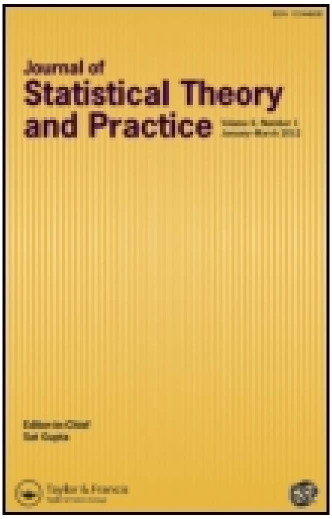

Journal of Statistical Theory and Practice

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ujsp20

Noncentralities Induced in Regression Diagnostics

D. R. Jensen[a] & D. E. Ramirez[b]

[a] Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA

[b] Department of Mathematics, University of Virginia, Charlottesville, Virginia, USA

Accepted author version posted online: 24 Feb 2014.Published online: 24 Mar 2014.

Journal of Statistical Theory and Practice, 8:141–165, 2014

ISSN: 1559-8608 print / 1559-8616 online
DOI: 10.1080/15598608.2014.847758

Noncentralities Induced in Regression Diagnostics

D. R. JENSEN[1] AND D. E. RAMIREZ[2]

[1]Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA
[2]Department of Mathematics, University of Virginia, Charlottesville, Virginia, USA

Anomalies persist in the use of deletion diagnostics in regression. Tests for outliers under subset deletions utilize the R–Fisher F_I statistics, each having a noncentral F-distribution with noncentrality parameter λ as a function of shifts only at deleted rows in the index set I. Numerous studies examine empirical outcomes of these diagnostics in random experiments. In contrast, studies here are probabilistic, examining distributions behind those empirical outcomes and tracking the effects of shifts at nondeleted rows. By allowing shifts at nondeleted rows in a set J, in addition to traditional shifts at deleted rows in I, F_I is shown to have a doubly noncentral F-distribution. By removing the unnecessary restriction that shifts occur only at deleted rows, these findings support constructs akin to power curves in tracking probabilities of masking or swamping as shifts evolve. In addition, "regression effects" among outliers may have unforeseen consequences. A dichotomy of shifts is discovered as projections into the "regressor" and "error" spaces of a model. Hidden shifts at nondeleted rows can obfuscate not only meanings ascribed to traditional outlier diagnostics, but also to subset influence diagnostics corresponding one-to-one with F_I. In short, despite wide usage abetted by software support, deletion diagnostics in current vogue no longer can be recommended to achieve objectives traditionally cited. Case studies illustrate the debilitating effects of these anomalies in practice, together with conclusions misleading to prospective users.

AMS Subject Classification: 62J05; 62J20.

Keywords: Subset leverages; Coleverages; Vector outliers; Regression diagnostics.

1. Introduction

Begin with $\boldsymbol{Y}_0 = \boldsymbol{X}_0\beta + \varepsilon_0$ of full rank having n observations, p regressors, and uncorrelated errors with variance σ^2, giving $\widehat{\beta}$ as Gauss–Markov solutions and S^2 as the residual mean square. Regression diagnostics seek *leverages* of regressors, *outlying* data, and observations deemed *influential* whose removal would alter essentials of the analysis. See Belsley et al. (1980), Cook and Weisberg (1982), Barnett and Lewis (1984), Atkinson (1985), Rousseeuw and Leroy (1987), Chatterjee and Hadi (1988), Myers (1990), Fox (1991), and others. Subset deletions follow on eliminating s rows $\{\boldsymbol{Y}_I, \boldsymbol{Z}, \varepsilon_I\}$ from $\{\boldsymbol{Y}_0, \boldsymbol{X}_0, \varepsilon_0\}$, leaving $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$ of full rank with $r = n - s > p$ rows, giving $(\widehat{\beta}_I, S_I^2)$ from the reduced data. Basic arrays include $\boldsymbol{H}_n = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'$; its diagonal elements $\{h_{ii} \in (0, 1);\ 1 \le i \le n\}$ are *leverages* attributed to rows $\{x_i';1 \le i \le n\}$ of $\boldsymbol{X}_0$; off-diagonal elements $\{h_{ij};\ i \ne j\}$ are *coleverages*; and elements of $(\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{Y}_0 = e_0 = [e_1, \ldots, e_n]'$ comprise the vector of

Received 10 May 2012; accepted 5 April 2013.
Address correspondence to: D. R. Jensen, Department of Statistics, Virginia Tech, 406-A Hutcheson Hall, Blacksburg, VA 24061-0439, USA. Email: djensen@vt.edu

residuals in the full data, to be partitioned as $e_0' = [e', e_I']$. Designs fully estimable after deletions are studied in Ghosh (1978).

For $s = 1$ the R–Student statistics t_i trace to Snedecor and Cochran (1968, 157) in testing for a single shift at x_i'; see also Beckman and Trussell (1974). Similarly, for I a subset of $\{1, 2, \ldots, n\}$, the R–Fisher statistic F_I serves to track a vector shift $\boldsymbol{Y}_I \to \boldsymbol{Y}_I + \delta$; see Gentleman and Wilk (1975). Here, t_i and F_I are given by

$$t_i = \frac{e_i}{S_i\sqrt{(1 - h_{ii})}}, \quad F_I = \frac{e_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I}{sS_I^2}$$

with $\boldsymbol{H}_{II} = \boldsymbol{Z}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{Z}'$. Powers of F_I to detect a shift $\boldsymbol{\delta}$ decrease with increasing leverages under Gaussian errors, since F_I then follows the noncentral Fisher distribution $F(s, n-p-s, \lambda_s)$ with noncentrality $\lambda_s = \boldsymbol{\delta}'(\boldsymbol{I}_s - \boldsymbol{H}_{II})\,\delta/\sigma^2$. It bears notice that if δ is concentrated in a subspace of $\mathbb{R}^s$ of dimension $t < s$, and having the same λ_s, then the power is greater for $F(t, n-p-t, \lambda_s)$ than $F(s, n-p-s, \lambda_s)$. This applies in context, a result of Das Gupta and Perlman (1974).

These facts apply for single shifts at designated loci. However, as specific loci seldom can be identified beforehand, all of $\{t_i;\ 1 \le i \le n\}$ are examined to isolate a suspected shift *somewhere* among the n observations. For subsets, Gentleman and Wilk (1975) defined the k most likely outliers as the subset of $s = k$ observations giving the largest reduction in the residual sum of squares when deleted, namely, $e_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I$. In practice, for $\mathcal{J}$ an ensemble of subsets I of $\{1, 2, \ldots, n\}$, the *Studentized* diagnostics $\{F_I; I \in \mathcal{J}\}$, adjusted for scale, are examined for selected subsets. Numerous studies, some by simulation, have examined empirical outcomes of these diagnostics based on observed $\boldsymbol{Y}_0$ in random experiments, tracing back over many decades of use. Unanswered are their workings in actual applications where shifts may occur in lieu of, or in addition to, the case Y_i or subset $\boldsymbol{Y}_I$ of current interest. These issues remain to be studied, as users deserve to know actual operating characteristics of these procedures in such circumstances.

To fill these gaps, studies here are probabilistic, allowing for shifts at nondeleted rows in J, in addition to traditional shifts at deleted rows in I, with F_I then having a doubly noncentral F-distribution. Specifically, we examine irregularities among distributions in an ensemble $\{\mathcal{L}(F_I); I \in \mathcal{J}\}$, where single and multiple shifts are shown to distribute as numerator and denominator noncentralities across $\{\mathcal{L}(F_I); I \in \mathcal{J}\}$, having structure intrinsic to and recovered from a given design. Effects on conventional useage are chaotic. Intended p-values may be rendered meaningless; masking and swamping may misdirect attention away from actual outlying observations; and users seldom are apprised that such anomalies have occurred. These findings in turn obscure the operating characteristics of outlier and influence diagnostics known to correspond one-to-one with F_I. Examples are given where nonoutliers are misidentified with the same likelihood that outliers are identified correctly, and similarly for ostensibly non–influential and influential observations. In short, subset deletion diagnostics generally fail to achieve the intended objectives and no longer can be recommended. An outline follows.

Section 2 surveys notation, distributions of note, models, and their error characteristics, and selected influence diagnostics. The principal findings of section 3 encompass irregularities in the workings of $\{F_I; I \in \mathcal{J}\}$ and related diagnostics, to include the masking and swamping of outliers. These findings are illustrated in section 4 through small but informative elementary examples, together with more comprehensive data from the literature. Specifically, for a data set held to be exemplary and studied since by various investigators, observations alleged to be outlying instead appear to be so as artifacts of swamping by shifts at other loci. These conclusions are supported by tools developed here. Essentials of the study are summarized in section 5. Further supporting facts are widely scattered; these are collected for reference in Appendix A.

2. Preliminaries

2.1. Notation

Spaces of interest include $\mathbb{R}^n$ as Euclidean n-space, $\mathbb{R}^n_+$ as its positive orthant, and $\mathbb{S}_n$ as the real symmetric $(n \times n)$ matrices. Vectors and matrices are set in bold type; the transpose, inverse, trace, and determinant of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\boldsymbol{A}^{-1}$, $tr(\boldsymbol{A})$, and $|\boldsymbol{A}|$; $\boldsymbol{I}_n$ is the $(n \times n)$ identity; and $Diag(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is a block-diagonal array. If $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k]$ is of order $(n \times k)$ and rank $k < n$, then $S_p(\boldsymbol{B})$ designates the column span of $\boldsymbol{B}$, that is, the k-dimensional subspace of $\mathbb{R}^n$ spanned by $[\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k]$. The ordered eigenvalues of $\boldsymbol{A} \in \mathbb{S}_n$ are $\{\lambda_i(\boldsymbol{A}) = \alpha_i;\ 1 \le i \le n\}$ with $\{\alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n\}$, and its spectral decomposition is $\boldsymbol{A} = \boldsymbol{P}\mathbf{D}_\alpha\boldsymbol{P}' = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i\boldsymbol{p}_i'$, where $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n]$ is orthogonal and $\boldsymbol{D}_\alpha = Diag(\alpha_1, \ldots, \alpha_n)$. Any *g-inverse* $\boldsymbol{A}^-$ of $\boldsymbol{A}$ satisfies $\boldsymbol{A}\mathbf{A}^-\boldsymbol{A} = \boldsymbol{A}$. The range and null spaces of $\boldsymbol{A}$ are designated as $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{N}(\boldsymbol{A})$. Specifically, if $\{\alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_r > \alpha_{r+1} = \cdots = \alpha_n = 0\}$ and if $\boldsymbol{P} = [\boldsymbol{P}_1, \boldsymbol{P}_2]$ with $\boldsymbol{P}_1 = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r]$ and $\boldsymbol{P}_2 = [\boldsymbol{p}_{r+1}, \ldots, \boldsymbol{p}_n]$, then $\mathcal{R}(\boldsymbol{A}) = S_p(\boldsymbol{P}_1)$ and $\mathcal{N}(\boldsymbol{A}) = S_p(\boldsymbol{P}_2)$.

Users long have focused on the masking and swamping of outliers: *masking*, when outliers remain undetected because others are present; *swamping*, when nonoutliers are wrongly identified owing to ambient outliers. Let I and J be subsets of observations, and λ a scalar measure of the outlyingness of observations in J. Then $\beta_P(\lambda)$ is the power curve at λ of a designated diagnostic to identify that J is outlying. Corresponding *swamping curves* are visualized as follows.

Definition 2.1.1. *Let λ measure the outlyingness of a subset J; let $E_{J \succ I}$ be the event that subset* I *is swamped by J, that is, $\boldsymbol{Y}_I$ is deemed to be outlying when it is not. Then a swamping curve is $\beta_S(\lambda) = P(E_{J \succeq I})$ as λ evolves.*

2.2. Special Distributions

The distribution of $\boldsymbol{Y} \in \mathbb{R}^n$, its characteristic function (*chf*), its mean vector, its dispersion matrix, and its generalized variance are denoted by $\mathcal{L}(\boldsymbol{Y})$, $\phi_Y(\boldsymbol{t})$, $\boldsymbol{E}(\boldsymbol{Y})$, $\boldsymbol{V}(\boldsymbol{Y}) = \boldsymbol{\Xi}$, say, and $G_V(\boldsymbol{Y}) = |\boldsymbol{\Xi}|$, with variance $Var(Y) = \sigma^2$ on $\mathbb{R}^1$. Specifically, $\mathcal{L}(\boldsymbol{Y}) = N_n(\mu, \boldsymbol{\Sigma})$ is Gaussian on $\mathbb{R}^n$ with $(\mu, \boldsymbol{\Sigma})$ as its mean and dispersion matrix. Distributions on $\mathbb{R}^1_+$ include $\chi^2(\nu, \lambda)$ as chi-squared having ν degrees of freedom, noncentrality parameter λ, and *chf* $\phi(t) = (1 - 2it)^{-\nu/2} exp\,[i\lambda t/(1 - 2it)]$; see Johnson and Kotz (1970, 132–133). In addition to $F(s, n - p - s, \lambda)$, with numerator noncentrality λ, we designate by $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ a doubly noncentral F-distribution having (ν_1, λ_1) and (ν_2, λ_2) as degrees of freedom and noncentralities in its numerator and denominator. This specializes at $s = 1$ to $F(1, \nu_2, \lambda_1, \lambda_2) = t^2(\nu_2, \lambda_1, \lambda_2)$ for the square of Student's t_i. Recall that $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ increases stochastically with increasing λ_1, and decreases stochastically with increasing λ, with other parameters held fixed. Moreover, if $\mathcal{L}(U) = F(\nu_1, \nu_2, \lambda_1, \lambda_2)$, then $\mathcal{L}(U^{-1}) = F(\nu_2, \nu_1, \lambda_2, \lambda_1)$. Identify $\{F_I > c_\alpha\}$ as the conventional α–level rejection rule based on $F(s, n - p - s, 0, 0)$.

2.3. The Models

We next put in place standard models under nonstandard assumptions. Take $\{Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i;\ 1 \le i \le n\}$ to model response Y_i to regressors $\{X_{i1}, \ldots, X_{ik}\}$ through parameters $\beta' = [\beta_0, \beta_1, \ldots, \beta_k]$. Arrayed as $\boldsymbol{Y}_0 = \boldsymbol{X}_0\beta + \varepsilon_0$, estimators $(\widehat{\beta},\ S^2)$ and $(\widehat{\beta}_I,\ S_I^2)$ are from the full and reduced data, where solutions are displayed as $\widehat{\beta} = [\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k]'$ and $\widehat{\beta}_I = [\widehat{\beta}_{I0}, \widehat{\beta}_{I1}, \ldots, \widehat{\beta}_{Ik}]'$. Properties of these and related constructs are scattered widely; essentials are assembled in Appendix A, to include conditions for inverting $(\boldsymbol{I}_s - \boldsymbol{H_{II}})$ as required of F_I.

To arrange elements of $\{\boldsymbol{Y}, \boldsymbol{X}, \varepsilon\}$ contiguously on deleting $\{\boldsymbol{Y_I}, \boldsymbol{Z}, \varepsilon_I\}$, we proceed in ordering $\{\boldsymbol{Y}_0, \boldsymbol{X}_0, \varepsilon_0\}$ so that $\{\boldsymbol{Y}_I, \boldsymbol{Z}, \varepsilon_I\}$ appear as the final s rows. Alternative choices to be deleted will require reordering. Accordingly, the residual vector $e_0 = \boldsymbol{G}\mathbf{Y}_0$, with $\boldsymbol{G} = (\boldsymbol{I}_n - \boldsymbol{H}_n)$, is now partitioned as

$$\begin{bmatrix} e \\ e_I \end{bmatrix} = \begin{bmatrix} (\boldsymbol{I}_r - \boldsymbol{H}_{00}) & -\boldsymbol{H}_{0I} \\ -\boldsymbol{H}_{I0} & (\boldsymbol{I}_s - \boldsymbol{H_{II}}) \end{bmatrix} \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y_I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y}_I \end{bmatrix} \tag{1}$$

where blocks are $\boldsymbol{H}_{00} = \boldsymbol{X}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}'$, $\boldsymbol{H}_{I0}' = \boldsymbol{H}_{0I} = \boldsymbol{X}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{Z}'$, and $\boldsymbol{H}_{II} = \boldsymbol{Z}(X_0'\boldsymbol{X}_0)^{-1}\boldsymbol{Z}'$.

Gauss–Markov assumptions on error moments, then distributions, are modified here as follows, where $\varepsilon_0 = [\varepsilon', \varepsilon_I']' \in \mathbb{R}^n$, $\varepsilon \in \mathbb{R}^r$, $\varepsilon_{\boldsymbol{I}} \in \mathbb{R}^s$, and $r + s = n$.

Assumptions A.

A1. $E(\varepsilon) = \gamma \in \mathbb{R}^r$ and $E(\varepsilon_I) = \delta \in \mathbb{R}^s$;
A2. $V(\varepsilon_0) = \sigma^2 I_n$; and
A3. $\mathcal{L}([(\varepsilon - \gamma)', (\varepsilon_I - \delta)']') = N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$.

Conventional outlier models take $\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta}$ at the deleted $\boldsymbol{Y_I}$ and $\gamma = \mathbf{0}$ elsewhere, a restriction unnecessary in concept and often unrealized in practice. To the contrary, this study allows unfettered additional, or alternative, shifts $\boldsymbol{Y} \to \boldsymbol{Y} + \gamma$ among the retained observations. Critical insight is gained on decomposing any $\gamma \in \mathbb{R}^r$ as in the following.

Definition 2.3.1.

(i) The decomposition $\gamma = \gamma_1 + \gamma_2$, with $\gamma_1 = \boldsymbol{H}_r\gamma$, $\gamma_2 = (\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma$, and $\boldsymbol{H}_r = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, entails projections $\gamma_1 \in \mathcal{R}(\boldsymbol{H}_r)$ and $\gamma_2 \in \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r)$ into the "Regressor" and "Error" spaces generated by $\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon$.
(ii) Angles between γ and these projections are $\theta_1 = \theta(\gamma, \gamma_1)$ and $\theta_2 = \theta(\gamma, \gamma_2)$.

2.4. Deletion Diagnostics: A Survey

Observations whose removal would alter essentials of the analysis have been called *influential*. For example, on deleting $\{\boldsymbol{Y}_i, \boldsymbol{x}_i', \varepsilon_i\}$ from $\{\boldsymbol{Y}_0, \boldsymbol{X}_0, \varepsilon_0\}$ and computing $DIFFIT_i = (\widehat{Y}_i - \widehat{Y_{i(i)}})/S_i\sqrt{h_{ii}}$ as a scaled divergence between predictors at x_i with and without Y_i, then Y_i is deemed *influential for prediction* at x_i provided that its removal alters $DIFFIT_i$ sufficiently; see Belsley et al. (1980). Many such diagnostics are deemed to be staples of regression, as cited in the opening paragraph, but with the critical disclaimers of Chatterjee and Hadi (1986) and discussants as in Appendix B.

Table 1
Subset deletion diagnostics

Diagnostic	Expression	Rule	Critical value
F_I	$\frac{e_I'(I_s - \boldsymbol{H}_{II})^{-1}e_I}{sS_I^2}$	$>$	c_α
OUT_I	$1 - \frac{S_I^2}{S^2}$	$>$	$\frac{[s(c_\alpha - 1)]}{[sc_\alpha + n - p - s]}$
AP_I	$\left[1 - \frac{(n-p-s)S_I^2\lvert \boldsymbol{X}'\boldsymbol{X}\rvert}{(n-p)S^2\lvert \boldsymbol{X}_0'\boldsymbol{X}_0\rvert}\right]$	$>$	$\frac{[sc_\alpha + (n-p-s)GL_I]}{[sc_\alpha + n - p - s]}$
CR_I	$\frac{\lvert S_I^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\rvert}{\lvert S^2(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\rvert}$	$<$	$\left[\frac{(n-p)}{(sc_\alpha + n - p - s)}\right]^p \frac{1}{\lvert G_{22}\rvert}$
FV_I	$\frac{\lvert S_I^2 Z(\boldsymbol{X}'\boldsymbol{X})^{-1}Z'\rvert}{\lvert S^2 Z(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}Z'\rvert}$	$<$	$\left[\frac{(n-p)}{(sc_\alpha + n - p - s)}\right]^s \frac{1}{\lvert G_{22}\rvert}$
D_I	$\frac{(\widehat{\beta} - \widehat{\beta}_I)'\boldsymbol{V}^-(\widehat{\beta} - \widehat{\beta}_I)}{sS_I^2}$	$>$	c_α

These and other diagnostics extend to encompass subset deletions, as given in part in Table 1. Here, F_I and OUT_I are intended as outlier diagnostics, the remainder to assess influence as the impact of subsets on essential features of the analysis. Rejection rules in Table 1 are consistent with F_I; the factor $GL_I = 1 - |\boldsymbol{G}_{22}|$ is the leverage diagnostic of Draper and John (1981); and $\boldsymbol{V}^-$ is a reflexive g–inverse of $\boldsymbol{V}(\widehat{\beta} - \widehat{\beta}_I)$. Further details are found in Jensen (2001) and Appendix B. The diagnostics of Table 1 are singled out here, precisely because each corresponds one-to-one with F_I and thus they comprise functionally equivalent tests. This underscores the need for further research regarding F_I : Anomalies in its distribution carry over directly to include the diagnostics of Table 1. These matters are undertaken next.

3. The Principal Findings

Our prime focus is $F_I = e_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I/sS_I^2$, specializing at $s = 1$ to t_i^2 against two-sided shifts. That $\mathcal{L}(F_I) = F(s, n-p-s, \lambda, 0)$, with $\lambda = \boldsymbol{\delta}'(\boldsymbol{I}_s - \boldsymbol{H}_{II})\boldsymbol{\delta}$, follows under Assumptions A for $E(\varepsilon_I) = \delta \in \mathbb{R}^s$, but restricted to $E(\varepsilon) = \boldsymbol{0} \in \mathbb{R}^r$. A critical reappraisal, to allow shifts anywhere, is undertaken next through the theory of quadratic forms, showing generally that $\mathcal{L}(F_I) = F(s, n-p-s, \lambda_1, \lambda_2)$, essentials then emerging via the null and range spaces of designated matrices. This analysis identifies (i) nonzero shifts that remain undetected; (ii) values $\gamma \neq \boldsymbol{0}$ nonetheless preserving $\mathcal{L}(F_I) = F(s, n-p-s, \lambda, 0)$ for some $\lambda > 0$; (iii) unanticipated "regression effects" among shifted outliers; and (iv) that F_I may be depressed or inflated stochastically due to outliers. In short, the uses of t_i^2 and F_I for outliers, and corresponding influence diagnostics, are fraught with heretofore unforeseen and debilitating consequences. Details follow, where both $\boldsymbol{H}_r = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{H}_n = \boldsymbol{X}_0(\boldsymbol{X}_0'X_0)^{-1}X_0'$ are idempotent of rank p, and we standardize $\sigma^2 = 1.0$ since other values may be reinstated as required.

To continue, technical details establishing the propagation of shifts as noncentrality parameters rest on distributions of quadratic forms. Details are given in Appendix A culminating in Lemma A.4. We next state the principal findings regarding outlier shifts and their effects on the diagnostic F_I. Recall from Eq. (1) that $(\boldsymbol{I}_n - \boldsymbol{H}_n) = [\boldsymbol{G}_{ij}]$ has blocks $\boldsymbol{G}_{11} = (\boldsymbol{I}_r - \boldsymbol{H}_{00})$, $\boldsymbol{G}'_{21} = \boldsymbol{G}_{12} = -\boldsymbol{H}_{0I}$, and $\boldsymbol{G}_{22} = (\boldsymbol{I}_s - \boldsymbol{H}_{II})$. It deserves emphasis that $\gamma_1 = \boldsymbol{H}_r\gamma$ and $\gamma_2 = (\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma$, *i.e.*, $\gamma_1 \in \mathcal{R}(\boldsymbol{H}_r)$ and $\gamma_2 \in \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r) = \mathcal{N}(\boldsymbol{H}_r)$.

Theorem 3.1. *Given Assumptions A, such that* $E(\varepsilon_0) = \omega = [\gamma', \boldsymbol{\delta}']'$; *decompose* $\gamma = \gamma_1 + \gamma_2$ *with* $\gamma_1 = \boldsymbol{H}_r\gamma$ *and* $\gamma_2 = (\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma$. *Then the distribution of* F_I *is the doubly noncentral* $\mathcal{L}(F_I) = F(s, n-p-s, \lambda_1, \lambda_2)$ *with*

i. $\lambda_1 = (\boldsymbol{G}_{21}\gamma_1 + \boldsymbol{G}_{22}\boldsymbol{\delta})'\boldsymbol{G}_{22}^{-1}(\boldsymbol{G}_{21}\gamma_1 + \boldsymbol{G}_{22}\boldsymbol{\delta})$;
ii. $\lambda_2 = \gamma_2'(\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma_2$.

Proof. The proof is given in Appendix A for Lemma A.4.

Under shifts $\gamma \neq \boldsymbol{0}$, an innovation of the present study, distributions of F_I emerge as doubly noncentral, or as singly noncentral in either the numerator or denominator, depending on γ. Generally speaking, the occurrence $\gamma \in \mathcal{R}(\boldsymbol{H}_r)$ gives $F(s, \nu, \lambda_1, 0)$, whereas $\gamma \in \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r)$ gives $F(s, \nu, \lambda_1, \lambda_2)$, for various values of λ_1. Details follow as in Lemma A.4, where we subsequently identify $\boldsymbol{G}_{121} = \boldsymbol{G}_{21}\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$.

Theorem 3.2. *Consider* $\mathcal{L}(F_I) = F(s, n-p-s, \lambda_1, \lambda_2)$ *as in Theorem 3.1 under Assumptions A. Essential properties emerge under further structures as follows:*

i. If $\gamma \in \mathcal{R}(\boldsymbol{H}_r)$, *then* $\mathcal{L}(F_I) = F(s, \nu, \lambda_1, 0)$ *with* λ_1 *as in Theorem 3.1, and at* $\boldsymbol{\delta} = \boldsymbol{0}$, *then* $\lambda_1 = \gamma'\boldsymbol{G}_{121}\gamma$.
ii. If $\gamma \in \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r)$, *then* $\mathcal{L}(F_I) = F(s, \nu, \lambda_1, \lambda_2)$ *with* $\lambda_1 = \boldsymbol{\delta}'(\boldsymbol{I}_s - \boldsymbol{H}_{II})\boldsymbol{\delta}$ *and* $\lambda_2 = \gamma'(\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma > 0$; *and at* $\boldsymbol{\delta} = \boldsymbol{0}$, $\mathcal{L}(F_I) = F(s, \nu, 0, \lambda_2)$.
iii. If $s < r$, $\boldsymbol{\delta} = \boldsymbol{0}$, *and* $\gamma \in \mathcal{R}(\boldsymbol{G}_{121})$, *then* $\mathcal{L}(F_I) = F(s, \nu, \lambda_1, 0)$ *with* $\lambda_1 = \gamma'\boldsymbol{G}_{121}\gamma > 0$; *and*
iv. If $s < r$, $\boldsymbol{\delta} = \boldsymbol{0}$, *and* $\gamma \in \mathcal{N}(\boldsymbol{G}_{121})$, *then* $\mathcal{L}(F_I) = F(s, \nu, 0, \lambda_2)$ *with* $\lambda_2 = \gamma'(\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma > 0$.
v. For $\gamma \in \mathcal{R}(\boldsymbol{H}_r)$, *the "Regression effect,"* $\boldsymbol{\delta} = \boldsymbol{R}'\gamma$ *with* $\boldsymbol{R}' = -\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$, *yields* $\lambda_1 = 0$.

Proof. As before, $r + s = n$ and $r > p$. Take λ_1 and λ_2 as in Theorem 3.1; then $\gamma \in \mathcal{N}(\boldsymbol{I}_r - \boldsymbol{H}_r)$ implies $(\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma = \boldsymbol{0}$ and $\lambda_2 = 0$ as in conclusion (i); otherwise, $\gamma \in \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r)$ implies $\lambda_2 > 0$ as in (ii). Because $(\boldsymbol{I}_r - \boldsymbol{H}_r)$ is $(r \times r)$ idempotent of rank $t = (r - p)$, its spectral decomposition is $(\boldsymbol{I}_r - \boldsymbol{H}_r) = \boldsymbol{P}_1\boldsymbol{P}_1'$ with $\boldsymbol{P}_1 = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_t]$ as eigenvectors corresponding to its t unit eigenvalues, such that $\boldsymbol{P}_1\boldsymbol{P}_1'$ is idempotent and $\boldsymbol{P}_1'\boldsymbol{P}_1 = I_t$. From (A.5) it follows that $\boldsymbol{H}_{I0}\boldsymbol{P}_1\boldsymbol{P}_1' = \boldsymbol{0}$, so that $\boldsymbol{H}_{I0}\boldsymbol{P}_1\boldsymbol{P}_1'\boldsymbol{P}_1 = \boldsymbol{H}_{I0}\boldsymbol{P}_1 = \boldsymbol{0}$. This in turn implies that $-\boldsymbol{H}_{I0}\gamma = \boldsymbol{G}_{21}\gamma = \boldsymbol{0}$ for any $\gamma = \sum_{i=1}^{t} a_i\boldsymbol{p}_i \in S_p(\boldsymbol{P}_1) = \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r)$, to give λ_1 as in conclusion (ii). In terms of $\boldsymbol{G} = [\boldsymbol{G}_{ij}]$, (A.5) further implies that $(\boldsymbol{G}_{121})(\boldsymbol{I}_r - \boldsymbol{H}_r) = \boldsymbol{0}$, and, since the factors commute, infer that $\mathcal{R}(\boldsymbol{G}_{121}) \subset \mathcal{N}(I_r - \boldsymbol{H}_r)$ and $\mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r) \subset \mathcal{N}(\boldsymbol{G}_{121})$. That $\mathcal{N}(\boldsymbol{G}_{121})$ is nonempty follows since $\boldsymbol{G}_{121}$ is $(r \times r)$ of rank $s < r$, to establish conclusion (iv). Conclusion (v) follows directly from (i), to complete our proof.

An immediate consequence is that, for $\boldsymbol{Y}_I$ not outlying, then $\mathcal{L}(F_I)$ nonetheless may be doubly noncentral whenever $\gamma \neq \boldsymbol{0}$; *i.e.* for $\boldsymbol{\delta} = \boldsymbol{0}$, $\mathcal{L}(F_I|\gamma) = F(s, \nu, \lambda_1, \lambda_2)$. Despite such irregularities, stochastic bounds may be constructed independently of the particular value of γ, depending only on its length and the structure of the design. To these ends, recall that *cdf*s $F(\cdot)$ and $G(\cdot)$ on $\mathbb{R}^1$ are ordered stochastically as $F \succeq_{st} G$ if and only if $F(u) \geq G(u)$ for every $u \in \mathbb{R}^1$. A stochastic envelope for $\mathcal{L}(F_I|\gamma)$ may be constructed as follows.

Theorem 3.3. *Consider* $\mathcal{L}(F_I|\gamma) = F(s, \nu, \lambda_1, \lambda_2)$ *under Assumptions A with* $\nu = n - p - s$, $\boldsymbol{\delta} = \mathbf{0}$, $\lambda_1 = \gamma' \boldsymbol{G}_{121}\gamma$ *and* $\lambda_2 = \gamma'(\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma$. *Let* $\boldsymbol{G}_{121} = \sum_{i=1}^{s} \xi_i \boldsymbol{q}_i \boldsymbol{q}_i'$ *be its spectral decomposition of rank s, with* $\{\xi_1 \geq \xi_2 \geq \ldots \geq \xi_s > 0\}$. *Then*

$$F(s, \nu, \lambda_m, \gamma'\gamma) \preceq_{st} F(s, \nu, \lambda_m, \lambda_2) \preceq_{st} F(s, \nu, \lambda_1, \lambda_2)$$

$$\preceq_{st} F(s, \nu, \lambda_M, \lambda_2) \preceq_{st} F(s, \nu, \lambda_M, 0),$$

where $\lambda_m = \gamma'\gamma\, \xi_s$ *and* $\lambda_M = \gamma'\gamma\, \xi_1$.

Proof. Variational properties of the Rayleigh quotient $\xi_s \leq \gamma' \boldsymbol{G}_{121}\gamma / \gamma'\gamma \leq \xi_1$, give $\gamma'\gamma\, \xi_s \leq \lambda_1 \leq \gamma'\gamma\, \xi_1$. Moreover, since $(\boldsymbol{I}_r - \boldsymbol{H}_r)$ is $(r \times r)$ idempotent of rank $t = (r - p)$ with spectral decomposition $\boldsymbol{P}_1\boldsymbol{P}_1'$ having eigenvalues in $\{0, 1\}$, the Rayleigh quotient again ensures that $\{0 \leq \gamma'(I_r - H_r)\gamma \leq \gamma'\gamma\}$. The string of inequalities now follows since $F(s, \nu, \lambda_1, \lambda_2)$ is stochastically increasing in λ_1 and stochastically decreasing in λ_2.

Irreparable defects are wrought in the actual workings of $\{F_I; I \in \mathcal{J}\}$ when there are shifts at nondeleted rows with $\gamma \neq 0$. Confounding effects on outlier detection are the masking and swamping of outliers as in section 2.1. Indeed, masking "is an important problem in influence analysis which deserves further study" (Hoaglin and Kempthorne 1986, 410). Our tools offer further insight. A survey follows, where elements of $\boldsymbol{G} = [\boldsymbol{G}_{ij}]$, as in Eq. (1) and Theorems 3.1 and 3.2, often are reexpressed in terms of $(\boldsymbol{I}_n - \boldsymbol{H}_n)$.

Properties of F_I.

P1. Convention claims to test $H_0 : \boldsymbol{\delta} = \mathbf{0}$ *vs.* $H_1 : \boldsymbol{\delta} \neq \mathbf{0}$ at level α using $\{F_I > c_\alpha\}$ with c_α as the upper critical value from $F(s, n-p-s, 0, 0)$.

a. For $\gamma \in \mathcal{N}(\boldsymbol{I}_r - \boldsymbol{H}_r)$, this executes a factual test for the "Regression effect" $H_0 : \boldsymbol{\delta} = \boldsymbol{R}'\gamma$ *vs.* $H_1 : \boldsymbol{\delta} \neq \boldsymbol{R}'\gamma$, from Theorem 3.2(v), with $\lambda_1 = (-\boldsymbol{R}'\gamma + \boldsymbol{\delta})'(\boldsymbol{I}_s - \boldsymbol{H}_{II})(-\boldsymbol{R}'\gamma + \boldsymbol{\delta})$ and "coefficient matrix" $\boldsymbol{R}' = (\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}\boldsymbol{H}_{I0}$.

b. In the event that (i) $\boldsymbol{\delta} \approx \boldsymbol{R}'\gamma$ holds approximately, (ii) λ_1 is small, and (iii) $F_I < c_\alpha$, then the conventional but mistaken inference that $\boldsymbol{\delta} = \mathbf{0}$ would serve to mask $\boldsymbol{\delta} \neq \mathbf{0}$ through $\boldsymbol{\gamma} \neq \mathbf{0}$.

c. A test for $H_0 : \boldsymbol{\delta} = \mathbf{0}$ *vs.* $H_1 : \boldsymbol{\delta} \neq \mathbf{0}$, as usually intended, is afforded by Theorem 3.2(ii) at $\gamma \in \mathcal{R}(\boldsymbol{I}_r - \boldsymbol{H}_r)$, but at the expense of $\lambda_2 > 0$. The resulting test is conservative, with $P(F_I > c_\alpha|H_0) < \alpha$ since the actual null distribution $F(s, n-p-s, 0, \lambda_2)$ is stochastically smaller than the nominal $F(s, n-p-s, 0, 0)$, as appropriate for $\gamma = \mathbf{0}$.

d. A "significant" outcome $F_I > c_\alpha$ may be misattributed to $H_1 : \boldsymbol{\delta} \neq \mathbf{0}$, when in fact $\boldsymbol{\delta} = \mathbf{0}$ but $\gamma \in \mathcal{R}(\boldsymbol{H}_{0I}(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}\boldsymbol{H}_{I0})$, as seen from Theorem 3.2(iii). This in turn effects a venue for the swamping of $\boldsymbol{\delta} = \mathbf{0}$ by $\boldsymbol{\gamma} \neq \mathbf{0}$.

P2. With $\boldsymbol{\delta} = \mathbf{0}$, Theorem 3.2(i) shows that F_I can have large values with $\mathcal{L}(F_I) = F(s, n-p-s, \lambda_1, 0)$ skewed to the right of the null distribution $F(s, n-p-s, 0, 0)$, supporting the false conclusion that $\boldsymbol{\delta} \neq \mathbf{0}$. Similarly, with $\boldsymbol{\delta} = \mathbf{0}$, Theorem 3.2(iv) shows that F_I can have small values with $\mathcal{L}(F_I) = F(s, n-p-s, 0, \lambda_2)$ skewed to the left of the null distribution. Additionally, with $\boldsymbol{\delta} \neq \mathbf{0}$, intermediate values of λ_1 and λ_2 can have $\mathcal{L}(F_I)$ approximating the null distribution.

P3. Several influence diagnostics are seen to decrease with increasing F_I through $S_I^2/S^2 = (n-p)/[sF_I + (n-p-s)]$; see Table 1, for example. Whatever the cutoff rules, these are altered irrevocably by anomalies in F_I owing to outliers. Accordingly, influence ascribed to subsets of observations $\boldsymbol{Y_I}$ in $\boldsymbol{I}$, using diagnostics in Table 1, may actually be the result of unexamined outliers in the nondeleted rows J, rather than of some vaguely posed notion of "influence" from the literature.

P4. Conversely, data points having outliers but deemed *noninfluential* might have been declared *influential* had there been no outliers. That outliers and influence, in its many guises, are irrevocably entangled clearly thwarts much of the meaning ascribed historically to influence diagnostics.

P5. The condition $\gamma \in \mathcal{N}(\boldsymbol{I}_r - \boldsymbol{H}_r)$ asserts that γ is invariant under projection $\boldsymbol{P}_X(\gamma) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\gamma = \gamma$ onto the space of the regressors $\boldsymbol{X}$. Since $-\boldsymbol{G}_{21} = \boldsymbol{Z}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}'$ and $Z(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}'P_X(\gamma) = \boldsymbol{Z}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}'\gamma$, it follows that λ_1 of Theorem 3.2(i), together with the implied tests, are invariant under $\gamma \to P_X(\gamma)$.

These issues are best exemplified numerically. Case studies to these ends are undertaken next.

4. Case Studies

Properties of $\mathcal{L}(F_I)$, culminating in section 3, are characteristic of $\boldsymbol{X}_0$ independently of $\boldsymbol{Y}_0$, and often may be accessed before undertaking an experiment. To identify subsets of outlying or influential observations, users typically examine F_I statistics, or influence diagnostics, for all or portions of subsets of size s comprising an ensemble $\{F_I; I \in \mathcal{J}\}$. Numerous studies, some by simulation, have examined empirical outcomes of these diagnostics based on $\boldsymbol{Y}_0$ in random experiments. To the contrary, our studies are probabilistic as noted, examining the driving forces behind those outcomes, namely, their doubly noncentral distributions themselves. To enable a full but concise accounting, to include manageable intermediate displays, we first consider a small data set in a single regressor. A more comprehensive example from the literature then suppresses cumbersome intermediate details. Computations utilize the Minitab and Maple software packages, and without comment we report noncentrality parameters in units of σ^2, as if $\sigma^2 = 1.0$, but reinstating other values on occasion as needed.

4.1. *Case Study 1*

Take $\{Y_i = 10 + 3X_i - 2X_i^2 + \varepsilon_i;\ 1 \leq i \leq 7\}$, with design points augmenting a Central Composite Design (*CCD*) of Box and Wilson (1951). The first two rows of Table 2 comprise $\{(Y_i, X_i);\ 1 \leq i \leq 7\}$, where $\{\varepsilon_i; 1 \leq i \leq 7\}$ are generated as $\mathcal{L}(\varepsilon_i) = N(0, 0.5)$. Here, $[1, X_i, X_i^2]$ comprise the rows of $\boldsymbol{X}_0(7 \times 3)$, and leverages, as diagonals of $\boldsymbol{H}_7 = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'$, are listed as $\{h_{ii}; 1 \leq i \leq 7\}$ in Table 2.

We reiterate conventional properties. Given an excluded subset I having $\boldsymbol{Y}_I \to \boldsymbol{Y}_I + \boldsymbol{\delta}$, and no outliers elsewhere, the R–Fisher statistic $F_I = e_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I/sS_I^2$, together with

Table 2
Values $\{(Y_i, X_i, h_{ii}); 1 \leq i \leq 7\}$ comprising responses, design points, and corresponding leverages for a modified single regressor *CCD*

i	1	2	3	4	5	6	7
Y_i	−2.00020	0.50754	6.13322	10.86978	10.15305	9.91780	9.48890
X_i	−1.73205	−1.50000	−1.00000	0.00000	1.00000	1.50000	1.73205
h_{ii}	0.57051	0.35138	0.30237	0.55149	0.30237	0.35138	0.57051

the rejection rule $F_I > c_\alpha$, tests $H_0 : \boldsymbol{\delta} = \mathbf{0}$ against $H_1 : \boldsymbol{\delta} \neq \mathbf{0}$ at level α. For the single shift $\boldsymbol{\delta}$ at $\boldsymbol{Y}_I$ and subsets of dimension $s = 2$, we have $\mathcal{L}(F_I|\,\boldsymbol{\delta}) = F(2, 2, \lambda_1, 0)$, since $n - p - s = 7 - 3 - 2 = 2$ and $\lambda_1 = \delta'(\boldsymbol{I}_s - \boldsymbol{H}_{II})\boldsymbol{\delta}$. Accordingly, under Gaussian errors the critical value at $\alpha = 0.05$ is $c_\alpha = 19.00$ from $F(2, 2, 0, 0)$, and the power at $\boldsymbol{\delta} \neq \mathbf{0}$ is $\beta_P(\lambda_1) = P(F_I > 19.00)$ under $\mathcal{L}(F_I) = F(2, 2, \lambda_1, 0)$.

Remark 4.1. *Recall that* $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ *increases stochastically with increasing* λ_1, *and decreases stochastically with increasing* λ_2. *Nonetheless, a given* $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ *may be compared against* $F(\nu_1, \nu_2, 0, 0)$ *as reference. This follows on computing* $\beta = P(F_I > c_\alpha)$ *where* $\mathcal{L}(F_I) = F(\nu_1, \nu_2, \lambda_1, \lambda_2)$. *Then* $\beta < \alpha$ *identifies* $\mathcal{L}(F_I)$ *as skewed to the left of* $F(\nu_1, \nu_2, 0, 0)$, *whereas* $\beta > \alpha$ *identifies* $\mathcal{L}(F_I)$ *as skewed to its right. Further usage is explained later in context.*

For subsequent reference we list the matrix

$$\boldsymbol{I}_7 - \boldsymbol{H}_7 = \begin{bmatrix} 0.42949 & -0.42290 & -0.15913 & 0.13501 & 0.11794 & -0.00730 & -0.09311 \\ -0.42290 & 0.64862 & -0.21671 & -0.03661 & 0.02449 & 0.01042 & -0.00730 \\ -0.15913 & -0.21671 & 0.69763 & -0.32265 & -0.14157 & 0.02449 & 0.11794 \\ 0.13501 & -0.03661 & -0.32265 & 0.44851 & -0.32265 & -0.03661 & 0.13501 \\ 0.11794 & 0.02449 & -0.14157 & -0.32265 & 0.69763 & -0.21671 & -0.15913 \\ -0.00730 & 0.01042 & 0.02449 & -0.03661 & -0.21671 & 0.64862 & -0.42290 \\ -0.09311 & -0.00730 & 0.11794 & 0.13501 & -0.15913 & -0.42290 & 0.42949 \end{bmatrix}$$

Example 1. To fix ideas and to illustrate the computations explicitly, we delete rows $I = (6, 7)$ from $[\boldsymbol{Y}_0, \boldsymbol{X}_0, \varepsilon_0]$, retaining other elements in their natural order, to be identified as $ID(12345 \cdot 67)$. The shift vector $\boldsymbol{\omega}$ is given by $\boldsymbol{E}(\boldsymbol{Y}_0) = \boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\omega}$, partitioned as $\boldsymbol{\omega}' = [\boldsymbol{\gamma}', \boldsymbol{\delta}']$. If there are no shifts other than $\boldsymbol{\delta}' = [\delta_6, \delta_7]$, then $F_{(6,7)}$ has distribution $\mathcal{L}(F_{(6,7)}|\delta_6, \delta_7) = F(2, 2, \lambda_1, 0)$ as noted in Theorem 3.1, with $\lambda_1 = [\delta_6, \delta_7]\boldsymbol{G}_{22}[\delta_6, \delta_7]'$ and $\boldsymbol{G}_{22}$ as the lower right (2×2) principal block of $(\boldsymbol{I}_7 - \boldsymbol{H}_7)$.

However, if shifts γ occur at nondeleted rows, the curious fact from section 3 is that $\mathcal{L}(F_{(6,7)})$ is doubly noncentral, even if (Y_6, Y_7) are not outlying with $[\delta_6, \delta_7] = [0, 0]$. For case (i) $\boldsymbol{\omega}' = [\mathbf{0}', \boldsymbol{\delta}']$ with $\mathcal{L}(F_{(6,7)})$ as a singly noncentral F-distribution, or case (ii) $\omega' = [\gamma', \delta']$ with $\mathcal{L}(F_{(6,7)})$ as doubly noncentral; in either case $F_{(6,7)}$ is compared to $F(2, 2, 0, 0)$ with critical value $c_\alpha = 19.00$.

To examine this further, as in section 3 we require the matrix $\boldsymbol{G}_{121} = \boldsymbol{G}_{21}\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$ as

$$G_{121} = \begin{bmatrix} 0.06238 & 0.00021 & -0.08551 & -0.07033 & 0.19367 \\ 0.00021 & 0.00017 & -0.00005 & -0.00096 & -0.00249 \\ -0.08551 & -0.00005 & 0.11754 & 0.09541 & -0.26982 \\ -0.07033 & -0.00096 & 0.09541 & 0.08240 & -0.20492 \\ 0.19367 & -0.00249 & -0.26982 & -0.20492 & 0.65940 \end{bmatrix} \tag{2}$$

together with $\boldsymbol{H}_5 = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and

$$I_5 - H_5 = \begin{bmatrix} 0.36711 & -0.42311 & -0.07362 & 0.20534 & -0.07573 \\ -0.42311 & 0.64845 & -0.21666 & -0.03566 & 0.02698 \\ -0.07362 & -0.21666 & 0.58009 & -0.41806 & 0.12825 \\ 0.20534 & -0.03566 & -0.41806 & 0.36611 & -0.11773 \\ -0.07573 & 0.02698 & 0.12825 & -0.11773 & 0.03823 \end{bmatrix} \tag{3}$$

To continue, we next examine how pairs of shifts (γ_i, γ_j) at nondeleted rows $J = (i, j)$ serve to induce noncentralities in $\mathcal{L}(F_{(6,7)} \mid \gamma_i, \gamma_j) = F(2, 2, \lambda_1, \lambda_2)$, even when (Y_6, Y_7) remain unshifted and $[\delta_6, \delta_7] = [0, 0]$. Recall from section 3 that $\omega' = [\gamma', \delta'] \in \mathbb{R}^7$ with $\gamma \in \mathbb{R}^5$ and $\delta \in \mathbb{R}^2$. First consider $\{Y_1 \to Y_1 + \gamma_1, Y_2 \to Y_2 + \gamma_2\}$. Then Theorem 3.1(i) applies to give λ_1 with $\gamma' = [\gamma_1, \gamma_2, 0, 0, 0]$ and $\delta' = [0, 0]$. Similarly, Theorem 3.1(ii) applies to give λ_2. Both are quadratic forms in principal blocks $\boldsymbol{G}_{[12]}$ of $\boldsymbol{G}_{121}$ and $\boldsymbol{E}_{[12]}$ of $\boldsymbol{E} = (\boldsymbol{I}_5 - \boldsymbol{H}_5)$, namely, $\lambda_1 = [\gamma_1, \gamma_2]\boldsymbol{G}_{[12]}[\gamma_1, \gamma_2]'$ and $\lambda_2 = [\gamma_1, \gamma_2]\boldsymbol{E}_{[12]}[\gamma_1, \gamma_2]'$. These matrix pairs are listed explicitly in Table 3, corresponding to outlier pairs $\{(\gamma_1, \gamma_2), (\gamma_3, \gamma_4)\}$, found as principal blocks $\boldsymbol{G}_{[ij]}$ of $\boldsymbol{G}_{121}$ and $\boldsymbol{E}_{[ij]}$ of $\boldsymbol{E} = (\boldsymbol{I}_5 - \boldsymbol{H}_5)$.

Remark 4.2. *It is noteworthy that all submatrices in Table 3 are positive definite. Accordingly, apart from* $[\gamma_i, \gamma_j] = [0, 0]$, *all noncentralities in* $\mathcal{L}(F_{(6,7)} \mid \gamma_i, \gamma_j)$, *as induced through* $\{Y_i \to Y_i + \gamma_i, Y_j \to Y_j + \gamma_j; (i, j) \notin \{6, 7\}\}$, *will be positive.*

For fixed pairs (γ_i, γ_j) the computations proceed from Eqs. (2) and (3) as in

$$\lambda_1(\gamma_1, \gamma_2) = 0.06238\,\gamma_1^2 + 2(0.00021)\gamma_1\gamma_2 + 0.00017\,\gamma_2^2 \tag{4}$$

$$\lambda_2(\gamma_1, \gamma_2) = 0.36711\,\gamma_1^2 - 2(0.42311)\gamma_1\gamma_2 + 0.64845\,\gamma_2^2. \tag{5}$$

These noncentrality parameters deserve further study. For example, with $\gamma_1 = -\gamma_2$ we have $\mathcal{L}(F_{(6,7)}|\gamma_1 = -\gamma_2) = F(2, 2, 0.06213\,\gamma_1^2, 1.86178\,\gamma_1^2)$; further computations proceed similarly. Table 3 reports the special case $\gamma_i = \gamma_j$ for nondeleted rows $J = (i, j)$. We set the common shifts by $[\gamma_i, \gamma_j] = \gamma_{ij}[1, 1] \in \mathbb{R}^2$. For varying $J = (i, j)$, the noncentrality parameters for the distributions $\mathcal{L}(F_{(6,7)}|[\delta_6, \delta_7] = [0, 0])$ are shown in Table 3. Since (λ_1, λ_2) depend on γ_{ij}^2, λ_1 and λ_2 are both symmetric under reflection of γ_{ij} about zero. The β values have been computed for $\gamma_{ij} = \pm 2$, and although there are no shifts at the deleted rows $I = (6, 7)$ as $[\delta_6, \delta_7] = [0, 0]$, the values of $\beta = Pr(F_{(6,7)} > 19.00)$ are shown to vary from 0.01370 to 0.07055, demonstrating skewness of $\mathcal{L}(F_{(6,7)}|[\delta_6, \delta_7] = [0, 0])$ in both the left and right directions in comparison with $F(2, 2, 0, 0)$.

Table 3

Principal blocks $G_{[ij]}$ of G_{121} and blocks $E_{[ij]}$ of $E = (I_5 - H_5)$, as matrices of quadratic forms determining noncentralities for $\mathcal{L}\left(F_{(6,7)} \mid \gamma_i, \gamma_j\right) = F(2, 2, \lambda_1(i,j), \lambda_2(i,j))$, with $[\gamma_6, \gamma_7] = [0, 0]$, together with noncentralities of type $F_{(6,7)}\left(2, 2, a\,\gamma_{ij}^2, b\,\gamma_{ij}^2\right)$ when $[\gamma_i, \gamma_j] = \gamma_{ij}[1, 1]$, and β values for $\gamma_{ij} = \pm 2$ and $c_\alpha = 19.00$

ID	$G_{[ij]}$	$E_{[ij]}$
$\begin{bmatrix}\gamma_1\\ \gamma_2\end{bmatrix}$	$\begin{bmatrix}0.06238 & 0.00021\\ 0.00021 & 0.00017\end{bmatrix}$	$\begin{bmatrix}0.36711 & -0.42311\\ -0.42311 & 0.64845\end{bmatrix}$
$F_{(6,7)}(2, 2, 0.0630\,\gamma_{12}^2, 0.1693\,\gamma_{12}^2)$; $\beta = 0.04065$		
$\begin{bmatrix}\gamma_3\\ \gamma_4\end{bmatrix}$	$\begin{bmatrix}0.11754 & 0.09541\\ 0.09541 & 0.08240\end{bmatrix}$	$\begin{bmatrix}0.58009 & -0.41806\\ -0.41806 & 0.36611\end{bmatrix}$
$F_{(6,7)}(2, 2, 0.3908\,\gamma_{34}^2, 0.1101\,\gamma_{34}^2)$; $\beta = 0.07055$		

Table 4

Shifts $\{\gamma_A, \gamma_B, \gamma_C\}$ from (6); their projections $\gamma = \gamma_1 + \gamma_2$; distributions $\mathcal{L}\left(F_{(6,7)}|\gamma\right) = F\,(2,\,2,\,\lambda_1,\,\lambda_2)$ under $Y \to Y + \gamma$ for $ID = (12345{\cdot}67)$ with $\delta_6 = \delta_7 = 0$; and corresponding β's at $c_\alpha = 19.00$

Case A		Case B		Case C	
γ_{A1}	γ_{A2}	γ_{B1}	γ_{B2}	γ_{C1}	γ_{C2}
0.632886	0.367114	0.078054	0.921946	0.564463	0.435537
0.423110	−0.423110	0.323877	−1.323877	0.225587	−1.225587
0.073619	−0.073619	0.694931	0.305069	−0.269444	1.269444
−0.205341	0.205341	0.810952	0.189048	−0.339090	−0.660910
0.075727	−0.075727	0.092186	−0.092186	0.818484	0.181516
$F(2, 2, 0.0624, 0.3671)$		$F(2, 2, 0.1432, 2.7399)$		$F(2, 2, 0.9610, 3.7730)$	
$\beta = 0.04325$		$\beta = 0.01460$		$\beta = 0.01253$	
$\theta_1 = 37.3°, \theta_2 = 52.7°$		$\theta_1 = 55.9°, \theta_2 = 31.1°$		$\theta_1 = 60.3°, \theta_2 = 29.7°$	

Decompostion of $\boldsymbol{\gamma}$. Definition 2.3.1 has introduced the decomposition of $\gamma \in \mathbb{R}^r$ as $\gamma_1 = \boldsymbol{H}_r\gamma$ in the "regressor" space and $\gamma_2 = (\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma$ in the "error" space of $\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$. To illustrate, consider

$$\gamma_A' = [1, 0, 0, 0, 0],\ \gamma_B' = [1, -1, 1, 1, 0],\ \gamma_C' = [1, -1, 1, -1, 1] \tag{6}$$

which determine the three cases A, B, and C in Table 4. Details, as listed in Table 4, include projections (γ_1, γ_2) for each γ, together with $\mathcal{L}(F_{(6,7)}|\gamma) = F(2, 2, \lambda_1, \lambda_2)$ under $\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{\gamma}$, and corresponding β's. Recall from section 3 for $\boldsymbol{\delta} = \mathbf{0}$ that λ_1 and λ_2 are quadratic forms in γ_1 and γ_2, namely $\lambda_1 = \gamma_1' G_{121}\gamma_1$ and $\lambda_2 = \gamma_2'(\boldsymbol{I}_5 - \boldsymbol{H}_5)\gamma_2$ with the matrices from Eqs. (2) and (3).

The angles (θ_1, θ_2) between each vector γ and its projections (γ_1, γ_2), as in Definition 2.3.1(ii), give a measure of the propensity for the shift vector $\gamma \in \mathbb{R}^r$ to be in the "regressor" space when θ_1 is small, or in the "error" space when θ_1 is large, equivalently when θ_2 is small. Table 4 reports for Case C that $\theta_2 = 29.7°$ favoring the "error" space, together with $\beta = 0.0125$, less than the nominal value of 5%. Here, $F(2, 2, 0.9610, 3.7730)$ is skewed to the left of $F(2, 2, 0, 0)$.

Remark 4.3. *Theorem 3.1(i) gives* $\lambda_1 = \gamma'\boldsymbol{G}_{121}\boldsymbol{\gamma} + 2\boldsymbol{\delta}'\boldsymbol{G}_{21}\boldsymbol{\gamma} + \boldsymbol{\delta}'\boldsymbol{G}_{22}\boldsymbol{\delta}$. *To control the number of parameters,* $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ *may be varied along the equiangular lines* $\gamma = c\mathbf{1}_r \in \mathbb{R}^r$ *and* $\boldsymbol{\delta} = d\mathbf{1}_s \in \mathbb{R}^s$. *In this case, the restricted* λ_1 *is* $\lambda_1(c, d) = c^2\mathbf{1}'_r\boldsymbol{G}_{121}\mathbf{1}_r + 2cd\,\mathbf{1}'_s\boldsymbol{G}_{21}\mathbf{1}_r + d^2\mathbf{1}'_s\boldsymbol{G}_{22}\mathbf{1}_s$.

We continue Case A of Table 4. on adding shifts $\{Y_6 \to Y_6 + \delta, Y_7 \to Y_7 + \delta\}$ and scaling $c\gamma_A \to \gamma$. In keeping with Remark 4.3, on varying $\delta' = d[1, 1]$, λ_1 becomes $\lambda_1(c, d) = 0.062378\,c^2 - 0.200822\,cd + 0.232305\,d^2$. Accordingly, for Case A in the final row of Table 4, the listed distribution instead becomes $\mathcal{L}(F_{(6,7)}\,|\,\gamma, \delta) = F(2, 2, \lambda_1(c, d), 0.36711\,c^2)$, where the denominator remains unchanged apart from scaling by c^2.

To continue, Theorem 3.2(v) acknowledges possible "regression effects" relating δ to $\gamma = c\gamma_A$; in particular, $\boldsymbol{\delta} = \boldsymbol{R}'\gamma$ with $\boldsymbol{R}' = -\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$. Here we have

$$\boldsymbol{R}' = \begin{bmatrix} 0.42627 & -0.01391 & -0.60558 & -0.41484 & 1.60806 \\ 0.63652 & 0.00331 & -0.87089 & -0.72282 & 1.95389 \end{bmatrix},$$

so that $\boldsymbol{R}'\gamma = c[0.426274, 0.636521]'$. Substituting this for $\boldsymbol{\delta}$ gives $\mathcal{L}(F_{(6,7)} | \boldsymbol{\delta} = \boldsymbol{R}'\gamma) = F(2, 2, 0, 0.36711\,c^2)$. An outcome $F_{(6,7)}$ from this distribution would be likely to mask that both $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are outlying.

Example 2. Let δ denote a generic shift in $\boldsymbol{\omega}' = [\boldsymbol{\gamma}', \boldsymbol{\delta}']$ for either deleted rows in I or nondeleted rows in J. Consider outlier shifts $\{Y_1 \to Y_1 + \delta, Y_2 \to Y_2 + \delta\}$ exclusively at $I = \{1, 2\}$ and $ID(34567 \cdot 12)$. The distribution $\mathcal{L}(F_{(1,2)}|\delta)$ is $F(2, 2, \lambda_1, 0)$ with $\lambda_1 = 0.23231\,\delta^2$ and with power $\beta_P(\delta) = 0.07182$ at $\delta = \pm 2$ in testing $\{H_0 : [\delta_1, \delta_2] = [0, 0]\}$ against $\{H_1 : [\delta_1, \delta_2] = \delta\,[1, 1]; \delta \neq 0\}$.

To seek outlying subsets, users typically examine values $\{F_I; I \in \mathcal{J}\}$ for selected subsets of size s. In keeping with our probabilistic focus, Table 5. exhibits for $s = 2$ the driving forces behind empirical outcomes that the user might "see:" Specifically, doubly noncentral distributions $\mathcal{L}(F_{(i,j)})$ for the deleted pair $I = (i, j)$ under $ID(abcde \cdot ij)$, as induced by $\{Y_1 \to Y_1 + \delta, Y_2 \to Y_2 + \delta\}$. These deleted pairs are varied systematically; no pairs other than (Y_1, Y_2) have been shifted. Noncentralities are reported in Table 5, together with probabilities $\beta(\delta)$ at $\delta = \pm 2$, where the left-hand side of the table has $J = (1, 2)$ and I varying over $\{I(i, j); \{i, j\} \notin \{1, 2\}\}$, whereas the right-hand side of the table has one shift from $(1, 2)$ in J, and the other in I.

The reader is reminded that $ID(12567 \cdot 34)$ entails reordering $\boldsymbol{X}_0$ in keeping with Eq. (1), and recomputing $\boldsymbol{G}_{121}$ and $(\boldsymbol{I}_5 - \boldsymbol{H}_5)$ anew, and similarly for each case as listed. Moreover, ω of section 3 is the same for each row in Table 5, so that $\lambda_3 = \lambda_1 + \lambda_2$ is fixed; and the cells are shown in descending order of λ_1 and of β. These distributions in turn generate empirical outcomes as observed values of $\{F_{(i,j)}\}$.

Table 5
Outliers $\{Y_1 \to Y_1 + \delta, Y_2 \to Y_2 + \delta\}$ exclusively; their propagation as noncentralities across distributions $\mathcal{L}(F_{(i,j)})$ given $I = (i, j)$ under $ID(abcde \cdot ij)$; and probability $\beta = P(F_I > 19.00)$ at $\delta = \pm 2$

ID	λ_1/δ^2	λ_2/δ^2	β	ID	λ_1/δ^2	λ_2/δ^2	β
(12567·34)	0.22057	0.01124	0.06926	(14567·23)	0.22266	0.00915	0.06972
(12467·35)	0.20817	0.02365	0.06657	(24567·13)	0.21821	0.01360	0.06874
(12456 · 37)	0.20486	0.02695	0.06586	(13567·24)	0.10672	0.12510	0.04748
(12457·36)	0.20210	0.02971	0.06528	(13467·25)	0.10461	0.12720	0.04714
(12367·45)	0.11864	0.11317	0.04948	(13456·27)	0.10080	0.13101	0.04652
(12356·47)	0.06432	0.16750	0.04088	(13457·26)	0.07904	0.15278	0.04309
(12345·67)	0.06235	0.16947	0.04059	(23467·15)	0.02969	0.20212	0.03601
(12346·57)	0.04019	0.19162	0.03744	(23456·17)	0.02355	0.20827	0.03520
(12347·56)	0.03249	0.19932	0.03639	(23567·14)	0.02312	0.20869	0.03514
(12357·46)	0.02175	0.21006	0.03496	(23457·16)	0.00007	0.23174	0.03220

Each of the 10 rows in the left half of Table 5 has $\gamma' = [2, 2, 0, 0, 0]$. The decomposition $\gamma = \gamma_1 + \gamma_2$, with varying I, allows for the angle θ_1 between the vectors (γ, γ_1). As before, this measures the propensity for the shift vector γ to be in either the "regressor" space when θ_1 is small, or in the "error" space when θ_1 is large. The values in degrees for the angles θ_1 for these ten rows on the left in Table 5 are

$$\{4.3, 6.2, 6.7, 7.0, 13.8, 16.8, 16.9, 18.0, 18.4, 18.9\},$$

in agreement with the probabilities β, showing θ_1 to increase with decreasing β.

Remark 4.4. *Definition 2.1.1 anticipates a swamping curve $\beta_S(\cdot)$ as the evolution of the probability that (Y_i, Y_j) are deemed to be outlying when they are not, that is, $P(F_{(i,j)} | \delta_u, \delta_v) > c_\alpha)$; $\{i, j\} \notin \{u, v\}$, with this in correspondence with a power curve.*

Some Properties. *The left half of Table 5, having no outliers, supports the following:*

- For $I = (3, 4)$, the probability that (δ_1, δ_2) swamps (δ_3, δ_4) is $\beta_S(\delta) = P(F_{(3,4)} > c_\alpha)$; its value at $\delta = \pm 2$ is $\beta_S(\delta) = 0.06926$ from Table 5. This point on our Swamping curve is the probability $P(E_{J \geq I})$ of Definition 2.1.1, that is, that $I(3, 4)$ are deemed incorrectly to be outlying owing to $J(1, 2)$.
- For $I \in \{(3, 4), (3, 5), (3, 6), (3, 7)\}$, the swamping probabilities range over the set $\{0.06528 \leq \beta_S \leq 0.06926\}$. Of the probabilities that nonzero (δ_1, δ_2) will swamp $\{\delta_3, \delta_4, \delta_5, \delta_6, \delta_7\}$ when zero, these probabilities approximate from below the power $\beta_P(\delta) = 0.07182$ of $F_{(1,2)}$ at $\delta = \pm 2$.
- In short, the likelihoods that $(Y_3, Y_4, Y_5, Y_6, Y_7)$ are deemed to be outlying when they are not, are essentially the likelihood that (Y_1, Y_2) are correctly identified to be outlying. This in turn abrogates realistic prospects for correctly partitioning $\{Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7\}$ into outlying and non-outlying subsets.

In the right half of Table 5 either Y_1 or Y_2 is outlying, not both, to the following effects:

- Deletions $\{(1, 4), (1, 5), (1, 6), (1, 7)\}$ have β's in $\{0.03220 \leq \beta \leq 0.03601\}$, all skewed to the left of $F(2, 2, 0, 0)$.
- These appear to support that $\{Y_2 \rightarrow Y_2 + \delta\}$ serves to mask $\{Y_1 \rightarrow Y_1 + \delta\}$ when the latter is coupled on deletion with $\{Y_4, Y_5, Y_6, Y_7\}$.
- On the other hand, for deletions $\{(1, 3), (2, 3)\}$, their β's of $\{0.06874, 0.06972\}$ at $\delta = \pm 2$ are approximately the power $\beta_P(\delta) = 0.07182$ that $F_{(1,2)}$ correctly identifies (Y_1, Y_2) as outlying. At the same time, this is evidence that (δ_1, δ_2) separately swamp $\delta_3 = 0$.

Variational Results. In keeping with Remark 4.3, we have varied multiple shifts along equiangular lines. In contrast, we next demonstrate directions that outliers may take so as to maximize the F_I statistic. We consider F_I for $I(3, 4)$ when $\{\delta_3 = \delta_4 = 0\}$, but $J(1, 2)$ has nonzero (γ_1, γ_2). Accordingly, rearrange rows of $\boldsymbol{X}_0$ in the order $ID(12567 \cdot 34)$, and compute the corresponding matrices $\boldsymbol{G}_{12}\boldsymbol{G}_{22}^{-1}\boldsymbol{G}_{21}$ and $\boldsymbol{I}_5 - \boldsymbol{H}_5$. The upper left (2×2) blocks of these are matrices of quadratic forms, $\boldsymbol{H}$ determining λ_1 and $\boldsymbol{E}$ determining λ_2, as follows:

$$\boldsymbol{H} = \begin{bmatrix} 0.048899 & 0.021351 \\ 0.021351 & 0.129885 \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} 0.380953 & -0.444253 \\ -0.444253 & 0.518734 \end{bmatrix}.$$

Accordingly, we seek a direction $u' = [u_1, u_2]$ in the (γ_1, γ_2)-plane so as to maximize the ratio of noncentrality parameters, namely, $\lambda_1/\lambda_2 = \boldsymbol{u}'\boldsymbol{H}\mathbf{u}/\boldsymbol{u}'\boldsymbol{E}\mathbf{u}$. Variational properties of the Rayleigh quotient give

$$\xi_2 \leq \frac{\boldsymbol{u}'\boldsymbol{H}\mathbf{u}}{\boldsymbol{u}'\boldsymbol{E}\mathbf{u}} = \frac{\boldsymbol{v}'\boldsymbol{E}^{-\frac{1}{2}}\boldsymbol{H}\mathbf{E}^{-\frac{1}{2}}\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}} \leq \xi_1 \tag{7}$$

with $\boldsymbol{v} = \boldsymbol{E}^{\frac{1}{2}}\boldsymbol{u}$, where $\boldsymbol{E}^{-\frac{1}{2}}$ is the inverse spectral square root of $\boldsymbol{E}$, and $(\xi_2 < \xi_1)$ are roots of the determinantal equation $|\boldsymbol{H} - \xi\,\boldsymbol{E}| = 0$. The maximizing vector $\boldsymbol{v}_0$ in the (γ_1, γ_2)–plane is proportional to the eigenvector $\boldsymbol{q}_1' = [0.757044, 0.653363]$ corresponding to ξ_1, *i.e.*, $\boldsymbol{v}_0 = \gamma\boldsymbol{q}_1$. Substituting in $\lambda_1 = \boldsymbol{u}'\boldsymbol{H}\mathbf{u}$ and $\lambda_2 = \boldsymbol{u}'\boldsymbol{E}\mathbf{u}$ gives $\lambda_1 = 0.104592\,\gamma^2$ and $\lambda_2 = 0.000086\,\gamma^2$. Accordingly, $\mathcal{L}(F_I(3,4)\,|[\gamma_1, \gamma_2] = \gamma\boldsymbol{q}_1') = F(2, 2, 0.1046\,\gamma^2, 0.0001\,\gamma^2)$.

This example serves two purposes: not only extremal properties, but to show that $\{[\gamma_1, \gamma_2] = \gamma q_1'\}$ swamps $\{\delta_3 = \delta_4 = 0\}$ with probability β_S essentially the power β_P of $F_I(1,2)\,|\,\gamma_1, \gamma_2)$ to reject $H_0 : [\gamma_1, \gamma_2] = [0, 0]$ against $H_1 : [\gamma_1, \gamma_2] = \gamma q_1'$. This is because $\mathcal{L}(F_I(1,2)\,|[\gamma_1, \gamma_2] = \gamma q_1') = F(2, 2, 0.104677\,\gamma^2, 0)$ from $ID(34567 \cdot 12)$ is nearly the same distribution as $F(2, 2, 0.1046\,\gamma^2, 0.0001\,\gamma^2)$. In short, (Y_3, Y_4) will be labeled incorrectly as outlying using $F_{(3,4)}$, with essentially the power of $F_{(1,2)}$ to correctly identify (Y_1, Y_2) as outlying. Such prospects again are troublesome to users seeking to isolate outlying from non-outlying subsets.

4.2. *Case Study 2*

To illustrate subsets of size $s = 3$, we expand Table 2 to encompass the case $n = 8$, retaining quadratic responses with $p = 3$, and having design points and leverages as in the following display:

$$\begin{bmatrix} \boldsymbol{X}_i : & -1.00000 & -0.75000 & -0.50000 & -0.25000 & 0.25000 & 0.50000 & 0.75000 & 1.00000 \\ \boldsymbol{h}_{ii} : & 0.67171 & 0.28372 & 0.23915 & 0.30543 & 0.30543 & 0.23915 & 0.28372 & 0.67171 \end{bmatrix}.$$

To proceed, we adopt the shorthand $J_{j_1j_2j_3}$ as loci subject to outliers $\{(Y_{j_1}, Y_{j_2}, Y_{j_3}) \rightarrow (Y_{j_1}, Y_{j_2}, Y_{j_3}) + (\delta_{j_1}, \delta_{j_2}, \delta_{j_3})\}$, whereas $I_{i_1i_2i_3}$ indicates the deleted set $I(i_1i_2i_3)$. Details are summarized in Table 6, where, in keeping with Remark 4.3, outlying shifts are varied along the equiangular line as $[\delta_{j_1}, \delta_{j_2}, \delta_{j_3}] = c\,[1, 1, 1]$ in $\mathbb{R}^3$. For reference, note that $c_\alpha = 19.164$ from $F(3, 2, 0, 0)$.

If $\{(Y_4, Y_5, Y_6) \rightarrow (Y_4, Y_5, Y_6) + (c, c, c)\}$ comprise the only shifts, then Table 6. demonstrates the manner in which these are distributed across diagnostics for subsets $\{I_{123}, I_{237}, I_{378}\}$. Specifically, the probability that (Y_2, Y_3, Y_7) are incorrectly deemed to be outlying, that is, that $(\delta_4, \delta_5, \delta_6)$ will swamp the zero values $(\delta_2, \delta_3, \delta_7)$, is found from $\mathcal{L}(F_{237}\,|\,J_{456})$ as listed in Table 6. Swamping probabilities $\beta_S(c)$ are given in Table 7 for $c \in \{\pm 1, \pm 2, \pm 4\}$. Corresponding values for the power $\beta_P(c)$ of F_{456} to reject $H_0 : [\delta_4, \delta_5, \delta_6] = [0, 0, 0]$ against $H_1 : [\delta_4, \delta_5, \delta_6] = c\,[1, 1, 1]$ are listed in Table 7. This illustrates Definition 2.1.1 on taking $E_{J \succeq I}$ with $J = (4, 5, 6)$ and $I = (2, 3, 7)$. In short, these findings demonstrate a pitfall for users screening outcomes using F_I for subsets of size $s = 3$. Specifically, the non-outlying (Y_2, Y_3, Y_7) are about as likely to be decreed to be outlying, as are (Y_4, Y_5, Y_6) to be so decreed correctly.

Table 6
Noncentrality parameters for diagnostics $\mathcal{L}(F_{i_1i_2i_3}|J_{j_1j_2j_3})$, given outliers $(Y_{j_1}, Y_{j_2}, Y_{j_3}) + (c, c, c)$ at $J_{j_1j_2j_3}$, for varying deleted subsets $I_{i_1i_2i_3}$ and values for β with $c = \pm 2$, $c = \pm 4$ for the case $c_\alpha = 19.164$

$I_{i_1i_2i_3}$	$J_{j_1j_2j_3}$	λ_1/c^2	λ_2/c^2	$\beta\,(c = \pm 2)$	$\beta\,(c = \pm 4)$
I_{456}	J_{456}	0.753101	0.000000	0.09691	0.22418
I_{123}	J_{123}	0.361240	0.000000	0.07280	0.13796
	J_{456}	0.529242	0.223859	0.05476	0.03627
	J_{468}	0.607197	0.842415	0.01831	0.00052
I_{237}	J_{237}	1.807360	0.000000	0.15871	0.41572
	J_{456}	0.725659	0.027442	0.09051	0.18102
	J_{468}	0.642510	0.807103	0.02008	0.00071
I_{378}	J_{378}	1.294570	0.000000	0.12919	0.32932
	J_{456}	0.629605	0.123496	0.07102	0.08392
	J_{146}	0.614238	1.102040	0.01134	0.00008

Table 7
Probabilities $\beta_S(c)$ that $(\delta_4, \delta_5, \delta_6)$ will swamp the zero values $(\delta_2, \delta_3, \delta_7)$, from $\mathcal{L}(F_{237}|J_{456})$, and powers $\beta_P(c)$ from $\mathcal{L}(F_{456}|J_{456})$, giving points on the swamping and power curves at $c \in \{\pm 1, \pm 2, \pm 4\}$

Distribution	$\beta(c)$	$c = \pm 1$	$c = \pm 2$	$c = \pm 4$
$F(3, 2, 0.725659c^2, 0.0274415c^2)$	$\beta_S(c)$	0.06072	0.09051	0.18102
$F(3, 2, 0.753101c^2, 0.000000)$	$\beta_P(c)$	0.06195	0.09691	0.22418

4.3. *Case Study 3: Drill Data From Cook and Weisberg*

We revisit the classical Drill Data set from Cook and Weisberg (1982, 149), long held to be instructive and studied on various occasions since. Specifically, we show that observations traditionally alleged to be outlying may, in fact, not be outlying when shifts in nondeleted rows are taken into account.

The Drill Data are modeled as a second-order response surface in three regressors (speed of rotation, feed rate, and diameter of the drill bit) having intercept, three linear, and six second-order terms with response variable $logY$, where Y is the axial load on the drill bit during the drilling process. The sample size is $n = 31$, and the number of regressors is $p = 10$.

Using standard software such as Minitab, the single-case deletion diagnostic t_i^2 identifies three rows as potential outliers with p-values less than 0.025, namely, rows $(9, 28, 31)$ with p-values $(0.00312, 0.01100, 0.01900)$, respectively. Cook and Weisberg (1982, 152) noted, "Cases 9 and 31 have the largest potential and the largest influence." We consider these two cases in detail. The reader is reminded that the listed p-values have been computed under the assumption that there are no shifts in the non-deleted rows.

The Drill Data set was studied in Jensen and Ramirez (1996) using the subset deletion diagnostic D_I of Table 1, which is functionally equivalent to F_I as noted. They found that

D_I identified 16 pairs of potential outliers, all containing at least one row from $(9, 28, 31)$ except the pair $(5, 26)$. Although neither row 5 nor 26 was detected as outlying using t_i^2, their joint outlier diagnostic is $F_{5,26} = 14.32$ having p-value 0.000161 from $\mathcal{L}(F_{5,26}) = F(2, 19, 0, 0)$, offering compelling evidence that (Y_5, Y_{26}) are outlying. Moreover, on deleting $I = (9, 31)$ from X_0 instead, the R–Fisher diagnostic is $F_{9,31} = 7.564$ with p-value 0.0040, ostensible evidence that the subset (Y_9, Y_{31}) is outlying, in further support of the Cook and Weisberg assertion.

We continue our earlier investigation into the role of the outlying pair $(5, 26)$ using tools developed here. This in turn leads to some unprecedented conclusions.

Since rows $(9, 31)$ have been screened as outlying using t_i^2, the removal of rows $(5, 26)$ would appear not to alter the status of $(9, 31)$ as prospective outliers. In short, actual shifts in (Y_9, Y_{31}), having occurred during the course of the experiment, would remain embedded in the data, whether or not rows $(5, 26)$ are excluded in subsequent analyses. However, from the p-values reported in Table 8, this is not the case. Indeed, in the reduced model $X_{[5,26]}$, t_i^2 now reports that rows $(9, 31)$ offer negligible evidence as outlying, their p-values now being (0.3820, 0.2560), respectively. From this the claim that rows $(9, 31)$ are outlying at best is dubious.

Developments reported here help to explain this anomaly. To properly determine whether (Y_9, Y_{31}) are outlying, we must take into account that rows $(5, 26)$, nondeleted from $\boldsymbol{X}_0$ in traditional diagnostics for $(9, 31)$, have been identified as outlying. Possible consequences are twofold. First, shifts in (Y_5, Y_{26}) generally effect disturbances in all residuals serving as building blocks for t_i^2 and F_I. Second, the apparent significance of $F_{9,31}$ may owe instead to swamping by outliers at (Y_5, Y_{26}), effectively transforming the distribution $\mathcal{L}(F_{9,31}) = F(2, 19, 0, 0)$, assumed to be central under no shifts at $(9, 31)$, into a doubly noncentral $F(2, 19, \lambda_1, \lambda_2)$ to be determined. We follow these two leads in subsequent paragraphs. To these ends consider shift vectors $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']' \in \mathbb{R}^n$ as in section 2.3, where elements of γ now take common values $\{\gamma_5 = \gamma_{26} \equiv \gamma_{5,26}\}$ in keeping with Remark 4.3, with $\gamma_i = 0$ otherwise.

Effects of Shifts on Expected Residuals. Both t_i^2 and F_I tend to grow stochastically as elements of $e_0' = [e_1, \ldots, e_n]$ themselves grow stochastically in magnitude. Accordingly, it is instructive to examine disturbances in properties of the observed residuals exerted by shifts at (Y_5, Y_{26}), specifically, through their altered expectations, in contrast to non–shifted data where these expectations are all zero.

To these ends consider $E(\varepsilon_0) = \boldsymbol{\omega} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']' \in \mathbb{R}^{31}$ as in Assumptions A on reordering $\boldsymbol{X}_0$ to have the former $(9, 31)$ as its final rows. As in Definition 2.3.1 decompose $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ with $\boldsymbol{\omega}_1 = \boldsymbol{H}_n\boldsymbol{\omega}$ and $\boldsymbol{\omega}_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$, the "regressor" and "error" spaces for $\boldsymbol{Y}_0 = \boldsymbol{X}_0\boldsymbol{\beta} + \varepsilon_0$. That $E(e_0) = \omega_2$ is shown in Lemma A.2(i). Accordingly, all 31 elements of

Table 8
Selected p-values from $\mathcal{L}(t_i^2)$ for single row deletions for the full model X_0 and for the reduced model $X_{[5,26]}$ with $I = [5, 26]$ as the deleted rows

$I = (i)$	X_0	$X_{[5,26]}$
(9)	0.003120	0.382000
(28)	0.011000	0.000253
(31)	0.019000	0.256000

Table 9

Elements $\{\omega(i) = \omega_1(i) + \omega_2(i)\}$ scaled by $\gamma_{5,26}$ with $\omega_2(i)$ as expected residuals from nonzero shifts of $\gamma_{5,26}$ at $(5, 26)$, and the proportion $\|\omega_2(i)\|^2 / \|\omega_2\|^2$ with $\|\omega\|^2 = 2\gamma_{5,26}^2$, $\|\omega_1\|^2 = 1.3089\gamma_{5,26}^2$, and $\|\omega_2\|^2 = 0.6911\gamma_{5,26}^2$

i	$\omega(i)/\gamma_{5,26}$	$\omega_1(i)/\gamma_{5,26}$	$\omega_2(i)/\gamma_{5,26}$	$\|\omega_2(i)\|^2 / \|\omega_2\|^2$
5	1	0.654459	0.345541	17.28%
26	1	0.654459	0.345541	17.28%
6	0	0.135180	−0.135180	2.64%
9	0	0.415088	−0.415088	24.93%
12	0	0.248480	−0.248480	8.93%
28	0	0.068783	−0.068783	0.68%
31	0	0.248480	−0.248480	8.93%

$E(e_0) = \omega_2$ are computed in terms of $\gamma_{5,26}$ as the shift common to (Y_5, Y_{26}), and $\{\gamma_i = 0, \delta_i = 0\}$ otherwise, thereby identifying nonzero shifts only for rows $(5, 26)$ in the original array.

A partial list of the 31 cases is given in Table 9, to include $\{\omega(i), \omega_1(i), \omega_2(i)\}$ as elements of $\{\omega, \omega_1, \omega_2\}$, respectively, so that $\|\omega\|^2 = 2\gamma_{5,26}^2$. The row indicators in the table revert back to those of the original data. Also listed are the values $\|\omega_2(i)\|^2 / \|\omega_2\|^2$, namely, portions of the squared length of the disturbed expected vector ω_2 that can be attributed to individual rows. These ratios are free of the scale parameter $\gamma_{5,26}$ and thus hold for any such shifts; they sum to unity over the 31 cases; and for cases in Table 9 these sum to 80.67%. The maximal percentage in column 5 for cases not listed is 1.56%. The table shows that shifts in rows $(5, 26)$ affect the expectation of the residual in the non-shifted row 9 even more than expectations of the residuals in either of the shifted rows $(5, 26)$. Owing to this induced and excessive shift in $E(e_9)$, this in turn suggests potential swamping at row 9 by shifts at rows $(5, 26)$, with lesser effect at row 31. This is considered next in some detail.

Evidence for Swamping. To study prospects for swamping, we proceed as in section 3 with $n = 31$ and $r = 29$, retaining $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']' \in \mathbb{R}^{31}$ such that $\boldsymbol{\gamma} \in \mathbb{R}^{29}$ has the common values $\{\gamma_5 = \gamma_{26} \equiv \gamma_{5,26}\}$ as before, and $\gamma_i = 0$ otherwise, so that $\|\gamma\|^2 = 2\gamma_{5,26}^2$. The vectors $\boldsymbol{\gamma}_1 = \boldsymbol{H}_r\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_2 = (\boldsymbol{I}_r - \boldsymbol{H}_r)\boldsymbol{\gamma}$ are projections into the "regressor" and "error" spaces of the reduced model $\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon$, as used throughout section 3. From Definition 2.3.1(ii), the angle $\theta_2 = arcsin\,(\|\gamma_2\| / \|\gamma\|)$ between (γ, γ_2) is $\theta_2 = 16.9$ degrees, showing that γ tends to skew $\mathcal{L}(F_{9,31})$ to the right of $F(2, 19, 0, 0)$. Accordingly, the swamping of rows $I = (9, 31)$ by $J = (5, 26)$ would help to explain that (Y_9, Y_{31}) are not outlying, but instead appear to be so through swamping.

To confirm that shifts at the nondeleted rows $(5, 26)$ have skewed $\mathcal{L}(F_{9,31})$ to the right of $F(2, 19, 0, 0)$ on deleting $I = (9, 31)$, we use Theorem 3.1 to compute the swamping probabilities for the doubly noncentral F having $\lambda_1 = \|\gamma_1\|^2 = 1.8315\gamma_{5,26}^2$ and $\lambda_2 = \|\gamma_2\|^2 = 0.16851\gamma_{5,26}^2$, for $\gamma_{5,26}$ varying over $\{0, 1, 2, 3, 4\}$. The $\alpha = 5\%$ critical value for $F(2, 19, 0, 0)$ is $c_\alpha = 3.5219$. Table 10 reports the corresponding $\alpha = 5\%$ values d_α solving $Pr(F > d_\alpha) = 0.05$, together with the swamping probabilities $\beta_S = Pr(F > 3.5219)$. Both probabilities are computed for the case $\mathcal{L}(F_{9,31}) = F(2, 19, 1.8315\gamma_{5,26}^2, 0.16851\gamma_{5,26}^2)$. For example, with $\gamma_{5,26} = 2$, the value for the doubly noncentral $F(2, 19, 7.3259, 0.6741)$

Table 10
Values d_α solving $Pr(F > d_\alpha) = 0.05$, and the swamping probabilities $\beta_S = Pr(F > 3.5219)$, are tabulated for the case $\mathcal{L}(F) = F(2, 19, \lambda_1, \lambda_2)$ with $\lambda_1 = 1.8315\gamma_{5,26}^2$, $\lambda_2 = 0.16851\gamma_{5,26}^2$, as $\gamma_{5,26}$ varies over $\{0, 1, 2, 3, 4\}$

$\gamma_{5,26}$	d_α	β_S
0	3.5219	0.0500
1	6.2096	0.1818
2	12.1889	0.5851
3	20.3684	0.9124
4	30.1703	0.9934

is $d_\alpha = 12.1889$, and the swamping probability is $\beta_S = Pr(F > 3.5219) = 0.5851$ as reported in Table 10, that is, the probability of declaring (Y_9, Y_{31}) to be outlying when they are not.

In short, the conventional deletion diagnostics ignore the critical role of shifts at nonexcluded rows. When these are taken into account, the resulting swamping probabilities support the view that observations (Y_9, Y_{31}) are not true outliers but instead, contrary to convention, appear so as artifacts of shifts at $(5, 26)$. These in turn are abetted by obscure relationships among rows $(9, 31)$ and $(5, 26)$ as imbedded in $\boldsymbol{X}_0$. These findings serve to update, but run contrary to, the Cook and Weisberg (1982, 152) assertion that "Cases 9 and 31 have the largest potential and the largest influence."

4.4. Case Study 3 Continued: Estimating Shifts in the Drill Data

Our findings support the further discovery of linear relations among the shifts themselves. Specifically, for $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']' \in \mathbb{R}^n$ with γ fixed, Lemma A.3 under Assumptions A gives $\tilde{\boldsymbol{\delta}}(\boldsymbol{\gamma}) = (\boldsymbol{Y_I} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_I) + \boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\gamma}$ as unbiased for $\boldsymbol{\delta}$ at fixed γ, with dispersion matrix $\boldsymbol{V}(\tilde{\boldsymbol{\delta}}) = \sigma^2[\boldsymbol{I}_s + \boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}']$ not depending on γ.As an example, we again set the rows to be deleted as $I = (9, 31)$ and the nondeleted rows as $J = (5, 26)$, with the only nonzero shifts occurring in these four rows, namely, $\{\gamma_5, \gamma_{26}, \delta_9, \delta_{31}\}$. From these we get the linear equations

$$\tilde{\delta}_9 = 0.6274 - 0.5792\gamma_5 - 0.5792\gamma_{26}$$

$$\tilde{\delta}_{31} = 0.3509 - 0.0840\gamma_5 - 0.0840\gamma_{26}$$

as estimates for (δ_9, δ_{31}) in terms of (γ_5, γ_{26}), with solutions in Table 11 for varying values of the shifts (γ_5, γ_{26}). These in turn reflect the interdependence between the pairs $(5, 26)$ and $(9, 31)$.

Earlier we considered possible shifts at rows $(5, 9, 26, 28, 31)$. To continue, we next suppose that shifts occur only at those loci, all others having zero shifts. Designate these shifts as the generic quantities $\{\xi_5, \xi_9, \xi_{26}, \xi_{28}, \xi_{31}\}$; in what follows these will identify in turn with the (γ_i, δ_j) of earlier usage. For each of these cases we set $I = (i)$ in Lemma A3(iv)

Table 11
The moment estimators $(\widetilde{\delta}_9, \widetilde{\delta}_{31})$ for the shifts (δ_9, δ_{31}) for varying values of shifts (γ_5, γ_{26})

γ_5	γ_{26}	$\widetilde{\delta}_9$	$\widetilde{\delta}_{31}$
−2	−2	2.9442	0.6867
−2	0	1.7858	0.5188
−2	2	0.6274	0.3509
0	−2	1.7858	0.5188
0	0	0.6274	0.3509
0	2	−0.5310	0.1829
2	−2	0.6274	0.3509
2	0	−0.5310	0.1829
2	2	−0.6894	0.0150

to compute a moment estimator for its shift as a function of the other shifts using $\widetilde{\xi}_i = (Y_i - z_i'\widehat{\beta}_i) + z_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\xi}$. Here, z_i' is the row deleted from $\boldsymbol{X}_0$; $\widehat{\boldsymbol{\beta}}_i$ is from the nondeleted data; and $\boldsymbol{\xi}$, of order (30×1), consists of zeros together with elements of $\{\xi_5, \xi_9, \xi_{26}, \xi_{28}, \xi_{31}\}$ excluding the test case ξ_i. Letting $i \in \{5, 9, 26, 28, 31\}$ and ξ_i range over $\{\xi_5, \xi_9, \xi_{26}, \xi_{28}, \xi_{31}\}$ in succession gives five linear equations in five unknowns, namely,

$$\xi_5 = 0.4507 - 0.3679\xi_9 + 0.3875\xi_{26} + 0.2244\xi_{28} - 0.2202\xi_{31}$$

$$\xi_9 = 0.7829 - 0.6164\xi_5 - 0.6164\xi_{26} + 0.0578\xi_{28} - 0.4433\xi_{31}$$

$$\xi_{26} = 0.4126 + 0.3875\xi_5 - 0.3679\xi_9 - 0.1025\xi_{28} - 0.2202\xi_{31}$$

$$\xi_{28} = 0.5080 + 0.2028\xi_5 + 0.0312\xi_9 - 0.0926\xi_{26} + 0.2001\xi_{31}$$

$$\xi_{31} = 0.5584 - 0.2760\xi_5 - 0.3316\xi_9 - 0.2760\xi_{26} + 0.2776\xi_{28}$$

with solutions $\widetilde{\xi}_5 = 0.7284$, $\widetilde{\xi}_9 = -0.2255$, $\widetilde{\xi}_{26} = 0.8005$, $\widetilde{\xi}_{28} = -0.4226$, $\widetilde{\xi}_{31} = 0.0938$, showing that the dominant outliers occur at rows $(5, 26)$ and not at rows $(9, 31)$. Here, $\widetilde{\xi}_i$ serves to remind that all such quantities are determined empirically from the data. To convert the moment estimators of the shifts into standard units, we use the estimator for σ^2 from the full model, namely, $S^2 = 0.0272$ with $S = 0.1649$. The moment estimators in standard units are thus $\widetilde{\xi}_5 = 4.4162$, $\widetilde{\xi}_9 = -1.3674S$, $\widetilde{\xi}_{26} = 4.8534$, $\widetilde{\xi}_{28} = -2.5625$, and $\widetilde{\xi}_{31} = 0.5689$.

As noted earlier, with p-value 0.0040 from the presumed $\mathcal{L}(F_{9,31}) = F(2, 19, 0, 0)$, the conventional R-Fisher diagnostic $F_{9,31} = 7.564$ is taken as evidence that (Y_9, Y_{31}) are outlying. Using the foregoing moment estimates, we combine these with developments in Section 3 to gauge the probability that unshifted observations at $I = (9, 31)$ are swamped by outliers at $J = (5, 26)$.

Throughout section 3, noncentrality parameters are reported in standard units as if $\sigma^2 = 1.0$, other values to be reinstated in context. Specifically, take $F(\nu_1, \nu_2, \lambda_1^*, \lambda_2^*)$ with $\lambda_1^* = \lambda_1/\sigma^2$ and $\lambda_2^* = \lambda_2/\sigma^2$ to adjust for scale. Accordingly, on relegating rows (9,31) to the last of $\boldsymbol{X}_0$, we have $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']' \in \mathbb{R}^{31}$ such that $\boldsymbol{\delta} = \boldsymbol{0} \in \mathbb{R}^2$ and $\gamma \in \mathbb{R}^{29}$ has shifts (ξ_5, ξ_{26}) at $J = (5, 26)$, with zero elements otherwise. Given shifts estimated as $\widetilde{\xi}_5 = 4.4162$ and $\widetilde{\xi}_{26} = 4.8534$ in standard units, and $\gamma_i = 0$ otherwise, we proceed as in Theorem 3.1 to compute $\gamma'\gamma = 4.4162^2 + 4.8534^2 = 43.0583$; $\lambda_1^* = \gamma'\boldsymbol{H}_r\gamma = 39.3637$; and $\lambda_2^* = 43.0583 - 39.3637 = 3.6946$. Using these values, we now approximate the probability that unshifted observations at $I = (9, 31)$ are swamped by outliers at $J = (5, 26)$. The result is $\beta_S \approxeq Pr(F_{9,31} > 3.5219) = 0.9630$ from the approximating distribution $\mathcal{L}(F_{9,31}) \approxeq F(2, 19, 39.3637, 3.6946)$. This offers persuasive evidence that, contrary to convention, (Y_9, Y_{31}) are not true outliers, but instead appear so as an artifact of swamping and relationships among the rows $(9, 31)$ and $(5, 26)$.

5. Conclusions

The statistics literature is replete with diagnostics for influential and outlying observations. An important fact, to be restated, is that the Table 1 diagnostics are functionally equivalent to F_I. In contrast to numerous experimental and simulation studies reporting these diagnostics, the present work concerns distributions of F_I and irregularities induced by shifted outliers. These irregularities are given in section 3 as doubly noncentral distributions, their parameters as functions of the design and shifts in observations. An informative dichotomy emerges as projections of shifts into the "regressor" and "error" spaces of a model.

A large body of known results, assuming outliers only at data to be deleted, is extended here to include shifts anywhere. The induced distributions $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ may be skewed either to the left or right of $F(\nu_1, \nu_2, 0, 0)$ as reference, accounting on occasion for masking or swamping effects. Case studies in section 4 illustrate the basic concepts, to include swamping curves as evolving probabilities that a non-outlying subset is deemed incorrectly to be outlying. Of particular interest is the Drill Data set of Cook and Weisberg (1982, 149), long a benchmark for deletion diagnostics. Cases 9 and 31, singled out by those authors as having "the largest potential and the largest influence," are now refuted, appearing instead to be so through swamping. Moreover, our tools support moment estimation of selected shifts themselves in the Drill Data set.

Anomalies uncovered here emphasize difficulties intrinsic to identifying outliers in regression. Noncentralities correspond to shifts; masking and swamping may misidentify shifts; and these typically are hidden from the user. Accordingly, prospects for correcting p-values to account for shifts using our methods typically are not feasible. These anomalies serve to abrogate realistic prospects for correctly distinguishing outlying from non-outlying, or equivalently, influential from noninfluential subsets. Despite wide and continuing usage, abetted by available if ill–understood software support, conventional deletion diagnostics no longer can be recommended to achieve the objectives traditionally cited. Nonetheless, our tools may support future projects using, for example, moment estimates for shifts as in Lemma A.3(iv). Current work in progress by the authors seeks alternatives to deletions for identifying outlying data.

"Theory," as set forth over many decades, claims to offer a succession of new and effective methods, some preferred in comparison to others. On revisiting foundations supporting that methodology, we demonstrate much to be flawed and offer ours as a much-needed corrective paradigm. Rather than offering yet more layers of things to be done, that is, even newer methods, this study is cautionary regarding unintended consequences that may befall a body of literature that has devolved "for the most part . . . based on ad hoc reasoning," but for "a more complete understanding of past results, ad hoc reasoning no longer seems sufficient" (Cook 1986).

References

Andrews, D. F., and T. Pregibon. 1978. Finding outliers that matter. *J. R. Stat. Soc. B*, 40, 85–93.

Atkinson, A. C. 1985. *Plots, transformations, and regression*. Oxford, U.K: Oxford University Press.

Barnett, V., and T. Lewis. 1984. *Outliers in statistical data, 2nd ed*. New York, NY: Wiley.

Beckman, R. J., and H. J. Trussell. 1974. The distribution of an arbitrary Studentized residual and the effects of updating in multiple regression. *J. Am. Stat. Assoc.* 69, 199–201.

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression diagnostics: Identifying influential data and sources of collinearity*. New York, NY: Wiley.

Box, G. E. P., and K. B. Wilson. 1951. On the experimental attainment of optimum conditions. *J. R. Stat. Soc. B*, 13, 1–45.

Bulgren, W. 1971. On representations of the doubly non-central F distribution. *J. Am. Stat. Assoc.*, 66, 184–186.

Chatterjee, S., and A. S. Hadi. 1986. Influential observations, high leverage points, and outliers in linear regression. *Stat. Sci.*, 1, 379–393.

Chatterjee, S., and A. S. Hadi. 1988. *Sensitivity analysis in linear regression*. New York, NY: Wiley.

Cook, R. D. 1977. Detection of influential observations in linear regression. *Technometrics*, 19, 15–18.

Cook, R. D. 1986. [Influential observations, high leverage points, and outliers in linear regression]: Comment. *Stat. Sci.*, 1, 393–397.

Cook, R. D., and S. Weisberg. 1982. *Residuals and influence in regression*. London, UK: Chapman and Hall.

Draper, N. R., J. A. John. 1981. Influential observations and outliers in regression. *Technometrics*, 23, 21–26.

Ennis, D., and N. Johnson. 1993. Noncentral and central chi-square, F and beta distribution functions as special cases of the distribution function of an indefinite quadratic form. *Commun. Stat. Theory Methods*, 22, 897–905.

Fox, J. 1991. *Regression diagnostics*. Newbury Park, CA: Sage.

Gentleman, J. F., and W. B. Wilk. 1975. Detecting outliers. II. Supplementing the direct analysis of residuals. *Biometrics*, 31, 387–410.

Ghosh, S. 1978. On robustness of designs against incomplete data. *Sankhyā Ser. B*, 40, 204–208.

Das Gupta, S., and M. D. Perlman. 1974. Power of the noncentral F-test: Effect of additional variates on Hotelling's T^2–test. *J. Am. Stat. Assoc.*, 69, 174–180.

Hoaglin, D. C., and P. J. Kempthorne. 1986. [Influential observations, high leverage points, and outliers in linear regression]: Comment. *Stat. Sci.*, 1, 408–412.

Imhof, J. 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419–426.

Jensen, D. R. 2000. The use of Studentized diagnostics in regression. *Metrika*, 52, 213–223.

Jensen, D. R. 2001. Properties of selected subset diagnostics in regression. *Stat. Prob. Lett.*, 51, 377–388.

Jensen, D. R., and D. E. Ramirez. 1996. Computing the CDF of Cook's D_I statistic. In *Proceedings of the 12th Symposium in Computational Statistics* ed. A. Prat, and E. Ripoll, 65–66. Barcelona, Spain: Institut d'Estadistica de Catalunya.

Johnson, N. L., and S. Kotz. 1970. *Distributions in statistics: Continuous univariate distributions—2*. Boston, MA: Houghton Mifflin.

LaMotte, L. R. 1999. Collapsibility hypotheses and diagnostic bounds in regression analysis. *Metrika*, 50, 109–119.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India*, 12, 49–55.

Myers, R. H. 1990. *Classical and modern regression with applications*, 2nd ed. Boston, MA: PWS-KENT.

Rousseeuw, P. J., and A.M. Leroy. 1987. *Robust regression and outlier detection*. New York, NY: Wiley.

Snedecor, G. W., and W. G. Cochran. 1968. *Statistical methods*, 6th ed. Ames, IA: Iowa State University Press.
Welsch, R. E. 1982. Influence functions and regression diagnostics. In *Modern data analysis*, ed. R. L. Launer and A. F. Siegel, 149–169. New York, NY: Academic Press.
Welsch, R. E., and E. Kuh. 1977. Linear regression diagnostics. Technical Report 923–77, Cambridge, MA: Sloan School of Management, Massachusetts Institute of Technology.

Appendix A: Foundations

The construction $F_I = \boldsymbol{e}_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}\boldsymbol{e}_I/sS_I^2$ requires that $\mathcal{L}(e_I)$ be of full rank s, that is, that $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ be invertible, as assumed in Gentleman and Wilk (1975). This need not hold, in which case F_I is undefined. This matter is covered analytically through ranks $R(\cdot)$ with matrix argument as follows.

Lemma A.1. *Consider $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$, of order $(s \times s)$, as a principal submatrix of the idempotent $(\boldsymbol{I}_n - \boldsymbol{H}_n)$. Then*

i. *In order that $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ be invertible, it is necessary that $s \leq n - p$.*
ii. *$R(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ is s diminished by the number of unit eigenvalues of $\boldsymbol{H}_{II}$.*
iii. *In particular, $R(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ is s diminished by the number of unit leverages appearing on the diagonal of $\boldsymbol{H}_{II}$.*

Proof. Clearly $(\boldsymbol{I}_n - \boldsymbol{H}_n)$ is idempotent of rank $r = n - p$, having r linearly independent rows and, by symmetry, r linearly independent columns. Accordingly, every principal $(s \times s)$ submatrix of order $s > r$ is necessarily deficient in rank, to establish conclusion (i) by contradiction. From the eigenvalue identities $\{\lambda_i(\boldsymbol{I}_s - \boldsymbol{H}_{II}) = 1.0 - \lambda_i(\boldsymbol{H}_{II});\ 1 \leq i \leq s\}$, it follows that as many eigenvalues of $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ are zero as there are unit eigenvalues of $\boldsymbol{H}_{II}$, to verify conclusion (ii). Conclusion (iii) follows from (ii) since each unit leverage on the diagonal of $\boldsymbol{H}_{II}$ generates a unit eigenvalue of $\boldsymbol{H}_{II}$. For suppose that the leading element of $\boldsymbol{H}_{II}$ is unity, say $h_{r+1,r+1} = 1$. Since $\boldsymbol{H}_n$ is symmetric idempotent, every other element in the row and column containing $h_{r+1,r+1}$ must vanish, so that $\boldsymbol{H}_{II} = Diag(1.0, B)$ has one unit eigenvalue, and $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ has rank $s - 1$. Applied recursively, this gives (iii) and our proof.

To continue, properties of F_I in turn rest on those of the subvector $\mathbf{e}_I$ of ordinary residuals. For brevity rewrite expression (1) as

$$\begin{bmatrix} e \\ e_I \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y}_I \end{bmatrix} \tag{A.1}$$

with $\boldsymbol{G}_{11} = (\boldsymbol{I}_r - \boldsymbol{H}_{00})$, $\boldsymbol{G}_{21}' = \boldsymbol{G}_{12} = -\boldsymbol{H}_{0I}$, and $\boldsymbol{G}_{22} = (\boldsymbol{I}_s - \boldsymbol{H}_{II})$.

Lemma A.2. *Consider the ordinary residuals $e_0' = [e', e_I']$ under Assumption A1: $E(\varepsilon_0) = \omega$, with $\omega' = [\gamma', \delta']$, and A2: $V(\varepsilon_0) = \sigma^2\boldsymbol{I}_n$. As in Definition 2.3.1 decompose $\omega = \omega_1 + \omega_2$ with $\omega_1 = \boldsymbol{H}_n\omega$ and $\omega_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\omega$. Then*

i. *$E(e_0) = \omega_2$;*
ii. *$E(e_I) = \boldsymbol{G}_{21}\gamma + \boldsymbol{G}_{22}\delta$; and*
iii. *$V(e_I) = \sigma^2\boldsymbol{G}_{22} = \sigma^2(\boldsymbol{I}_s - \boldsymbol{H}_{II})$.*

Moreover, under Assumption A3: $\mathcal{L}([(\varepsilon - \gamma)', (\varepsilon_I - \delta)']') = N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$, it follows that

iv. *$\mathcal{L}(\mathbf{e}_I) = N_s(\boldsymbol{G}_{21}\gamma + \boldsymbol{G}_{22}\boldsymbol{\delta}, \sigma^2\boldsymbol{G}_{22})$.*

Proof. Assumption A1 gives $\mathrm{E}(e_0) = (\boldsymbol{I}_n - \boldsymbol{H}_n)(X_0\boldsymbol{\beta} + \boldsymbol{\omega}) = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega} = \boldsymbol{\omega}_2$ as in (i) since $(\boldsymbol{I}_n - \boldsymbol{H}_n)X_0 = \boldsymbol{0}$. In partitioned form this is

$$\mathrm{E}\begin{bmatrix} e \\ e_I \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{bmatrix}\begin{bmatrix} (\boldsymbol{X}\beta + \boldsymbol{\gamma}) \\ (\boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\delta}) \end{bmatrix} = \begin{bmatrix} (\boldsymbol{G}_{11}\boldsymbol{\gamma} + \boldsymbol{G}_{12}\boldsymbol{\delta}) \\ (\boldsymbol{G}_{21}\boldsymbol{\gamma} + \boldsymbol{G}_{22}\boldsymbol{\delta}) \end{bmatrix}. \tag{A.2}$$

giving conclusion (ii). Moreover, under Assumption A2 with $\mathrm{V}(e_0) = \sigma^2 G$, the marginal values $\mathrm{E}(e_I) = (\boldsymbol{G}_{21}\boldsymbol{\gamma} + \boldsymbol{G}_{22}\boldsymbol{\delta})$ and $\mathrm{V}(e_I) = \sigma^2\boldsymbol{G}_{22} = \sigma^2(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ are immediate, as in (ii) and (iii). Conclusion (iv) follows directly, to complete our proof.

Theorem 3.1 rests on a version of the Fisher–Cochran theorem and the decomposition of quadratic forms. Accordingly, we turn next to working expressions equivalent to $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, $(\widehat{\beta} - \widehat{\beta}_I)$, and the reduced residual sum of squares $RSS(Y) = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta}_I)'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\beta}_I)$. In view of Lemma A.1 we take $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ to have rank s, that is, that no eigenvalue of $\boldsymbol{H}_{II}$ is unity. Recurring matrices include $\boldsymbol{R} = [\boldsymbol{R}_1, \boldsymbol{R}_2]$, with $\boldsymbol{R}_1 = (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}'$ and $\boldsymbol{R}_2 = (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{Z}'$.

Lemma A.3. *Consider* $\boldsymbol{X}_0' = [\boldsymbol{X}', \boldsymbol{Z}']$, $\boldsymbol{R} = [\boldsymbol{R}_1, \boldsymbol{R}_2]$, *and* $\boldsymbol{H}_n = [\boldsymbol{H}_{ij}]$, *together with the partitioned vector* $\boldsymbol{\omega}' = [\boldsymbol{\gamma}', \boldsymbol{\delta}']$ *in* $\{\boldsymbol{Y}_0 \to \boldsymbol{Y}_0 + \boldsymbol{\omega}\}$. *Then*

i. $(\boldsymbol{X}'\boldsymbol{X})^{-1} = (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1} + \boldsymbol{R}_2(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}\boldsymbol{R}_2'$;
ii. $(\widehat{\beta} - \widehat{\beta}_I) = \boldsymbol{R}_2(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I$; and
iii. $RSS(Y) = e_0'e_0 - e_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I$.

Under Assumptions A, with $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ *fixed we have*

iv. $E(\boldsymbol{Y}_I - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_I) = \boldsymbol{\delta} - \boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\gamma} \overset{def}{=} \boldsymbol{\theta}$;
v. $V(\boldsymbol{Y}_I - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_I) = \sigma^2[\boldsymbol{I}_s + \boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}'] \overset{def}{=} \sigma^2\boldsymbol{\Xi}$;
vi. $\mathcal{L}(\boldsymbol{Y}_I - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_I) = N_s(\boldsymbol{\theta}, \sigma^2\boldsymbol{\Xi})$.

Proof. For (i)–(iii) see the proof for Lemma A.1 of Jensen (2001). For conclusion (iv) observe that $\mathrm{E}(\boldsymbol{Y}_I - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_I) = (\boldsymbol{\beta} + \boldsymbol{\delta}) - \boldsymbol{Z}[\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\gamma}]$ since $\mathrm{E}(\widehat{\boldsymbol{\beta}}_I) = [\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\gamma}]$. Conclusion (v) follows since $(\boldsymbol{Y}, \boldsymbol{Y}_I)$ are uncorrelated, and conclusion (vi) follows directly.

To continue, we put in place nonstandard distributions of quadratic forms in Gaussian vectors. In the notation of Eq. (1), Lemma A.3(iii) supplies the needed Fisher–Cochran expansion

$$e_I'(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}e_I + (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_I)'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_I) = e_0'e_0, \tag{A.3}$$

where invertibility of $(\boldsymbol{I}_s - \boldsymbol{H}_{II})$ is covered in Lemma A.1. As quadratic forms in $\boldsymbol{Y}_0$, (A.3) becomes $\boldsymbol{Y}_0'\boldsymbol{A}_1\boldsymbol{Y}_0 + \boldsymbol{Y}_0'\boldsymbol{A}_2 Y_0 = \boldsymbol{Y}_0'\boldsymbol{A}_3\boldsymbol{Y}_0$, where $\boldsymbol{A}_3 = (\boldsymbol{I}_n - \boldsymbol{H}_n)$, $\boldsymbol{A}_2 = Diag(\boldsymbol{I}_r - \boldsymbol{H}_r, \boldsymbol{0})$ with $r = n - s > p$, and $\boldsymbol{A}_1 = \boldsymbol{A}_3 - \boldsymbol{A}_2$ in partitioned form is

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{H}_{0I}(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}\boldsymbol{H}_{I0} & -\boldsymbol{H}_{0I} \\ -\boldsymbol{H}_{I0} & (\boldsymbol{I}_s - \boldsymbol{H}_{II}) \end{bmatrix}. \tag{A.4}$$

Its leading element, namely, $[(\boldsymbol{I}_r - \boldsymbol{X}](\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}') - (\boldsymbol{I}_r - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')]$, is

$$\boldsymbol{X}[(\boldsymbol{X}'\boldsymbol{X})^{-1} - (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}]\boldsymbol{X}' = \boldsymbol{H}_{0I}(\boldsymbol{I}_s - \boldsymbol{H}_{II})^{-1}\boldsymbol{H}_{I0}$$

using Lemma A.3(i). Observe further that

$$\boldsymbol{H}_{I0}(\boldsymbol{I}_r - \boldsymbol{H}_r) = \boldsymbol{Z}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}'[\boldsymbol{I}_r - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] = \boldsymbol{0}. \tag{A.5}$$

To continue, for $\boldsymbol{U} \in \mathbb{R}^n$ random having $\mathrm{E}(\boldsymbol{U}) = \boldsymbol{\mu}$, the noncentrality parameter for $\boldsymbol{U}'\boldsymbol{A}\mathbf{U}$ is the quadratic form $\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$ in its expectation. Accordingly, write $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{Y}_0) = \boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\omega}$ with $\boldsymbol{\omega}' = [\boldsymbol{\gamma}', \boldsymbol{\delta}']$ under Assumptions A. Then the corresponding noncentralities are $\lambda_3 = \boldsymbol{\mu}'\boldsymbol{A}_3\boldsymbol{\mu} = \boldsymbol{\omega}'\boldsymbol{A}_3\boldsymbol{\omega} = \boldsymbol{\omega}'(\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$ since $(\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{X}_0\boldsymbol{\beta} = \boldsymbol{0}$; similarly $\lambda_2 = (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\gamma})'(\boldsymbol{I}_r - \boldsymbol{H}_r)(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\gamma}) = \boldsymbol{\gamma}'(\boldsymbol{I}_r - \boldsymbol{H}_r)\boldsymbol{\gamma}$; and $\lambda_1 = (\boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\omega})'\boldsymbol{A}_1(\boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\omega})$, to be reexamined subsequently. Accordingly, consider the quadratic forms $Q_1 = \boldsymbol{Y}_0'\boldsymbol{A}_1\boldsymbol{Y}_0$, $Q_2 = \boldsymbol{Y}_0'\boldsymbol{A}_2\boldsymbol{Y}_0$, and $Q_3 = \boldsymbol{Y}_0'\boldsymbol{A}_3\boldsymbol{Y}_0$, such that $Q_1 + Q_2 = Q_3$. Further essentials follow; normality assumes a central role; $(\boldsymbol{I}_n - \boldsymbol{H}_n) = [\boldsymbol{G}_{ij}]$ has blocks $\boldsymbol{G}_{11} = (\boldsymbol{I}_r - \boldsymbol{H}_{00})$, $\boldsymbol{G}_{21}' = \boldsymbol{G}_{12} = -\boldsymbol{H}_{0I}$, and $\boldsymbol{G}_{22} = (\boldsymbol{I}_s - \boldsymbol{H}_{II})$ as in Eq. (1); and $F_I = (Q_1/Q_2)[(n-p-s)/s]$.

Lemma A.4. *Given Assumptions A, such that* $E(\varepsilon_0) = \boldsymbol{\omega} = [\boldsymbol{\gamma}', \boldsymbol{\delta}']'$; *decompose* $\gamma = \gamma_1 + \gamma_2$ *with* $\gamma_1 = \boldsymbol{H}_r\gamma$ *and* $\gamma_2 = (\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma$. *Then the respective distributions of* Q_1, Q_2, *and* Q_3 *are the following:*

i. $\mathcal{L}(Q_1) = \chi^2(s, \lambda_1)$, *with* $\lambda_1 = (\boldsymbol{G}_{21}\gamma_1 + \boldsymbol{G}_{22}\delta)'\boldsymbol{G}_{22}^{-1}(\boldsymbol{G}_{21}\gamma_1 + G_{22}\delta)$;
ii. $\mathcal{L}(Q_2) = \chi^2(n-p-s, \lambda_2)$, *with* $\lambda_2 = \gamma_2'(\boldsymbol{I}_r - \boldsymbol{H}_r)\gamma_2$;
iii. $\mathcal{L}(Q_3) = \chi^2(n-p, \lambda_3)$, *with* $\lambda_3 = \omega'(\boldsymbol{I}_n - \boldsymbol{H}_n)\omega$.

Moreover, the forms $\{Q_1, Q_2, F_I\}$ *have the properties*

iv. Q_1 and Q_2 are independent; and
v. $\mathcal{L}(F_I) = F(s, n-p-s, \lambda_1, \lambda_2)$.

Proof. Let $\{\phi_i(t);\ 1 \le i \le 3\}$ be the *chf*'s of $\{Q_i;\ 1 \le i \le 3\}$ as quadratic forms in the elements of $\boldsymbol{Y}_0$. This result is based in part but goes beyond Lemma A.2 of Jensen (2001). We work backward from (iii), (ii), and (iv) to (i) and (v). Since $Q_3 = \boldsymbol{Y}_0'\boldsymbol{A}_3\boldsymbol{Y}_0$ and $\boldsymbol{A}_3$ is $(n \times n)$ idempotent of rank $n-p$, it follows directly that $\mathcal{L}(Q_3) = \chi^2(n-p, \lambda_3)$, with *chf* $\phi_3(t) = (1-2it)^{-(n-p)/2}\,exp\,[i\lambda_3 t/(1-2it)]$ and $\lambda_3 = \boldsymbol{\omega}'\boldsymbol{A}_3\boldsymbol{\omega}$, as claimed in conclusion (iii). Similarly, since $Q_2 = \boldsymbol{Y}_0'\boldsymbol{A}_2\boldsymbol{Y}_0$ and $\boldsymbol{A}_2$ is idempotent of rank $(r-p) = (n-p-s)$, it follows that $\mathcal{L}(Q_2) = \chi^2(n-p-s, \lambda_2)$, with *chf* $\phi_2(t) = (1-2it)^{-(n-p-s)/2}\,exp\,[i\lambda_2 t/(1-2it)]$ and $\lambda_2 = \boldsymbol{\omega}'\boldsymbol{A}_2\boldsymbol{\omega}$, as asserted in conclusion (ii). That $\boldsymbol{A}_1\boldsymbol{A}_2 = \boldsymbol{0}$ is verified directly, assuring both the independence of Q_1 and Q_2 as in (iv), and the factorization $\phi_1(t)\phi_2(t) = \phi_3(t)$ as in Jensen (2001). Substituting for $\phi_2(t)$ and $\phi_3(t)$ and solving gives

$$\phi_1(t) = (1-2it)^{-s/2}\,exp\ [i\lambda_1 t/(1-2it)]$$

and $\lambda_1 = \lambda_3 - \lambda_2$, so that $\mathcal{L}(Q_1) = \chi^2(s, \lambda_1)$ from uniqueness of *chf*'s. The equivalent form $\lambda_1 = (\boldsymbol{G}_{21}\gamma + \boldsymbol{G}_{22}\delta)'\boldsymbol{G}_{22}^{-1}(\boldsymbol{G}_{21}\gamma + \boldsymbol{G}_{22}\delta)$, as listed in conclusion (i), derives from $\mathcal{L}(e_I)$ in Lemma A.2(iii), the quadratic form $Q_1 = e_I'[\mathrm{V}(e_I)]^{-1}e_I$ in e_I having $\lambda_1 = [\mathrm{E}(e_I)]'[\mathrm{V}(e_I)]^{-1}[\mathrm{E}(e_I)]$ as given. Conclusion (v) follows directly.

Appendix B: Deletion Diagnostics

In a sweeping survey by Chatterjee and Hadi (1986) and discussants, deletion diagnostics have been labeled as bewildering, excessive, and largely redundant; as "chaotic," begging "distillation" to an "integrated set of procedures"; and as devolving through "ad hoc reasoning," explaining "the diversity of recommendations" for cutoff values, with the latter seen as "vague and . . . contradictory," not to be "sanctified," but instead to be "guided by statistical theory." One contrived and contradictory rule among many (see Jensen (2000)) is Myers's (1990, 261) claim that "a yardstick of approximately ±2 on *DIFFIT*s and *DFBETA*s may be reasonable." Despite those caveats, the misguided use of these diagnostics and their benchmarks continues apace, abetted in turn by software support.

Those concerns since have been vindicated; they trace to the mistaken advocacy of diagnostics *standardized,* but not *Studentized* using correct standard deviations. Moreover, these concerns mostly have been resolved through the findings that many single-case deletion diagnostics correspond one-to-one with t_i or t_i^2, as in LaMotte (1999) and Jensen (2000), and that the Table 1 diagnostics correspond one-to-one with F_I, Jensen (2001).

A number of *influence distance diagnostics* purport to be patterned on the metrics of Mahalanobis (1936). These include diagnostics of Cook (1977), Welsch and Kuh (1977), Welsch (1982), and D_I as in Table 1. Apart from D_I, their matrices are ad hoc, diverse, subjective, often contradictory, and devoid of rational bases; none is properly Studentized; none accounts for the singularity of $\mathcal{L}(\widehat{\beta} - \widehat{\beta}_I) \in \mathbb{R}^p$ of rank $s < p$; none is a genuine Mahalanobis (1936) metric; and, unlike proper metrics on $\mathbb{R}^p$, the diagnostic of Cook (1977) has bounded range.

Appendix C

We turn to computing probabilities for doubly noncentral F-distributions. Bulgren (1971) has given a series representation for the *cdf* of $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ in terms of incomplete Beta functions. Building on the earlier work of Imhof (1961), Ennis and Johnson (1993) have expressed the *cdf* for $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ as a one-dimensional integral using trigonometric functions. This result is easy to code, for example, in Maple, and the Ennis and Johnson representation for the *cdf* was used to compute the probabilities shown in the tables of section 4. The National Institute of Standards and Technology makes available Dataplot as a public-domain software system. Dataplot contains the FORTRAN subroutine DNFCDF based on the algorithm of Bulgren (1971). With modern software such as Mathematica 8.0, probabilities for the doubly noncentral F-distributions are readily accessible.