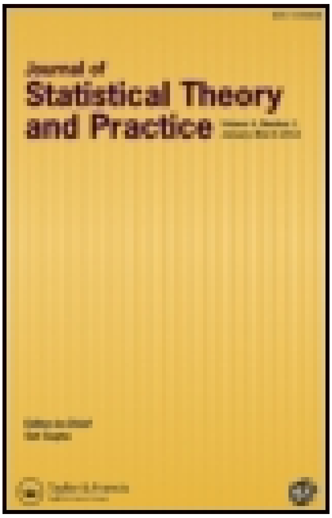

Journal of Statistical Theory and Practice
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ujsp20

Outlier Detection Under Star-Contoured Errors

D. R. Jensen[a] & D. E. Ramirez[b]
[a] Department of Statistics, Virginia Tech, Blacksburg, Virginia, USA
[b] University of Virginia, Charlottesville, Virginia, USA
Accepted author version posted online: 28 Apr 2015.Published online: 28 Apr 2015.

Journal of Statistical Theory and Practice, 00:1–18, 2015

ISSN: 1559-8608 print / 1559-8616 online
DOI: 10.1080/15598608.2015.1034386

Outlier Detection Under Star-Contoured Errors

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

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

Consider $[Y_1, \ldots, Y_n]$ as Gaussian observations with common mean μ and dispersion matrix $\boldsymbol{\Sigma}$. Approaches for detecting outlying observations include the R–Student statistics in regression diagnostics, as well as tests due to Grubbs, Dixon, and Ferguson using order statistics. All are known to be valid under $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$; Grubbs's test also holds under an equicorrelated matrix $\boldsymbol{\Sigma}(\rho)$ and the more general structure $\boldsymbol{\Sigma}\ (\boldsymbol{\xi}) = \sigma^2[\boldsymbol{I}_n + 1_n\boldsymbol{\xi}' + \boldsymbol{\xi}\mathbf{1}_n' - \bar{\boldsymbol{\xi}}1_n\mathbf{1}_n']$. Dispersion mixtures of Gaussian errors having $\boldsymbol{\Sigma}(\rho)$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi})$ are studied in detail; their densities have star-shaped contours as encountered on occasion in practice. Under these mixtures, the aforementioned diagnostics all are shown to be exact in significance level and in power as for the case where $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$. This expands considerably their range of applicability in practice. Case studies serve to illustrate essentials of the findings.

AMS Subject Classification: 62E10; 62F04.

Keywords: Outlying data; Deletion diagnostics; Order statistics; Structured errors.

1. Introduction

Given observations $\boldsymbol{Y}_0' = [Y_1, ..., Y_n]$ with common mean μ, a recurring question in statistical practice is to identify outlying data. Here the data are modeled as coming from $\{\boldsymbol{Y}_0 = \mu\mathbf{1}_n + \boldsymbol{\varepsilon}_0\}$ with errors having mean $\mathrm{E}\,(\boldsymbol{\varepsilon}_0) = \mathbf{0}$ and dispersion matrix $\mathrm{V}(\boldsymbol{\varepsilon}_0) = \boldsymbol{\Sigma}$. An outlier in row i of $\boldsymbol{Y}_0$ is modeled as a shift $\{Y_i \to Y_i + \delta\}$ in the response Y_i. Standard outlier diagnostics for testing H_0: $\delta = 0$ have been derived under Gaussian errors having $\boldsymbol{\Sigma}= \sigma^2\boldsymbol{I}_n$. These include the R-Student deletion test of Snedecor and Cochran (1968) and tests due to Grubbs (1950), Dixon (1950), and Ferguson (1961) based on the order statistics. Grubbs's (1950) test has been found to remain exact in level of significance and in power, under the intraclass correlation model $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\rho)$, as shown in Srivastava (1980). Young et al. (1989) show the more general structure $\boldsymbol{\Sigma}\ (\boldsymbol{\xi}) = \sigma^2[\boldsymbol{I}_n + \mathbf{1}_n\boldsymbol{\xi}' + \boldsymbol{\xi}\mathbf{1}_n' - \bar{\boldsymbol{\xi}}\mathbf{1}_n\mathbf{1}_n']$ to be sufficient for validity under Gaussian errors, while Baksalary and Puntanen (1990) show it is also necessary. Accordingly, Grubbs's test applies much more widely in practice, whereas the R-Student test and those of Dixon (1950) and Ferguson (1961) so far only apply strictly under $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$.

In this study we show that the R-Student test and the tests of Dixon (1950) and Ferguson (1961) apply also under $\boldsymbol{\Sigma}(\rho)$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi})$. In addition, the aforementioned tests

Received 14 May 2014; accepted 23 March 2015.
Address correspondence to: D. E. Ramirez, Department of Mathematics, University of Virginia, PO Box 400137, Charlottesville, VA 22904, USA. Email: der@virginia.edu

all are shown to apply under scale mixtures of Gaussian families, with each having the dispersion matrix $\boldsymbol{\Sigma}(\rho)$ or $\boldsymbol{\Sigma}(\boldsymbol{\xi})$. Gaussian mixtures arise in practice, to include data collected from subsamples as in Box and Tiao (1968) and Aitken and Wilson (1980). Their density functions as studied here have star-shaped contours.

For motivation, DasGupta (2013) highlights Grubbs's (1950) test as among 215 influential developments in statistics. Accordingly, users will value its substantially enhanced capability. At the same time, our tools also support the wider applicability of the *R*–Student diagnostic and those of Dixon (1950) and Ferguson (1961). An outline of the article follows.

Preliminaries in section 2 extend the conventional models to include mixture distributions. Section 3 develops nonstandard properties of the residuals under these models, and of tests based on the order statistics and the *R*–Student diagnostic. Uses are illustrated in section 4. Essential supporting topics are attached as an appendix, including distributions of quadratic forms under dependent errors.

2. Preliminaries

2.1. Notation

Spaces 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 $\boldsymbol{A}'$, $\boldsymbol{A}^{-1}$, $\text{tr}(\boldsymbol{A})$, and $|\boldsymbol{A}|$; $\boldsymbol{I}_n$ is the $(n\times n)$ identity; and $\text{Diag}(\boldsymbol{A}_1,\ldots,\boldsymbol{A}_k)$ is a block-diagonal array. If $\boldsymbol{B} = [\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k]$ is of order $(n\times k)$ and rank $R_k(\boldsymbol{B}) = k < n$, then $S_p(\boldsymbol{B})$ designates the column span of $\boldsymbol{B}$, that is, the k-dimensional subspace of $\mathbb{R}^n$ spanned by $[\boldsymbol{b}_1,\ldots,\boldsymbol{b}_k]$. Throughout, we designate the idempotent matrices $\boldsymbol{B}_n = \left(\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n'\right)$ and $\boldsymbol{B}_r = \left(\boldsymbol{I}_r - \frac{1}{r}\boldsymbol{1}_r\boldsymbol{1}_r'\right)$.

A random $\boldsymbol{Y} \in \mathbb{R}^n$ has distribution $\mathcal{L}(\boldsymbol{Y})$; mean vector $\text{E}(\boldsymbol{Y})$; dispersion matrix $\text{V}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$, say, with variance $\text{Var}(Y) = \sigma^2$ on $\mathbb{R}^1$; a density function (probability distribution function, *pdf*) $g(\boldsymbol{y})$; and a cumulative distribution function (*cdf*) $G(\boldsymbol{y})$. Specifically, $\mathcal{L}(\boldsymbol{Y}) = N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is Gaussian on $\mathbb{R}^n$ with mean and dispersion matrix $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Distributions on $\mathbb{R}^1_+$ include $\chi^2(\nu;\lambda)$ as chi-squared having ν degrees of freedom and noncentrality λ, and the noncentral Student's t^2 distribution $t^2(\nu;\lambda_1)$, with numerator noncentrality λ_1. More generally, $t^2(\nu;\lambda_1,\lambda_2)$ is doubly noncentral, having ν degrees of freedom and numerator and denominator noncentralities (λ_1,λ_2). Specifically, if $t_i^2 = \nu U/V$, such that (U, V) are independent, where $\mathcal{L}(U) = \chi^2(1,\lambda_1)$, and $\mathcal{L}(V) = \chi^2(\nu;\lambda_2)$, then $\mathcal{L}\left(t_i^2\right) = t^2(\nu,\lambda_1,\lambda_2)$. Moreover, $t^2(\nu;\lambda_1,\lambda_2)$ increases stochastically with increasing λ_1, and decreases stochastically with increasing λ_2, with other parameters held fixed. See Nandi and Choudhury (2002) and references cited there in. Identify $\left\{t_i^2 > c_\alpha\right\}$ as the conventional α-level rejection rule based on $t^2(\nu;0,0)$.

2.2. The Model

Taking $\{\boldsymbol{Y}_0 = \mu\boldsymbol{1}_n + \boldsymbol{\varepsilon}_0\}$ as reference is a special case of $\{\boldsymbol{Y}_0 = \boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0\}$. Several essential properties in the present study follow directly. The pivotal matrix $\boldsymbol{H}_n = \boldsymbol{X}_0\left(\boldsymbol{X}_0'\boldsymbol{X}_0\right)^{-1}\boldsymbol{X}_0'$ from linear inference here becomes $\boldsymbol{H}_n = \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}'_n$, so that diagonal *leverages,* and off-diagonal *coleverages,* are all equal to $\left\{h_{ij} = \frac{1}{n}; \forall i,j\right\}$. Further designate by $\boldsymbol{e}' = [e_1,\ldots,e_n]$ the residual vector.

Deletion diagnostics follow on rearranging $\boldsymbol{Y}_0' = \left[\boldsymbol{Y}', Y_i\right]$ and $\boldsymbol{\varepsilon}' = \left[\boldsymbol{\varepsilon}', \boldsymbol{\varepsilon}_j\right]$; on deleting $[Y_i, 1, \varepsilon_i]$ from $\{\boldsymbol{Y}_0 = \mu\boldsymbol{1}_n + \boldsymbol{\varepsilon}_0\}$; on retaining $\{\boldsymbol{Y} = \mu\boldsymbol{1}_r + \boldsymbol{\varepsilon}\}$ with $\boldsymbol{Y} \in \mathbb{R}^r$ and $r = n - 1$;

and letting S_i^2 be the residual mean square from $\boldsymbol{Y}$. Central to this study are the *R*–Student statistics $\left\{t_i^2 = e_i^2/S_i^2\,(1-h_{ii})\right\}$ of Snedecor and Cochran (1968, 157) for identifying a single shift, as considered also in Beckman and Trussell (1974). These specialize here to $\left\{t_i^2 = ne_i^2/rS_i^2\right\}$, known to have Student's t^2 distribution under Gaussian errors with $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$ as default. In addition, for arbitrary shifts $\{\boldsymbol{Y}_0 \to \boldsymbol{Y}_0 + \boldsymbol{\omega}\}$ corresponding to $\boldsymbol{Y}_0 = \left[\boldsymbol{Y}', Y_i\right]'$, the shift vector ω is partitioned as $\boldsymbol{\omega} = \left[\boldsymbol{\gamma}', \delta_i\right]'$, of order $(n \times 1)$, with $\boldsymbol{\gamma} \in \mathbb{R}^r$ and $\delta_i \in \mathbb{R}^1$. Gauss–Markov assumptions are extended here as follows, with matrices $\boldsymbol{\Sigma}\,(\theta) = \sigma^2\left(\boldsymbol{I}_n + \theta\boldsymbol{1}_n\boldsymbol{1}'_n\right)$ and $\boldsymbol{\Sigma}\,(\boldsymbol{\xi}) = \sigma^2\left(\boldsymbol{I}_n + \boldsymbol{1}_n\boldsymbol{\xi}' + \boldsymbol{\xi}\,\boldsymbol{1}'_n - \bar{\xi}\boldsymbol{1}_n\boldsymbol{1}'_n\right)$ as in section 2.3, often standardized to $\sigma^2 = 1$.

Assumptions A. The following hold:

A_1. $\mathrm{E}(\boldsymbol{\varepsilon}_0) = \boldsymbol{\omega}$; *i.e.* $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\gamma} \in \mathbb{R}^r$ and $\mathrm{E}(\boldsymbol{\varepsilon}_i) = \delta_i \in \mathbb{R}^1$.
A_2. $\mathrm{V}(\boldsymbol{\varepsilon}_0) = \boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}\,(\theta), \boldsymbol{\Sigma}\,(\boldsymbol{\xi})\}$
A_3. $\mathcal{L}\,(\boldsymbol{\varepsilon}_0) = N_n\,(\boldsymbol{\omega}, \boldsymbol{\Sigma})$ for $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}\,(\theta), \boldsymbol{\Sigma}\,(\boldsymbol{\xi})\}$.

Conventional deletion diagnostics model an outlying entry in row *i* as a shift in the response $\{Y_i \to Y_i + \delta_i\}$ as noted, and allowing no other shifts, that is, $\gamma = 0$ with $\omega' = \left[0', \delta_i\right]$. Our procedures allow for shifts in any row with $\{\boldsymbol{Y} \to \boldsymbol{Y} + \gamma\}$ and $\gamma \neq 0$. A key feature of this study is to decompose $\boldsymbol{\gamma}$ into its fundamental components, as in the following.

Definition 1. *(i) The decomposition $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2$ entails projections $\boldsymbol{\gamma}_1 = \frac{1}{r}\boldsymbol{1}_r\boldsymbol{1}'_r\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_2 = \left(\boldsymbol{I}_r - \frac{1}{r}\boldsymbol{1}_r\boldsymbol{1}'_r\right)\boldsymbol{\gamma}$ into the "Regressor" and "Error" spaces generated by $\{\boldsymbol{Y} = \mu\boldsymbol{1}_r + \boldsymbol{\varepsilon}\}$. (ii) Angles between $\boldsymbol{\gamma}$ and these projections are $\theta_1 = \theta(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1)$ and $\theta_2 = \theta(\boldsymbol{\gamma}, \boldsymbol{\gamma}_2)$.*

2.3. *The Matrices* Ξ

Structured dispersion matrices arise in studies of validity in linear inference. Three cases are considered where, for $\boldsymbol{\xi}' = [\xi_1, ..., \xi_n]$, we have $\tau_1 = \xi_1 + ... + \xi_n = n\bar{\xi}$ and $\tau_2 = \sum_{i=1}^n \left(\xi_i - \bar{\boldsymbol{\xi}}\right)^2$. Details follow, where

$$\alpha_1 = \frac{1}{2}\left[\tau_1 + \left(\tau_1^2 + 4n\tau_2\right)^{\frac{1}{2}}\right] \text{ and } \alpha_n = \frac{1}{2}\left[\tau_1 - \left(\tau_1^2 + 4n\tau_2\right)^{\frac{1}{2}}\right]. \tag{1}$$

Lemma 2.1. *(i) Let $\boldsymbol{\Sigma}\,(\theta) = \sigma^2\left(\boldsymbol{I}_n + \theta\boldsymbol{1}_n\boldsymbol{1}'_n\right)$; its eigenvalues are 1.0, with multiplicity $n-1$, and $1 + n\theta$, so that $\boldsymbol{\Sigma}(\theta)$ is positive definite if and only if $\theta \in \Gamma_1 = \left\{\theta : \theta > -\frac{1}{n}\right\}$.*

(ii) Let $\boldsymbol{\Sigma}(\boldsymbol{\xi}) = \sigma^2\left(\boldsymbol{I}_n + \boldsymbol{1}_n\boldsymbol{\xi}' + \boldsymbol{\xi}\boldsymbol{1}'_n - \bar{\boldsymbol{\xi}}\boldsymbol{1}_n\boldsymbol{1}'_n\right)$ with $\boldsymbol{0} \neq \boldsymbol{\xi} \neq \theta\boldsymbol{1}_n$; its ordered eigenvalues are $\{\kappa_1 = 1 + \alpha_1, \kappa_2 = \ldots = \kappa_{n-1} = 1, \kappa_n = 1 + \alpha_n\}$ as in Eq. (1); then $\boldsymbol{\Sigma}(\boldsymbol{\xi})$ is positive definite if and only if $\boldsymbol{\xi} \in \Gamma_2 = \{\boldsymbol{\xi} \in \mathbb{R}^n : \tau_1 > n\tau_2 - 1\}$.

(iii) Let $\boldsymbol{\Sigma}(\rho) = \sigma^2\left[(1-\rho)\,\boldsymbol{I}_n + \rho\boldsymbol{1}_n\boldsymbol{1}'_n\right]$, the equicorrelated case; then $\boldsymbol{\Sigma}(\rho)$ is positive definite if and only if $\rho \in \Gamma_3 = \left\{\rho : -\frac{1}{n-1} < \rho < 1\right\}$.

Proof. Details are given in Jensen (1996). □

Clearly $\boldsymbol{\Sigma}(\theta)$ and $\boldsymbol{\Sigma}(\rho)$ are equivalent. For if $\mathcal{L}(Z) = N_n\left(\frac{\mu}{k}\boldsymbol{1}_n, \boldsymbol{\Sigma}\,(\theta)\right)$, make the change of scale $\{\boldsymbol{Z} \to \boldsymbol{Y} = k\boldsymbol{Z}\}$; then $\mathrm{V}(\boldsymbol{Y}) = k^2\left(\boldsymbol{I}_n + \theta\boldsymbol{1}_n\boldsymbol{1}'_n\right)$. Next taking $k^2 = (1-\rho)$ and $k^2\theta = \rho$, infer that $\theta = \rho/(1-\rho)$, and $\boldsymbol{\Sigma}(\theta) = \frac{1}{1-\rho}\boldsymbol{\Sigma}(\rho)$. We take $\boldsymbol{\Sigma}(\theta)$ for convenience, despite the fact that $\boldsymbol{\Sigma}(\theta)$ occurs prominently in practice, to include calibrated measurements as in Jensen and Ramirez (2009; 2012). Accordingly, the collections

$\Xi_1 = \{\boldsymbol{\Sigma}(\theta); \theta \in \Gamma_1\}$ and $\Xi_2 = \{\boldsymbol{\Sigma}(\boldsymbol{\xi}); \boldsymbol{\xi} \in \Gamma_2\}$ are ensembles of positive definite matrices, and take $\Xi = \Xi_1 \cup \Xi_2$. For further details see Jensen (1996).

2.4. Mixture Distributions

Take $\boldsymbol{Y} \in \mathbb{R}^n$ having some distribution $\mathcal{L}(\boldsymbol{Y}) = \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; it suffices to center $\mathcal{L}(\boldsymbol{Y} - \boldsymbol{\mu})$ at $\mathbf{0} \in \mathbb{R}^n$. We have the following.

Definition 2. (i) A set $\mathcal{A} \subset \mathbb{R}^n$ is said to be *symmetric about* $\mathbf{0} \in \mathbb{R}^n$ if $x \in \mathcal{A}$ implies $-x \in \mathcal{A}$ for each $x \in \mathcal{A}$.

(ii) A set $\mathcal{S} \subset \mathbb{R}^n$ containing $\mathbf{0}$ is called *star-shaped* about $\mathbf{0}$ if, for every $x \in \mathcal{S}$, the line segment joining $\mathbf{0}$ to $\mathbf{x}$ is in $\mathcal{S}$.

(iii) A distribution $\mathrm{P}(\cdot)$ on $\mathbb{R}^n$ is called *symmetric star-unimodal* about $\mathbf{0}$ if it belongs to the closed convex hull of probability measures uniform on sets symmetric and star-shaped about $\mathbf{0}$; we designate by $P_n(\mathbf{0})$ the class of these distributions.

Essential properties follow. If $P(\cdot)$ has a continuous density $f(\cdot)$ on $\mathbb{R}^n$, then $P(\cdot) \in P_n(\mathbf{0})$ if and only if the level sets $B_t = \{x \in \mathbb{R}^n : f(x) > t > 0\}$ are either symmetric star-shaped about $\mathbf{0} \in \mathbb{R}^n$, or are empty. Kanter's (1977) class $K_n(\mathbf{0})$ consists of mixtures on $\mathbb{R}^n$ generated as the closed convex hull of measures uniform on convex bodies that are symmetric about $\mathbf{0} \in \mathbb{R}^n$. See Dharmadhikari and Joag-Dev (1988, 38ff), who demonstrate that the classes $P_n(\mathbf{0})$ and $K_n(\mathbf{0})$ coincide.

To continue, Gaussian densities $g_n = (\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ on $\mathbb{R}^n$ generate ensembles as $\boldsymbol{\Sigma}$ ranges over Ξ. These are

$$E_1(\Gamma_1) = \{g_n(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}(\theta)); \theta \in \Gamma_1\} \tag{2}$$

$$E_2(\Gamma_2) = \{g_n(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\xi})); \boldsymbol{\xi} \in \Gamma_2\}, \tag{3}$$

both unimodal in the sense of Sherman (1955). Taking $E_1(\Gamma_1)$ and $E_2(\Gamma_2)$ to have mixing parameters θ and $\boldsymbol{\xi}$, mixtures $f_i(x, \mu, G_i) = \int_{\Gamma_i} g_n(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}(i))\, dG_i(\cdot)$ emerge with $G_i \in \{G_1, G_2\}$ as a *cdf* on $\Gamma_i \in \{\Gamma_1, \Gamma_2\}$, and with $\boldsymbol{\Sigma}(i) \in [\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})]$. In particular, the densities $f_1(\boldsymbol{x}; \boldsymbol{\mu}, G_1)$ and $f_2(\boldsymbol{x}; \boldsymbol{\mu}, G_2)$ are dispersion mixtures of elliptical Gaussian distributions on $\mathbb{R}^n$ centered at $\mu \in \mathbb{R}^n$. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ comprise all *cdf*s on Γ_1 and Γ_2, respectively; these in turn generate the collections

$$\mathrm{M}_1 = \{f_1(\boldsymbol{x}; \boldsymbol{\mu}, G_1); G_1 \in \mathcal{G}_1\} \tag{4}$$

$$\mathrm{M}_2 = \{f_2(\boldsymbol{x}; \boldsymbol{\mu}, G_2); G_2 \in \mathcal{G}_2\} \tag{5}$$

comprising all dispersion mixtures of the referenced types, and all belonging to the shifted class $P_n(\boldsymbol{\mu})$.

Note that nonstandard joint distributions arise in a variety of applications. For example, see Verhoeven and McAleer (2004), with applications in the actuarial sciences.

3. The Principal Findings

3.1. Overview

In regard to the R–Student statistics, shifts are propagated as noncentrality parameters in $t^2(\nu, \lambda_1, \lambda_2)$. This rests on distributions of quadratic forms as detailed in Appendix A. We next establish essentials regarding outlier shifts themselves and their effects on various diagnostics.

3.2. Properties of Residuals

We next show properties of the observed residuals under Assumptions A. It remains to evaluate $E(e_i)$, $\mathrm{Var}(e_i)$ and $\mathcal{L}(e_i)$ as special cases. Details follow, where again $\boldsymbol{B}_n = (\boldsymbol{I}_n - \boldsymbol{H}_n) = \left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\right)$.

Lemma 3.1. *Consider the ordinary residuals* $\boldsymbol{e}'_0 = [\boldsymbol{e}', e_i]$ *under Assumptions* $\boldsymbol{A}_1 : \mathrm{E}(\boldsymbol{\varepsilon}_0) = \boldsymbol{\omega}$, *with* $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \delta_i]'$, *and* $\boldsymbol{A}_2 : V(\boldsymbol{\varepsilon}_0) = \boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$. *As in Definition 1 decompose* $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ *with* $\boldsymbol{\omega}_1 = \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\boldsymbol{\omega}$ *and* $\boldsymbol{\omega}_2 = \boldsymbol{B}_n\boldsymbol{\omega}$; *and let* $\mathrm{T}(\boldsymbol{e}_0)$ *be a mapping to a linear space* $\mathcal{V}$. *Then the following properties hold independently of* $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$:

i. $\mathrm{E}(\boldsymbol{e}_0) = \boldsymbol{\omega}_2$.
ii. $\mathrm{V}(\boldsymbol{e}_0) = \sigma^2\boldsymbol{B}_n$.
iii. $\mathrm{E}(e_i) = \frac{r}{n}(-\bar{\gamma} + \delta_i)$ *and* $\mathrm{Var}(e_i) = \frac{r}{n}\sigma^2$.
Moreover, under Assumption $\boldsymbol{A}_3 : \mathcal{L}(\boldsymbol{\varepsilon}_0) = N_n(\boldsymbol{\omega}, \boldsymbol{\Sigma})$, *it follows for* $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$ *that*
iv. $\mathcal{L}(\boldsymbol{e}_0|\boldsymbol{\Sigma}) = N_n(\boldsymbol{\omega}_2, \sigma^2\boldsymbol{B}_n)$.
v. $\mathcal{L}(T(\boldsymbol{e}_0)|\boldsymbol{\Sigma}) = \mathcal{L}(T(\boldsymbol{e}_0)|\sigma^2\boldsymbol{I}_n)$.
vi. In particular, $\mathcal{L}(e_i|\boldsymbol{\Sigma}) = \mathcal{L}(e_i|\sigma^2\boldsymbol{I}_n) = N_1\left(\frac{r}{n}(-\bar{\gamma} + \delta_i), \frac{r}{n}\sigma^2\right)$.

Proof. Assumption A_1 gives $\mathrm{E}(\boldsymbol{e}_0) = \boldsymbol{B}_n(\mu\mathbf{1}_n + \boldsymbol{\omega}) = \boldsymbol{B}_n\boldsymbol{\omega} = \boldsymbol{\omega}_2$ as in (i), since $\boldsymbol{B}_n\mathbf{1}_n = \mathbf{0}$. Assumption A_2 implies $\mathrm{V}(\boldsymbol{e}_0) = \boldsymbol{B}_n\boldsymbol{\Sigma}(\cdot)\boldsymbol{B}_n = \boldsymbol{B}_n$, since $\boldsymbol{B}_n\boldsymbol{\Sigma}(\cdot)\boldsymbol{B}_n$ annihilates terms beyond the first in $\boldsymbol{\Sigma}(\theta)$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi})$, and $\boldsymbol{B}_n$ is idempotent. This establishes conclusion (ii) independently of $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$. Expressing $\mathrm{E}(\boldsymbol{e}_0) = \left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\right)\omega$ in partitioned form is

$$\mathrm{E}\begin{bmatrix} e \\ e_i \end{bmatrix} = \frac{1}{n}\begin{bmatrix} (n\boldsymbol{I}_r - \mathbf{1}_r\mathbf{1}'_r) & -\mathbf{1}_r \\ -\mathbf{1}'_r & r \end{bmatrix}\begin{bmatrix} \boldsymbol{\gamma} \\ \delta_i \end{bmatrix} = \frac{1}{n}\begin{bmatrix} (n\boldsymbol{I}_r - \mathbf{1}_r\mathbf{1}'_r)\boldsymbol{\gamma} - \mathbf{1}_r\delta_i \\ -\mathbf{1}'_r\boldsymbol{\gamma} + r\delta_i \end{bmatrix}. \tag{6}$$

Thus $\mathrm{E}(e_i) = \frac{1}{n}\left(-\mathbf{1}'_r\boldsymbol{\gamma} + r\delta_i\right) = \frac{r}{n}(-\bar{\boldsymbol{\gamma}} + \delta_i)$, and $\mathrm{Var}(e_i) = \frac{r}{n}\sigma^2$ as the (n, n) element of $\sigma^2\left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\right)$, giving conclusion (iii). Finally, conclusions (iv)–(vi) follow directly on combining earlier results with Assumption A_3, to complete our proof. □

3.3. Tests: Order Statistics

Consider shifts $\{\boldsymbol{Y}_0 \to \boldsymbol{Y}_0 + \delta\boldsymbol{K}\}$ with $\boldsymbol{K} \in \mathbb{R}^n$ consisting of $n - 1$ zeros and a one. Begining with $\boldsymbol{Y}'_0 = [Y_1, \ldots, Y_n]'$, let $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $(n-1)S^2 = \sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2$. Corresponding to the ordered observations, let $\left\{e_{(1)} \leq e_{(2)} \leq \ldots \leq e_{(n)}\right\}$ be the ordered residuals, with $\left\{e_{(i)} = \left(Y_{(i)} - \bar{Y}\right); 1 \leq i \leq n\right\}$. Further, let $\left\{Z_{(i)} = e_{(i)}/S;\ 1 \leq i \leq n\right\}$, and

$Z_{max} = Z_{(n)}$. If $Y_{(n)}$ is outlying, it behaves as $\mathcal{L}\left(Y_{(n)}\right) = N_1\left(\mu + \lambda\sigma, \sigma^2\right)$ with $\lambda > 0$; see Bendre and Kale (1987, 893).

We next survey three classical approaches to tests for outliers based on the order statistics as follow.

i. Grubbs's (1950) test rejects H_0: $\lambda = 0$ in favor of H_1: $\lambda > 0$ when $Z_{max} \geq c_\alpha$, its Type I error at level α. Tables of critical values are reported in Grubbs (1950; 1969) and Grubbs and Beck (1972). Grubbs's test is highlighted in DasGupta (2013) as among 215 influential developments in statistics, as noted earlier.
ii. Dixon (1950) proposed the statistic $D = \left(Y_{(n)} - Y_{(n-1)}\right) / \left(Y_{(n)} - Y_{(1)}\right)$, declaring $Y_{(n)}$ to be outlying at level α if $D > d_{n,\alpha}$. Extensive tables of critical values were reported recently in Verma and Quiroz-Ruiz (2006).
iii. Ferguson's (1961) test takes $F = \sum_{i=1}^{n} e_{(i)}^3 / \left[\sum_{i=1}^{n} e_{(i)}^2\right]^{\frac{3}{2}}$ as a coefficient of skewness, rejecting at level α for $F > f_{n,\alpha}$, and shown to be locally most powerful invariant against outliers $\{\mu_i \to \mu_i + \lambda_i\sigma;\ \lambda_i > 0\}$ unspecified in number.

That these diagnostics may be validated under star-contoured errors is the subject of the following, where $\boldsymbol{\tau} = \mu\mathbf{1}_n + \boldsymbol{\omega}$.

Theorem 3.1. *Given the model* $\{\boldsymbol{Y}_0 + \boldsymbol{\tau} + \boldsymbol{\varepsilon}_0\}$ *having a Gaussian mixture density* $f_1\left(\boldsymbol{y}_0; \boldsymbol{\tau}, G_1\right)$ *belonging to* M_1 *as in Eq. (4), or a density* $f_2\left(\boldsymbol{y}_0; \boldsymbol{\tau}, G_2\right)$ *belonging to* M_2 *as in Eq. (5). Consider the following, all initially derived from* $\mathcal{L}\left(\boldsymbol{Y}_0\right) = N_n\left(\mu\mathbf{1}_n, \sigma^2\boldsymbol{I}_n\right)$*:*

i. Grubbs's (1950) test rejecting for $Z_{max} \geq c_\alpha$.
ii. Dixon's (1950) test rejecting for $D > d_{n,\alpha}$.
iii. Ferguson's (1961) test rejecting for $F > f_{n,\alpha}$.
iv. These tests remain exact in level and power for all mixtures in M_1 *and* M_2 *in Eqs. (4) and (5), the same as for* $\mathcal{L}\left(\boldsymbol{Y}_0\right) = N_n\left(\mu\mathbf{1}_n, \sigma^2\boldsymbol{I}_n\right)$.

Proof. Returning to section 2.4 and

$$f_i\left(\mathbf{y}_0, \boldsymbol{\tau}, G_i\right) = \int_{\Gamma_i} g_n\left(\mathbf{y}_0; \boldsymbol{\tau}, \boldsymbol{\Sigma}\,(i)\right) dG_i\,(\cdot), \tag{7}$$

we argue conditionally on fixing $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}\,(\theta), \boldsymbol{\Sigma}\,(\boldsymbol{\xi})\}$, then making the change of variables behind the integral, so that the derived unconditional error density is

$$f_i\left(\boldsymbol{e}_0, \boldsymbol{\omega}, G_i\right) = \int_{\Gamma_i} g_n\left(\boldsymbol{e}_0; \boldsymbol{\omega}, \sigma^2\boldsymbol{B}_n\right) dG_i\,(\cdot) = g_n\left(\boldsymbol{e}_0; \boldsymbol{\omega}, \sigma^2\boldsymbol{B}_n\right) \tag{8}$$

independently of G_i since $\int_{\Gamma_i} dG_i = 1$. For Grubbs's (1950) test the residuals map into

$$\boldsymbol{e}_0 \to \left\{e_{(1)} \leq e_{(2)} \leq \ldots \leq e_{(n)}, S^2\right\} \to T\left(\boldsymbol{e}_0\right) = \left\{Z_{(1)} \leq Z_{(2)} \leq \ldots \leq Z_{(n)}\right\}. \tag{9}$$

Lemma 3.1(v) establishes that $\mathcal{L}\left(Z_{max}\,|\boldsymbol{\Sigma}\right) = \mathcal{L}\left(Z_{max}\,\middle|\sigma^2\boldsymbol{I}_n\right)$, as asserted in conclusion (i), independently of $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}\,(\theta), \boldsymbol{\Sigma}\,(\boldsymbol{\xi})\}$. Invariance for conclusion (ii) follows, as $D = \left[e_{(n)} - e_{(n-1)}\right]/\left[e_{(n)} - e_{(1)}\right]$ is equivalent to Dixon's (1950) statistic, and conclusion (iii) follows along similar lines. □

It is noteworthy that these results are complementary to and extend considerably beyond the findings of Srivastava (1980), Young et al. (1989), and Baksalary and Puntanen (1990).

3.4. Tests: Deletion Diagnostics

Deletion diagnostics have a rich history, including not only the *R*-Student statistics, but also numerous influence diagnostics intended to track changes in the regression output as incurred on deleting observations. References include Belsley et al. (1980), Cook and Weisberg (1982), Barnett and Lewis (1994), Atkinson (1985), Rousseeuw and Leroy (1987), Chatterjee and Hadi (1988), Myers (1990), Fox (1991), and, more recently, Woodward and Sain (2003), Ullah and Pasha (2009), and Martin et al. (2010).

The *R*–Student diagnostic t_i^2 historically ignores shifts in rows other than i. See Snedecor and Cochran (1968, 157) for identifying a single shift, as considered also in Beckman and Trussell (1974). When other shifts occur, these confound the true descriptive level α, introducing masking and swamping of outliers as in Jensen and Ramirez (2014). Indeed, masking "is an important problem in influence analysis which deserves further study" (Hoaglin and Kempthorne 1986, 410).

Our tools offer further insight. That these diagnostics may be validated under star-contoured errors, and thus that anomalies carry forward beyond Gaussian errors to include mixtures, is the subject of the following, where $\boldsymbol{\tau} = \mu\mathbf{1}_n + \boldsymbol{\omega}$.

Theorem 3.2. *Given the model* $\{\boldsymbol{Y}_0 = \boldsymbol{\tau} + \varepsilon_0\}$ *having a Gaussian mixture density* $f_1\left(\boldsymbol{y}_0; \boldsymbol{\tau}, G_1\right)$ *belonging to* M_1 *as in Eq. (4), or a density* $f_2\left(\boldsymbol{y}_0; \boldsymbol{\tau}, G_2\right)$ *belonging to* M_2 *as in Eq. (5). Then the distribution* $\mathcal{L}\left(t_i^2\right)$ *is the doubly noncentral* $\mathcal{L}\left(t_i^2\right) = t^2\left(\nu; \lambda_1, \lambda_2\right)$ *with*

i. $\lambda_1 = \frac{r}{n}\left(-\bar{\gamma} + \delta_i\right)^2$; $\lambda_2 = \sum_{i=1}^{r}\left(\gamma_i - \bar{\gamma}\right)^2$*; and other properties as in Theorem A.2.*
ii. Diagnostics using t_i^2 *remain exact in level and power for all mixtures in* M_1 *and* M_2*, the same as for* $\mathcal{L}\left(\boldsymbol{Y}_0\right) = N_n\left(\mu\mathbf{1}_n, \sigma^2\boldsymbol{I}_n\right)$.

Proof. As in the proof for Theorem 3.1, begin with the mixture of Eq. (7). We proceed conditionally on fixing $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}\left(\theta\right), \boldsymbol{\Sigma}\left(\boldsymbol{\xi}\right)\}$ as before, then making the change of variables behind the integral, so that the derived unconditional density is

$$f_i\left(t_i^2; \lambda_1, \lambda_2, G_i\right) = \int_{\Gamma_i} g\left(t_i^2; \lambda_1, \lambda_2 \left|\sigma^2\boldsymbol{I}_n\right.\right) dG_i\left(\cdot\right) = t^2\left(\nu; \lambda_1, \lambda_2\right) \tag{10}$$

from Appendix Theorem A.2(vi), since $\int_{\Gamma_i} dG_i = 1$. Accordingly, its properties carry forward unconditionally as in Theorem A.2, independently of the mixing distributions $G_1 \in \mathcal{G}_1$ and $G_2 \in \mathcal{G}_2$ of Eqs. (4) and (5), and thus for all such mixtures. □

It is noteworthy that these results extend considerably beyond the classical model for deletion diagnostics, where $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$. An immediate consequence is that for Y_i not outlying, $\mathcal{L}\left(t_i^2\right)$ nonetheless may be doubly noncentral whenever $\boldsymbol{\gamma} \neq \mathbf{0}$. That is, for $\delta_i = 0$, it may happen that $\mathcal{L}\left(t_i^2 \left|\gamma\right.\right) = t^2\left(\nu, \lambda_1, \lambda_2\right)$. Moreover, our mixture distributions extend beyond t_i^2 to include every influence and deletion diagnostic in Table 1 of Jensen (2000). This follows since these are known to correspond one-to-one with t_i or t_i^2; their doubly noncentral distributions then derive from $t^2\left(\nu, \lambda_1, \lambda_2\right)$ as if $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$, despite the fact that $\boldsymbol{\Sigma} \in \left[\boldsymbol{\Sigma}\left(\rho\right), \boldsymbol{\Sigma}\left(\boldsymbol{\xi}\right)\right]$.

Table 1
Tabulated c_α from Grubbs (1969) and empirical values for $Z_{max} \geq c_\alpha$ with $\mathbf{\Sigma} = \mathbf{\Sigma}(\rho), n = 5, \sigma^2 = 1, N = 40{,}000$ runs, and $\rho \in [0.0, 0.5, 0.8]$

α	10%	5%	2.5%	1%
Tabulated c_α	1.620	1.672	1.715	1.749
$\rho = 0.0$	1.603	1.672	1.715	1.748
$\rho = 0.5$	1.600	1.671	1.716	1.750
$\rho = 0.8$	1.605	1.675	1.716	1.750

3.5. Related Studies

Normal-theory tests for location and scale often are exact in level, and sometimes power, for joint distributions having spherical or elliptical contours. Examples are given in early work of Jensen (1985), Kariya and Sinha (1989), Fang et al. (1990), Fang and Zhang (1990), and subsequently. Such distributions fill a conspicuous void in practice, especially for data having excessive tails, even without first or second moments. Table 1 of Jensen (1985) lists a variety of such distributions on $\mathbb{R}^n$, to include Gaussian, Pearson Type II and Type VII, the Student's t family, spherical Cauchy distributions, scale mixtures, and the symmetric stable laws on $\mathbb{R}^n$. Inferences for these often have the same critical values as for Gaussian models. Nor are problems with outliers restricted to Gaussian data.

Tests for outliers due to Dixon (1950), Grubbs (1950), and Ferguson (1961) figure prominently as noted. That these tests remain exact at level α for every spherical error distribution, with or without moments, is a consequence of Theorem 1 of Jensen and Good (1981), and therefore these are genuinely nonparametric. On the other hand, non-null distributions in tests for location typically depend on the particular spherical distribution at hand. Power properties accordingly remain obscure, to be considered case-by-case along the lines of Jensen (1981) for topics in linear inference. Similar comments apply for the doubly noncentral $\mathcal{L}\left(t_i^2\right) = t^2\left(\nu, \lambda_1, \lambda_2\right)$ of the present study. These further complications are preempted here on taking conditionally Gaussian models.

4. Case Studies

4.1. Darwin Example

As noted earlier, Gaussian mixtures have modeled data collected from subsamples. For example, Box and Tiao (1968) and Aitken and Wilson (1980) used a two-component Gaussian mixture, with common mean and unequal variances, to model Darwin's data comprising 15 differences of heights of cross-fertilized and self-fertilized plants as discussed in Fisher (1960). The data are:

$$\begin{array}{cccccccc} -67 & -48 & 6 & 8 & 14 & 16 & 23 & 24 \\ 28 & 29 & 41 & 49 & 56 & 60 & 75. & \end{array}$$

The first two negative observations appear to differ from the other values, and Fisher (1960) was concerned that the heights of the plants were affected by selection of seeds, soil fertility, sun light, evaporation, and so on.

In view of latent correlations among contiguous plots, we model the data instead as a Gaussian mixture with subsamples satisfying Assumption A_2 in the equivalent form A_2: $\mathrm{V}(\boldsymbol{\varepsilon}_0) = \boldsymbol{\Sigma}(\rho)$ as in Lemma 2.1. The importance of Theorem 3.1 is that we may proceed with the standard outlier tests (Grubbs, Dixon, Ferguson) invoking the critical values as would be appropriate for independent normal data. For example, with Grubbs's test, $|\mathrm{Z}_{\min}| = |(-67 - 20.93)/37.74| = 2.32$, which is less than the 5% critical value 2.41 from Grubbs (1969), to conclude that $Y(_1) = -67$ is not an outlier. Similarly, with Dixon's gap-over-range test, $D = |(-67 + 48)/(75 + 67)| = 0.134$ is less than the one-sided $\alpha = 5\%$ critical value $c_\alpha = 0.338$ with a one-sided p-value greater than 20% from tables in Dixon (1951) and Rorabacher (1991).

4.2. Grubbs and Dixon Tests

To demonstrate Lemma 2.1, $N = 40,000$ random samples with sample size $n = 5$ were generated from a multivariate normal distribution with mean $\boldsymbol{\mu} = \mathbf{0}$ and dispersion matrix $\boldsymbol{\Sigma}(\rho)$, with $\sigma^2 = 1$ and varying ρ as in Lemma 2.1. Minitab was used for the simulations. Table 1 reports the empirical critical values for Grubbs's (1950) test using Z_{max} for $\rho \in [0.0, 0.5, 0.8]$, with the corresponding tabulated values from Grubbs (1969). Lemma 3.1(v) assures that the empirical values and the tabular values will be the same for varying ρ, as confirmed in the table apart from simulation errors.

Similarly, Table 2 reports the empirical critical values for Dixon's (1950) test using $D = (Y_{(n)} - Y_{(n-1)})/(Y_{(n)} - Y_{(1)})$ for $\rho \in [0.0, 0.5, 0.8]$, with the corresponding tabulated values from Dixon (1951) or Rorabacher (1991), who report critical values for $\alpha \in [1\%, 2\%, 5\%, 10\%]$.

More importantly, a similar result in Table 3 reports the empirical critical values for Grubbs's test using a two-component Gaussian mixture with (50%, 50%) weights and equicorrelated dispersion matrices $\boldsymbol{\Sigma}(0.5), \boldsymbol{\Sigma}(0.8)$. We set $\sigma^2 = 1$, the sample size $n = 5$ for each mixture component, to encompass $N = 40,000$ repetitions. The empirical critical values for the mixture agree with the Grubbs (1969) tabular values as supported by Theorem 3.1(i).

Additionally, Theorem 3.1(i) assures, for distributions satisfying Assumptions A, that the power of the standard tests (Grubbs, Dixon, Ferguson) will be the same as the power for independent normal data. This is demonstrated in Table 4 for Grubbs's test invoking Z_{max} with N = 40,000 repetitions, sample size $n = 10$, and $\sigma^2 = 1$. Nominal powers at $\alpha = 5\%$ are compared empirically by perturbing $\{Y_{10} \to Y_{10} + \delta_{10}\}$, with varying $\delta_{10} \in [0.0, 0.5, 1.0, 1.5, 2.0, 2.5]$ for equicorrelated Gaussian models with $\boldsymbol{\Sigma}(0)$ and $\boldsymbol{\Sigma}(0.8)$.

Table 2

Tabulated c_α from Dixon (1951) or Rorabacher (1991) and empirical values for $D \geq c_\alpha$ with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\rho), n = 5, \sigma^2 = 1, N = 40{,}000$ runs, and $\rho \in [0.0, 0.5, 0.8]$

α	10%	5%	2.5%	1%
Tabulated c_α	0.557	0.642	0.729	0.780
$\rho = 0.0$	0.560	0.644	0.734	0.774
$\rho = 0.5$	0.557	0.638	0.727	0.779
$\rho = 0.8$	0.558	0.640	0.726	0.774

Table 3
Tabulated c_α from Grubbs (1969) and Empirical Values for $Z_{max} \geq c_\alpha$ in a 50%–50% Gaussian Mixture having $(\boldsymbol{\Sigma}(\rho_1), \boldsymbol{\Sigma}(\rho_1))$, with $n = 5, \sigma^2 = 1$, and $N = 40{,}000$

α	10%	5%	2.5%	1%
Tabulated c_α	2.036	2.176	2.290	2.410
$\rho_1 = 0, \rho_2 = 0$	2.038	2.172	2.286	2.409
$\rho_1 = 0.5, \rho_2 = 0.8$	2.012	2.163	2.281	2.417

Table 4
Empirical power for Z_{max} at level 5% and varying δ_{10}, N = 40,000 runs, $n = 10$, $\sigma^2 = 1$ for $\boldsymbol{\Sigma}(\rho)$ evaluated at $\rho \in (0.0, 0.8)$

δ_{10}	$\Sigma(0.0)$	$\Sigma(0.8)$
0.0	0.049	0.051
0.5	0.067	0.069
1.0	0.197	0.197
1.5	0.483	0.490
2.0	0.778	0.780
2.5	0.991	0.942

For each of 40,000 runs of $n = 10$ observations, Z_{max} was computed, and the fraction of cases having $Z_{max} \geq c_\alpha$ is the reported empirical power. Assumptions A allow for shifts $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \delta_i]'$ with $\delta_i \neq 0$. The empirical powers for Grubbs's test, under both the independent and correlated data, are seen empirically to coincide up to simulation errors as in Lemma 3.1(iv).

To visualize the mixture of normal distributions, we display in Figure 1 a cross-sectional level surface for the 50%–50% mixture of bivariate normal distributions with correlations $\rho \in \{0.0, 0.8\}$ which is "lemon-shaped," and with $\rho \in \{-0.95, 0.95\}$, which is "star-shaped."

4.3. *Doubly Noncentral t^2 $(\nu, \lambda_1, \lambda_2)$ Test*

Tests invoking the doubly noncentral t^2 $(\nu, \lambda_1, \lambda_2)$ serve to generalize conventional outlier tests, to include nonzero shifts at arbitrary positions in the shift vector $\boldsymbol{\omega} = \left[\boldsymbol{\gamma}', \delta_i\right]'$ of section 3.4. By convention the R–Student t_i^2 test for $H_0 \colon \delta_i = 0$ is restricted in allowing no additional shifts elsewhere, that is, $\boldsymbol{\gamma} = \mathbf{0}$. Theorem A.2 using t^2 $(\nu, \lambda_1, \lambda_2)$ circumvents this restriction, enabling studies where $\boldsymbol{\gamma} \neq \mathbf{0}$. This is essential for examining the swamping and masking of outliers, and these findings carry forward where $\boldsymbol{\Sigma} \neq \sigma^2 \boldsymbol{I}_n$, and mixtures of these as in section 3.4.

To illustrate, consider a data set from Barnett and Lewis (1994, 109) with $\boldsymbol{Y}_0'$ as

$$\boldsymbol{Y}_0' = \left[3, 4, 7, 8, 10, y_6, 951\right]$$

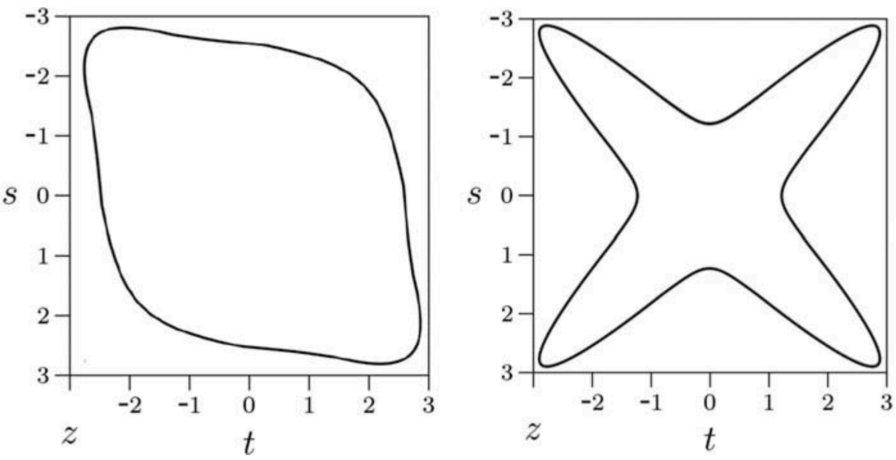

Figure 1. Level surfaces for mixtures with $\rho = 0.0, 0.8$, and $-0.95, 0.95$.

and assigning y_6 to have varying values. For the case that $y_6 = 13$, $y_{(n)} = y_7 = 951$ is a clear outlier, while with $y_6 = 949$, the maximum value $y_7 = 951$ will now fail the Grubbs and Dixon outlier tests. We expand the range to include $y_6 \in \{13, \pm 733, \pm 949\}$ and show how to apply Theorem A.2 as y_6 is varied.

The notation for the full model is $\{\boldsymbol{Y}_0 = \mu \mathbf{1}_n + \boldsymbol{\varepsilon}_0\}$ with $n = 7$, giving $\left(\bar{Y}, S^2\right)$. The reduced model on eliminating y_7 has $r = n - 1 = 6$ entries, and values $\left(\bar{Y}_6, S_6^2\right)$. The shift under consideration supports $H_0\colon\ \delta_7 = 0$, allowing for an additional shift of $\gamma_6 \neq 0$. Thus, $\boldsymbol{\omega} = \left[\boldsymbol{\gamma}', \delta_7\right]'$ with $\boldsymbol{\gamma}' = \left[0, 0, 0, 0, 0, \gamma_6\right]$ such that both (γ_6, δ_7) are nonzero. The vector $\boldsymbol{\gamma}$ may contain multiple nonzero entries, but for this example we have restricted to only one such entry.

The resolution $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2$ of Definition 1 serves here to decompose the shift vector $\gamma = \left[0, 0, 0, 0, \gamma_6\right]'$ into orthogonal vectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ lying in the “Regressor” and “Error” spaces, respectively, with $\gamma_6 > 0$. The angles between γ and its components are given by $\theta_1 = \theta\,(\gamma, \gamma_1) = 65.9^\circ$ and $\theta_2 = \theta\left(\boldsymbol{\gamma}, \boldsymbol{\gamma}_2\right) = 24.1^\circ$.

For the doubly noncentral $\mathcal{L}\left(t_i^2\right) = t^2\,(\nu, \lambda_1, \lambda_2)$ we have $\nu = n - 2 = 5$. We proceed to estimate the noncentrality parameters (λ_1, λ_2) using the moment estimators from Lemma A3 of Jensen and Ramirez (2014) under Assumptions A. For our example, the more general moment equations from that reference reduce to

$$\gamma_6 = \left[y_6 - (y_1 + y_2 + y_3 + y_4 + y_5 + y_7)/6\right] + \delta_7/6$$

$$\delta_7 = \left[y_7 - (y_1 + y_2 + y_3 + y_4 + y_5 + y_6)/6\right] + \gamma_6/6$$

with solutions $\left\{\widetilde{\gamma}_6, \tilde{\delta}_7\right\}$. The moment estimators are scaled in standard σ units on dividing by S_6, and the noncentrality parameters in standard units as

$$\lambda_1 = \frac{r}{n}\left(-\bar{\gamma} + \tilde{\delta}_7\right)^2 \Big/ S_6^2, \qquad \lambda_2 = \sum_{i=1}^{r} (\gamma_i - \bar{\gamma})^2 \Big/ S_6^2. \tag{11}$$

Table 5

In testing $H_0: \delta_7 = 0$, p values $p_{(\cdot)}$ using D, Z_{max}, t_ν^2 $t^2(\nu, \lambda_1, \lambda_2)$ for cases $\boldsymbol{Y}_0' = [3, 4, 7, 8, 10, y_6, 951]$ as y_6 is varied, together with other supporting quantities

y_6	13	−733	733	−949	949
Z_{max}	2.27	1.87	1.71	1.72	1.47
$p_{(G)}$	0.0094	0.1497	0.2515	0.2446	0.4999
D	0.989	0.559	0.230	0.495	0.002
$p_{(D)}$	(0, 0.005)	(0.02, 0.05)	(0.30, 0.40)	(0.05, 0.10)	(0.95, 1.00)
t_7^2	54893.66	10.73	6.61	6.86	3.59
$p_{(t)}$	0.0000	0.0221	0.0500	0.0471	0.1167
$\widetilde{\gamma}_6/S_r$	1.77	−2.45	2.45	−2.45	2.45
δ_7/S_r	253.36	3.13	3.18	2.42	2.45
λ_1/S_r^2	54893.66	10.73	6.61	6.86	3.59
λ_2/S_r^2	2.61	5.00	5.00	5.00	5.00
λ_3/S_r^2	54896.27	15.73	11.61	11.86	8.59
$p_{(\lambda)}$	0.0000	0.2348	0.2577	0.2557	0.2924

For varying $y_6 \in \{13, \pm 733, \pm 949\}$, Table 5 shows the p values for testing against H_0 : $\delta_7 = 0$ that $\boldsymbol{Y}_{(7)} = 951$ is outlying. The tests under consideration are (1) Grubbs's test; (2) Dixon's test; (3) the R-Student test t_7^2, which assumes no nonzero entries for ω other than δ_7; and (4) the noncentral $t^2(\nu, \lambda_1, \lambda_2)$ test from Theorem A.2, which allows for nonzero entries in γ.

In the notation of Theorem A.2 with $\mathcal{L}(\boldsymbol{Y}_0 - \mu\mathbf{1}_n)$ centered at $\mathbf{0} \in \mathbb{R}^n$, then $\lambda_3 = \boldsymbol{\omega}'\boldsymbol{B}_n\boldsymbol{\omega}$, so with $\boldsymbol{\omega} = [0, 0, 0, 0, 0, \gamma_r, \delta_n]'$, we have $\lambda_3 = \lambda_1 + \lambda_2$, with $\lambda_1 = \frac{r}{n}(-\bar{\gamma}+\delta_i)^2$ and $\lambda_2 = \boldsymbol{\Sigma}_{i=1}^r(\gamma_i - \bar{\gamma})^2$ from Theorem A.2. On standardizing these parameters to $\{\lambda_i \to \lambda_i/S_r^2;, 1 \le i \le 3\}$ as in Eq. (11), their values are reported in Table 5.

This empirical standardization results in the curious fact that the R-Student statistics t_7^2 for $H_0 : \delta_7 = 0$ are identical to the standardized first noncentral parameter λ_1/S_r^2 for the corresponding doubly noncentral $t^2(\nu, \lambda_1, \lambda_2)$ as reported in Table 5. This follows from

$$\lambda_1/S_r^2 = \frac{r}{n}\left(-\bar{\gamma}+\widetilde{\delta}_n\right)^2\Big/S_r^2 = \frac{1}{nr}\left[-\widetilde{\gamma}_n + r\left(y_n - \bar{Y}_r\right) + \widetilde{\gamma}_r\right]^2\Big/S_r^2$$

$$= \left(y_n - \bar{Y}_r\right)^2\Big/\left(nS_r^2/r\right) = re_n^2/\left(nS_r^2\right) = t_n^2$$

from standard results for Studentized deleted residuals.

Grubbs's and Dixon's tests are one-sided, whereas tests based on $t^2(\nu, \lambda_1, \lambda_2)$ are necessarily two-sided. For Grubbs's test the 5% critical value for Z_{max} with $n = 7$ is 1.94 with upper bound $Z_{max} \le (n-1)/\sqrt{n} = 2.267$ (Shiffler 1988). For Dixon's test with $n = 7$ the 5% critical value for D is 0.507. It is important to note that all reported p values are valid when $\mathcal{L}(\boldsymbol{Y}_0) = N_n(\boldsymbol{\tau}, \boldsymbol{\Sigma})$ with $\boldsymbol{\tau} = \mu\mathbf{1}_n + \boldsymbol{\omega}$ and $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\gamma), \boldsymbol{\Sigma}(\rho)\}$, and for dispersion mixtures over these. For notational convenience, the p values are denoted by $p_{(G)}$ for Grubbs's test; $p_{(D)}$ for Dixon's test; $p_{(t)}$ for the R–Student test based on $t^2(\nu)$; and $p_{(\lambda)}$ from Theorem A.2 based on the noncentral $t^2(\nu, \lambda_1, \lambda_2)$, with $(\lambda_1, \lambda 2)$ as functions of the moment estimators $(\widetilde{\gamma}_6, \widetilde{\delta}_7)$ shown in rows 8 and 9 of Table 5.

In Appendix A.2, we outline our computations for the p-values $\{p_{(G)}, p_{(t)}, p_{(\lambda)}\}$ as listed in Table 5, to include ranges of values for $p_{(D)}$. For example, for $y_6 = -733$, the observed $D = 0.559$ is between the critical values c_α for $n = 7$ from Dixon (1951) of 0.586, and 0.507, which correspond to the one-sided levels for $\alpha = 2\%$ and 5%.

As expected, for the case $y_6 = 13$ all tests identify $y_7 = 951$ as an outlier, with Z_{max} nearly achieving the Shiffler (1988) bound; when $y_6 = 949$, none of the tests flag $y_7 = 951$ as an outlier. However, with $y_6 = -949$, the R–Student test reverses the conclusion and flags $y_7 = 951$ as an outlier, while the test invoking the noncentral $t^2(\nu, \lambda_1, \lambda_2)$ exhibits little change in the p value. The value $y_6 = 733$ has been chosen so that the p value for the R–Student test will be the nominal value of 0.05. Switching the sign so that $y_6 = -733$, the p value drops to 0.0221 for the R–Student test. The test based on $t^2(\nu, \lambda_1, \lambda_2)$ shows no outliers for $y_7 = 951$ when $y_6 = \pm 733$, having respective p values 0.2348 and 0.2577.

It is essential to note that the R-Student test has flagged $y_7 = 951$ as an outlier for the values for y_6 in columns 2, 3, 4, and 5 of Table 5. The test based on the doubly noncentral $t^2(\nu, \lambda_1, \lambda_2)$ from Theorem A.2, which allows for shifts in the nondeleted rows, is in agreement with Grubbs test in flagging $y_7 = 951$ only as an outlier when $y_6 = 13$, as shown in column 2 of Table 5. In regard to $H_0: \delta_7 = 0 \,|\, \delta_6 = 0$, that is, $y_7 = 951$ is not outlying, with no other outlying values, a reviewer notes that the p values for the R–Student t^2 test have the range (0.0500, 0.1167) as y_6 has the range (733, 949). Thus, extreme values for y_6 can mask the extreme value of $y_7 = 951$. The doubly noncentral $t^2(\nu, \lambda_1, \lambda_2)$ test that allows for y_6 to be outlying (i.e., $H_0: \delta_7 = 0 \,|\, \delta_6 \neq 0$) has p values ranging between (0.2577, 0.2924).

5. Conclusions

Given Gaussian observations $\boldsymbol{Y}_0 = [Y_1, \ldots Y_n]'$, normal-theory tests for location and scale often have critical values exact at level α, as derived under classical assumptions such as $\mathrm{V}(\boldsymbol{Y}_0) = \sigma^2 \boldsymbol{I}_n$. Exact normal-theory tests on occasion hold also for spherical error distributions, thus enabling the researcher to apply standard outlier tests in heavy-tailed data not necessarily having first or second moments. We explore such results for Gaussian observations with dispersion matrix $\boldsymbol{\Sigma}$ taken from $[\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi}), \boldsymbol{\Sigma}(\rho)]$, and extend these to include dispersion mixtures of such distributions, where the matrices are positive definite in well-defined regions given in Lemma 2.1. For such data the ordinary residuals $\boldsymbol{e}_0 = (\boldsymbol{Y}_0 - \bar{Y}\boldsymbol{1}_n)$ have dispersion matrix $\mathrm{V}(\boldsymbol{e}_0) = \sigma^2 (\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}'_n)$ independently of $\boldsymbol{\Sigma} \in [\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi}), \boldsymbol{\Sigma}(\rho)]$ for the data in hand. Accordingly, the outlier tests of Grubbs, Dixon, and Ferguson, as functions of these residuals, will remain exact in level and power for such data, and for mixtures over these as in Theorem 3.1. In consequence, these tests have been updated here for use in data having structures substantially beyond the classical assumptions, as required in some contemporary experiments.

Case studies are reported demonstrating for $\boldsymbol{\Sigma}(\rho)$ that the critical values for the tests of Grubbs and Dixon are independent of the value of ρ. Moreover, the critical values for Grubbs's test are seen to be independent of values for the pair (ρ_1, ρ_2) in a two-component mixture using $\{\boldsymbol{\Sigma}(\rho_1), \boldsymbol{\Sigma}(\rho_2)\}$. Figure 1 shows that a bivariate normal distribution can have "star-shaped" contours for its level surfaces, and hence the title for this work.

The R-Student deletion diagnostic t_i^2 supports conventional tests for outliers and for shifts in linear models in continuing wide usage. Under the standard model where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$, Jensen and Ramirez (2014) have shown that the R-Student t^2 distribution extends to include shifts in nondeleted rows, resulting in doubly noncentral distributions $t^2(\nu, \lambda_1, \lambda_2)$ of note in the masking and swamping of outliers. Here we extend those results to include data having a dispersion matrix from $\left[\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\boldsymbol{\xi}), \mathbf{\Sigma}(\rho)\right]$, and for mixtures over these. Supporting work, referred to an appendix, verifies the chi-squared character of quadratic forms, and their independence, under the highly irregular $\mathbf{\Sigma} \in \left[\mathbf{\Sigma}(\theta), \mathbf{\Sigma}(\boldsymbol{\xi}), \mathbf{\Sigma}(\rho)\right]$. This in turn validates the use of t_i^2 in data going substantially beyond the classical assumptions. A case study is given showing that the doubly noncentral $t^2(\nu, \lambda_1, \lambda_2)$ is in agreement with the popular Grubbs's test, and since it allows for shifts in nondeleted rows, t_i^2 will have protection against masking effects arising through nondeleted shifts, in contrast to Grubbs's test.

References

Aitkin, M., and G. Wilson. 1980. Mixture models, outliers, and the EM algorithm. *Technometrics*, 22, 325–331.

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

Baksalary, J. K., and S. Puntanen. 1990. A complete solution to the problem of robustness of Grubbs' test. *Can. J. Stat.*, 18, 285–287.

Barnett, V., and T. Lewis. 1994. *Outliers in statistical data*, 3rd 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.

Bendre, S. M., and B. K. Kale. 1987. Masking effect on tests for outliers in normal samples. *Biometrika*, 74, 891–896.

Box, G. E. P., and G. C. Tiao. 1968. A Bayesian approach to some outlier problems. *Biometrika*, 55, 119–129.

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

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

DasGupta, A. 2013. Anirban's angle: 215 Influential developments in statistics. *IMS Bull.*, 42(8), 10–11.

Dharmadhikari, S., and K. Joag-Dev. 1988. *Unimodality, convexity, and applications*. New York, NY: Academic Press.

Dixon, W. J. 1950. Analysis of extreme values. *Ann. Math. Stat.*, 21, 488–506.

Dixon, W. J. 1951. Ratios involving extreme values. *Ann. Math. Stat.*, 22, 68–78.

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.

Fang, K. T., S. Kotz, and K. W. Ng. 1990. *Symmetric multivariate and related distributions*. London, UK: Chapman and Hall.

Fang, K. T., and Y. T. Zhang. 1990. *Generalized multivariate analysis*. New York, NY: Springer-Verlag.

Ferguson, T. S. 1961. On the rejection of outliers. *Proc. 4th Berkeley Symp.*, 1, 253–287.

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

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

Grubbs, F. E. 1950. Criteria for testing outlying observations. *Ann. Math. Stat.*, 21, 27–58.

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics*, 11, 1–21.

Grubbs, F. E., and G. Beck. 1972. Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14, 847–854.

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. 1981. Power of invariant tests for linear hypotheses under spherical symmetry. *Scand. J. Stat.*, 8, 169–174.

Jensen, D. R. 1985. Multivariate distributions. In *Encyclopedia of statistical sciences, vol. 6*, ed. N. L. Johnson, S. Kotz, and C. B. Read, 43–55. New York, NY: Wiley.

Jensen, D. R., 1996. Structured dispersion and validity in linear inference. *Linear Algebra Appl.*, 249, 189–196.

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 I. J. Good. 1981. Invariant distributions associated with matrix laws under structural symmetry. *J. R. Statist. Soc. B*, 43, 327–332.

Jensen, D. R., and D. E. Ramirez. 2009. Anomalies in the analysis of calibrated data. *J. Stat. Comput. Sim.*, 79, 299–314.

Jensen, D. R., and D. E. Ramirez. 2012. Irregularities in X(Y) from Y(X) in linear regression. *J. Stat. Comput. Sim.*, 83, 1807–1828.

Jensen, D. R., and D. E. Ramirez. 2014. Noncentralities induced in regression diagnostics. *J. Statist. Theory & Pract.*, 8, 1–25.

Kanter, M. 1977. Unimodality and dominance for symmetric random vectors. *Trans. Am. Math. Soc.*, 229, 65–85.

Kariya, T., and B. K. Sinha. 1989. *Robustness of statistical tests*. New York, NY: Academic Press.

Martin, M. A., S. Roberts, and L. Zheng. 2010. Delete-2 and delete-3 jackknife procedures for unmasking in regression. *Aust. NZ J. Stat.*, 52, 45–60.

Mathai, A. M., and S. B. Provost. 1992. *Quadratic forms in random variables: Theory and applications*. New York, NY: Marcel Dekker.

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

Nair, K. R., 1948. The distribution of the extreme deviate from the sample mean and its Studentized form. *Biometrika*, 35, 118–144.

Nandi, S. B., and S. Choudhury. 2002. Series representation of doubly noncentral t distribution by use of the Mellin integral transform. *Commun. Stat. Theory Methods*, 31, 2139–2149.

Rorabacher, D. 1991. Statistical treatment for rejection of deviant values: Critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level. *Anal. Chem.*, 63, 139–146.

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

Sherman, S. 1955. A theorem on convex sets with applications. *Ann. Math. Stat.*, 26, 763–766.

Shiffler, R. 1988. Maximum Z scores and outliers. *Am. Stat.*, 42, 79–80.

Snedecor, G. W., and W. G. Cochran. 1968. *Statistical methods*, 6th ed. Ames, IA: Iowa State University Press.

Srivastava, M. S. 1980. Effect of equicorrelation in detecting a spurious observation. *Can. J. Stat..*, 8, 249–251.

Thompson, W. R. 1935. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Ann. Math. Stat.*, 6, 214–219.

Ullah, M. A., and G. R. Pasha. 2009. The origin and developments of influence measures in regression. *Pak. J. Stat.*, 25, 295–307.

Verhoeven, P., and M. McAleer. 2004. Fat tails and asymmetry in financial volatility models. *Math. Comput. Simulat.*, 64, 351–361.

Verma, S. P., and A. Quiroz-Ruiz. 2006. Critical values for six Dixon tests for outliers in normal samples up to sizes 100, and applications in science and engineering. *Rev. Mex. Cienc. Geol.*, 23, 133–161.

Woodward, W., and S. R. Sain. 2003. Testing for outliers from a mixture distribution when some data is missing. *Comput. Stat. Data Anal.*, 44, 193–210.
Young, D. M., R. Pavur, and V. R. Marco. 1989. On the effect of correlation and unequal variances in detecting a spurious observation. *Can. J. Stat.*, 17, 103–105.

Appendix

A.1. Foundations

We draw on and extend the work of Jensen and Ramirez (2014), on specializing their model $\{Y_0 = X_0\beta + \varepsilon_0\}$ to $\{\boldsymbol{Y}_0 = \mu\mathbf{1}_n + \boldsymbol{\varepsilon}_0\}$, to be partitioned as $\boldsymbol{Y}_0' = [\boldsymbol{Y}', Y_i]$ and deleting $[Y_i, 1, \varepsilon_i]$ yet retaining $\{\boldsymbol{Y} = \mu\mathbf{1}_r + \boldsymbol{\varepsilon}\}$. The matrix $\boldsymbol{B}_n$ is partitioned subsequently as

$$\boldsymbol{B}_n = \left(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\right) = \frac{1}{n}\begin{bmatrix} \left(n\boldsymbol{I}_r - \mathbf{1}_r\mathbf{1}_r'\right) & -\mathbf{1}_r \\ -\mathbf{1}_r' & r \end{bmatrix} \tag{A.1}$$

and $\boldsymbol{\Sigma}(\theta)$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi})$ are partitioned conformably as

$$\boldsymbol{\Sigma}(\theta) = \begin{bmatrix} \left(\boldsymbol{I}_r + \theta\mathbf{1}_r\mathbf{1}_r'\right) & \theta\mathbf{1}_r \\ \theta\mathbf{1}_r' & (1+\theta) \end{bmatrix} \tag{A.2}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\xi}) = \begin{bmatrix} \boldsymbol{I}_r & \mathbf{0} \\ \mathbf{0}' & 1 \end{bmatrix} + \begin{bmatrix} \mathbf{1}_r \\ 1 \end{bmatrix}\boldsymbol{\xi}' + \boldsymbol{\xi}\left[\mathbf{1}_r', 1\right] - \bar{\xi}\begin{bmatrix} \mathbf{1}_r\mathbf{1}_r' & \mathbf{1}_r \\ \mathbf{1}_r' & 1 \end{bmatrix}. \tag{A.3}$$

A critical step entails quadratic forms in Gaussian vectors, now having dispersion matrix $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$. Let $\bar{Y}_r = \frac{1}{r}\sum_{i=1}^{r} Y_i$; with $(1 - h_{ii}) = r/n$, take

$$\frac{ne_i^2}{r} + \sum_{i=1}^{r}\left(Y_i - \bar{Y}_r\right)^2 = e_0'\boldsymbol{e}_0 \tag{A.4}$$

as the requisite Fisher–Cochran expansion; see Lemma A.1(iii) of Jensen (2001). As quadratic forms in $\boldsymbol{Y}_{\mathbf{0}}$, Eq. (A.4) becomes $\boldsymbol{Y}_0'\boldsymbol{A}_1\boldsymbol{Y}_0 + \boldsymbol{Y}_0'\boldsymbol{A}_2\boldsymbol{Y}_0 = \boldsymbol{Y}_0'\boldsymbol{A}_3\boldsymbol{Y}_0$, with $\boldsymbol{A}_3 = \boldsymbol{B}_n$, $\boldsymbol{A}_2 = Diag\,(\boldsymbol{B}_r, 0)$ with $\boldsymbol{B}_r = \left(\boldsymbol{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r\right)$, and $\boldsymbol{A}_1 = \boldsymbol{A}_3 - \boldsymbol{A}_2 = \frac{1}{n}\begin{bmatrix} \frac{1}{r}\mathbf{1}_r\mathbf{1}_r' & -\mathbf{1}_r \\ -\mathbf{1}_r' & r \end{bmatrix}$ in partitioned form.

Properties of these quadratic forms are essential. Under the conventional assumption that $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$, this would amount to demonstrating that the matrices are idempotent and evaluating their ranks. A principal result is the following, instead drawing heavily on the work of Mathai and Provost (1992, 201), taking into account the special structure of $\boldsymbol{\Sigma} \neq \sigma^2\boldsymbol{I}_n$. To continue, recall that $\boldsymbol{B}_n\mathbf{1}_n = 0$ and $\boldsymbol{B}_r\mathbf{1}_r = 0$.

Theorem A.1. *Suppose that* $\mathcal{L}(\boldsymbol{Y}_0) = N_n(\boldsymbol{\tau}, \boldsymbol{\Sigma})$ *for* $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$; *take* $\boldsymbol{e}_0 = \boldsymbol{B}_n\boldsymbol{Y}_0$; *consider the quadratic forms* $\{Q_1, Q_2, Q_3\}$ *such that* $Q_1 + Q_2 = Q_3$; *and let* $t_i^2 = (n-2)\,Q_1/Q_2$. *Then for each* $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$ *we have:*

i. $\boldsymbol{A}_i\boldsymbol{\Sigma}\boldsymbol{A}_i = \boldsymbol{A}_i$ *for* $i \in \{1, 2, 3\}$.
ii. $\boldsymbol{A}_1\boldsymbol{\Sigma}\boldsymbol{A}_2 = 0$.
iii. $\mathcal{L}(Q_i) = \chi^2(\nu_i, \lambda_i)$ *with* $\nu_i = tr\,(\boldsymbol{A}_i\boldsymbol{\Sigma})$ *and* $\lambda_i = \tau'\boldsymbol{A}_i\tau$ *for* $i \in \{1, 2, 3\}$.
iv. *Degrees of freedom for* $\{Q_1, Q_2, Q_3\}$ *are* $\nu_i \in \{1, n-2, n-1\}$, *respectively.*

v. Q_1 *and* Q_2 *are distributed independently.*
vi. $\mathcal{L}\left(t_i^2\right) = t^2\,(n-2, \lambda_1, \lambda_2)$.
vii. *Properties (iii)–(vi) hold independently of* $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$, *and are identical to the case that* $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$.

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 $\boldsymbol{A}_3\boldsymbol{\Sigma}(\theta) = \mathrm{B}_n\left(\boldsymbol{I}_n + \theta\mathbf{1}_n\mathbf{1}_n'\right) = \mathrm{B}_n$, so that $\mathrm{tr}(\boldsymbol{A}_3\boldsymbol{\Sigma}(\theta)) = n-1$ as in conclusion (iv). Moreover, $\boldsymbol{A}_3\boldsymbol{\Sigma}(\theta)\boldsymbol{A}_3 = \boldsymbol{A}_3^2 = \boldsymbol{A}_3$ as in conclusion (i) for $\{i=3\}$ since $\mathbf{A_3}$ is idempotent. Similarly,

$$\boldsymbol{A}_3\boldsymbol{\Sigma}(\boldsymbol{\xi}) = \boldsymbol{B}_n\left(\boldsymbol{I}_n + \mathbf{1}_n\boldsymbol{\xi}' + \boldsymbol{\xi}\mathbf{1}_n' - \bar{\xi}\mathbf{1}_n\mathbf{1}_n'\right) = \boldsymbol{B}_n + \boldsymbol{B}_n\boldsymbol{\xi}\mathbf{1}_n' \tag{A.5}$$

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

Case 2: For $\boldsymbol{A}_2 = Diag\,(\boldsymbol{B}_r, 0)$, the partitioned form (A.2) gives

$$\boldsymbol{A}_2\boldsymbol{\Sigma}(\theta) = \begin{bmatrix} \boldsymbol{B}_r & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix}\begin{bmatrix} \left(\boldsymbol{I}_r + \theta\mathbf{1}_r\mathbf{1}_r'\right) & \theta\mathbf{1}_r\mathbf{1}_s' \\ \theta\mathbf{1}_s\mathbf{1}_r' & \left(\boldsymbol{I}_s + \theta\mathbf{1}_s\mathbf{1}_s'\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_r & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} = \boldsymbol{A}_2 \tag{A.6}$$

and $\mathrm{tr}(\boldsymbol{A}_2\boldsymbol{\Sigma}(\theta)) = n-2$ as in conclusion (iv). In parallel with Case 1, $\boldsymbol{A}_2\boldsymbol{\Sigma}(\theta)\boldsymbol{A}_2 = \boldsymbol{A}_2$ to give conclusion (i) for $(i=2)$. To continue, in the partitioned form (A.3) we have

$$\begin{aligned}\boldsymbol{A}_2\boldsymbol{\Sigma}(\boldsymbol{\xi}) &= \begin{bmatrix} \boldsymbol{B}_r & 0 \\ 0' & 0 \end{bmatrix}\left\{\begin{bmatrix} \boldsymbol{I}_r & 0 \\ 0' & 1 \end{bmatrix} + \begin{bmatrix} \mathbf{1}_r \\ 1 \end{bmatrix}\boldsymbol{\xi}' + \boldsymbol{\xi}\,[\,\mathbf{1}_r' \quad 1\,] - \bar{\xi}\begin{bmatrix} \mathbf{1}_r\mathbf{1}_r' & \mathbf{1}_r \\ \mathbf{1}_r' & 1 \end{bmatrix}\right\} \\ &= \boldsymbol{A}_2 + \boldsymbol{A}_2\boldsymbol{\xi}\mathbf{1}_n'\end{aligned} \tag{A.7}$$

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

Case 3: Next consider $\boldsymbol{A}_1 = \boldsymbol{A}_s - \boldsymbol{A}_2$. From the foregoing developments infer that $\mathrm{tr}\,(\boldsymbol{A}_1\boldsymbol{\Sigma}(\theta)) = (n-1) - (n-2) = 1$ as in conclusion (iv), whereas $\boldsymbol{A}_1\boldsymbol{\Sigma}(\theta)\boldsymbol{A}_1 = \boldsymbol{A}_1$ to verify conclusion (i) for $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta)$. Similarly, $\mathrm{tr}\,(\boldsymbol{A}_1\boldsymbol{\Sigma}(\boldsymbol{\xi})) = (n-1)-(n-2) = 1$ as in conclusion (iv), and $\boldsymbol{A}_1\boldsymbol{\Sigma}(\boldsymbol{\xi})\boldsymbol{A}_1 = \boldsymbol{A}_1$ as in conclusion (i) for (i=1) and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\xi})$.

Case 4: To verify conclusion (ii) separately for $\boldsymbol{\Sigma}(\theta)$ and $\boldsymbol{\Sigma}(\boldsymbol{\xi})$, take $\boldsymbol{A}_3\boldsymbol{\Sigma}(\theta) = \boldsymbol{A}_3$ from Case 1 and $\boldsymbol{A}_2\boldsymbol{\Sigma}(\theta) = \boldsymbol{A}_2$ from (A.7), so that $\boldsymbol{A}_1\boldsymbol{\Sigma}(\theta)\boldsymbol{A}_2 = (\boldsymbol{A}_3 - \boldsymbol{A}_2)\boldsymbol{A}_2$, namely,

$$\boldsymbol{A}_2\boldsymbol{\Sigma}(\theta)\boldsymbol{A}_2 = (\boldsymbol{A}_3 - \boldsymbol{I}_n)\boldsymbol{A}_2 = -\frac{1}{n}\begin{bmatrix} \mathbf{1}_r\mathbf{1}_r' & \mathbf{1}_r \\ \mathbf{1}_r' & 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{B}_r & \mathbf{0} \\ \mathbf{0} & 0) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{A.8}$$

To continue, write $\boldsymbol{A}_3\boldsymbol{\Sigma}(\boldsymbol{\xi}) = \mathrm{B}_n + \mathrm{B}_n\boldsymbol{\xi}\mathbf{1}_n'$ from Eq. (A.6), and $\boldsymbol{A}_2\boldsymbol{\Sigma}(\boldsymbol{\xi}) = \boldsymbol{A}_2 + \boldsymbol{A}_2\boldsymbol{\xi}\mathbf{1}_n'$ from Eq. (A.8), so that $\boldsymbol{A}_1\boldsymbol{\Sigma}(\boldsymbol{\xi})\boldsymbol{A}_2 = (\boldsymbol{A}_3 - \boldsymbol{A}_2)\boldsymbol{A}_2 + (\boldsymbol{A}_3 - \boldsymbol{A}_2)\,\boldsymbol{\xi}\,[\,\mathbf{1}_r' \quad 1\,]\boldsymbol{A}_2$. But $[\,\mathbf{1}_r' \quad 1\,]\boldsymbol{A}_2 = 0$, and $(\boldsymbol{A}_3 - \boldsymbol{A}_2)\boldsymbol{A}_2 = 0$ from Eq. (A.8). Conclusion (v) follows as a consequence of (ii) and Craig's Theorem, as in Mathai and Provost (1992, 209).

Case 5: The noncentralities in (iii) use that $\mathcal{L}(Q_i) = \chi^2(\nu_i, \lambda_i)$ independently of $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$ and, if $\boldsymbol{Y}_0 \in \mathbb{R}^n$ is random having $E(\boldsymbol{Y}_0) = \tau$, the noncentrality parameter for $\boldsymbol{Y}_0'\boldsymbol{A}\boldsymbol{Y}_0$ is the quadratic form $\lambda = \boldsymbol{\tau}'\boldsymbol{A}\boldsymbol{\tau}$ in its expectation. □

Under shifts $\boldsymbol{\gamma} \neq \mathbf{0}$, an innovation of the present study, distributions of t_i^2 emerge as doubly noncentral, or as singly noncentral in either the numerator or denominator, depending on $\boldsymbol{\gamma}$. We next specialize Theorem A.1 for the case of reference models $\{\boldsymbol{Y}_0 = \mu\mathbf{1}_n + \boldsymbol{\varepsilon}_0\}$ for which shifts are allowed in both deleted and nondeleted rows; that is, $\boldsymbol{\gamma} \neq \mathbf{0}$ with $\omega' = [\gamma', \delta_i]$. By decomposing $\boldsymbol{\gamma}$ into its canonical components, formulas for the parameters of $t^2(\nu, \lambda_1, \lambda_2)$ are derived. Details follow.

Theorem A.2. *Given $\mathcal{L}(\boldsymbol{Y}_0) = N_n(\boldsymbol{\tau}, \boldsymbol{\Sigma})$ with $\tau = \mu\mathbf{1}_n + \boldsymbol{\omega}$ and $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$, with $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \delta_i]'$. Decompose $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2$ with $\boldsymbol{\gamma}_1 = \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_2 = \left(\boldsymbol{I}_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}'_r\right)\boldsymbol{\gamma}$ as in Definition 1; andlet $\nu = n - 2$. Then the distribution $\mathcal{L}(t_i^2)$ is the doubly noncentral $\mathcal{L}(t_i^2) = t^2(\nu; \lambda_1, \lambda_2)$ with:*

i. $\lambda_1 = \frac{r}{n}(-\bar{\gamma} + \delta_i)^2$.
ii. $\lambda_2 = \boldsymbol{\gamma}'_2\boldsymbol{\gamma}_2 = \sum_{i=1}^{r}(\gamma_i - \bar{\gamma})^2$.
iii. *If $\boldsymbol{\gamma} \in \mathcal{R}(\mathbf{1}_r\mathbf{1}'_r)$, then $\mathcal{L}(t_i^2) = t^2(\nu, \lambda_1, 0)$ with λ_1 as in (i); and at $\delta_i = 0$, then $\lambda_1 = \frac{r}{n}\bar{\gamma}_1^2$.*
iv. *If $\boldsymbol{\gamma} \in \mathcal{R}(\boldsymbol{I}_r - \mathbf{1}_r\mathbf{1}'_r)$, then $\mathcal{L}(t_i^2) = t^2(\nu, \lambda_1, \lambda_2)$ with $\lambda_1 = \frac{r}{n}(-\bar{\gamma}_2 + \delta_i)^2$ and $\lambda_2 = \boldsymbol{\gamma}'_2\boldsymbol{\gamma}_2 > 0$.*
v. *The "regression effect," $\delta_i = \bar{\gamma}$, yields $\lambda_1 = 0$.*
vi. *The foregoing properties hold for $\mathcal{L}(t_i^2|\boldsymbol{\Sigma}) = \mathcal{L}(t_i^2|\sigma^2\boldsymbol{I}_n)$, independently of $\boldsymbol{\Sigma} \in \{\boldsymbol{\Sigma}(\theta), \boldsymbol{\Sigma}(\boldsymbol{\xi})\}$.*

Proof. In keeping with Theorem A.1(iii), identify $\tau = \mathrm{E}(\boldsymbol{Y}_0) = \mu\mathbf{1}_n + \boldsymbol{\omega}$ with $\boldsymbol{\omega} = [\boldsymbol{\gamma}', \delta_i]'$. A direct evaluation from $\boldsymbol{A}_1 = (\boldsymbol{A}_3 - \boldsymbol{A}_2)$, as displayed earlier in partitioned form following Eq. (A.4), gives $\lambda_1 = \boldsymbol{\tau}'\boldsymbol{A}_1\boldsymbol{\tau} = \frac{r}{n}(-\bar{\gamma} + \delta_i)^2$ as in conclusion (i). Similarly, $\lambda_2 = \boldsymbol{\tau}'\boldsymbol{A}_2\boldsymbol{\tau} = \tau'\,Diag(\boldsymbol{B}_r, 0)\,\boldsymbol{\tau} = \boldsymbol{\gamma}'\boldsymbol{B}_r\boldsymbol{\gamma}$ as in conclusion (ii). Conclusion (iii) follows since $\boldsymbol{\gamma} \in \mathcal{R}(\mathbf{1}_r\mathbf{1}'_r)$ implies that $\boldsymbol{\gamma}_2 = \mathbf{0}$, and conclusion (iv) since $\gamma \in \mathcal{R}(\boldsymbol{I}_r - \mathbf{1}_r\mathbf{1}'_r)$ implies $\boldsymbol{\gamma}_1 = \mathbf{0}$. Conclusion (v) follows from (i), and (vi) from Theorem A.1(vii). □

A.2. *Calculations of* p *values*

Building on the earlier work of Imhof (1961), Ennis and Johnson (1993) have expressed the *cdf* for the doubly noncentral $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ 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 Table 5.

For the p values $p_{(G)}$ of Grubbs's test, we follow Eq. (7) of Nair (1948) and use the Thompson (1935) procedure that relates the probability for a deviate Z_i to the probability from the Student t_ν distribution with degrees of freedom $\nu = n - 2$ as

$$\Pr(Z_i = (y_i - \bar{y})/S < \tau) = \Pr(t < t_*|\mathcal{L}(t) = t_\nu)$$

$$\tau = (n-1)\,t_*/\sqrt{nt_*^2 + n(n-2)};$$

and thus $Pr(Z_{max} = (y_{\max} - \bar{y})/S < \tau) = [Pr(t < t_*|\mathcal{L}(t) = t_\nu)]^n$.

Regarding p values $p_{(D)}$ for Dixon's test, we use the tabulated critical values from Dixon (1951) and Rorabacher (1991), who report c_α values for α in

$$[0.5\%, 1\%, 2\%, 5\%, 10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 95\%].$$