \hat{\sigma}^{-2} \mathbf{r}^T \mathbf{D}_k \mathbf{r}$$

$$\left.\frac{\partial^2 \lambda(\boldsymbol{\omega})}{\partial \omega_k \, \partial \sigma_i^2}\right|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \mathbf{r}^T \mathbf{D}_k \mathbf{U}_i \mathbf{U}_i^T \hat{\mathbf{V}}^{-1} \mathbf{r} - \frac{1}{2} TR\{\mathbf{U}_i^T \mathbf{D}_k \mathbf{U}_i\}, \qquad i = 1, \ldots, c, \quad (7)$$

where $\mathbf{D}_k = \partial V^{-1}(\boldsymbol{\omega}) / \partial \omega_k \, |_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$. Equations (7) are quite general in that they apply whenever the perturbation scheme involves only elements of the variance matrix. For perturbation of the error variances, we show in Appendix B that

$$\mathbf{D}_k = -\hat{\sigma}^2 \mathbf{V}_k^{-1} (\hat{\mathbf{V}}_k^{-1})^T, \qquad k = 1, 2, \ldots, n,$$

where $\hat{\mathbf{V}} = \sum_{i=1}^{m} \mathbf{U}_i \mathbf{U}_i^T \hat{\sigma}_i^2 + \hat{\sigma}^2 \mathbf{I}$. Since $\mathbf{Q}^{-1}$ is available upon convergence of the estimation procedure, construction of $\mathbf{F}$ via (2) is immediate.

Plots of the direction of maximum curvature vectors for the one-way ANOVA example are presented in Figure 2. In this figure indexes for panels in rows III and IV refer to group $i = 1, 10$ as shown in Table 2. Indexes for the panels in rows I, II, V, and VI refer to the corresponding data-point numbers given in Table 2.

In Figure 2, panel I(a), since the data points in Table 2 are ordered from smallest to largest within each group of three, maximal changes in the estimate of the full parameter vector $\hat{\boldsymbol{\theta}}$ are obtained by either decreasing or increasing (remembering that the sign of the direction vector is not determined) the variances associated with the largest and smallest values in each group. In contrast, rapid changes in the estimate of the mean are obtained by changing the assumed variances of the largest and smallest observations in *opposite* directions. This will tend to move the estimate toward the data point with the smallest variance. The similarity of Figure 2 panels I(a), I(c), and I(d) leads to the conclusion that, in this example, the direction associated with changes to the full parameter vector is driven more by sensitivity to the estimates of the variance and variance component than to the estimate of the mean.

Row II of Figure 2 is simply a reconstruction of row I, after the largest data point ($y_{10,3}$, case 30) had been increased by $2\sigma = 2$. Panels II(a), II(c), and II(d) all present a similar message: Small changes in the assumed variance of the 30th data point will have the most influence on the estimates of σ^2 and σ_A^2. A similar but less dramatic change in $\hat{\mu}$ is indicated by panel II(b). We learn from this row that if Case 30 is an outlier and if we make a minor modification in the assumed error variance for that case, it will strongly influence the outcome of the analysis.

To illustrate the effect of an outlying random effect, $2\sigma_A = 6$ was added to each element of the 10th group (Cases 28, 29, 30). The resulting plots were nearly indistinguishable from those of Figure 2, row I, and are not shown. In retrospect, of course, this makes sense: the e_{ij}'s are still as originally generated in the idealized set. The basic assumptions with regard to the error variance are still valid.

3.4 Perturbation of Random-Effects Variances

An important assumption in mixed-model ANOVA is that the q_i realizations of the ith random factor are drawn from the same normal population. In this section we suggest a perturbation scheme for studying the effects of departures from such assumed conditions and give equations necessary for construction for $\boldsymbol{\Delta}$.

We assume that the hth element of the random vector $\mathbf{b}_k$ was drawn from an $N(0, \omega_{kh} \sigma_k^2)$ population. Under the usual assumptions, $\omega_{kh} = 1$ implies that $\boldsymbol{\omega}_0 = \mathbf{1}$. The $\boldsymbol{\Delta}$ matrix is given by (7) with $\boldsymbol{\omega}$ and $\mathbf{D}$ doubly subscripted:

$$\{D_{kh}\}_{ij} = -\hat{\sigma}_k^2 \sum_{\substack{u=1 \\ u \le v}}^{n} \sum_{v=1}^{n} U_{uh}^k U_{vh}^k \times [\hat{V}_{iu}^{-1} \hat{V}_{jv}^{-1} + \hat{V}_{iv}^{-1} \hat{V}_{ju}^{-1}](1 + \delta_{uv})^{-1},$$

where δ_{uv} is the Kronecker δ (see App. B). With this scheme, large modulus entries in $\mathbf{h}_{\max}$ will indicate treatment groups that have large relative local influence.

Figure 2, row III, depicts directions of maximum curvature for the idealized data set, again when interest centers on the full model, the mean, the error variance, and the random effects variance, respectively. The index for the panels in this row and row IV correspond to the index number i for the A_i effect. Note again the similarities in panels III(a), III(c), and III(d). Whereas it is clear that simultaneously increasing the assumed variances of the smaller groups and reducing those of the larger groups (or vice versa) will effect changes in the estimate of μ, maximal changes to the estimates of σ^2 and σ_A^2 are brought about by decreasing (or increasing) the assumed variances of the outlying groups. Adding 2σ to the 30th case has little effect on the directions and is not shown. The effect of a $2\sigma_A$ increase to A_{10}, however, is clear from an inspection of Figure 2, row IV. In every case the 10th group is visually isolated as influ-

ential. The panels of row IV tell the analyst that the A_i's may not have been drawn from the same normal population and in particular that the results of the analysis are most sensitive to A_{10}.

3.5 Perturbation of the Response Vector

The notion of leverage in fixed-effect linear models is useful for understanding how an individual response contributes to the fitted values. In a recent study, Emerson, Hoaglin, and Kempthorne (1984) used perturbations in the response to study leverage in additive-plus-multiplicative fits for two-way tables.

In the present context, perturbations of the responses can be introduced as follows. In general, for the ith observation y_i of a data set, let $y_i(\omega) = y_i + s\omega$ denote the ith perturbed response, where s is a chosen scale that is used to convert the generic perturbation ω to the required size and units. In this case, $\boldsymbol{\omega}_0 = \mathbf{0}$, and it is straightforward to show that the $\boldsymbol{\Delta}$ matrix is given by

$$\boldsymbol{\Delta}^T = s\hat{\mathbf{V}}^{-1}[\mathbf{X},\ \hat{\sigma}^{-2}\mathbf{r},\ \hat{\sigma}^{-2}\hat{\mathbf{V}}^{-1}\mathbf{U}_1\mathbf{U}_1^T\hat{\mathbf{V}}^{-1}\mathbf{r},\ \ldots,\ \hat{\sigma}^{-2}\hat{\mathbf{V}}^{-1}\mathbf{U}_c\mathbf{U}_c^T\hat{\mathbf{V}}^{-1}\mathbf{r}].$$

From the preceding, it is clear that only the magnitude of the curvature will depend on s; $\mathbf{h}_{\max}$ will not.

Diagnostic plots of the idealized data set for the full model and for individual parameters are presented in Figure 2, row V. For maximal local effect of the parameter estimates, panel V(a) indicates that the low value in each group should be decreased and the high value in each group should be increased. Moreover, a very slight decreasing trend in the groups is evident. Therefore, in addition to separating the data points within a group, the groups must also be slightly separated to obtain the maximal local effect on the parameter estimates. The plot corresponding to the mean indicates the obvious. To move the sample mean, all data points should be either increased or decreased simultaneously. As with the previous schemes, panels V(c) and V(d) are quite similar to V(a), although, as V(d) indicates, maximum change to the estimated variance component is effected by increasing the magnitude of observations in the larger groups and decreasing those in the smaller groups.

In Figure 2, row VI, similar plots are presented after $2\sigma = 2$ has been added to $y_{10,3}$ (Case 30). We note that Case 30 is highlighted in the VI(a), VI(c), and VI(d) panels, but not nearly as dramatically as in row II for perturbation of error variances. In fact, the change had no effect at all in panel VI(b). When the 10th group is perturbed, the resulting diagnostic displays (not shown here) are virtually those of row V. This suggests that diagnostics obtained via perturbation of the y values may be less sensitive than diagnostics derived from perturbations to the model. In Section 5, we elaborate on this point in connection with the aerosol data.

If all effects are fixed, $\hat{\mathbf{V}}^{-1} = \hat{\sigma}^{-2}\mathbf{I}$ and the last c columns of $\boldsymbol{\Delta}^T$ drop out to give

$$\boldsymbol{\Delta}^T = s[\hat{\sigma}^{-2}\mathbf{X},\ \hat{\sigma}^{-4}\mathbf{r}].$$

Since

$$\mathbf{Q}^{-1} = \begin{bmatrix} (\mathbf{X}^T\mathbf{X})^{-1}\hat{\sigma}^2 & \mathbf{0} \\ \mathbf{0} & 2\hat{\sigma}^4/n \end{bmatrix},$$

we have

$$\mathbf{F} = s^2[\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\hat{\sigma}^{-2} + 2\hat{\sigma}^{-4}\mathbf{r}\mathbf{r}^T/n].$$

The first term of $\mathbf{F}$ reflects the sensitivity of $\hat{\boldsymbol{\beta}}$ to the imposed perturbations, whereas the second term reflects the sensitivity of $\hat{\sigma}^2$. When σ^2 is known, $\mathbf{F} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^Ts^2/\sigma^2$ and the notion of leverage in linear models emerges. Similarly, when σ^2 is unknown and only $\boldsymbol{\beta}$ is of interest, $\mathbf{F} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^Ts^2/\hat{\sigma}^2$. In general, the diagonals of $\mathbf{F}$ may be viewed as extended leverage values that depend on the size of the perturbations as measured by s^2/σ^2. The diagnostic value of the diagonal elements of $\mathbf{F}$ for arbitrary perturbation schemes is more fully discussed in the following section.

3.6 Discussion of the Idealized Example

The idealized example illustrates several useful properties of the operating characteristics of the direction of maximum curvature. First, relatively large or small elements of the maximum-curvature direction vector indicate that the corresponding component of the data/model is relatively influential for the parameter estimate under consideration. Thus, as shown in Figure 2, panel II(c), when Case 30 is modified to be an outlier it contributes substantially to the estimate of σ^2. Second, outlying elements in the direction-of-maximum-curvature vector should always be followed by further investigation, as they do not necessarily indicate an aberration in the data. For example, in Figure 2, panel III(c), the 1st and 10th groups have the largest impact on $\hat{\sigma}^2$ because the sample size is small, not because of any aberration. For an uncomplicated analysis, such sensitivity is to be avoided. This may require additional data as implied by row III of Figure 2 or an understanding of the "null" model such as that given in Figure 2, panel III(c). Third, as illustrated in rows II, IV, and VI of Figure 2, it is usually unnecessary to monitor changes in each of the individual parameters; for this model the full parameter vector, which contains information on each parameter, seems to tell the essential story. Finally, based on this simple

example, perturbations of the response seem to be of little value.

4. DISCUSSION

Many other schemes, of course, are possible. Sensitivity to the errors-in-variables problem may be assessed via perturbations of the columns of $\mathbf{X}$. Similarly, sensitivity to the assumption of independence could be investigated by perturbations to the elements of the covariance matrix. One could also consider perturbations of caseweights. For fixed-effects models, this is equivalent to perturbation of the error variances. The meaning of a case weight, however, for random-effects models is not so clear, and, for this reason, we have chosen to omit a detailed treatment.

Computation of $C_{\max}$ and $\mathbf{h}_{\max}$ can at times be both difficult and computationally expensive. For example, the error-variance perturbation scheme in Section 3.3 requires computation of n $n \times n$ derivative matrices, $\mathbf{D}_k$. Once the $n \times n$ acceleration matrix $\mathbf{F}$ is computed, eigenvectors and eigenvalues must be obtained. For large n, available precision will sometimes preclude a full eigenanalysis of the matrix. Fortunately, since only the largest eigenvalue and its associated eigenvector are required, a much simpler computation is usually sufficient. Although $\mathbf{F}$ may be quite large, $\mathbf{Q}^{-1}$ is generally quite small. Assuming that $\mathbf{Q}$ is positive semidefinite, we may write

$$\mathbf{Q}^{-1} = \mathbf{\Gamma}^T\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{\Gamma} = \mathbf{B}^T\mathbf{B},$$

where $\mathbf{\Gamma}$ is orthogonal, $\mathbf{D}$ is a diagonal matrix with diagonal elements composed of the eigenvalues of $\mathbf{Q}^{-1}$, and $\mathbf{B} = \mathbf{D}^{1/2}\mathbf{\Gamma}$. Then $\mathbf{F} = \mathbf{A}\mathbf{A}^T$, where $\mathbf{A} = \mathbf{\Delta}^T\mathbf{B}^T$ is $q \times p$. Now the first p eigenvalues of $\mathbf{F}$ are those of

$$\mathbf{A}^T\mathbf{A} = \mathbf{D}^{1/2}\mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Delta}^T\mathbf{\Gamma}^T\mathbf{D}^{1/2}.$$

Let ξ denote the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$, and assume that $\mathbf{z}$ is the associated eigenvector. It is straightforward to show that

$$\mathbf{h}_{\max} = \mathbf{A}\mathbf{z}/\sqrt{\xi}.$$

Thus it is only necessary to decompose the $p \times p$ matrices $\mathbf{Q}^{-1}$ and $\mathbf{A}^T\mathbf{A}$, and to store the $n \times p$ matrix $\mathbf{A}$, to determine the maximum curvature and its associated direction.

When time or precision is a problem, single-

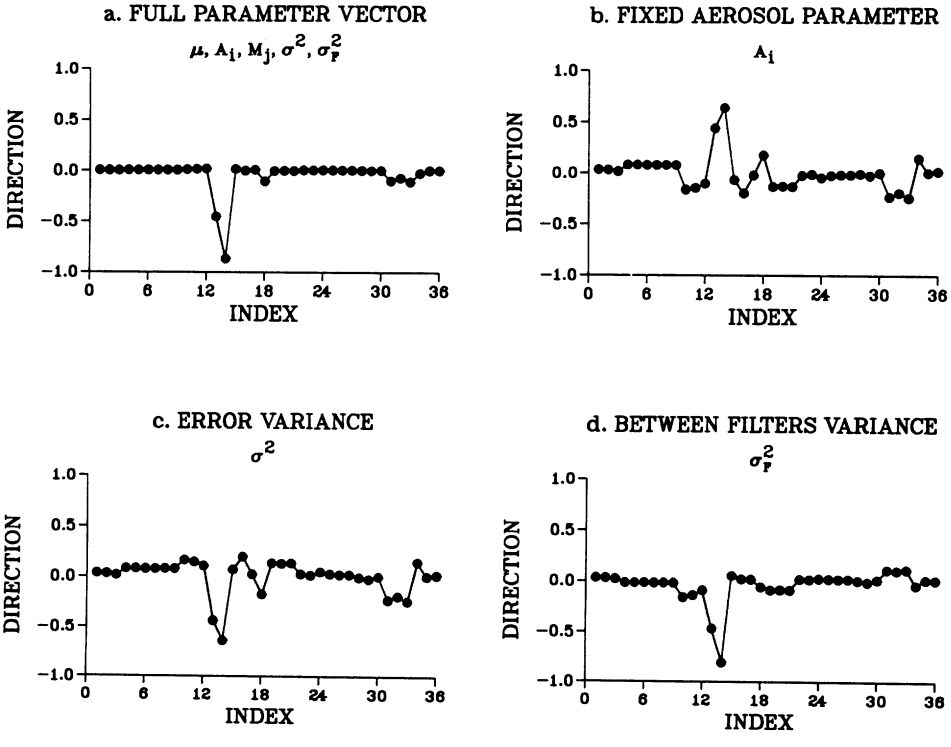

Figure 3. Direction Vector of Maximum Curvature Corresponding to Perturbation of the Error Variances for Aerosol Data: (a) Likelihood Displacement for the Entire Parameter Vector; (b) Aerosol Parameter Only; (c) Error Parameter Only; (d) Between-Filter Random-Effects Parameter.

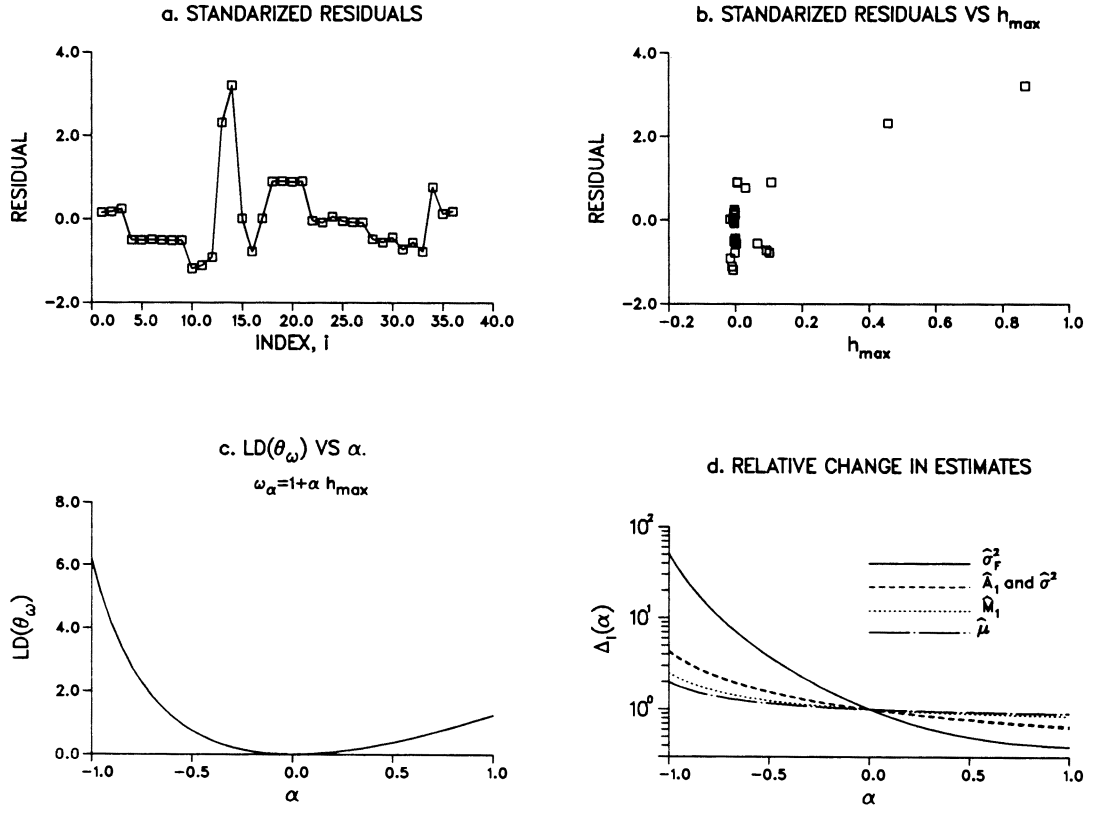

Figure 4. Diagnostics for the Aerosol Example: (a) Standardized Residuals by Data Index; (b) Standardized Residuals as a Function of the Direction of Maximum-Curvature Vector h_{max}; (c) the Lifted Line in the Direction of Maximum Curvature; (d) Relative Change in the Parameter Estimates in the Direction of Maximum Curvature.

variable analogies of the discussed perturbation schemes would be useful. For example, one might consider n separate schemes such that in the ith perturbation scheme only the ith error variance is perturbed multiplicatively. Then

$$\Delta = \partial^2 \lambda(\omega_i)/\partial\theta\partial\omega_i \Big|_{\omega_i = 1, \theta = \hat{\theta}}.$$

But this quantity is the ith column of Δ as defined in Section 3.1. Further, $\mathbf{F}$ in this case is just a scalar, $\Delta^T \mathbf{Q}^{-1} \Delta$, so $C_{max} = \mathbf{F}$. Thus, in general, the n curvatures corresponding to the n separate single-variable perturbation schemes are given by the diagonal of $\mathbf{F}$ in the full scheme. Hence the diagnostic usefulness of the diagonal of $\mathbf{F}$ is not restricted to perturbations of the responses. Index plots of the diagonal of $\mathbf{F}$ may be of value, particularly if accurate computation of the corresponding eigenvalues and eigenvectors is either prohibitively expensive or impossible.

We must emphasize that, without modification, the method is inappropriate whenever any first derivatives of the log-likelihood, evaluated at the MLE, are nonzero. This, of course, can happen, because the variance and variance-component estimates are constrained to be nonnegative. Two alternative approaches in such cases might be suggested. In the first, one might simply remove the nonnegativity constraints and allow the Newton–Raphson procedure to find the unconstrained MLE's. Use of the methods described herein may shed light on the "cause" of the negative estimates of variance. Alternatively, one might decide to assess the influence of the perturbation scheme of interest on the positive variance components and fixed effects only. The methodology for assessment of influence of perturbations on subsets of the parameter vector $\boldsymbol{\theta}$, as described in Section 2, would then be applicable.

5. ANALYSIS OF THE AEROSOL EXAMPLE

We now consider the use of these three diagnostics schemes in conjunction with the aerosol data as described in Section 2. MLE's are $\hat{\mu} = .992$, $\hat{A}_1 = .197$, $\hat{M}_1 = .597$, $\hat{\sigma}_F^2 = .136$, and $\hat{\sigma}^2 = .633$.

We first consider diagnostics for the full model, for perturbation of error variances. For this scheme, the nonzero eigenvalues of $\mathbf{F}$ (4.11, .83, .34, .10, and .07) indicate that the first eigenvalue is quite large in relation to the rest. Thus, in this case, the most critical diagnostic information will be associated with $\mathbf{h}_{max}$. The index plot of $\mathbf{h}_{max}$, presented in Figure 3(a), is

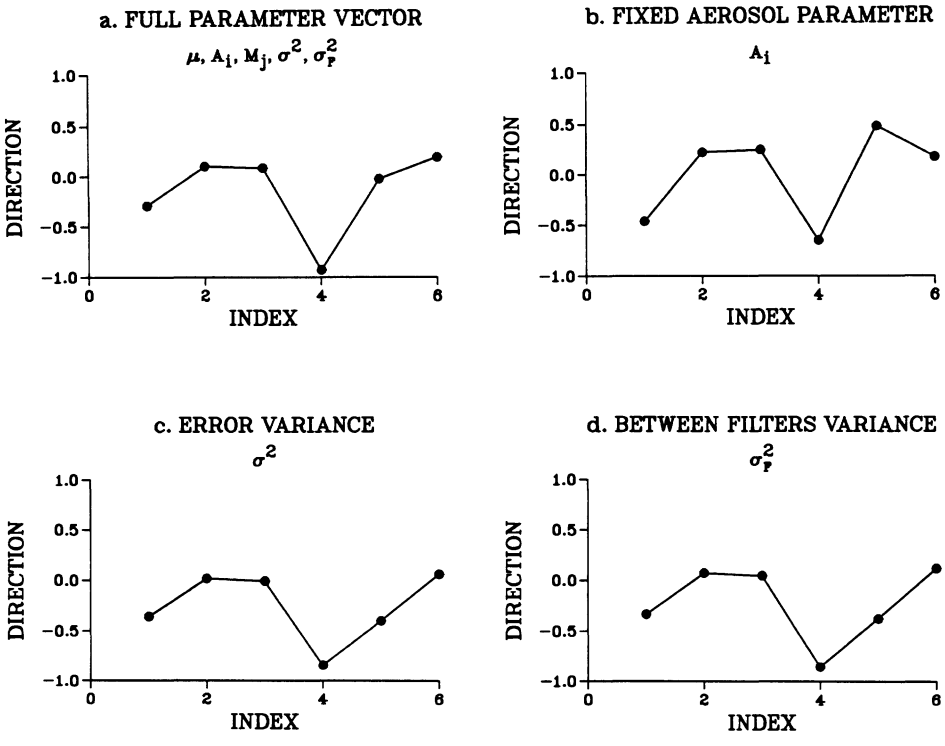

Figure 5. Direction of the Maximum-Curvature Vector for Perturbation of the Random-Effects Variances for Aerosol Data: (a) Likelihood Displacement for the Entire Parameter Vector; (b) Aerosol Parameter Only; (c) Error Parameter Only; (d) Between-Filter Random-Effects Parameter.

revealing. It indicates that perturbations in assumptions leading to either a simultaneous increase or decrease in the variances of the two filter-5 observations 13 and 14 (y_{1221} and y_{1222}) will lead to rapid changes in $\hat{\boldsymbol{\theta}}$, as measured by the likelihood displacement. Plots of $\mathbf{h}_{max}$ when interest centers on the individual parameters A, σ^2, and σ_F^2 are presented in Figure 3(b, c, d). In each part, Cases 13 and 14 again stand out. Locally these observations are highly influential.

This conclusion might have been surmised from inspection of the index plot of standardized residuals [Fig. 4(a)]. Although not so prominent, the 13th and 14th observations again stand out, having standardized residuals of 2.3 and 3.2, respectively. But note, for example, that the entry in $\mathbf{h}_{max}$ corresponding to the third largest modulus residual (Case 10) is essentially 0. $\mathbf{h}_{max}$ is plotted against the residual vector in Figure 4(b), further underscoring the point that different information is provided by these two vectors.

To further assess the influence of the 13th and 14th observations, we studied the change in the likelihood and parameter estimates as the perturbations moved in the direction of $\pm\mathbf{h}_{max}$. Let $\boldsymbol{\omega}_\alpha = \mathbf{1} + \alpha\mathbf{h}_{max}$, and let $\Delta_i(\alpha) = \hat{\theta}_i^\alpha/\hat{\theta}_i$, where $\hat{\theta}_i^\alpha$ is the MLE for θ_i under perturbation $\boldsymbol{\omega}_\alpha$. $\Delta_i(\alpha)$ measures the relative change in $\hat{\theta}_i^\alpha$. A plot of the likelihood displacement $LD(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}_\alpha})$ is presented in Figure 4(c) for $-1 \le \alpha \le 1$. From the figure, it is evident that movement in the direction $-\mathbf{h}_{max}$ will lead to more substantive changes in $\hat{\boldsymbol{\theta}}$ than will perturbations in the positive direction. The corresponding changes, $\Delta_i(\alpha)$, in each of the parameters are pictured in Figure 4(d). It is clear from the figure that the estimated variance component, $\hat{\sigma}_F^2$, is most susceptible to changes in $\boldsymbol{\omega}_\alpha$, as its changes completely dominate the relative changes in the other parameters. The coordinate of $\mathbf{h}_{max}$ corresponding to observation 14 has the value .87. Thus, at $\alpha = -1$ the assumed variance has been reduced from σ^2 to $.13\sigma^2$. This leads roughly to a 50-fold increase in $\hat{\sigma}_F^2$ and to 2-fold to 4-fold increases in the remaining estimates. On the other hand, movement in the direction $+\mathbf{h}_{max}$ will lead to an increase in the variances of these two points and, in turn, to reductions (though less dramatic) in the parameter estimates.

The direction vectors for perturbation of random-effects variances are presented in Figure 5. Since there are only three random effects for each of the two manufacturers, $\mathbf{h}_{max}$ in each case is a 6 vector. Surprisingly, the fifth filter, which contains the two anomalous observations identified previously, is not identified as influential. Rather, the fourth filter

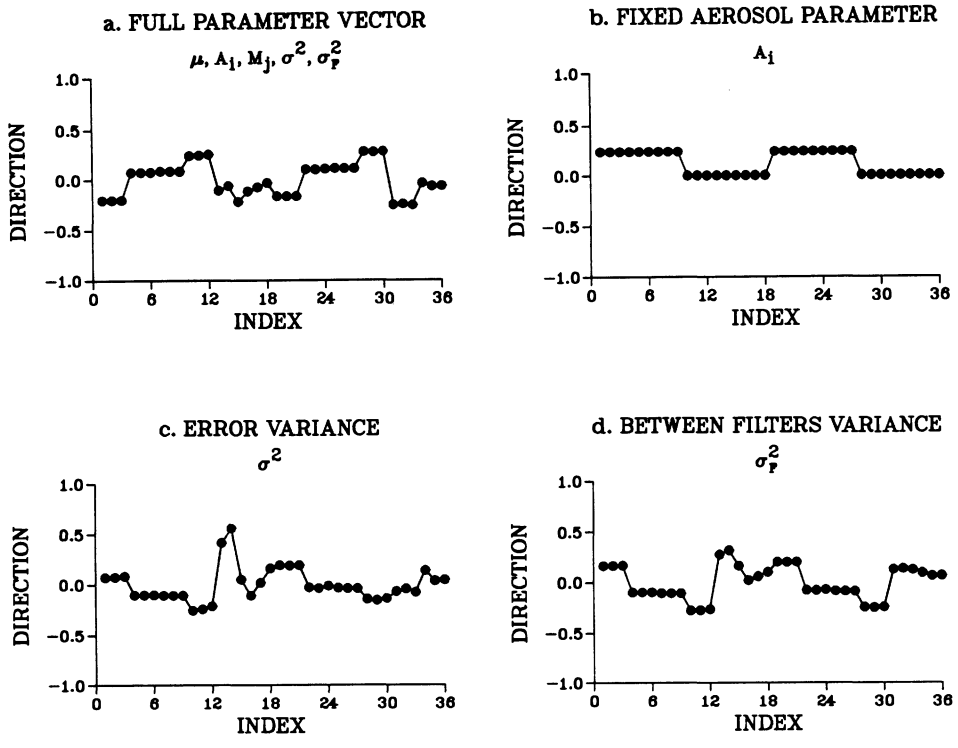

Figure 6. Direction of the Maximum-Curvature Vector for Perturbation of the y Values for Aerosol Data: (a) Likelihood Displacement for the Entire Parameter Vector; (b) Aerosol Parameter Only; (c) Error Parameter Only; (d) Between-Filter Random-Effects Parameter.

stands out. This was somewhat disconcerting at first, but it was later corroborated by an examination of the average residuals, r_i, for each filter. These were $r_1 = .55$, $r_2 = -.26$, $r_3 = -.29$, $r_4 = -.78$, $r_5 = .57$, and $r_6 = .43$. Notice that the largest modulus average residual ($r_4 = -.78$) corresponds to the fourth filter.

Results for perturbation of the y values are given in Figure 6. Very surprisingly, Cases 13 and 14 are not identified as basically influential for effects on the full parameter vector [Fig. 6(a)]. Only when interest is limited to effects on the error variance does a pattern resembling that previously seen in Figure 3(a) emerge. From these results and from those previously reported in connection with the one-way ANOVA example, it appears that diagnostics produced from data perturbations are not sufficiently informative. We conclude that an analysis is best served by a routine inspection of the $\mathbf{h}_{max}$ for perturbation of the error variance and for perturbation of the random-effects variances.

6. CONCLUSIONS

Diagnostic methods are useful for assessing the adequacy of assumptions underlying a modeling process and for identifying unexpected characteristics of the data that may seriously influence conclusions or require special attention. Such methods may also serve as the initial step in the determination of the robustness of a sample.

In recent years, deleting observations has become a popular and extremely useful basis for studying sensitivity in statistical problems. The deletion of observations, however, is just one possible way in which a postulated model might be perturbed meaningfully. The methodology discussed in this article is designed to allow an assessment of the sensitivity of a mixed-model analysis to perturbations in many of the standard assumptions. Results from sensitivity analyses must be treated with caution, and the sensitivity itself must become part of the conclusions. When perturbing error variances, the cause of the sensitivity can often be traced via $\mathbf{h}_{max}$ to a few influential observations, and in such cases the proposed methodology agrees well with case deletion. The usefulness of case deletion, however, seems limited to providing an understanding only of certain aspects of the error component of a mixed model. For example, case deletion will not generally lead to the identification of influential groups associated with a variance component. As briefly described in Section 3, the proposed methodology can be adapted easily to handle such concerns.

In general, we have found the proposed methodology to be a useful aid for assessing the dependence of a mixed-model analysis on standard assumptions. The local nature of the methodology may be a limitation, however. For example, a few observations that are jointly but not individually influential can interact to produce a likelihood that is very flat around the MLE. In such cases the maximum curvature for error-variance perturbation schemes will tend to be small, correctly indicating little or no *local* sensitivity. To identify the observations in question, more difficult global sensitivity analyses, such as that provided by group deletion, are required. Fortunately, our experience indicates that, even when the maximum curvature is small, global sensitivity is often indicated by large modulus entries in $\mathbf{h}_{\max}$, reinforcing the recommendation that $\mathbf{h}_{\max}$ should be inspected regardless of the size of $C_{\max}$.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant DMS-860 1314 and by Nuclear Regulatory Commission Grant RES 60-83-176. We would like to thank Mark Johnson and Luis Escobar for their comments on an earlier draft of this article. We also thank A. Juan.

APPENDIX A: MATRIX EQUALITY

Consider the matrix $\mathbf{A} = \mathbf{Q}^{-1} - \mathbf{B}_{22}$, where

$$\mathbf{B}_{22} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{22}^{-1} \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} q_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix}.$$

Assume that $\mathbf{A}$ is partitioned as $\mathbf{Q}$, and let $\mathbf{G} = \mathbf{Q}_{12}/q_{11}$ and $\mathbf{E} = \mathbf{Q}_{22} - \mathbf{Q}_{21}\mathbf{Q}_{12}/q_{11}$. Then, using standard matrix equalities (see Rao 1965, p. 29),

$$\mathbf{E}^{-1} = \mathbf{Q}_{22}^{-1} + \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}/c,$$

where $c = q_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}$. Then $a_{11} = q_{11}^{-1} + \mathbf{G}\mathbf{E}^{-1}\mathbf{G}$, which after simplification is $a_{11} = 1/c$. By similar arguments, $\mathbf{A}_{12} = -\mathbf{G}\mathbf{E}^{-1} = -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}/c$, and $\mathbf{A}_{22} = \mathbf{E}^{-1} - \mathbf{Q}_{22}^{-1} = \mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}/c$. Now, $\mathbf{A}$ may be written in the form $\mathbf{H}\mathbf{H}^T$, where $\mathbf{H}^T = (1, -\mathbf{Q}_{12}\mathbf{Q}_{22}^{-1})/\sqrt{c}$.

APPENDIX B: DERIVATION OF $\mathbf{D}_k$ FOR PERTURBATIONS IN THE VARIANCE MATRIX

We assume that perturbations, $\boldsymbol{\omega}$, enter only through the variance matrix. That is, $\mathbf{V} = \mathbf{V}(\boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is an m vector of perturbations. Here ω_k perturbs some subset of the elements of the assumed covariance matrix, $\mathbf{V}$. We develop a general expression for the kth derivative matrix

$$\mathbf{D}_k = \mathbf{D}_k(\boldsymbol{\omega}_0) = \left.\frac{\partial \mathbf{V}^{-1}(\boldsymbol{\omega})}{\partial \omega_k}\right|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}$$

for $k = 1, \ldots, m$. For $i \le j$, let $\{\mathbf{V}^{-1}(\boldsymbol{\omega})\}_{ij} = f_{ij}(\mathbf{V}(\boldsymbol{\omega}))$, where

$$\mathbf{f}^T = (f_{11}(\mathbf{V}(\boldsymbol{\omega})), f_{12}(\mathbf{V}(\boldsymbol{\omega})), \ldots, f_{1n}(\mathbf{V}(\boldsymbol{\omega})), f_{22}(\mathbf{V}(\boldsymbol{\omega})), \ldots, f_{nn}(\mathbf{V}(\boldsymbol{\omega}))).$$

Since $\mathbf{V}(\boldsymbol{\omega})$ is composed of $n(n+1)/2$ unique elements, we denote by $\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))$ the $[n(n+1)/2] \times [n(n+1)/2]$ Jacobian of $f(\mathbf{V}(\boldsymbol{\omega}))$. In our notation, an element in this matrix occupies the (i, j)th row and (u, v)th column ($i, j, u, v = 1, \ldots, n$; $i \le j$, $u \le v$). Similarly, writing $\{\mathbf{V}(\boldsymbol{\omega})\}_{ij} = g_{ij}(\boldsymbol{\omega})$, with

$$\mathbf{g}^T = (g_{11}(\boldsymbol{\omega}), \ldots, g_{nn}(\boldsymbol{\omega})),$$

the $[n(n+1)/2] \times m$ Jacobian of $\mathbf{g}(\boldsymbol{\omega})$ is $\mathbf{J}_g(\boldsymbol{\omega})$. Here again, an element of this matrix occupies the (i, j)th row and the kth column. Letting $\mathbf{J}_g(\boldsymbol{\omega})_k$ denote the column of $\mathbf{J}_g(\boldsymbol{\omega})$ corresponding to ω_k and $\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))^{ij}$ denote the (i, j)th row of $\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))$, elements of the derivative matrix $\mathbf{D}_k(\boldsymbol{\omega})$ are specified by

$$\{\mathbf{D}_k\}_{ij} = \mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))^{ij}\mathbf{J}_g(\boldsymbol{\omega})_k. \tag{B.1}$$

Using a result in matrix differentiation (see, e.g., Graybill 1969), it follows that

$$\{\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))\}_{(i,j),(u,v)} = -\{\mathbf{V}^{-1}(\boldsymbol{\omega})\boldsymbol{\Gamma}_{uv}\mathbf{V}^{-1}(\boldsymbol{\omega})\}_{ij}, \tag{B.2}$$

where $\boldsymbol{\Gamma}_{uv}$ is an $n \times n$ matrix with ones in the (u, v)th and (v, u)th positions and zeros elsewhere. Notice that, after evaluation at $\boldsymbol{\omega}_0$, we have

$$\{\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}_0))\}_{(i,j),(u,v)} = -\{V_{iu}^{-1}V_{jv}^{-1} + V_{iv}^{-1}V_{ju}^{-1}\}[1 + \delta_{uv}]^{-1}, \tag{B.3}$$

where $\delta_{uv} = 1$ if $u = v$ and 0 otherwise. Thus, for any scheme involving perturbations of the variance matrix only, $\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}_0))$ is as specified previously. Changes in $\mathbf{D}_k$ enter only through changes in $\mathbf{J}_g(\boldsymbol{\omega}_0)$, which is usually simple to compute. In the subsections that follow, we evaluate (B.1) for each of the perturbation schemes discussed in the article.

B.1 Derivation of the $\mathbf{D}_k$ Matrix for Perturbation of the Error Variances

For perturbation of the error variances, $\mathbf{J}_g(\boldsymbol{\omega}_0)_k$ is an $n(n+1)/2$ vector with (k, k)th entry equal to $\hat{\sigma}^2$ and all others 0. From (B.1) and (B.2),

$$\begin{aligned}\{D_{kk}\}_{ij} &= \left.\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))^{ij}\mathbf{J}_g(\boldsymbol{\omega})_{kk}\right|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \\ &= -\hat{\sigma}^2\{\hat{\mathbf{V}}_k^{-1}\boldsymbol{\Gamma}_{kk}\hat{\mathbf{V}}_k^{-1}\}_{ij} \\ &= -\hat{\sigma}^2\{\hat{\mathbf{V}}_k^{-1}(\hat{\mathbf{V}}_k^{-1})^T\}_{ij}.\end{aligned}$$

It follows that $\mathbf{D}_k = -\hat{\sigma}^2\hat{\mathbf{V}}_k^{-1}(\hat{\mathbf{V}}_k^{-1})^T$.

B.2 Derivation of the D_k Matrix for Perturbations of the Variances of the Random Effects

Here we have assumed that

$$\mathbf{V}(\boldsymbol{\omega}) = \sum_{i=1}^{c} \sum_{j=1}^{q_i} \mathbf{U}_j^i \mathbf{U}_j^{iT} \sigma_i^2 \omega_{ij} + \sigma^2 \mathbf{I},$$

so the perturbation vector, $\boldsymbol{\omega}$, is written as

$$\boldsymbol{\omega}^T = (\omega_{11}, \ldots, \omega_{1q_1}, \ldots, \omega_{c1}, \ldots, \omega_{cq_c}).$$

The dimension of $\boldsymbol{\omega}$ is $q = \sum_{i=1}^{c} q_i$. In this case, the (k, h)th column of $\mathbf{J}_g(\boldsymbol{\omega}_0)$ has $U_{uh}^k U_{vh}^k \sigma_k^2$ in the (u, v)th position, where U_{jh}^k is the jth element of $\mathbf{U}_h^k$. With $\mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}_0))$ as specified in (B.2), it follows that

$$\begin{aligned}\{D_{kh}\}_{ij} &= \mathbf{J}_f(\mathbf{V}(\boldsymbol{\omega}))^{ij} \mathbf{J}_g(\boldsymbol{\omega})_{kh} \Big|_{\boldsymbol{\omega} = \boldsymbol{\omega}_0, \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \\ &= -\hat{\sigma}_k^2 \sum_{\substack{u=1 \\ u \leq v}}^{n} \sum_{v=1}^{n} U_{uh}^k U_{vh}^k \\ &\quad \times [\hat{V}_{iu}^{-1} \hat{V}_{jv}^{-1} + \hat{V}_{iv}^{-1} \hat{V}_{uj}^{-1}][1 + \delta_{uv}]^{-1}.\end{aligned}$$

[*Received February 1986. Revised April 1987.*]

REFERENCES

Beckman, R. J., and Cook, R. D. (1983), "Outlier ... s," *Technometrics*, 25, 119–149.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: John Wiley.

Cook, R. D. (1986), "Assessment of Local Influence" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 48, 133–169.

Cook, R. D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman & Hall.

Emerson, H. D., Hoaglin, D. C., and Kempthorne, P. J. (1984), "Leverage in Least Squares Additive-Plus-Multiplicative Fits for Two-Way Tables," *Journal of the American Statistical Association*, 79, 329–335.

Fellner, W. H. (1986), "Robust Estimation of Variance Components," *Technometrics*, 28, 51–60.

Graybill, F. A. (1969), *Introduction to Matrices With Applications in Statistics*, Belmont, CA: Wadsworth.

Hocking, R. R., Green, J. W., and Bremer, R. H. (1986), "Estimation of Variance Components in Mixed Factorial Models Including Model-based Diagnostics," unpublished paper presented at the annual Joint Statistical Meetings, Chicago, August.

Jennrich, R. I., and Sampson, P. F. (1976), "Newton–Raphson and Related Algorithm for Maximum Likelihood Variance Component Estimation," *Technometrics*, 18, 11–17.

Kerschner, H. F., Ettinger, H. J., DeField, J. D., and Beckman, R. J. (1984), "A Comparative Study of HEPA Filter Efficiencies When Challenged With Thermal and Air-Jet Generated Di-2-Ethylhexyl Sebecate, Di-2-Ethylhexyl Phthalate and Sodium Chloride," Laboratory Report LA-9985-MS, Los Alamos National Laboratory, Los Alamos, NM.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*, New York: John Wiley.