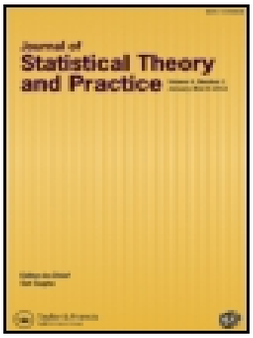

Journal of Statistical Theory and Practice

ISSN: 1559-8608 (Print) 1559-8616 (Online) Journal homepage: http://www.tandfonline.com/loi/ujsp20

# Seeking outlying subsets under star-contoured errors

Donald R. Jensen & Donald E. Ramirez

**To cite this article:** Donald R. Jensen & Donald E. Ramirez (2018) Seeking outlying subsets under star-contoured errors, Journal of Statistical Theory and Practice, 12:4, 757-775, DOI: 10.1080/15598608.2018.1473183

**To link to this article:** https://doi.org/10.1080/15598608.2018.1473183

Accepted author version posted online: 10 May 2018.
Published online: 12 Jun 2018.

JOURNAL OF STATISTICAL THEORY AND PRACTICE
2018, VOL. 12, NO. 4, 757–775
https://doi.org/10.1080/15598608.2018.1473183

# Seeking outlying subsets under star-contoured errors

Donald R. Jensen[a] and Donald E. Ramirez [b]

[a]Department of Statistics,Virginia Tech, Blacksburg, Virginia, USA; [b]Department of Mathematics, University of Virginia, Charlottesville, Virginia, USA

**ABSTRACT**
Given observations $\{Y_i; 1 \leq i \leq n\}$ with dispersion matrix $\boldsymbol{\Sigma}$, a pervading issue is whether shifts have occurred in designated subsets of the observations. Early work on single shifts used order statistics or the R–Student $t_i$ statistics as diagnostics, initially derived under i.i.d. Gaussian assumptions. These diagnostics recently have been shown to remain exact in level and power under equicorrelation and more general dispersion structures, and under star-contoured mixtures supplanting Gaussian errors, with an accounting for irregularities engendered by shifts at other than the designated cases. Extensions here pertain to outlying subsets using the *R*-Fisher diagnostics $F_I$, showing invariance of its distribution and of related diagnostics under more general dispersion structures and mixtures over these. Shifts occurring at cases other than those designated induce doubly noncentral *F* distributions. These elicit profound disturbances in operating characteristics of the diagnostics, serving in turn to explain masking and swamping, and the discovery of hidden "regression effects" among outliers. Evidence for anomalies arising from denominator noncentralities rests on two-sided rejection rules to be given. Numerical studies serve to illuminate the essence of the findings in practice.

**ARTICLE HISTORY**
Received 23 January 2017
Revised 4 February 2018
Accepted 2 May 2018

**KEYWORDS**
Outlying subsets; structured errors; mixtures; invariance; misdirected inferences

**AMS SUBJECT CLASSIFICATION**
62E10; 62F04

## 1. Introduction

Given observations $\boldsymbol{Y}_0' = [Y_1, \ldots, Y_n]$ having a vector mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$, much research has focused on a pervading issue: to assess whether a single shift, $\{Y_i \to Y_i + \delta_i\}$, or a subset shift, $\{\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta_I}\}$, has occurred. Consequences are profound: whether to discard irregular observations as compromised or anomalous, or to pursue heretofore unknown features that may have been revealed in explanation. It is instructive to enumerate the typical assumptions from the literature as follow.

**Assumptions A.** $A_1$, $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$; $A_2$, $\boldsymbol{Y}_0$ has a Gaussian distribution; $A_3$, the shifts $\{Y_i \to Y_i + \delta_i\}$ or $\{\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta_I}\}$ occur exclusively at the designated observations.

Against this background we survey precedents as well as objectives of the present study. Looking ahead, assumptions $A_1$ and $A_2$ are unnecessarily limiting, while $A_3$ largely has been overlooked, to the detriment of the essential objectives of many studies.

Procedures to identify $\{Y_i \to Y_i + \delta_i\}$ trace to Dixon (1950), Grubbs (1950), and Ferguson (1961) based on order statistics, and to the *R*-Student $t_i^2$ of Snedecor and

**CONTACT** Donald R. Jensen djensen@vt.edu Department of Statistics,Virginia Tech, 406-A Hutcheson Hall, Blacksburg, VA 24061-0439, USA

Cochran (1968, p. 157) and Beckman and Trussell (1974), all predicated on Assumptions A. Nonetheless, these diagnostics remain exact in level and power under the equicorrelation matrix $\boldsymbol{\Sigma}(\rho)$ and $\boldsymbol{\Sigma}(\xi)$ to be identified, both in lieu of $A_1$, and for star-contoured Gaussian mixtures supplanting $A_2$. Moreover, irregularities in the Student $t_i$ diagnostics owing to the failure of $A_3$ have been documented. See Jensen and Ramirez (2015). In addition, those findings encompass both single-case influence and outlier diagnostics corresponding one-to-one with $t_i$ or $t_i^2$.

Subset deletions are undertaken here when each of Assumptions A fails; these studies in parallel with the preceding paragraph. Accordingly, partition $\boldsymbol{Y}'_0 = [\boldsymbol{Y}', \boldsymbol{Y}'_I]$, and the ordinary residuals conformably as $e'_0 = [e', e'_I]$ of orders $(1 \times r)$ and $(1 \times s)$, $r + s = n$. The means initially are taken to be constant through $Y_r$ but arbitrary thereafter, to be modeled as $\{\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta}\}$. The shift diagnostic of Gentleman and Wilk (1975) takes the form

$$\mathrm{F_I} = \frac{e'_I(\mathbf{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)e_I}{s\mathrm{S}_\mathrm{I}^2} \tag{1}$$

with $\mathrm{S}_\mathrm{I}^2$ as the sample variance from $\boldsymbol{Y}$. The distribution then is $\mathcal{L}(\mathrm{F_I}) = F(u; s, r-1, \lambda)$ in Gaussian data such that $\lambda = \delta'(\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)\delta/\sigma^2$. Details and extensions are given subsequently, where specializing to $s = 1$ gives $t_i^2$. Moreover, if $\delta$ is concentrated in a subspace of $\mathbb{R}^s$ of dimension $k < s$ and having the same $\lambda$, then a result of Das Gupta and Perlman (1974) shows the power to be greater for $F(u; k, n-k-1, \lambda)$ than for $F(u; s, r-1, \lambda)$. Designs fully estimable after deletions are reported in Ghosh (1978).

In perspective, regression diagnostics appear widely in the literature, applying as needed for the model $\{\boldsymbol{Y_0} = \mu + \epsilon_0\}$ featured here. Events referred to as the masking and swamping of outliers serve to obfuscate meanings ascribed to outlier diagnostics, where masking "is an important problem in influence analysis which deserves further study" Hoaglin and Kempthorne (1986, p. 410). See also Bendre and Kale (1987). Our technical tools offer further insight, as reported subsequently.

This study goes beyond the classical paradigm since Assumptions A often fail. In regard to $A_1$, correlated data arise in direct and inverse calibrations as in Jensen and Ramirez (2009, 2012) and numerous other settings. In addition, Box and Tiao (1968) and Aitkin and Wilson (1980) have modeled data from subsamples as Gaussian mixtures. In addition, mixtures are central in clustering models as seen in Punzo, Browne, and McNicholas (2016). These mixtures effectively stipulate that Assumption $A_2$ be replaced in their respective fields of application. In short, that dependent data and mixtures do emerge in practice serves to motivate their inclusion here, whereas outliers continue to arise in the expanded models. It follows that $\mathrm{F_I}$, traditionally viewed among normal-theory parametric procedures, is seen from this study to be genuinely nonparametric, holding exactly for Gaussian mixture distributions. This expands substantially their ranges of applicability beyond those known to date. An outline follows.

The developments of section 2 include notation and properties of the ensembles $\{\boldsymbol{\Sigma}(\rho)\}$ and $\{\boldsymbol{\Sigma}(\xi)\}$ together with essential properties of the residuals. Irregularities in the use of $\mathrm{F_I}$ are identified in section 3 for the case where $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$. Extensions to

include $\mathbf{\Sigma} \in \{\mathbf{\Sigma}(\rho), \mathbf{\Sigma}(\xi)\}$, and for mixture distributions over these ensembles, are given in section 4. Some consequences are detailed through numerical studies in section 5, and critical supporting topics are attached for completeness as an appendix.

## 2. Preliminaries

### 2.1. Notation

Spaces of note include Euclidean $n$-space $\mathbb{R}^n$, its positive orthant $\mathbb{R}^n_+$, and the real symmetric $(n \times n)$ matrices $\mathbb{S}_n$. Vectors and matrices are set in bold type; the transpose, inverse, trace, and determinant of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\boldsymbol{A}^{-1}$, $\mathrm{tr}(\boldsymbol{A})$, and $|\boldsymbol{A}|$; $\boldsymbol{I}_n$ is the $(n \times n)$ identity; and $\mathrm{Diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is a block-diagonal array. Throughout we designate $\boldsymbol{A}_n = \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$ and $\boldsymbol{B}_n = (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n)$ as idempotent matrices of ranks 1 and $n$ – 1, respectively.

A random $\boldsymbol{Y} \in \mathbb{R}^n$ has distribution $\mathcal{L}(\boldsymbol{Y})$; mean $\mathrm{E}(\boldsymbol{Y})$; dispersion matrix $V(\boldsymbol{Y}) = \mathbf{\Sigma}'$ taking the value $\mathrm{Var}(Y) = \sigma^2$ on $\mathbb{R}^1$; and a density function $(pdf)$ $g(y)$. Specifically, $\mathcal{L}(\boldsymbol{Y}) = N_n(\mu, \mathbf{\Sigma})$ is Gaussian on $\mathbb{R}^n$ with mean $\mu$ and dispersion matrix $\Sigma$. Distributions on $\mathbb{R}^1_+$ of note include $\chi^2(v; \lambda)$ as chi-squared having $v$ degrees of freedom and noncentrality $\lambda$, and the doubly noncentral Snedecor–Fisher $F(u; v_1, v_2, \lambda_1, \lambda_2)$ having $(v_1, v_2)$ degrees of freedom, and numerator and denominator noncentrality parameters $(\lambda_1, \lambda_2)$. Identify $\{\mathrm{F_I} > c_\alpha\}$ as the conventional upper $\alpha$-level rejection rule based on $F(u; v_1, v_2, 0, 0)$. Essential ordering properties of $F(u, v_1, v_2, \lambda_1, \lambda_2)$ are known as follow.

**Proposition 2.1.** *Given $\mathcal{L}(\mathrm{F_I}) = F(u; v_1, v_2, \lambda_1, \lambda_2)$. Then this distribution (i) increases stochastically with increasing $\lambda_1$, and (ii) decreases stochastically with increasing $\lambda_2$, with other parameters held fixed.*

### 2.2. The model

The general linear model of reference is $\{\boldsymbol{Y}_0 = \boldsymbol{X}_0\beta + \epsilon_0\}$ where $(\boldsymbol{Y}_0, \epsilon_0) \in \mathbb{R}^n$; $\boldsymbol{X}_0$ is of order $(n \times k)$ having rank $k < n$, and $\mathbf{H}_n = \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0' = [h_{ij}]$ has leverages $\{h_{ii}; 1 \leq i \leq n\}$ on its diagonal. That model specializes here to

$$\left\{\boldsymbol{Y}_0 = \mu + \epsilon_0; \quad \boldsymbol{H}_n = [h_{ij}] = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n{}'\right\}. \tag{2}$$

To continue, partition $\mathbf{Y}'_0 = [\mathbf{Y}', \mathbf{Y}'_I]$ and $\epsilon'_0 = [\epsilon', \epsilon'_I]$ conformably, with $\boldsymbol{Y} \in \mathbb{R}^r$ and $\boldsymbol{Y}_I \in \mathbb{R}^s$, $r + s = n$. This study deviates from convention, as noted, in allowing shifts $\{Y_0 \to Y_0 + \omega\}$ anywhere such that $\omega' = [\gamma', \delta'] \in \mathbb{R}^n$, with $\gamma \in \mathbb{R}^r$ and $\delta \in \mathbb{R}^s$, to include but transcend the conventional shifts $\{\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta}\}$.

Accordingly, Gauss–Markov assumptions are recast as follows, where $V(\boldsymbol{Y}_0) = \sigma^2\boldsymbol{I}_n$ often is standardized to $\sigma^2 = 1.0$.

**Assumptions B.** The following hold with $\epsilon_0' = [\epsilon', \epsilon_I']$ and $\omega' = [\gamma', \delta']$.

$\mathrm{B}_1$. $E(\epsilon) = \gamma \in \mathbb{R}^r$ and $E(\epsilon_I) = \delta \in \mathbb{R}^s$.

$\mathrm{B}_2$. $V(\epsilon_0) = \sigma^2 I_n$.

$\mathrm{B}_3$. $\mathcal{L}(\epsilon_0) = N_n(\omega, \sigma^2 \boldsymbol{I}_n)$.

**Definition 2.1.** *Decompose $\omega$ as $\omega = \omega_1 + \omega_2 \in \mathbb{R}^n$ with $\omega_1 = \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\omega$ and $\omega_2 = (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n)\omega$ as projections into the "mean" and "error" spaces of $\{\boldsymbol{Y}_0 = \mu + \epsilon_0\}$. Similarly, take $\gamma = \gamma_1 + \gamma_2 \in \mathbb{R}^r$, with $\gamma_1 = \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r\gamma$ and $\gamma_2 = (\boldsymbol{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r)\gamma$.*

In addition, users long have been cognizant of critical misdiagnoses of outliers, as in the following.

**Definition 2.2.** *Masking occurs when outliers remain undetected; swamping occurs when non-outliers are mistakenly identified as outliers.*

## 2.3. Properties of residuals

This section pertains to violating Assumption $\mathrm{A}_3$ of the Introduction, where shifts now may occur at cases other than the designated $\boldsymbol{Y}_I$. For consistency arrange $\mathbf{Y}_0 = [\mathbf{Y}', \mathbf{Y}_I']'$ so as to delete $\boldsymbol{Y}_I$ in this configuration; other choices for deletion then will require reconfiguring. Recalling $\boldsymbol{B}_n = [\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n]$, take the residual vector $e_0 = \mathbf{B}_n\mathbf{Y}_0$ to be partitioned as

$$\begin{bmatrix} e \\ e_I \end{bmatrix} = \begin{bmatrix} (\boldsymbol{I}_r - \frac{1}{n}\mathbf{1}_r\mathbf{1}'_r) & -\frac{1}{n}\mathbf{1}_r\mathbf{1}'_s \\ -\frac{1}{n}\mathbf{1}_s\mathbf{1}'_r & (\boldsymbol{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s) \end{bmatrix} \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{Y}_I \end{bmatrix}. \tag{3}$$

First essentials regarding $\{e_I, \mathrm{S}_\mathrm{I}^2, \mathrm{F}_\mathrm{I}\}$ are assembled next, where the matrix of the quadratic form in the numerator for $\mathrm{F}_\mathrm{I} = \frac{e'_I(\mathbf{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)e_I}{s\mathrm{S}_\mathrm{I}^2}$ follows from the identity $(\boldsymbol{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s)^{-1} = (\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)$.

**Lemma 2.1.** *Properties of $\{e_I, \mathrm{S}_\mathrm{I}^2, \mathrm{F}_\mathrm{I}\}$ under Assumptions* B, *where $Q_1 = e_I'(\mathbf{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)e_I$ and $Q_2 = (r-1)\mathrm{S}_\mathrm{I}^2$, may be categorized as:*

*(i)* $\mathrm{E}(e_I) = (\delta - \bar{\omega}\mathbf{1}_s) = \mu_{\mathbf{e}_I}$, *where $\bar{\omega}$ is the mean of $\omega$ in Assumptions* B;
*(ii)* $\mathrm{V}(e_I) = \sigma^2(I_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s)$;
*(iii)* $\mathcal{L}(e_I) = N_s(\mu_{\mathbf{e}_I}, \sigma^2(I_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s))$;
*(iv)* $\mathcal{L}(e_I'(I_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)e_I) = \chi^2(s, \lambda_1)$ *with* $\lambda_1 = \mu_{e_I}'(I_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)\mu_{\mathbf{e}_I}$;
*(v)* $\mathcal{L}((r-1)\mathrm{S}_\mathrm{I}^2) = \chi^2(r-1, \lambda_2)$ *with* $\lambda_2 = \gamma' B_r \gamma = \gamma'_2\gamma_2$;
*(vi)* $Q_1$ *and* $Q_2$ *are distributed independently;*
*(vii)* $\mathcal{L}(\mathrm{F}_\mathrm{I}) = F(u; s, r-1, \lambda_1, \lambda_2)$.

Proof. (i) $E(e_0)$ is found on substituting $E(\boldsymbol{Y}) = \gamma$ and $E(\boldsymbol{Y}_I) = \delta$ on the right of expression (3). Since $V(e_0) = \sigma^2 B_n$, the marginal dispersion matrix follows as $\mathrm{V}(e_I) =$

$\sigma^2(\boldsymbol{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s)$ as the lower right submatrix in Eq. (3), giving conclusion (ii). Conclusion (iii) follows directly from Assumption $B_3$. To continue, the theory of quadratic forms asserts that $e'_I(\mathbf{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s)^{-1}e_I$ has a noncentral $\chi^2(s,\lambda_1)$ distribution with $\lambda_1 = \mu'_{\mathbf{e}_I}(\boldsymbol{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s)^{-1}\mu_{\mathbf{e}_I}$, giving conclusion (iv) on noting that $(\boldsymbol{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s)^{-1} = (\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)$. In a similar manner the quadratic form $(r-1)\,\mathrm{S}_\mathrm{I}^2 = \boldsymbol{Y}'\boldsymbol{B}_r\boldsymbol{Y}$ is distributed as $\chi^2(r-1,\lambda_2)$ with $\lambda_2 = \gamma'\boldsymbol{B}_r\gamma = \gamma'_2\gamma_2$ from Definition 1, to establish conclusion (v). Conclusion (vi) follows on identifying $Q_1 = \boldsymbol{Y}'_0\boldsymbol{A}_1\boldsymbol{Y}_0$ and $Q_2 = \boldsymbol{Y}'_0\boldsymbol{A}_2\boldsymbol{Y}_0$ from Appendix expression (6) and verifying that $\boldsymbol{A}_1\boldsymbol{A}_2 = \mathbf{0}$. Conclusion (vii) follows since $\mathrm{F_I} = (r-1)Q_1/sQ_2$, to complete our proof.

## 3. Inferences based on $\mathrm{F_I}$

Two essential topics emerge: (i) to examine disturbances in the use of $\mathrm{F_I}$ owing to $\lambda_2 > 0.$, and (ii) to garner sample evidence regarding $\lambda_2$, historically taken to be zero without comment. A first look in section 3.1 establishes that $\mathcal{L}(\mathrm{F_I})$ is doubly noncentral under $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$, inducing disturbances in the use of $\mathrm{F_I}$ as in section 3.2. Topic (ii) is reexamined in section 3.3. Those findings in turn are extended in section 4 for ensembles of type $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\rho), \boldsymbol{\Sigma}(\xi)\}$ and for mixtures over these ensembles.

### 3.1. *Properties of $\mathrm{F_I} : \boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$*

Outliers and their effects on $\mathrm{F_I}$ are considered next, in continuation of violating Assumption $A_3$. Recall that $\gamma_1 = \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r\gamma$ and $\gamma_2 = (\boldsymbol{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r)\gamma$ are projections of the shift $\{\boldsymbol{Y} \to \boldsymbol{Y} + \gamma\}$ into the respective location and error spaces of the model $\{Y = \mu\mathbf{1}_r + \epsilon\}$. A principal finding is the following.

**Theorem 3.1.** *Given Assumptions* B *with* $\mathrm{E}(\epsilon_0) = \omega = [\gamma', \delta']'$, *together with* $\gamma = \gamma_1 + \gamma_2$. *Then the distribution of* $\mathrm{F_I}$ *is* $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, \lambda_1, \lambda_2)$ *such that*

*(i)* $\lambda_1 = (\delta - \bar{\omega}\mathbf{1}_s)'(\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)(\delta - \bar{\omega}\mathbf{1}_s)$, *where* $\bar{\omega} = \frac{r\bar{\gamma}+s\bar{\delta}}{n}$.
*(ii)* $\lambda_2 = \gamma'B_r\gamma = \sum_{i=1}^{r}(\gamma_i - \overline{\gamma})^2 = \gamma'_2\gamma_2$.
*(iii) For* $\gamma = \gamma_1$, *then* $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, \lambda_1, 0)$ *with* $\lambda_1$ *as in (i); and at* $\boldsymbol{\delta} = \mathbf{0}$, *then* $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, \lambda_1, 0)$ *with* $\lambda_1 = rs\overline{\gamma}^2/n$.
*(iv) For* $\gamma = \gamma_2$, *then* $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, \lambda_1, \lambda_2)$ *with* $\lambda_1 = (\delta - \frac{s\bar{\delta}}{n}\mathbf{1}_s)'(\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)(\delta - \frac{s\bar{\delta}}{n}\mathbf{1}_s)$ *and* $\lambda_2$ *as in (ii); and at* $\boldsymbol{\delta} = \mathbf{0}$, *then* $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, 0, \lambda_2)$.

*Proof.* Conclusion (i) follows from Lemma 1(i),(iv), and conclusion (ii) from Lemma 1 (v). Taking $\gamma = \gamma_1$ assigns $\gamma_2 = \mathbf{0}$ and $\lambda_2 = 0$ from (ii); at $\boldsymbol{\delta} = \mathbf{0}$, then $\overline{\omega} = r\overline{\gamma}/n$ to give conclusion (iii). On the other hand, taking $\gamma = \gamma_2$ implies $\overline{\gamma} = 0$ and $\overline{\omega} = s\overline{\delta}/n$, to establish conclusion (iv).

### 3.2. Anomalies in the use of $F_I$

As in expression (1), conventional usage takes $\mathcal{L}(\mathrm{F_I}) = F(u; s, r-1, \lambda_1)$, rejecting for $\{\mathrm{F_I} > c_\alpha\}$ as an $\alpha$-level unbiased test for $H_0 : \delta = \mathbf{0}$ against $H_1 : \delta \neq \mathbf{0}$, with $c_\alpha$ from $F(u; \nu_1, \nu_2, 0)$ and with power increasing in $\lambda_1$. This is contra-indicated as follows for $\mathcal{L}(\mathrm{F_I}) = F(u; s, r-1, \lambda_1, \lambda_2)$ as in Theorem 3.1, where $\gamma = \gamma_1 + \gamma_2$ as before. Departures from convention include the following properties of $\mathrm{F_I}$ owing exclusively to the violation of Assumption $\mathrm{A}_3$.

$\mathrm{P}_1$. (i) For $\mathcal{L}(\mathrm{F_I})$ as in Theorem 3.1(iii), tests using $\{\mathrm{F_I} > c_\alpha\}$ are anticonservative, rejecting with probability greater than $\alpha$.
(ii) For $\mathcal{L}(\mathrm{F_I})$ as in Theorem 3.1(iv), tests using $\{\mathrm{F_I} > c_\alpha\}$ are conservative, rejecting with probability less than $\alpha$.

$\mathrm{P}_2$. The rule $\{\mathrm{F_I} > c_\alpha\}$ serves instead to test the "regression effect" $H_0 : \delta = \bar{\omega}\mathbf{1}_s$ *vs* $H_1 : \delta \neq \bar{\omega}\mathbf{1}_s$ as in Theorem 3.1(i). This serves to mask that $\delta \neq \mathbf{0}$ through $\overline{\omega}\ \mathbf{1}_s \neq \mathbf{0}$.

$\mathrm{P}_3$. The test intended for $H_0 : \delta = \mathbf{0}$ is achieved at $\gamma = \gamma_2$, in the sense that $\bar{\gamma} = 0$ and $\lambda_1 = 0$ under $H_0$, but at the expense of $\lambda_2 > 0$ as in Theorem 3.1(iv) and $\mathrm{P}_1(ii)$.

$\mathrm{P}_4$. An outcome $\{\mathrm{F_I} > c_\alpha\}$, misattributed to $\delta \neq \mathbf{0}$ when in fact $\delta = \mathbf{0}$, accounts for $\delta = \mathbf{0}$ to be swamped by $\gamma$. For the case $\gamma_2 = \mathbf{0}$, the swamping probability is $P(\mathrm{F_I} > c_\alpha | \lambda_1)$ as determined from $F(u; s, r-1, \lambda_1, 0)$ with $\lambda_1 = rs\overline{\gamma}^2/n$ as in Theorem 3.1(iii).

$\mathrm{P}_5$. The event $\gamma = \gamma_2 \neq \mathbf{0}$ serves to mask $\delta \neq \mathbf{0}$ since $F(u; s, r-1, 0, \lambda_2)$ is stochastically smaller than $F(u; s, r-1, 0, 0)$ by Proposition 2.1(ii). Accordingly, the value $\lambda_1^\dagger$ such that $P(\mathrm{F_I} > c_\alpha) = \alpha$ exactly, under $F(u; s, r-1, \lambda_1^\dagger, \lambda_2)$, serves to gauge the "minimal" element $\lambda_1^\dagger$ to be discerned at level $\alpha$, as the power increases monotonically from the value $\alpha$ for $\lambda_1 > \lambda_1^\dagger$. In particular, all parameters on the boundary and interior of the ellipsoid $\{\delta^{\dagger\prime}(I_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)\boldsymbol{\delta}^\dagger \leq \lambda_1^\dagger\}$ fail to be rejected with probability $1 - \alpha$, and thus comprise the parameters masked by $\gamma_2 \neq \mathbf{0}$. For example, for $F(u; 4, 20, \lambda_1^\dagger, 2)$, the value at $\alpha = 0.05$ is $\lambda_1^\dagger = 0.4152$.

$\mathrm{P}_6$. Subset diagnostics entrenched in the genre include those of Appendix A.2 as they pertain to models of the present study. Unfortunately, these are thrown into disarray, altered irrevocably by anomalies in $\mathrm{F_I}$. In particular, if subsets are ascribed to be outlying or influential, or not, using the diagnostics of Table A.1, these may have been misidentified owing to unexamined outliers in the nondeleted elements of $Y_0$.

### 3.3. Two-sided rejection rules

That the intended properties of $\mathrm{F_I}$ are thwarted for $\lambda_2 > 0$ is evident. Empirical evidence to this effect may be garnered through two-sided rejection rules. Indeed, large $\mathrm{F_I}$ is classical evidence against $H_0 : \delta = \mathbf{0}$. If instead $\mathrm{F_I}$ is deemed to be small, then from Proposition 2.1(ii) this would be consistent with a stochastically smaller distribution, specifically $F(u; \nu_1, \nu_2, 0, \lambda_2)$ with $\lambda_2 > 0$. Accordingly, two-sided rejection rules for $\mathrm{F_I}$ appear to offer plausible evidence in regard to $[\lambda_1, \lambda_2]$. Since $\mathrm{F_I} = (Q_1/\nu_1)/(Q_2/\nu_2)$ is the ratio of scaled and independent chi-squared variables, further insight is gained from the equivalence of $\mathcal{L}(\mathrm{F_I}) = F(u; \nu_1, \nu_2, \lambda_1, \lambda_2)$ and $\mathcal{L}(\mathrm{F_I^{-1}}) = F(u; \nu_2, \nu_1, \lambda_2, \lambda_1)$. This in turn

**Table 3.1.** Lower and upper $(1-\alpha) = 0.95$-level acceptance limits for $\mathrm{F_I}$ in selected cases.

| $(\nu_1, \nu_2)$ | $1/c_1$ | $c_2$ | $(\nu_1, \nu_2)$ | $1/c_1$ | $c_2$ |
|---|---|---|---|---|---|
| (2, 2) | 0.0256 | 39.0000 | (3, 8) | 0.0688 | 5.4160 |
| (2, 3) | 0.0255 | 16.0040 | (10, 10) | 0.0269 | 3.7169 |

prompts the following heuristic approach for testing $H_0 : [\lambda_1, \lambda_2] = [0, 0]$ against designated alternatives.

**Algorithm 3.1.** *Consider $\mathcal{L}(\mathrm{F_I}) = F(u; \nu_1, \nu_2, \lambda_1, \lambda_2)$ as alternative to $F(u; \nu_1, \nu_2, 0, 0)$. Take $c_1$ and $c_2$ as the upper $(\alpha/2)$-fractiles of $F(u; \nu_2, \nu_1, 0, 0)$ and $F(u; \nu_1, \nu_2, 0, 0)$, respectively. Then:*

*(i) The interval $[1/c_1 \le \mathrm{F_I} \le c_2]$ serves as a $(1-\alpha)$-level acceptance region for $H_0$.*
*(ii) $\mathrm{F_I} > c_2$ is taken as evidence against $H_0$ in favor of $\lambda_1 > 0$.*
*(iii) $\mathrm{F_I} < 1/c_1$, is taken as evidence against $H_0$ in favor of $\lambda_2 > 0$.*

To illustrate, lower and upper acceptance limits are given in Table 3.1 for $\alpha = 0.05$ in selected cases.

**Remark 3.1.** As the reader will note, this algorithm itself departs from convention, where tests for $H_0 : [\lambda_1, \lambda_2] = [0, 0]$ would rest on sample estimates for $[\lambda_1, \lambda_2]$, not undertaken here, with rejection rules from their joint distribution. But this would entail concepts of isotonic regression as in Barlow et al. (1972), since alternatives to $H_0$ are necessarily the one-sided $H_1 : \{\lambda_1 > 0 \,\text{and/or}\, \lambda_2 > 0\}$.

## 4. Extended models

This section outlines dispersion matrices $\mathbf{\Sigma} \in \{\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\xi), \mathbf{\Sigma}(\rho)\}$ as alternatives to Assumption $\mathrm{A}_1 : \mathbf{\Sigma} = \sigma^2 I_n$ in the Introduction.

### 4.1. *Properties of* $\mathrm{F_I} : \mathbf{\Sigma} \in \{\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\xi), \mathbf{\Sigma}(\rho)\}$

Structured dispersion matrices alternative to $\mathbf{\Sigma} = \sigma^2 \boldsymbol{I}_n$ arise on occasion in practice, while nonetheless supporting validity in linear inference. Three ensembles of such matrices are considered next, together with conditions that these be positive definite.

**Lemma 4.1.** *(i) $\Xi(\theta) := \{\mathbf{\Sigma}(\theta) = \sigma^2(I_n + \theta \mathbf{1}_n \mathbf{1}'_n); \theta \in \Gamma_1\}$, where $\Gamma_1 = \{\theta > -\frac{1}{n}\}$.*

*(ii) $\Xi(\xi) := \{\mathbf{\Sigma}(\xi) := \sigma^2(I_n + \mathbf{1}_n \xi' + \xi \mathbf{1}'_n - \bar{\xi} \mathbf{1}_n \mathbf{1}_n'); \xi \in \Gamma_2\}$, where $\Gamma_2 = \{\xi \in \mathbb{R}^n : \tau_1 > n\tau_2 - 1\}$ with $\tau_1 = \xi_1 + \ldots + \xi_n = n\bar{\xi}$ and $\tau_2 = \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2$.*

*(iii) $\Xi(\rho) := \{\mathbf{\Sigma}(\rho) = \sigma^2[(1-\rho)I_n + \rho \mathbf{1}_n \mathbf{1}'_n]; \rho \in \Gamma_3\}$, where $\Gamma_3 = \{-\frac{1}{n-1} < \rho < 1\}$.*
*(iv) $\Xi(\theta)$ and $\Xi(\rho)$ are equivalent; take $k^2\theta = \rho$, then $\mathbf{\Sigma}(\theta) = \frac{1}{1-\rho}\mathbf{\Sigma}(\rho)$.*
Proof. See Jensen (1996).

To continue, Assumptions B are revised next in keeping with $\mathbf{\Sigma} \in \{\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\xi), \mathbf{\Sigma}(\rho)\}$.

**Assumptions B$^\dagger$.** The following hold:

$\mathrm{B}_1{}^\dagger$. $\mathrm{E}(\epsilon_0) = \omega = [\gamma', \delta']'$; i.e., $E(\epsilon) = \gamma \in \mathbb{R}^r$ and $E(\epsilon_I) = \delta \in \mathbb{R}^s$;

$\mathrm{B}_2{}^\dagger$. $\mathrm{V}(\epsilon_0) = \mathbf{\Sigma} \in \{\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\xi), \mathbf{\Sigma}(\rho)\}$;

$\mathrm{B}_3{}^\dagger$. $\mathcal{L}(\epsilon_0) = N_n(\omega, \mathbf{\Sigma})$ for $\mathbf{\Sigma} \in \{\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\xi), \mathbf{\Sigma}(\rho)\}$.

These in turn support invariance properties of $\mathcal{L}(\mathrm{F_I})$ for dispersion matrices in the larger class.

**Theorem 4.1.** *Given Assumptions $B^\dagger$ with* $\mathrm{E}(\epsilon_0) = \omega' = [\gamma', \delta']$, *such that* $\gamma = \gamma_1 + \gamma_2$, *and with* $\mathbf{\Sigma} \in \{\mathbf{\Xi}(\theta), \mathbf{\Xi}(\xi), \mathbf{\Xi}(\rho)\}$ *as in Lemma 4.1:*

*(i) Then the distribution of* $\mathrm{F_I}$ *is* $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, \lambda_1, \lambda_2)$ *precisely as in Theorem 3.1, independently of* $\mathbf{\Sigma} \in \{\mathbf{\Xi}(\theta), \mathbf{\Xi}(\xi), \mathbf{\Xi}(\rho)\}$

*(ii) Anomalies for* $\mathrm{F_I}$ *listed as* $\mathrm{P}_1$ *through* $\mathrm{P}_6$ *in section 3.2 continue to hold.*

*Proof.* As in the proof for Lemma 2.1 write $Q_1 = Y_0' A_1 Y_0$ and $Q_2 = Y_0' A_2 Y_0$. Theorem 5.1.4 of Mathai and Provost (1992, 201) identifies the conditions (a) $\{A_i \Sigma A_i = A_i; i = 1, 2\}$ to be necessary and sufficient for (b) $\{\mathcal{L}(Q_i) = \chi^2(\nu_i; \lambda_i); i = 1, 2\}$ with $\nu_i = \mathrm{tr}(A_i \Sigma)$ and $\lambda_i = \tau' A_i \tau$, where $\tau = E(Y_0)$. Moreover, $\{A_1 \Sigma A_2 = \mathbf{0}\}$ is necessary and sufficient so that $Q_1$ and $Q_2$ are independent. Verifying these for $\Sigma \in \{\Sigma(\theta); \Sigma(\xi)\}$, and for $\mathbf{\Sigma}(\rho)$ by equivalence, is carried out in Theorem A.1 of Appendix A.1.

## 4.2. Properties of $\mathrm{F_I}$ under mixtues

In keeping with our goal to replace each of Assumptions A, we finally determine that $\mathrm{A}_2$ is sufficient but not necessary for our principal findings. Accordingly, consider mixtures over Gaussian dispersion ensembles as in Lemma 4.1, taking $g_n(y; \mu, \Sigma)$ as the Gaussian density corresponding to $N_n(\mu, \Sigma)$.

**Definition 4.1.** Take $\Xi := \{\mathbf{\Sigma}(\psi); \psi \in \Gamma\}$ as a typical ensemble of Lemma 4.1, considered to have mixing parameters $\psi \in \Gamma$ with mixing cdf $G$ on $\Gamma$, for each $\psi \in \{\theta, \xi, \rho\}$. Further consider the dispersion mixtures

$$\mathrm{M}(\psi) = \{f(y; \mu, G) = \int_\Gamma g_n(y; \mu, \mathbf{\Sigma}(\psi))\, dG;\ G \in \mathcal{G}\} \tag{4}$$

where $\mathcal{G}$ consists of all *cdf*s on $\Gamma$.

**Remark 4.1.** In particular, $\{f(y; \mu, G); G \in \mathcal{G}\}$ are dispersion mixtures of elliptical Gaussian densities on $\mathbb{R}^n$ centered at $\mu \in \mathbb{R}^n$. These are symmetric star-unimodal densities as

identified in Jensen ansd Ramirez (2015). In particular, if $\mathbf{\Sigma} = \sigma^2 \sum_{1=1}^{q} w_i R_i$ with $R_i = [(1 - \rho_i)\boldsymbol{I}_n + \rho_i \mathbf{1}_n \mathbf{1}'_n]$ and $\{w_i \geq 0\}$ such that $\sum_{1=1}^{q} w_i = 1$, then this is a finite mixture.

A result of Jensen (1989) has established invariance of the $R$-Student distribution $\mathcal{L}(t_i^2)$ under Gaussian mixtures of the types considered here. A principal finding next extends that result to encompass $\mathcal{L}(\mathrm{F_I})$ for each of the ensembles of Lemma 4.1.

**Theorem 4.2.** *Given Assumptions $B^\dagger$ with $\mathbf{E}(\epsilon_0) = \omega = [\gamma', \delta']'$, such that $\gamma = \gamma_1 + \gamma_2$, and with $\mathcal{L}(Y)$ as a mixture in the class $\mathrm{M}(\psi)$ of Definition 4.1 for each $\psi \in \{\theta, \xi, \rho\}$.*

*(i) Then the distribution of $\mathrm{F_I}$ is $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, \lambda_1, \lambda_2)$ precisely as in Theorem 3.1 for the case $\mathbf{\Sigma} = \sigma^2 \boldsymbol{I}_n$, independently of the collection $\mathcal{G}$ of mixing distributions on $\Gamma$.*

*(ii) Anomalies regarding $\mathrm{F_I}$ listed as $\mathrm{P_1}$ through $\mathrm{P_6}$ of section 3.2 continue to hold for these mixtures.*

*(iii) Distributtions of the diagnostics $\{\mathrm{OUT_I}, \mathrm{AP_I}, \mathrm{CR_I}, \mathrm{FV_I}\}$ of Table A.1 are identical to those under Assumption $\mathrm{B_3}$. $\mathcal{L}(\epsilon_0) = N_n(\omega, \sigma^2 I_n)$, independently of the collection $\mathcal{G}$ of mixing distributions.*

Proof. Begin with a density in $\mathrm{M}(\psi)$, namely, $f(\mathrm{y}_0; \mu, G) = \int_\Gamma g_n(\mathrm{y}_0; \mu, \mathbf{\Sigma}(\psi))\, dG$, where $\mathbf{y}_0$ is the argument for $\boldsymbol{Y}_0 = [\boldsymbol{Y}', \boldsymbol{Y}_I']'$. Making the change of variables to $\mathrm{F_I} = \frac{e_I'(I_s + \frac{1}{r}\mathbf{1}_s \mathbf{1}_s')\, e_I}{s\mathrm{S}_\mathrm{I}^2}$ behind the integral gives $\mathcal{L}(\mathrm{F_I}|\mathbf{\Sigma}(\psi)) = \int_\Gamma F(u; s, r-1, \lambda_1, \lambda_2)\, dG$. But this is independent of $\Sigma(\psi)$ by Theorem 4.1, so that $\mathcal{L}(\mathrm{F_I}) = f(u; s, r-1, \lambda_1, \lambda_2)$ unconditionally since $\int_\Gamma dG = 1$. Parallel arguments apply for each $\psi \in \{\theta, \xi, \rho\}$. Conclusion (iii) follows since these diagnostics correspond one-to-one with $\mathrm{F_I}$ as seen in Appendix A.2, to complete the proof.

It is noteworthy that these findings extend substantially beyond the classical venues for deletion diagnostics, to include nonstandard dispersion matrices and mixtures. Moreover, our findings are complementary to and extend considerably beyond the work of Srivastava (1980), Young, Pavur, and Marco (1989), and Baksalary et al. (1992).

## 5. Numerical studies

Case studies follow in which dimensions are kept small in the interest of brevity. Note that if $\gamma = \gamma_1$ in $\mathbb{R}^r$ with $\gamma_2 = \mathbf{0}$, then $\gamma = c\mathbf{1}_r$ holds necessarily, so that $\{Y \to Y + c\mathbf{1}_r\}$ now comprises their common shift. Moreover, if $\gamma = \gamma_2$ with $\gamma_1 = \mathbf{0}$, then $\bar{\gamma} = 0$.

### 5.1. *Overview*

Irregularities tracing to $F(u; \nu_1, \nu_2, \lambda_1, \lambda_2)$ include $\beta = P(\mathrm{F_I} > c_\alpha)$ as common reference points. Small sample sizes are considered, beginning with $(r = 3, s = 2)$ such that $(\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s \mathbf{1}'_s) = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$, together with numerator noncentralities of the type

$$\lambda_1 = \phi'(\boldsymbol{I}_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}'_s)\phi = \frac{1}{3}[\phi_1 \quad \phi_2]\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \frac{1}{3}[4\phi_1^2 + 2\phi_1\phi_2 + 4\phi_2^2]$$

with $\phi = (\delta - \bar{\omega}\mathbf{1}_s)$ from Theorem 3.1(i). Accordingly, for fixed $\lambda_1 = \lambda_1^\dagger$, the collection $R(\lambda_1^\dagger) = \{(\phi_1, \phi_2) | \frac{1}{3}[4\phi_1^2 + 2\phi_1\phi_2 + 4\phi_2^2] = \lambda_1^\dagger\}$ is the boundary of an ellipse comprising an equivalence class of two-dimensional parameters having the same probability $\beta$. For example, if $[\phi_1, \phi_2] = [\phi_1, 0]$, then $R(\lambda_1^\dagger) = \{(\phi_1 | \frac{4\phi_1^2}{3} = \lambda_1^\dagger\}$, an even function of $\phi_1$, with a corresponding expression for $[\phi_1, \phi_2] = [0, \phi_2]$.

## 5.2. *Masking*

The critical masking of the shift $\{\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta}\}$ conveys erroneously that $\delta = \mathbf{0}$. One venue is property $P_5$ of section 3.2, attributing this to $\lambda_2 > 0$ since $F(u; \nu_1, \nu_2, \lambda_1, \lambda_2)$ is then stochastically smaller than $F(u; \nu_1, \nu_2, \lambda_1, 0)$. To these ends probabilities $\beta(\lambda_1, \lambda_2) = P(\mathrm{F_I} > c_\alpha)$ are reported in Table 5.1 for the case $F(u; 3, 2, \lambda_1, \lambda_2)$ with $\lambda_2$ zero or not.

In particular, the left portion where $\lambda_2 = 0$ gives points on the power curve from $F(u; 3, 2, \lambda_1, 0)$ with $\alpha = 0.05$. In contrast, entries on the right of Table 5.1 encompass (i) actual probabilities $\beta(\lambda_1, \lambda_2)$ from the doubly noncentral $F(u; 3, 2, \lambda_1, \lambda_2)$, together with (ii) $\beta(\lambda_1, 0)$ from the singly noncentral $F(u; 3, 2, \lambda_1, 0)$, were the experiment instead to have been carried out so as to ensure that $\gamma_2 = \mathbf{0}$. Accordingly, the difference $\beta(\lambda_1, \lambda_2) - \beta(\lambda_1, 0)$ serves to quantify the extent to which these upper tail probabilities are suppressed owing to $\lambda_2 > 0$.

## 5.3. *Swamping*

Dual to the problem that $\delta \neq \mathbf{0}$ may be masked is that $\delta = \mathbf{0}$ may be swamped by outliers $\gamma_1 \neq \mathbf{0}$ at nondeleted cases, leading to the erroneous and often critical misstatement that $\delta \neq \mathbf{0}$. Property $P_4$ of section 3.2 identifies these probabilities as $P(\mathrm{F_I} > c_\alpha | \lambda_1)$.

To illustrate, the swamping probabilities $P(\mathrm{F_I} > c_\alpha | \lambda_1)$, as attributed to $\lambda_1$ through $\gamma_1$, are listed in Table 5.2 for the case $(r = 4, s = 3)$ as $\bar{\gamma}$ varies, with $c_\alpha = 9.2766$ at $\alpha = 0.05$. In particular, the case $\bar{\gamma}^2 = 0$ reflects correctly that $\delta = \mathbf{0}$ is not swamped by $\gamma = \mathbf{0}$. On the other hand, Table 5.2 indicates that the propensity for swamping slightly exceeds $\alpha = 0.05$ for $\bar{\gamma}^2 = 1$, but escalates to the problematic value 0.25 over the course of the table.

## 5.4. *Outliers: the Darwin data*

Darwin's data comprise 15 differences in heights from cross-fertilized and self-fertilized plants, as discussed in Fisher (1960), who expressed concern that the heights may have

**Table 5.1.** Noncentrality parameters $(\lambda_1, \lambda_2)$ for the case $F(u; 3, 2, \lambda_1, \lambda_2)$, together with probabilities $\beta = P(\mathrm{F_I} > c_\alpha)$ with $c_\alpha = 19.164$ at $\alpha = 0.05$, and similarly for $F(3, 2, \lambda_1, 0)$.

| $\lambda_1$ | $\lambda_2$ | $\beta(\lambda_1, 0)$ | $\lambda_1$ | $\lambda_2$ | $\beta(\lambda_1, \lambda_2)$ | $\beta(\lambda_1, 0)$ |
|---|---|---|---|---|---|---|
| 3.01240 | 0.00000 | 0.09691 | 2.42879 | 3.36966 | 0.01831 | 0.08800 |
| 5.17828 | 0.00000 | 0.12919 | 2.57004 | 3.22841 | 0.02008 | 0.09017 |
| 7.22944 | 0.00000 | 0.15871 | 2.11697 | 0.89544 | 0.05476 | 0.08321 |

**Table 5.2.** Swamping probabilities $P(\mathrm{F_I} > c_\alpha|\lambda_1)$ attributed to $\gamma_1 \neq \mathbf{0}$ with $\lambda_1 = rs\bar{\gamma}^2/(r+s)$, for the case $(r = 4, s = 3)$ as $\bar{\gamma}$ varies, with $c_\alpha = 9.2766$ from $F(u; 3, 3, 0, 0)$ at $\alpha = 0.05$.

| $\bar{\gamma}^2$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P(\mathrm{F_I} > c_\alpha\|\lambda_1)$ | 0.0500 | 0.0890 | 0.1296 | 0.1709 | 0.2122 | 0.2531 |

been affected by latent variables such as seed selection, soil fertility, moisture, and sunlight. The data are

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| -67 | -48 | 6 | 8 | 14 | 16 | 23 | 24 |
| 28 | 29 | 41 | 49 | 56 | 60 | 75 | |

to be denoted in order as $[x_1, \ldots, x_{15}]$. Jensen and Ramirez (2015) have studied single-case outlier diagnostics, namely, the $R$-Student $t_i^2$ and tests based on order statistics due to Grubbs (1950), Dixon (1950), and Ferguson (1961). The first two negative entries in the Darwin data appear to differ from the other values and are possible outliers, but neither the parametric Student $t_i^2$ nor the nonparametric tests detected that the minimum value $x_1 = -67$ was an outlier, with all tests failing with $\alpha = 5\%$.

Tests for joint outliers clearly may proceed using $\mathrm{F_I}$, say for $\boldsymbol{I} = [x_u, x_v]$. First consider $\boldsymbol{I} = [x_1, x_2]$; other cases follow subsequently. The full and reduced models are $\{\boldsymbol{Y}_0 = \beta_0 \boldsymbol{I}_n + \epsilon_0\}$ and $\{\boldsymbol{Y} = \beta \boldsymbol{I}_r + \epsilon\}$. Core numerical values for Theorem 3.1 are $\{n = 15, r = 13, s = 2, \widehat{\beta} = 33.0000, \widehat{\beta}_0 = 20.9333\}$ with $Q_3 = e_0'e_0$, $Q_2 = e'e$, $\mathrm{S_I^2} = Q_2/(r-1)$, and related values as listed in the first row of the body of Table 5.3. Thus, from Theorem A.1 we have that $\mathrm{F_I} = (r-1)Q_1/sQ_2 = 15.49$ is greater than the critical value $c_{0.05} = 3.885$, where $Q_1 = Q_3 - Q_2$. The one-sided $p$ value for testing that $[x_1, x_2]$ are outliers is $p = 0.0004737$ from the null distribution $\mathcal{L}(\mathrm{F_I}) = F(u; 2, 12, 0, 0)$. This offers exceptionally strong evidence that $[x_1, x_2]$ in fact differ from the remaining data. Moreover, this example demonstrates that subset diagnostics may uncover pairs of outliers not found through any of several single-case diagnostics as cited in Jensen and Ramirez (2015).

To continue, we next examine the pair $I = [x_{14}, x_{15}]$ in the full data; details are given in Table 5.3. Specifically, $\mathrm{F_I} = 2.07$ with $p$ value $p = 0.1689$ fails to flag these as outlying. On the other hand, we have just seen pervading evidence that $[x_1, x_2]$ are outlying, that is, that there are nonzero shifts $[x_1, x_2] \to [x_1 + \gamma_1, x_2 + \gamma_2]$ that indeed may have served to mask shifts in $[x_{14}, x_{15}]$. Accordingly, we next remove $[x_1, x_2]$ from the data and proceed as in the third row of Table 5.3. The evidence, with $p$ value $p = 0.0240$, now strongly supports that these are outlying, in keeping with the conjecture that the shifted values $[x_1 + \gamma_1, x_2 + \gamma_2]$ indeed have masked a prospective shift in $[x_{14}, x_{15}]$ in the full data. Clearly the finding "$[x_1, x_2]$ are outliers" is robust in the following sense: The swamping probability is $P(\mathrm{F_I} > c_\alpha|\lambda_1)$ as determined from $F(u; s, r-1, \lambda_1, 0)$ with $\lambda_1 = rs\bar{\gamma}^2/n$ as in

**Table 5.3.** Values for selected subset cases $\boldsymbol{I} = [x_u, x_v]$ in the Darwin data.

| Item | $Q_2$ | $Q_3$ | $\mathrm{S_I^2}$ | $\mathrm{F_I}$ | $c_{0.05}$ | $p$ Value |
|---|---|---|---|---|---|---|
| $\{[x_1, x_2]\|x_1, \ldots, x_{15}\}$ | 5,568.0 | 19,944.9 | 464.0 | 15.49 | 3.885 | 0.0004737 |
| $\{[x_{14}, x_{15},]\|x_1, \ldots, x_{15}\}$ | 14,828.3 | 19,944.9 | 1,235.7 | 2.07 | 3.885 | 0.1689 |
| $\{[x_{14}, x_{15}]\|x_3, \ldots, x_{15}\}$ | 2,642.2 | 5,568.0 | 264.2 | 5.54 | 4.103 | 0.0240 |

Theorem 3.1(iii). This from property $P_4$ of section 3.2. Accordingly, here $\lambda_1^\dagger = 8.226$ is found to satisfy $P((F_I\lambda_1^\dagger, \lambda_2 = 0) > 15.49) = 0.05$, and then $\lambda_1^\dagger = 8.226 = rs\bar{\gamma}^2/n$ is solved for $\bar{\gamma}^2 = 4.746$. Since the outlier test is based on the assumption that $\lambda_1 = 0$, that is, the reduced model is specified correctly, then our conclusion will be correct even under a misspecified reduced model with $\bar{\gamma}^2 \leq 4.746$, thus assuring that the $F_I$ statistic has not been swamped.

### 5.5. Outliers in mixtures: the baseball data

Woodward (1970) studied the times for running the bases for $n = 22$ baseball players. Each player ran the bases three times and their average times are

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5.48 | 5.77 | 5.43 | 5.48 | 5.82 | 5.53 | 5.38 | 5.43 |
| 5.10 | 5.78 | 5.18 | 5.55 | 5.47 | 5.00 | 5.47 | 5.50 |
| 5.48 | 5.50 | 5.40 | 5.55 | 5.63 | 6.28 | | |

to be denoted as $[x_1, \ldots, x_{22}]$. These data have been reported by Morrison (2005) to test for outliers. The times that appear abnormal are those for Player 14 and Player 22 with times of 5.00 and 6.28, respectively. Beckett (1977) has identified the $n = 22$ data points as consisting of two clusters, namely, [2, 4, 5, 7–15, 17, 19–22] and [1, 3, 6, 16, 18], where each cluster consists of correlated data. We postulate that this in turn may be approximated by a mixture $\mathbf{\Sigma}(\alpha) = \alpha\,\mathbf{\Sigma}(\rho) + (1-\alpha)\,\mathbf{\Sigma}(\xi)$ with $\alpha \in [0, 1]$. As the classical test for outliers among i.i.d. observations fails to apply, Theorems 4.1 and 4.2 would extend the outlier test using $F_I$ to the extent that the mixture $\mathbf{\Sigma}(\alpha)$ approximates the actual dispersion matrix. As with the Darwin data, the full and reduced models are $\{\mathbf{Y}_0 = \beta_0\mathbf{I}_n + \epsilon_0\}$ and $\{\mathbf{Y} = \beta\mathbf{I}_r + \epsilon\}$. Core numerical values for Theorems 4.1 and 4.2 are $\{n = 22,\ s = 2,\ r = 20, \widehat{\beta_0} = 5.511, \hat{\beta} = 5.498\}$ with $Q_2$, $Q_3$, and other quantities as listed in Table 5.4. Thus, $F_I = 14.32$ is greater than the critical value $c_{0.05} = 3.522$ with $p$ value $p = 0.0001612$, strongly supporting that the pair $[x_{14}, x_{22}]$ differs from the remaining data.

To examine the operating characteristics further, apply property $P_4$ with $\gamma_2 = \mathbf{0}$; compute $\lambda_1^\dagger = 8.925$ to satisfy $P(F_I(u; s, r-1, \lambda_1^\dagger, 0) > 14.32) = 0.05$. Then $\lambda_1^\dagger = rs\bar{\gamma}^2/n$ and $\bar{\gamma}^2 = 4.909$, so $H_0$ will be rejected for any reduced model with $\{\gamma_2 = \mathbf{0},\ \bar{\gamma}^2 \leq 4.909\}$ with swamping probability $P(F_I > 3.522 | \lambda_1 = 8.925, \lambda_2 = 0) = 0.6937$.

**Table 5.4.** Values for selected subset cases $I = [x_{14}, x_{22}]$ in the baseball data.

| Item | $Q_2$ | $Q_3$ | $S_I^2$ | $F_I$ | $c_{0.05}$ | $p$ Value |
|---|---|---|---|---|---|---|
| $\{[x_{14}, x_{22}] \mid x_1, \ldots, x_{22}\}$ | 0.5713 | 1.4325 | 0.03007 | 14.32 | 3.522 | 0.0001612 |

### 5.6. Mixtures: the salinity data

Theorem 4.2 asserts that the distributions of $\mathrm{F_I}$ and related diagnostics, and thus their essential properties, are invariant under Gaussian mixtures as stipulated. A reviewer notes parallel work under spherical linear regression models on $\mathbb{R}^n$ with densities of type

$$f_{\mathrm{Y}}(y) = g((y - \mathbf{X}\beta)'(y - \mathbf{X}\beta)), \tag{5}$$

namely, the work of Galea, Riquelme, and Paula (2000) and others cited there. Those studies include a variety of single-case deletion diagnostics, shown to be invariant under spherical symmetry, and illustrated with the salinity data of Ruppert and Carroll (1980). The model (5) clearly includes the mean shift outlier model with $X\beta = \mu$ together with $\mathrm{F_I}$ as considered here. Note the following:

**Disclaimer**. Invariance as reported in Galea, Riquelme, and Paula (2000) for single-case deletions applies for null distributions exclusively. Although not noted by those authors, this constraint follows on examining the cited invariance theorem from the literature. In contrast, the invariance properties for $\mathrm{F_I}$ as in Theorem 4.2 for Gaussian mixtures apply for nonnull distributions as well.

**Proposition 5.1.** *Let $\mathcal{L}(Y_0)$ of section 2.2 instead have the density $f_{\mathrm{Y}_0}(y_0) = g(y_0 - \mu)'(y_0 - \mu))$. Then the null distribution of $\mathrm{F_I}$, namely, $\mathcal{L}(\mathrm{F_I}) = F(s, r-1, 0, 0)$, holds precisely as in Theorem 3.1 for every spherical linear regression model independently of $g(\cdot)$. In particular, this applies on specializing to the Student $t_i^2$ in single-case deletions.*

*Proof.* The conclusions follows directly from Theorem 1 of Jensen and Good (1981).

To complement earlier studies of the salinity data under single-case deletions as cited, we next undertake subset deletions, specifically, their normal-theory values, as these are invariant under both our mixture models and the linear spherical regression models. The salinity of water in Pamlico Sound, North Carolina, as reported in Ruppert and Caroll (1980), pertains to the linear model $\{Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon\}$ with $n = 28$ runs, taking $Y$ as the measured salinity, and explanatory variables $x_1$ as the salinity lagged by 2 weeks, $x_2$ as the river flow, and $x_3$ as a time trend variable. Specifically, we consider subsets of size $s = 2$ from $\{x_1, \ldots, x_{15}, x_{17}, \ldots, x_{28}\}$, having removed case 16 as reported consistently in the cited studies as outlying to excess. Specifically, $t_{16}^2 = 14.36$ with $p$ value $p = 0.0009$. Table 5.5

**Table 5.5.** Values for selected pairs from the salinity data $\{x_1, \ldots, x_{15}, x_{17}, \ldots, x_{28}\}$ with $Q_2 = 26.1516$ and $c_{0.05} = 3.47$.

| Subset | $Q_3$ | $S_{\mathrm{I}}^2$ | $\mathrm{F_I}$ | $p$ Value |
|---|---|---|---|---|
| $\{x_{15}, x_{17}\}$ | 14.5364 | 0.6922 | 8.39 | 0.0021 |
| $\{x_5, x_{15}\}$ | 15.3016 | 0.7286 | 7.45 | 0.0036 |
| $\{x_9, x_{15}\}$ | 17.5506 | 0.8357 | 5.15 | 0.0152 |
| $\{x_{13}, x_{15}\}$ | 17.8042 | 0.8478 | 4.92 | 0.0176 |
| $\{x_9, x_{17}\}$ | 17.8053 | 0.8479 | 4.92 | 0.0177 |
| $\{x_1, x_{15}\}$ | 17.9354 | 0.8541 | 4.81 | 0.0191 |
| $\{x_8, x_{15}\}$ | 17.9397 | 0.8543 | 4.81 | 0.0191 |
| $\{x_5, x_{17}\}$ | 18.1208 | 0.8629 | 4.65 | 0.0212 |
| $\{x_8, x_{17}\}$ | 18.6558 | 0.8842 | 4.22 | 0.0288 |
| $\{x_{15}, x_{28}\}$ | 18.7657 | 0.8936 | 4.13 | 0.0307 |
| $\{x_1, x_{17}\}$ | 18.7929 | 0.8949 | 4.11 | 0.0311 |
| $\{x_5, x_8\}$ | 19.0217 | 0.9058 | 3.94 | 0.0354 |

reports the twelve most influential pairs of data from the 351 possible pairs. Eleven of the 12 pairs contain either case 15 or case 17, both flagged under single-case deletions. An additional pair is $\{x_5, x_8\}$, although neither was detected using single-case deletions.

**Remark 5.1.** Proposition 5.1 extends to include a variety of single-case outlier and influence diagnostics that correspond one-to-one with $t_i^2$. These are established in Jensen (2000), and include many diagnostics set forth in Belsley, Kuh, and Welsch (1980) and related references.

In summary, the importance of the current findings rests heavily on invariance properties of diagnostics for outlying and influential observations. These apply for underlying distributions having correlated errors, for mixtures of these distributions, and for a large class of spherical regression models containing such heavy-tailed distributions as $n$-dimensional Cauchy errors.

### 5.7. *Further examples*

Additional cases are provided in Table 5.6. These include $(\lambda_1, \lambda_2, \beta)$ over a range of values of the parameters, where $\beta(\lambda_1, \lambda_2) = P(\mathrm{F_I} > c_\alpha)$ as before with $\alpha = 0.05$. These are arranged so that $\lambda_1$ decreases and $\lambda_2$ increases in both the left and right portions of Table 5.6.

It is seen that $F(u; \nu_1, \nu_2, \lambda_1, \lambda_2)$ increases stochastically with increasing $\lambda_1$, and decreases stochastically with increasing $\lambda_2$, with other parameters held fixed, in keeping with Proposition 2.1.

## 6. Conclusions

Gaussian observations $\boldsymbol{Y}_0 = [\boldsymbol{Y}', \boldsymbol{Y}_I']'$ subject to shifts $\{\boldsymbol{Y} \to \boldsymbol{Y} + \gamma\}$ and $\{\boldsymbol{Y_I} \to \boldsymbol{Y_I} + \boldsymbol{\delta}\}$ are considered. The use of the diagnostic $\mathrm{F_I}$ is reexamined in regard to $H_0 : \delta = \mathbf{0}$ against $H_1 : \delta \neq \mathbf{0}$. In addition, dispersion matrices $V(Y_0)$ of types $\{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\rho), \boldsymbol{\Sigma}(\xi)\}$ are taken in lieu of $\sigma^2 \boldsymbol{I}_n$, as are mixtures having star-shaped contours. These innovations substantially transcend the classical setting in which $\gamma = \mathbf{0}$ and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$, including the use of the $R$-Student diagnostic $t_i^2$ under single shifts $\{Y_i \to Y_i + \delta_i\}$ as in Jensen and Ramirez (2015).

**Table 5.6.** Noncentrality parameters $(\lambda_1, \lambda_2)$ together with $\beta(\lambda_1, \lambda_2) = P(\mathrm{F_I} > c_a)$ where $\lambda_1$ is decreasing and $\lambda_2$ increasing in each half of the table, and $a = 0.05$.

| $\lambda_1$ | $\lambda_2$ | $\beta$ | $\lambda_1$ | $\lambda_2$ | $\beta$ |
|---|---|---|---|---|---|
| Case $F(2, 2, \lambda_1, \lambda_2)$; $c_a = 19.000$ | | | | | |
| 0.88228 | 0.04496 | 0.06926 | 0.89064 | 0.03660 | 0.06972 |
| 0.83268 | 0.09460 | 0.06657 | 0.87284 | 0.05440 | 0.06874 |
| 0.81944 | 0.10780 | 0.06586 | 0.42688 | 0.50040 | 0.04748 |
| 0.80840 | 0.11884 | 0.06528 | 0.41844 | 0.50880 | 0.04714 |
| 0.47456 | 0.45268 | 0.04948 | 0.40320 | 0.52404 | 0.04652 |
| 0.25728 | 0.67000 | 0.04088 | 0.31616 | 0.61112 | 0.04309 |
| 0.24940 | 0.67788 | 0.04059 | 0.11876 | 0.80848 | 0.03601 |
| 0.16076 | 0.76648 | 0.03744 | 0.09420 | 0.83308 | 0.03520 |
| 0.12996 | 0.79728 | 0.03639 | 0.09248 | 0.83476 | 0.03514 |
| 0.08700 | 0.84024 | 0.03496 | 0.00028 | 0.92696 | 0.03220 |

The diagnostics $F_I$ are shown to be genuinely distribution-free, whereas irregularities trace to denominator noncentrality parameters stemming from $\{\boldsymbol{Y} \to \boldsymbol{Y} + \gamma\}$. Our models have been shown to explain the masking and swamping of outliers, and numerical studies serve to illustrate the essential findings. In short, $F_I$ now has been updated for use in data having structure substantially beyond the classical venue, as required on occasion in contemporary experimental settings under patently nonstandard conditions.

Structured dispersion matrices of earlier vintage include $\boldsymbol{\Sigma}(\lambda, \alpha) = (\lambda \mathbf{I}_n + \mathbf{1}_n \alpha' + \alpha 1'_n)$ of Baldessari (1966) in certain analysis-of-variance problems. In addition, Huynh and Feldt (1970) and Rouanet and Lepine (1970) characterized as $\boldsymbol{\Sigma}(\lambda) = [\lambda_i + \lambda_j + \delta_{ij}\lambda]$ the class of all within-subject dispersion matrices preserving validity of conventional $F$-tests in the analysis of $k$ repeated measurements on each of $n$ experimental subjects, with $\delta_{ij}$ as Kronecker's symbol. It is clear that the structures $\boldsymbol{\Sigma}(\lambda, \alpha)$, $\boldsymbol{\Sigma}(\lambda)$, and our $\boldsymbol{\Sigma}(\xi)$ are equivalent. Moreover, the Euclidean distance matrices of Gower (1982) have the structure $(D + \mathbf{1}_n \gamma' + \gamma \mathbf{1}'_n)$, with applications to linear inference as in Farebrother (1985).

## ORCID

Donald E. Ramirez http://orcid.org/0000-0002-2419-4538

## References

Aitkin, M., and G. Wilson. 1980. Mixture models, outliers, and the EM algorithm. *Technometrics* 22:325–31.

Andrews, D. F., and D. Pregibon. 1978. Finding outliers that matter. *Journal of the Royal Statistical Society, Series B* 40:85–93.

Baksalary, J. K., M. Nurhonen, and S. Puntanen. 1992. Effect of correlations and unequal variances in testing for outliers in linear regression. *Scandinavian Journal of Statistics* 19:91–95.

Baldessari, B. 1966. Analysis of variance of dependent data. *Statistica (Bologna)* 26:895–903.

Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. 1972. *Statistical Inference under Order Restrictions*. New York: Wiley Publishing Co.

Barnett, V., and T. Lewis. 1994. *Outliers in Statistical Data*, Third ed. New York: Wiley Publishing Co.

Beckett, J., 1977. Clustering rank preference data, in *ASA Proceedings of the Social Statist. Section*, pp. 983–86.

Beckman, R. J., and H. J. Trussell. 1974. The distribution of an arbitrary Studentized residual and the effects of updating in multiple regression. *Journal of the American Statistical Association* 69:199–201.

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

Bendre, S. M., and B. K. Kale. 1987. Masking effect on tests for outliers in normal samples. *Biometrika* 74:891–96.

Box, G. E. P., and G. C. Tiao. 1968. A Bayesian approach to some outlier problems. *Biometrika* 55:119–29.

Das Gupta, S., and M. D. Perlman. 1974. Power of the noncentral $F$–Test: Effect of additional variates on Hotelling's $T^2$–Test. *Journal of the American Statistical Association* 69:174–80.

Dixon, W. J. 1950. Analysis of extreme values. *Annals of Mathematical Statistics* 21:488–506.

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. Communications in statistics. *Theory and Methods* 22:897–905.

Farebrother, R. W. 1985. Linear unbiased approximators of the disturbances in the standard linear model. *Linear Algebra and Its Applications* 67:259–74.

Ferguson, T. S., 1961. On the rejection of outliers. In Proceedings of the 4th Berkeley Symposium, 1, 253–87.

Fisher, R. A. 1960. *The Design of Experiments*, *8th* ed. Edinburgh, UK: Oliver and Boyd.

Galea, M., M. Riquelme, and G. Paula. 2000. Diagnostic methods in elliptical linear regression models. *Brazilian Journal of Probability and Statistics* 14:167–84.

Gentleman, J. F., and M. 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. *Sankhya Series B* 40:204–08.

Gower, J. C. 1982. Euclidean distance geometry. *Mathematical Science* 7:1–14.

Grubbs, F. E. 1950. Criteria for testing outlying observations. *Annals of Mathematical Statistics* 21:27–58.

Hoaglin, D. C., and P. J. Kempthorne. 1986. Influential observations, high leverage points, and outliers in linear regression: Comment. *Statistical Science* 1:408–12.

Huynh, H., and L. S. Feldt. 1970. Conditions under which mean square ratios in repeated measurements designs have exact *F*–Distributions. *Journal of the American Statistical Association* 65:1582–89.

Imhof, J. 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* 48:419–26.

Jensen, D. R. 1989. Misspecified location and scale. I. Student's test. *Journal of Statistical Computation and Simulation* 32:115–18.

Jensen, D. R. 1996. Structured dispersion and validity in linear inference. *Linear Algebra and Its Applications* 249:189–96.

Jensen, D. R. 2000. `The use of Studentized diagnostics in regression. *Metrika* 52:213–23.

Jensen, D. R. 2001. Properties of selected subset diagnostics in regression. *Statistics and Probability Letters* 51:377–88.

Jensen, D. R., and I. J. Good. 1981. Invariant distributions associated with matrix laws under structural symmetry. *Journal of the Royal Statistical Society, B* 43:327–32.

Jensen, D. R., and D. E. Ramirez. 2009. Anomalies in the analysis of calibrated data. *Journal of Statistical Computation and Simulation* 79:299–314.

Jensen, D. R., and D. E. Ramirez. 2012. Irregularities in X(Y) from Y(X) in linear regression. *Journal of Statistical Computation and Simulation* 83:1807–28.

Jensen, D. R., and D. E. Ramirez. 2015. Outlier detection under star–contoured errors. *Journal of Statistical Theory and Practice* 9:850–67.

Mathai, A. M., and S. B. Provost. 1992. *Quadratic Forms in Random Variables: Theory and Applications*. New York: Marcel Dekker Publishing Co.

Morrison, D. F. 2005. *Multivariate Statistical Methods*, *Fourth* ed. Belmont, CA: Brooks/Cole Thomson Learning.

Punzo, A., R. P. Browne, and P. D. McNicholas. 2016. Hypothesis testing for model selection. *Journal of Statistical Computation and Simulation* 86:2797–818.

Rouanet, H., and D. Lépine. 1970. Comparison between treatments in a repeated–measurement design: ANOVA and multivariate methods. *British Journal of Mathematical Statistics and Psychology* 23:147–63.

Ruppert, D., and R. J. Caroll. 1980. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75:828–38.
Snedecor, G. W., and W. G. Cochran. 1968. *Statistical Methods*, Sixth ed. Ames, IA: Iowa State University Press.
Srivastava, M. S. 1980. Effect of equicorrelation in detecting a spurious observation. *Canadian Journal of Statistics* 8:249–51.
Woodward, W. F. (1970). *A Comparison of Base Running Methods in Baseball*, M.S. Thesis, Florida State University.
Young, D. M., R. Pavur, and V. R. Marco. 1989. On the effect of correlation and unequal variances in detecting a spurious observation. *Canadian Journal of Statistics* 17:103–05.

## Appendix A

### A.1. Foundations

As noted, the general model $\{\boldsymbol{Y}_0 = \boldsymbol{X}_0\beta + \epsilon_0\}$ is specialized here to $\{\boldsymbol{Y}_0 = \mu\mathbf{1}_n + \epsilon_0\}$, to be partitioned as $\boldsymbol{Y}_0' = [\boldsymbol{Y}', \boldsymbol{Y}_I']$, then deleting $[\boldsymbol{Y}_I, \mathbf{1}_s, \epsilon_I]$ while retaining $\{\boldsymbol{Y} = \mu\mathbf{1}_r + \epsilon\}$. Recall that $\boldsymbol{B}_n = (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n)$, then partition $\boldsymbol{B}_n$, $\boldsymbol{\Sigma}(\theta)$, and $\boldsymbol{\Sigma}(\xi)$ conformably as

$$\boldsymbol{B}_n = \begin{bmatrix} (\boldsymbol{I}_r - \frac{1}{n}\mathbf{1}_r\mathbf{1}_r') & -\frac{1}{n}\mathbf{1}_r\mathbf{1}_s' \\ -\frac{1}{n}\mathbf{1}_s\mathbf{1}_r' & (\mathbf{I}_s - \frac{1}{n}\mathbf{I}_s\mathbf{I}_s') \end{bmatrix}; \tag{6}$$

$$\Sigma(\theta) = \begin{bmatrix} (\mathbf{I}_r + \theta\mathbf{1}_r\mathbf{1}_r') & \theta\mathbf{1}_r\mathbf{1}_s' \\ \theta\mathbf{1}_s\mathbf{1}_r' & (\mathbf{I}_s + \theta I_s I_s') \end{bmatrix}; \tag{7}$$

$$\Sigma(\xi) = \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} \\ \mathbf{0}' & I_s \end{bmatrix} + \begin{bmatrix} \mathbf{1}_r \\ \mathbf{1}_s \end{bmatrix}\xi' + \xi[\,\mathbf{1}'_r, \quad \mathbf{1}'_s\,] - \bar{\xi}\begin{bmatrix} 1_r\mathbf{1}'_r & \mathbf{1}_r\mathbf{1}'_s \\ \mathbf{1}_s\mathbf{1}'_r & \mathbf{1}_s\mathbf{1}'_s \end{bmatrix}. \tag{8}$$

A critical step is the Fisher–Cochran expansion $\boldsymbol{Y}'_0\boldsymbol{A}_1\boldsymbol{Y}_0 + \boldsymbol{Y}'_0\boldsymbol{A}_2\boldsymbol{Y}_0 = \boldsymbol{Y}'_0\boldsymbol{A}_3\boldsymbol{Y}_0$ for quadratic forms in $\mathbf{Y}_0$, specifically $Q_1 + Q_2 = Q_3$, such that $\mathbf{A}_3 = \mathbf{B}_n$; $\mathbf{A}_2 = Diag(\mathbf{B}_r, \mathbf{0})$ and

$$\boldsymbol{A}_1 = \boldsymbol{A}_3 - \boldsymbol{A}_2 = \begin{bmatrix} \frac{s}{rn}\mathbf{1}_r\mathbf{1}'_r & -\frac{1}{n}\mathbf{1}_r\mathbf{1}'_s \\ -\frac{1}{n}\mathbf{1}_s\mathbf{1}'_r & (\mathbf{I}_s - \frac{1}{n}\mathbf{1}_s\mathbf{1}'_s) \end{bmatrix}; \tag{9}$$

see Lemma A.1(iii) of Jensen (2001). We draw from the work of Mathai and Provost (1992, 201), taking into account the special structure of $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\xi)\}$, where $\boldsymbol{\Sigma}(\rho)$ follows by equivalence with $\boldsymbol{\Sigma}(\theta)$. Essentials follow.

**Theorem A.1.** *Suppose that* $\mathcal{L}(\mathbf{Y}_0) = N_n(\tau, \Sigma)$; *consider the quadratic forms* $\{Q_1, Q_2, Q_3\}$ *such that* $Q_1 + Q_2 = Q_3$ *and* $\mathrm{F_I} = (r-1)Q_1/sQ_2$. *Then for each* $\Sigma \in \{\Sigma(\theta), \Sigma(\xi)\}$ *we have*

*(i)* $A_i\Sigma A_i = A_i$ *for* $i \in \{1, 2, 3\}$;
*(ii)* $\mathbf{A}_1\Sigma\mathbf{A}_2 = \mathbf{0}$;
*(iii)* $\mathcal{L}(Q_i) = \chi^2(v_i; \lambda_i)$ *with* $v_i = tr(\mathbf{A}_i\Sigma)$ *and* $\lambda_i = \tau'\mathbf{A}_i\tau$ *for* $i \in \{1, 2, 3\}$;
*(iv) Degrees of freedom for* $\{Q_1, Q_2, Q_3\}$ *are* $v_i \in \{s, r-1, n-1\}$, *respectively;*
*(v)* $Q_1$ *and* $Q_2$ *are distributed independently;*
*(vi)* $\mathcal{L}(\mathrm{F_I}) = F(u; s, r-1, \lambda_1, \lambda_2)$, *with* $\lambda_1$ *and* $\lambda_2$ *as in Theorem 1;*
*(vii) Properties (iii)–(vi) hold independently of* $\Sigma \in \{\Sigma(\theta), \Sigma(\xi)\}$, *and are identical to the case that* $\sum = \sigma^2\mathbf{I}_n$.

Proof. Theorem 5.1.4 of Mathai and Provost (1992, 201) identifies (i) to be necessary and sufficient for (iii), and similarly (ii) for (v).

Case 1: First consider $\boldsymbol{A}_3 = \boldsymbol{B}_n$. Then $\boldsymbol{A}_3\Sigma(\theta) = \boldsymbol{B}_n(\boldsymbol{I}_n + \theta\mathbf{1}_n\mathbf{1}_n{}') = \boldsymbol{B}_n$, so that $tr(\boldsymbol{A}_3\Sigma(\theta)) = n-1$ as in conclusion (iv). Moreover, $\boldsymbol{A}_3\Sigma(\theta)\boldsymbol{A}_3 = \boldsymbol{A}_3{}^2 = \boldsymbol{A}_3$ as in conclusion (i) for $\{i = 3\}$ since $\boldsymbol{A}_3$ is idempotent. Similarly,

$$\boldsymbol{A}_3\Sigma(\xi) = \boldsymbol{B}_n(\boldsymbol{I}_n + \mathbf{1}_n\xi' + \xi\mathbf{1}_n{}' - \bar{\xi}\mathbf{1}_n\mathbf{1}_n{}') = \boldsymbol{B}_n + \boldsymbol{B}_n\xi\mathbf{1}_n{}' \tag{10}$$

so that $tr(\boldsymbol{A}_3\Sigma(\xi)) = n-1$ follows as in conclusion (iv) since $\mathrm{tr}(B_n\xi\mathbf{1}'_n) = \mathrm{tr}(\xi\mathbf{1}'_n\boldsymbol{B}_n) = 0$. Moreover, $\boldsymbol{A}_3\Sigma(\xi)\boldsymbol{A}_3 = \boldsymbol{A}_3$ as in conclusion (i).

Case 2: For $A_2 = \mathrm{Diag}(B_r, \mathbf{0})$, the partitioned form (7) gives

$$\boldsymbol{A}_2\Sigma(\theta) = \begin{bmatrix} \boldsymbol{B}_r & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} \begin{bmatrix} (\mathbf{I}_r + \theta\mathbf{1}_r\mathbf{1}_r{}') & \theta\mathbf{1}_r\mathbf{1}'_s \\ \theta\mathbf{1}_s\mathbf{1}_r{}' & (\mathbf{I}_s + \theta\mathbf{1}_s\mathbf{1}'_s) \end{bmatrix} = \begin{bmatrix} \mathbf{B}_r & \mathbf{0} \\ \mathbf{0}' & \mathbf{0} \end{bmatrix} = \mathbf{A}_2 \tag{11}$$

and $tr(\boldsymbol{A}_2\Sigma(\theta)) = r-1$ as in conclusion (iv). In parallel with Case 1, $\boldsymbol{A}_2\Sigma(\theta)\boldsymbol{A}_2 = \boldsymbol{A}_2$ to give conclusion (i) for $(i = 2)$. To continue, in the partitioned form (8) we have

$$\boldsymbol{A}_2\Sigma(\xi) = \begin{bmatrix} \boldsymbol{B}_r & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{0} \end{bmatrix} \left\{ \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{I}_s \end{bmatrix} + \begin{bmatrix} \boldsymbol{1}_r \\ \boldsymbol{1}_s \end{bmatrix} \xi' + \xi[\boldsymbol{1}'_r \quad \boldsymbol{1}'_s] - \bar{\xi} \begin{bmatrix} \boldsymbol{1}_r\boldsymbol{1}'_r & \boldsymbol{1}_r\boldsymbol{1}'_s \\ \boldsymbol{1}_s\boldsymbol{1}'_r & \boldsymbol{1}_s\boldsymbol{1}'_s 1 \end{bmatrix} \right\}$$

$$= \boldsymbol{A}_2 + \boldsymbol{A}_2\xi\mathbf{1}_n{}' \tag{12}$$

and $tr(\boldsymbol{A}_2\Sigma(\xi)) = r-1$ as in conclusion (iv). Moreover, $A_2\Sigma(\xi)A_2 = A_2$ to give conclusion (i) for $(i = 2)$.

Case 3: Next consider $A_1 = A_3 - A_2$. From the foregoing developments infer that $A_1\Sigma(\theta) = (A_3 - A_2)\Sigma(\theta) = A_3 - A_2$ so that $tr(A_1\Sigma(\theta)) = (n-1) - (r-1) = n - r = s$ as in conclusion (iv), whereas $A_1\Sigma(\theta)A_1 = A_1$ to verify conclusion (i) for $\Sigma = \Sigma(\theta)$. Similarly $tr(A_1\Sigma(\xi)) = (n-1) - (r-1) = s$ as in conclusion (iv), and $\mathbf{A}_1\Sigma(\xi)\mathbf{A}_1 = \mathbf{A}_1$ as in conclusion (i) for $(i = 1)$ and $\Sigma = \Sigma(\xi)$.

Case 4: To verify conclusion (ii) separately for $\Sigma(\theta)$ and $\Sigma(\xi)$, take $\boldsymbol{A}_3\Sigma(\theta) = \boldsymbol{A}_3$ from Case 1 and $\boldsymbol{A}_2\Sigma(\theta) = \boldsymbol{A}_2$ from Eq. (11), so that $\boldsymbol{A}_1\Sigma(\theta)\boldsymbol{A}_2 = (\boldsymbol{A}_3 - \boldsymbol{A}_2)\boldsymbol{A}_2$, namely,

$$(\boldsymbol{A}_3 - \boldsymbol{I}_n)\boldsymbol{A}_2 = -\frac{1}{n} \begin{bmatrix} \mathbf{1}_r\mathbf{1}_r{}' & \mathbf{1}_r \\ \mathbf{1}_r{}' & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_r & 0 \\ 0 & 0) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{13}$$

To continue, from Eq. (12) write $\boldsymbol{A}_2\Sigma(\xi)\boldsymbol{A}_1 = (\boldsymbol{A}_2 + \boldsymbol{A}_2\xi\mathbf{1}_n{}')\boldsymbol{A}_1 = \boldsymbol{A}_2\boldsymbol{A}_1 = \mathbf{0}$ since $\mathbf{1}_n{}'A_1 = \mathbf{0}$ and $\boldsymbol{A}_2\boldsymbol{A}_1 = \mathbf{0}$ from Eq. (13). Conclusion (v) follows as a consequence of (ii) and Craig's Theorem, as in Mathai and Provost (1992, 209).

**Table A.1.** Subset deletion diagnostics pertaining to the model $\{Y_0 = \mu + \epsilon_0\}$, where $Y_0{}' = [Y', Y_I{}']$ are of orders $[(1 \times n), (1 \times r), (1 \times s)]$, respectively.

| Diagnostic | Expression | Rule | Critical value | Range |
|---|---|---|---|---|
| $F_I$ | $\frac{\mathbf{e}_I{}'(I_s + \frac{1}{r}\mathbf{1}_s\mathbf{1}_s{}')\mathbf{e}_I}{sS_I^2}$ | $>$ | $c_a$ | $[0, \infty)$ |
| $OUT_I$ | $1 - \frac{S_I^2}{S^2}$ | $>$ | $\frac{[s(c_a-1)]}{[sc_a+n-s-1]}$ | $[\frac{-s}{n-s-1}, 1]$ |
| $AP_I$ | $\left[1 - \frac{r(n-s-1)S_I^2}{n(n-1)S^2}\right]$ | $>$ | $\frac{[sc_a+(n-s-1)s/n]}{[sc_a+n-s-1]}$ | $[\frac{s}{n}, 1]$ |
| $CR_I$ | $\frac{nS_I^2}{rS^2}$ | $<$ | $\left[\frac{n(n-1)}{r(sc_a+n-s-1)}\right]$ | $[0, \frac{n\kappa}{r}]$ |
| $FV_I$ | $\frac{n}{r}\left[\frac{S_I^2}{S^2}\right]^s$ | $<$ | $\frac{n}{r}\left[\frac{(n-1)}{(sc_a+n-s-1)}\right]^s$ | $[0, \frac{n\kappa^s}{r}]$ |
| | | $\kappa = \frac{(n-1)}{(sc_a+n-s-1)}$ | | |

## A.2. Deletion diagnostics: a survey

Regression diagnostics for the model $\{\mathbf{Y}_0 = \mu + \epsilon_0\}$ are surveyed next for completeness. In addition to *outliers*, observations whose removal would alter essentials of the analysis are designated as influential. In particular, points with large $\mathrm{OUT_I}$ of Barnett and Lewis (1994) are tagged as outlying, and with large $\mathrm{AP_I}$ of Andrews and Pregibon (1978) as influential. Of further interest here are the influence diagnostics $\mathrm{CR_I}$ (also $\mathrm{COVRATIO}_I$) and $\mathrm{FV_I}$ (also $\mathrm{HVARATIO}_I$) of Belsley, Kuh, and Welsch (1980). These are listed in Table A.1 together with rejection rules and their critical values, as well as the range for each diagnostic. These entries are found on specializing from Table 2 of Jensen (2001) so as to apply to the model at hand. A significant finding is that these are intimately interlaced: Each corresponds one-to-one with $\mathrm{F_I}$, as shown in Jensen (2001).

## A.3. *Calculations of p values*

Building on the earlier work of Imhof (1961), Ennis and Johnson (1993) have expressed the *cdf* for $F(u; \nu_1, \nu_2, \lambda_1, \lambda_2)$ as a one-dimensional integral using trigonometric functions. This result is straightforward to code, for example, in Maple, and the Ennis and Johnson representation for the *cdf* was used to compute the doubly noncentral probabilities as reported. In addition, probabilities for the singly noncentral $F(u; \nu_1, \nu_2, \lambda_1, 0)$ made use of the following online source: Dr. Daniel Soper, *Free Statistical Calculator, Version 4.0*, © 2006–2016, at http://www.danielsoper.com.