Validating Hotelling’s $T^2$
Diagnostics in Mixtures

Revision of 06–01–2016

D. R. Jensen, Department of Statistics, Virginia Tech, Blacksburg, VA 24061.
Email: djensen@vt.edu

D. E. Ramirez, University of Virginia, P. O. Box 400137, Charlottesville, VA 22904.
Email: der@virginia.edu

Abstract
Let $Y_0 = [Y_1, \ldots, Y_n]'$ be $k$–dimensional Gaussian vectors having the common mean $\mu$ and dispersion matrix $\Sigma$. To identify outlying shifts of type $\{Y_i \rightarrow Y_i + \delta\}$, Hotelling’s $T^2_i$ diagnostics properly account for dependencies among the $k$ elements of $Y_i$. This study reconsiders the use of $T^2_i$ under failure of the classical venue that $[Y_1, \ldots, Y_n]$ themselves are mutually independent. Diagnostics for the case $k = 1$ include those of Dixon (1950), Grubbs (1950) and Ferguson (1961) based on order statistics, and the $R$–Student deletion diagnostics $t^2_i$, all predicated on independent observations. To relax the latter, Jensen and Ramirez (2015) recently showed these procedures to remain exact in level and power under specified dependencies among $[Y_1, \ldots, Y_n]$, and for scale mixtures of these dependent distributions. These findings are in keeping with correlated data, as found in the analysis of calibrated measurements, for example. Specifically, Hotelling’s $T^2_i$ diagnostics are shown here to remain exact in level and power for dispersion mixtures of matrix Gaussian errors having star–shaped contours, to considerably enhance their range of applicability.

AMS Subject Classification: 62E15, 62H15, 62J20
Keywords: Outlying data; deletion diagnostics; dependent errors; Hotelling’s $T^2_i$. 
1 Introduction

1.1 Overview

Let \( Y_0 = [Y_1, \ldots, Y_n]' \) comprise \( k \)-dimensional vector observations having the common mean \( E(Y_i) = \mu \) and dispersion matrix \( V(Y_i) = \Sigma \). To identify outlying vector shifts of type \( \{Y_i \rightarrow Y_i + \delta\} \), Hotelling’s \( T^2_i \) diagnostics are found on partitioning \( Y_0 = [Y', Y_i]' \), and the observed error matrix as \( E_0 = [E', e_i]' \). Then \( T^2_i \) may be expressed as

\[
T^2_i = \frac{n}{n-1} e_i'S_i^{-1}e_i, \tag{1.1}
\]

with \( S_i \) as the sample dispersion matrix from \( Y' \); and the \( R \)-Student \( t^2 \) emerges on specializing to \( k = 1 \). Early developments regarding \( T^2_i \) are found in Caroni (1987) and references cited, and more recently, Barrett (2003). Validating assumptions traditionally entail both normality and the mutual independence of \([Y_1, \ldots, Y_n]\), the latter giving \( V(Y_0) = I_n \otimes \Sigma \) as a Kronecker product.


Diagnostics for \( k = 1 \) were reassessed in Jensen and Ramirez (2015), to include those of Dixon (1950), Grubbs (1950), and Ferguson (1961) based on order statistics, and the \( R \)-Student \( t^2 \) statistics. These were shown to remain exact in level and power under specified dependencies among \([Y_1, \ldots, Y_n]\), and for dispersion mixtures of these distributions having star–shaped contours. Those distributions are extended here to encompass mixtures of matrix distributions supporting properties of \( T^2_i \) under models for dependency to be specified. For these models the dispersion matrix instead becomes \( V(Y_0) = \Omega \otimes \Sigma \). In addition, dispersion mixtures of these support that normality and independence among \([Y_1, \ldots, Y_n]\) are sufficient but not necessary for validity: \( T^2_i \) remains exact in level and power for all such mixture distributions.

Mixtures figure prominently in statistical practice. Univariate data from subsamples have been modeled as Gaussian mixtures in Box and Tiao (1968) and in Aitken and Wilson (1980). Moreover, calibrated data subject to errors of calibration typically are equicorrelated under both direct and inverse calibration; see Jensen and Ramirez (2009, 2012). These features continue for multivariate observations in those settings. In addition, mixtures are central in Gaussian clustering models; see Punzo, Browne and McNicholas (2016). Accordingly, normal-theory \( T^2_i \) diagnostics are seen to apply exactly in a large class of non-Gaussian matrix distributions, thus validating their usage more widely than heretofore, and to be genuinely nonparametric. To continue, it is useful to place in perspective the sense in which \( T^2_i \) properly accounts for dependencies among elements of a typical \( Y_i \in \mathbb{R}^k \).
1.2 Vector Diagnostics: A Critique.

A hallmark of multivariate data is that observations \( Y'_i = [Y_{i1}, \ldots, Y_{ik}] \) at each experimental unit are dependent. Arrays of these \( k \)-dimensional vectors thus comprise a constellation of points in \( \mathbb{R}^k \), as do the ordinary residuals \( \{e_1, \ldots, e_n\} \) with typical element \( e'_i = [e_{i1}, \ldots, e_{ik}] \). The object is to identify outlying units in the constellation of points in \( \mathbb{R}^k \), and it seems essential to account for dependent responses as in \( T^2_i \). An alternative, advocated by some, relies on scalar diagnostics marginally for each of the \( k \) responses. These distinct approaches merit further consideration here.

To fix ideas, consider residuals \( e'_i = [e_1, e_2] \) and the ellipse in Figure 1 centered at \([0, 0]\). In reference to equation (1), the boundary and interior of the ellipse correspond to \( \{T^2_i \leq c_\alpha\} \) and its complement to \( \{T^2_i > c_\alpha\} \) with probability \( \alpha = 0.10 \). On the other hand, the intervals \([-k_1, +k_1]\) on the \( e_1 \)-axis, and \([-k_2, +k_2]\) on the \( e_2 \)-axis, are marginal cutoff rules. Clearly the circled \( \odot \) to the upper left of center is a \( T^2_i \)-outlier but within both marginal cutoff intervals. On the other hand, the circled \( \oplus \) to the upper right of center is within the 90\% cluster but outside both marginal cutoff intervals. Clearly the marginal cutoff rules fail to capture the essence of whether or not a point \( e_i \in \mathbb{R}^2 \) is consistent with an extant Gaussian constellation. Such anomalies persist, and often are exacerbated, when considered over a variety of correlation structures, and especially over dimensions \( k \geq 2 \) where identities may be obscured. This matter is considered further in Section 4.1 in reference to data from the literature. An outline of the study follows.

Preliminary developments are given in Section 2. The principal findings follow in Section 3, and some consequences of these are detailed through examples in Section 4. Critical supporting topics, to include essential matrix distributions and a survey of univariate deletion diagnostics, are attached for completeness as an Appendix. A signal result characterizes Wishart matrix forms through chi–squared quadratic forms, of univariate deletion diagnostics, are attached for completeness as an Appendix. A random \( Y \in \mathbb{R}^n \) has distribution \( \mathcal{L}(Y) \); its mean vector \( \mathbb{E}(Y) \); its dispersion matrix \( \mathbb{V}(Y) = \Sigma \), say, with variance \( \text{Var}(Y) = \sigma^2 \) on \( \mathbb{R}^1 \); its density function (pdf) \( g(y) \); its cumulative distribution function (cdf) \( G(y) \); and its characteristic function \( (chf) \phi_Y(t) \). Specifically, \( \mathcal{L}(Y) = N_n(\mu, \Sigma) \) is Gaussian in \( \mathbb{R}^n \) with mean vector \( \mu \) and dispersion matrix \( \Sigma \), whereas \( \mathcal{L}(Y_0) = N_{nk}(M, \Omega \otimes \Sigma) \) is Gaussian in \( \mathbb{R}^{nk} \) with designated parameters. A random \( W \in \mathbb{S}^n_i \) is said to have the Wishart distribution
\(\mathbb{W}_\nu(\nu, \Sigma, \Theta)\) of order \(k\), with \(\nu\) degrees of freedom, the scale parameters \(\Sigma\), and the noncentrality matrix \(\Theta\). Further details are supplied in Appendix A.1.

Distributions on \(\mathbb{R}^d_+\) include \(\chi^2(\nu; \sigma^2, \lambda)\) as chi-squared with argument \(u\), having \(\nu\) degrees of freedom, the scale parameter \(\sigma^2\), and noncentrality \(\lambda\); and Hotelling’s (1931) \(T^2_\nu(u; \nu, \lambda)\) of order \(k\) having \(\nu\) degrees of freedom and noncentrality parameter \(\lambda\). Identify \(\{T^2_\nu > c_\alpha\}\) as the conventional \(\alpha\)-level rejection rule based on \(T^2_\nu(u; \nu, 0)\).

### 2.2 The Model

The model, often specialized here, is \(\{Y_0 = X_0B + E_0\}\) with \(Y'_0 = [Y_1, \ldots, Y_n]\), together with \(X'_0 = [x_1, \ldots, x_n]\) of rank \(d < n\) with \(x_i\) as design point \(i\). Single-case deletions preserve \(\{Y = XB + E\}\) of full rank after eliminating row \([Y'_i, x'_i, \epsilon'_i]\) and \(B = [\beta_1, \ldots, \beta_k]\). Here \(E_0 = [\epsilon', \epsilon_i']\) consists of random errors; \(E_0 = [\epsilon', \epsilon_i']\) are ordinary residuals; and \(H_n = X_0(X'_0X_0)^{-1}X'_0\). Gauss-Markov assumptions are extended here such that \(V(Y_0) = \Omega \otimes \Sigma\), where \(\Omega\) takes values \(\Omega(\rho)\), equivalently \(\Omega(\theta)\), and \(\Omega(\xi)\) as in Section 2.4.

**Assumptions A.** The following hold.

- **A.1.** \(E(E_0) = \Delta = [0, \delta_i]'\) such that \(E(E) = 0\) and \(E(\epsilon_i) = \delta_i';\)
- **A.2.** \(V(E_0) = \Omega \otimes \Sigma\) with \(\Omega \in \{\Omega(\rho), \Omega(\xi)\}\); and
- **A.3.** \(L(E_0) = N_{nxk}(\Delta, \Omega \otimes \Sigma)\) for \(\Omega \in \{\Omega(\rho), \Omega(\xi)\}\).

As in conventional deletion diagnostics, this represents a shift \(\{Y_i \rightarrow Y_i + \delta_i\}\) at the design point \(x_i\) in \(X_0\).

### 2.3 Vector Deletions

Diagonstics here gauge the impact of \(Y_i\) on deleting \([Y'_i, x'_i, \epsilon'_i]\) as in Section 2.2. In Table 1 are vector versions of selected univariate diagnostics, where details for entries marked with (*) are found in Hossain and Naik (1989); see also Hossain and Naik (1991) and Naik (2003). Here \((\vec{B}, \vec{B})\) are the OLS solutions, and \((S, S_i)\) the sample dispersion matrices, from \(Y_0\) and from \(Y\) after deleting \(Y_i\); and \(b_{ii}\) is the \((i, i)\) element of \(H_n\).

Vector versions are \(\text{OUT}_i\) of Barnett and Lewis (1994); \(\text{CR}_i\) of Belsley et al. (1980); \(C_i\) of Cook (1977); and \(\text{WKS}_i\) of Welsch and Kuh (1977); where \(\{D_1^2, D_2^2, D_3^2\}\) are generalized non–Euclidean distances. Another criterion, advocated by Cook and Weisberg (1982), is the likelihood displacement \(L_D(\theta) = 2[L(\theta) - L(\tilde{\theta})]\), where \(\tilde{\theta}\) and \(\tilde{\theta}\) are Maximum Likelihood estimators with and without \(Y_i\), and \(L(\cdot)\) is the likelihood function. See also Cook (1987). Details in the present context are given in Table 1, where \(L_D(\Sigma)\) and \(L_D(\Sigma|B)\) are likelihood displacements with other parameters held fixed. Hossain and Naik (1989) illustrated \(\text{CR}_i, C_i,\) and \(\text{WKS}_i\) using three data sets from the literature.

The Table 1 entries reflect diverse attributes of \(\{Y_0 = X_0B + E_0\}\) and its analysis. As in the case of univariate data, users may wish to report and interpret these diagnostics, contingent on benchmarks pertaining to each. As noted in Barrett and Gray (1997; p.41): “There are many different types of influence and influence measures. Cases that
are influential according to one influence measure may not be influential with respect to another measure.” In explanation, the Table 1 diagnostics drew from an excess of univariate diagnostics which had evolved in disarray through flawed intuition and ad hoc cutoff rules later shown to be inconsistent. See Appendix A.2. To avoid such pitfalls in the present study, it is essential that correct and objective external benchmarks be employed.

Accordingly, all Table 1 diagnostics are equivalent, relating one–to–one with $T^2_i$; their joint distribution is singular of unit rank; none contains essential evidence not contained in another, differing only in their representation of that evidence; each exceeds or fails its respective benchmark when applied consistently; diagnostics other than $T^2_i$ are redundant; and conclusions using $T^2_i$ are identical for all listed diagnostics. Details follow, where

2.4 The Matrices $\Xi$

Validity in linear inference rests in part on the structure of dispersion matrices. Three cases are considered, namely $\Omega(\theta)$, $\Omega(\xi)$, and $\Omega(\rho)$ where, for $\xi' = [\xi_1, \ldots, \xi_n]$, we have $\tau_1 = \xi_1 + \ldots + \xi_n = n\xi$ and $\tau_2 = \sum_{i=1}^{n}\xi_i^2$. Details follow, where

\begin{table}
\centering
\begin{tabular}{|c|c|c|}
\hline
Diagnostic & Definitions/Connections & Impact of $Y_i$ on $X$ \\
\hline
$T^2_i$ & $\frac{e_i'S^{-1}e_i}{(1-h_{ii})}$ & $H_0: \delta_i = 0$ vs $H_1: \delta_i \neq 0$ \text{Identity Relation} \\
$u_i^2$ & $\frac{e_i'S^{-1}e_i}{(1-h_{ii})}$ & $H_0: \delta_i = 0$ vs $H_1: \delta_i \neq 0$ \text{Identity Relation} \\
$\bar{u}_i^2$ & $\frac{T^2_i(n-k-1)}{(T^2_i+n-k-2)}$ & \text{Sample Dispersion Matrices} \\
$\text{OUT}_i$ & $1 - \frac{|S_i|}{|B_i|} = 1 - \frac{T^2_i+n-d-1}{n-d-1}$ & $|V(\bar{B})|/|V(B)|$ \\
$*\text{CR}_i$ & $\frac{1}{(1-h_{ii})} T^2_i$ & $\text{D}^2_i(\bar{B}, \bar{B})$ \\
$*\text{C}_i$ & $\frac{1}{(1-h_{ii})} T^2_i$ & $\text{D}^2_i(\bar{Y}_i, \bar{Y}_i(i))$ at $x_i$ \\
$\text{h}_i^2$ & $\frac{1}{(1-h_{ii})} T^2_i$ & $\text{D}^2_i(\bar{Y}_i, \bar{Y}_i(i))$ at $x_i$ \\
$\text{h}_i^2$ & $\frac{1}{(1-h_{ii})} T^2_i$ & $\text{D}^2_i(\bar{Y}_i, \bar{Y}_i(i))$ at $x_i$ \\
$*\text{LD}_i(B, \Sigma)$ & $-n \log\left(\frac{T^2_i+n-d-1}{n-d-1}\right) + \frac{T^2_i}{(1-h_{ii})} - k$ & $\text{LD}_i$, $\text{LD}_i(\Sigma)$ \\
$*\text{WK}_i$ & $\text{LD}_i(B, \Sigma) - \text{LD}_i(\Sigma, B)$ & $\text{LD}_i$, $\text{LD}_i$ \\
\hline
\end{tabular}
\caption{Single–case deletion diagnostics for the model $\{Y_0 = X_0B + E_0\}$ such that $\mathcal{L}(E_0) = N_{n\times k}(\Delta, I_n \otimes \Sigma)$, where $(\hat{Y}_i, \hat{Y}_i(i))$ are predicted values at design point $x_i$ with and without $Y_i$, and $\text{LD}_i(B, \Sigma)$ is the Likelihood Displacement.}
\end{table}
Lemma 2.1. (i) Let $\Omega(\theta) = \sigma^2(I_n + \theta \mathbf{1}_n \mathbf{1}_n')$; its eigenvalues are 1.0, with multiplicity $n-1$, and 1 + $n\theta$, so that $\Omega(\theta)$ is positive definite if and only if $\theta \in \Gamma_1 = \{ \theta : \theta > -\frac{1}{n} \}$. 

(ii) Let $\Omega(\xi) = \sigma^2(I_n + 1_\xi + \xi \mathbf{1}_n' - \mathbf{1}_n \mathbf{1}_n')$ with $0 \neq \xi \neq \theta \mathbf{1}_n$; its ordered eigenvalues are $\kappa_1 = 1 + \alpha_1, \kappa_2 = \ldots = \kappa_{n-1} = 1, \kappa_n = 1 + \alpha_n$ as in (2.1); then $\Omega(\xi)$ is positive definite if and only if $\xi \in \Gamma_2 = \{ \xi \in \mathbb{R}^n : \tau_1 > n\tau_2 - 1 \}$.

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

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

Observe that $\Omega(\theta)$ and $\Omega(\rho)$ are equivalent. On taking $\theta = \rho/(1 - \rho)$, it follows that $\Omega(\theta) = \frac{1}{1-\rho} \Omega(\rho)$. We often take $\Omega(\theta)$ for convenience, despite that $\Omega(\rho)$ occurs prominently in practice. Accordingly, the collections $\Xi_1 = \{ \Omega(\theta) : \theta \in \Gamma_1 \}$, equivalently, $\Xi_1 = \{ \Omega(\rho) : \rho \in \Gamma_1 \}$, and $\Xi_2 = \{ \Omega(\xi) : \xi \in \Gamma_2 \}$, comprise ensembles of positive definite matrices, to be amalgamated as $\Xi = \Xi_1 \cup \Xi_2$. For further details see Jensen (1996).

2.5 Mixture Distributions

To continue, on taking Assumption $A_1$ into account, let $\Lambda = 1_n \otimes \mu' + \Delta$ and consider $g_{rock}(y; \Lambda, \Omega \otimes \Sigma)$ in $F_{rock}$ as the Gaussian density corresponding to $\mathbb{N}_{n \times k}(\Lambda, \Omega \otimes \Sigma)$ as in Appendix A.1. These generate ensembles as $\Omega$ ranges over $\Xi$, namely

$$E_1 = \{ g_{rock}(y; \Lambda, \Omega(\theta) \otimes \Sigma); \theta \in \Gamma_1 \}$$  

$$E_2 = \{ g_{rock}(y; \Lambda, \Omega(\xi) \otimes \Sigma); \xi \in \Gamma_2 \}.$$  

Next visualize the ensemble $E_1$ to have mixing parameters $\theta$, and $E_2$ to have mixing parameters $\xi$. Then mixtures in $F_{rock}$ of type

$$f_i(y; \Lambda, G_i) = \int_{\Gamma_1} g_{rock}(y; \Lambda, \Omega(i) \otimes \Sigma)dG_i(\cdot)$$  

emerge with $G_i \in \{ G_1, G_2 \}$ as cdfs on $\Gamma_i \in \{ \Gamma_1, \Gamma_2 \}$, and with $\Omega(i) \in [\Omega(\theta), \Omega(\xi)]$. In particular, the densities $f_1(y; \Lambda, G_1)$ and $f_2(y; \Lambda, G_2)$ are dispersion mixtures of elliptical Gaussian distributions on $F_{rock}$ centered at $\Lambda \in F_{rock}$. Let $G_1$ and $G_2$ comprise all cdfs on $\Gamma_1$ and $\Gamma_2$, respectively; these in turn generate the collections

$$M_1 = \{ f_1(y; \Lambda, G_1); G_1 \in G_1 \}$$  

$$M_2 = \{ f_2(y; \Lambda, G_2); G_2 \in G_2 \}$$  

comprising all dispersion mixtures of the referenced types.

3 The Principal Findings

3.1 Overview

Specializing $\{ Y_0 = 1_n \otimes \mu' + E_0 \}$ from $\{ Y_0 = X_0 B + E_0 \}$ in $F_{rock}$ as reference, and noting that $H_n = \frac{1}{n} 1_n 1_n'$, we rearrange elements such that $Y_0 = [Y', Y_1]', E_0 = [E', \epsilon_i]'$,
and $E_0 = [E', e_i]'$, and consider arbitrary shifts $\{Y_i \rightarrow Y_i + \delta_i\}$. Nonstandard versions of $L(T^2_i)$ as in expression (1.1) are to be studied, but where the independence of $\{Y_i, \ldots, Y_n\}$ fails. Instead take $L(Y_0) = N_{n\times k}(M, \Omega \otimes \Sigma)$ with $\Omega \in \{\Omega(\rho), \Omega(\xi)\}$, equivalently $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$, as the basic model undergirding the $T^2_i$ diagnostics. Here $M = \Lambda = L_0 \otimes \mu + \Delta$ from Assumption $A_1$. It is seen that shifts $\{Y_i \rightarrow Y_i + \delta_i\}$ propagate into noncentrality parameters of $L(T^2_i)$. These findings in turn rest on matrices of quadratic and bilinear forms of type $Y_0'AY_0$ in the observed matrix array $Y_0 \in F_{n\times k}$ as in Appendix A.1.

### 3.2 Properties of Residuals

The observed residuals $E_0$ under Assumptions $A$ are germane, as $T^2_i$ is a function of these. In particular, it remains to evaluate $E(e_i)$, $\text{Var}(e_i)$ and $L(e_i)$ as special cases. Details follow, where $r = (n-1)$ and $E(E_0) = \Delta = [\delta', \delta_i]'$ as Assumption $A_1$.

**Theorem 1.** Consider the ordinary residuals $E_0 = [E', e_i]'$ under Assumptions $A$, with $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$, and let $T(E_0)$ be a mapping to a linear space $\mathcal{V}$. Then the following properties hold independently of $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$.

1. $\mathcal{V}(E_0) = B_n \otimes \Sigma$ and $L(E_0) = N_{n\times k}(\Delta, B_n \otimes \Sigma)$;
2. $E(e_i) = \frac{1}{n} \delta_i', \mathcal{V}(e_i) = \frac{1}{n} \Sigma$, and $L(e_i) = N_k(\frac{1}{n} \delta_i, \frac{1}{n} \Sigma)$;
3. $L(T(E_0|\Omega)) = L(T(E_0|I_n))$.

**Proof.** Observe from $E_0 = B_n Y_0$ and the conventions of Appendix A.1 that $\mathcal{V}(E_0) = B_n \Omega B_n + \Sigma = B_n \otimes \Sigma$ for $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$, since $B_n I_n = 0$ annihilates successive terms in the products following the first. This together with Assumption $A_3$ gives (i). The expected product $E(E_0) = (I_n - \frac{1}{n} I_n I_n') \Delta$ in partitioned form is

$$
    E \left[ e_i' \right] = \frac{1}{n} \begin{bmatrix} (n I_n - 1) & -1_r \\ -1_r & r \end{bmatrix} \begin{bmatrix} 0 \\ \delta_i' \end{bmatrix} = \frac{1}{n} \begin{bmatrix} -1_r \delta_i' \\ r \delta_i' \end{bmatrix}.
$$

(3.1)

Thus $E(e_i) = r \delta_i / n$ and $\mathcal{V}(e_i) = \frac{1}{n} \Sigma$ is the $(n, n)$ block of $B_n \otimes \Sigma$ which, together with normality, give conclusion (ii). Conclusion (iii) follows directly. \(\square\)

### 3.3 Nonstandard Matrix Forms

Generalizing from Lemma A.1(iii) of Jensen (2001a) and from Jensen and Ramirez (2014) in order to accommodate matrix arrays, the multivariate Fisher–Cochran expansion generating $T^2_i$ is

$$
    Y_0'A_1 Y_0 + Y_0'A_2 Y_0 = Y_0'A_3 Y_0;
$$

(3.2)

$$
    \frac{r}{n} e_i e_i' + (r-1) S_i = E_0' E_0;
$$

(3.3)

$\mathcal{W}_k(1; \Sigma, \delta, \delta_i')$; $\mathcal{W}_k(r-1; \Sigma, 0)$; $\mathcal{W}_k(r; \Sigma, \delta, \delta_i')$;

(3.4)

where $(e_i, \delta_i)$ are of order $(k \times 1)$; the second line explains the first; and the third line identifies their respective distributions under the classical assumption $L(Y_0) = N_{n\times k}(M, I_n \otimes \Sigma)$ since $(A_1, A_2, A_3)$ are idempotent. These matrices are given explicitly in Jensen and Ramirez (2015) as $A_3 = B_n = (I_n - \frac{1}{n} I_n I_n')$, $A_2 = \text{Diag}(B_r, 0)$.
Invariant Properties of $T^2$

with $B_i = (I_n - \frac{1}{n} 1 1')$, and $A_1 = A_3 - A_2$. To continue, designate the aforementioned matrix forms as $Q_1 = Y_0' A_1 Y_0$, $Q_2 = Y_0' A_2 Y_0$, and $Q_3 = Y_0' A_3 Y_0$, and recall that $M = \Lambda = 1_n \otimes \mu' + \Delta$.

**Theorem 2.** Given $\mathcal{L}(Y_0) = N_n(\Lambda, \Omega \otimes \Sigma)$ under Assumptions $A$: take $E_0 = B_i Y_0$; and consider the matrix forms $\{Q_1, Q_2, Q_3\}$ as in (3.2). Then despite dependencies among elements of $Y_0 = [Y_1, \ldots, Y_n]'$, we have for each $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$ the following.

(i) $\mathcal{L}(Q_1) = \mathcal{L}_k(1; \Sigma, \Theta_1)$, $\Theta_i = \frac{1}{n} \delta_i' \delta_i'$;

(ii) $\mathcal{L}(Q_2) = \mathcal{L}_k(r - 1; \Sigma, 0)$;

(iii) $\mathcal{L}(Q_3) = \mathcal{L}_k(r; \Sigma, \Theta_1)$;

(iv) $Q_1$ and $Q_2$ are distributed independently;

(v) $T^2_1 = (r - 1) \text{tr } Q_1 Q_2^{-1}$;

(vi) $\{\mathcal{L}(T^2_1) = T^2_k(u; r - 1, \lambda_i); \lambda_i = \frac{1}{n} \delta_i' \Sigma^{-1} \delta_i; 1 \leq i \leq n\}$.

**Proof.** Fix $u \in \mathbb{R}^k$ and let $\{Q_i^u = u' Y_0' A_i Y_0 u; i = 1, 2, 3\}$, such that $\mathcal{L}(u' Y_0) = N_n(u' \Lambda, \sigma^2_u \Omega)$ with $\sigma^2_u = u' \Sigma u$. Drawing on Mathai and Provost (1992; p.201), Jensen and Ramirez (2015) established conclusions (i)–(iv) for $(Q_1^u, Q_2^u, Q_3^u)$ in terms of the corresponding $\chi^2$ distributions on demonstrating that $\{A_i \Omega A_i = \Lambda_i; i = 1, 2, 3\}$ and that $A_1 \Omega A_2 = 0$ for $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$, as in Appendix A.1 of Jensen and Ramirez (2015). Theorem A.1 of the attached Appendix now lifts those results to encompass the Wishart distributions of conclusions (i)–(iv). Conclusion (v) follows directly from expressions (1.1) and (3.3), and conclusion (vi) as the Wishart analog of the noncentral properties of Theorem A.1 of Jensen and Ramirez (2015). $\square$

### 3.4 Invariance under Mixtures

That the $T^2_1$ diagnostics may be valid under star–contoured errors is the subject of the following, where $\Lambda = 1_n \otimes \mu' + \Delta$ as in Assumption $A_1$.

**Theorem 3.** Given the model $\{Y_0 = \Lambda + E_0\}$ having a Gaussian mixture density $f_1(y_0; \Lambda, G_1)$ in $M_1$ or $f_2(y_0; \Lambda, G_2)$ in $M_2$ as in Section 2.4.

(i) Tests using $T^2_1$ remain exact in level and power for all mixtures in $M_1$;

(ii) Tests using $T^2_1$ remain exact in level and power for all mixtures in $M_2$;

(iii) These properties follow for all Table 1 diagnostics for mixtures in $M_1$ and $M_2$;

(iv) These $T^2$ distributions are identical to those initially derived from $\mathcal{L}(Y_0) = N_n(k, \Lambda, I_n \otimes \Sigma)$.

**Proof.** Return to Section 2.4 and expression (2.4). We argue conditionally as follows: (i) Fix $\Omega \in \{\Omega(\theta), \Omega(\xi)\}$; (ii) note from Theorem 3.2(vi) that $\mathcal{L}(T^2_1)$ holds independently of $\Omega$; then (iii) make the change–of–variables behind the integral, to conclude
for \(i = 1, 2\) that

\[
\begin{align*}
 f_i(y_0, \Lambda, G_i) &= \int_{\Gamma_i} g_{\text{rock}}(y_0; \Lambda, \Omega \otimes \Sigma) dG_i(\cdot) \\
 f_i(y_0, \Lambda, G_i) &\rightarrow \int_{\Gamma_i} T_i^2(u; r-1, \lambda_i) dG_i(\cdot) = T_i^2(u; r-1, \lambda_i)
\end{align*}
\]

independently of \(\Omega \in \{\Omega(\theta), \Omega(\xi)\}\) and of \(G_i\), and since \(\int_{\Gamma_i} dG_i = 1\). □

### 3.5 Other Venues

Calibrated data often entail calibration curves, direct or indirect, both injecting dependencies among the calibrated measurements; see Jensen and Ramirez (2009, 2012). These apply in the analysis of univariate data. Another venue adjusts observations directly to a common standard, as in compensating for the tare weight of a scale, or in assessing yield increments relative to a control yield as in Jensen (2001b). Subsequent examples fall within the latter framework, which we develop next for multivariate data amenable to Hotelling’s \(T^2\) diagnostics.

In short, observations \(Y_0 = [Y_1, \ldots, Y_n]'\) from \(\{Y_0 = 1_n \mu' + E_0\}\) are sought. For Case I the user sees \(Z_0 = [Z_1, \ldots, Z_n]'\) having \(\{Z_j = Y_j + W \in \mathbb{R}^k; 1 \leq j \leq n\}\) with scalar shifts \(W = [w_1, \ldots, w_k]\) often themselves random. To model this we proceed as follows: (a) Append \(Y_0 = [Y_1, \ldots, Y_n, W]'\); (b) suppose that \(V(Y_0^r) = I_N \otimes \Sigma\) with \(N = n+1\); and (c) let \(A = [I_n, 1_n]\). Then \(Z_0 = AY_0^r \in \mathbb{R}^{n+k}\). For Case II, if rows of \(Y_0\) are to be adjusted instead against \(W\) as standard, then \(\{Z_j = (Y_j - W); 1 \leq j \leq n\}\) and \(A = [I_n, -1_n]\). To continue, in both Case I and Case II we have \(V(Z_0) = AI_N A' \otimes \Sigma\) from Remark A.1. In what follows we parallel steps given heretofore in working from \(Y_0\) to \(T_i^2\).

**Theorem 4.** Begin with \(Z_0\) as constructed; rearrange as \(Z_0 = [Z', Z_i]'\) with \(Z_i\) as the test case; and determine \(T_i^2\) from \(Z_0\) as before using \(Y_0\). Then

(i) The residuals \(R_0 = B_n Z_0\) have \(V(R_0) = B_n \otimes \Sigma\) independently of \(W\);

(ii) \(L(T_i^2) = T_i^2(u; r-1, \lambda_i)\) independently of \(W\), with \(\lambda_i = \frac{\pi}{6} \delta_i \Sigma^{-1} \delta_i\);

(iii) \(T_i^2\) remains exact in level and power for all mixtures in \(M_1\) of (2.5);

(iv) These properties hold for all Table 1 diagnostics and for all mixtures in \(M_1\);

(v) These \(T^2\) distributions are identical to those initially derived from the unadjusted \(L(Y_0) = N_{\text{rock}}(A, I_n \otimes \Sigma)\).

**Proof.** Taking \(V(Z_0) = AA' \otimes \Sigma\) from \(A = [I_n, \pm 1_n]\), it follows that \(B_n AA' B_n \otimes \Sigma = B_n \otimes \Sigma\) since \(B_n\) is idempotent, to give conclusion (i). Conclusion (ii) follows directly, setting the stage so that conclusions (iii)–(v) now follow from Theorem 3, to complete our proof. □

**Remark 3.1.** Observe that a fractional adjustment \(\{Z_j = (Y_j \pm \kappa W); 1 \leq j \leq n\}\) can be achieved on taking \(A = [I_n, \pm \kappa 1_n]\). The stated conclusions follow directly if so modified.
4 Case Studies

4.1 Dependent vs Marginal Diagnostics

Section 1.2 offers a qualitative distinction between $T_1^2$, taking into account dependence between $[Y_1, Y_2]$, and marginal cutoff rules not accounting for dependence. Olive et al. (2015) essentially take marginal outliers to be descriptive of multivariate data. In reference to Section 2.2 and the model $Y_0 = X_0B + E_0$, those authors reexamined Example 1 of Cook and Weisberg (1999, pp.351, 433, 447) giving data on 82 mussels sampled off the coast of New Zealand. Here $Y_1 = \log($shell mass$)$ and $Y_2 = \log($muscle mass$)$. The predictors are the length, $\log($width$)$ and height of each shell. Case 79 is seen by those authors to be prominent for $Y_1$, as are Cases $\{8, 25, 48\}$ for $Y_2$. These cases are seen to exceed the conventional if flawed threshold for Cook’s (1977) marginal distance functions.

Instead of the separate residual plots of Olive et al. (2015) marginally, in Figure 2 we plot the observed residual pairs as points in $\mathbb{R}^2$, namely $\{(e_i, e_i); 1 \leq i \leq n\}$, as anticipated in Figure 1. Superimposed is the 90% ellipse. It is germane to ask whether cases $\{8, 25, 48, 79\}$, as identified marginally, are genuine outliers with respect to the constellation of data points in $\mathbb{R}^2$. These cases are identified by number in Figure 2. It is seen that cases $\{8, 48, 79\}$, as well as cases $\{11, 37\}$ not flagged by those authors marginally, lie outside the 90% ellipse.

To further confirm or refute those marginal assessments, we compute the diagnostics

$$T_i^2 = \frac{e_i' S^{-1} e_i}{(1-h_{ii})}, \ 1 \leq i \leq 6$$

(4.1)

accounting for dependence in the six cases in question. Values for these diagnostics are reported in Table 2, together with their equivalent $F$ and $p$-values. In short, it is seen that cases $\{8, 48, 79\}$, as flagged marginally, are joint outliers, but case 25 apparently is not. On the other hand, cases $\{11, 37\}$, not flagged marginally, appear to be outlying in respect to the constellation of data points in $\mathbb{R}^2$.

Table 2: Tabulated $T^2$, the equivalent $F$-diagnostic, and corresponding $p$-values for eight cases from the Mussels data.

<table>
<thead>
<tr>
<th>Case</th>
<th>$T^2$</th>
<th>$F$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>24.72</td>
<td>12.21</td>
<td>0.0000240</td>
</tr>
<tr>
<td>11</td>
<td>7.76</td>
<td>3.83</td>
<td>0.0258</td>
</tr>
<tr>
<td>25</td>
<td>4.31</td>
<td>2.13</td>
<td>0.126</td>
</tr>
<tr>
<td>37</td>
<td>7.52</td>
<td>3.71</td>
<td>0.0288</td>
</tr>
<tr>
<td>48</td>
<td>33.40</td>
<td>16.49</td>
<td>0.0000104</td>
</tr>
<tr>
<td>79</td>
<td>15.30</td>
<td>7.56</td>
<td>0.000994</td>
</tr>
</tbody>
</table>

In summary, the issue is not whether coordinates are outlying, as if they stand alone, but whether vector observations $Y_i$ themselves are outlying with respect to the constellation of data points in $\mathbb{R}^k$. In essence, we find this portion of Olive et al. (2015)
to be misdirected and thus a disservice to users faced with outlying vector observations. Moreover, it goes against accepted usage of $T^2$ as found in the literature, and it serves to thwart the spirit of multivariate statistics to take joint dependencies properly into account.

4.2 Simulation Studies

As developments heretofore are tedious, convoluted, and unconventional, it is instructive to demonstrate Theorem 2 and then Theorem 3. Details follow.

(i) Induced Correlations.

Accordingly, $N = 40,000$ random samples $Y_0 \in \mathbb{F}_{n \times k}$ of size $n = 10$ and $k = 2$ were generated from $N_{n \times k}(0, I_n \otimes \Sigma)$ with rows as independent bivariate Gaussian vectors having zero means and dispersion matrix $\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$. Correlations among rows were induced through $Y_0 \rightarrow Z_0 = [\Omega(\theta)]^{1/2}Y_0$ using the spectral square root, so that $V(Z_0) = \Omega(\theta) \otimes \Sigma$ using Remark A.1. In consequence, rows of $Z_0$ are equicorrelated with $\rho = \theta/(1 + \theta)$ and, to illustrate, we vary $\rho \in [0.0, 0.2, 0.5, 0.8]$ with corresponding $\theta \in [0.00, 0.25, 1.00, 4.00]$. MINITAB was used for all the simulations.

Table 3 reports the empirical critical values for $T^2_i$ corresponding to tabulated critical values $c_{\alpha}$ from Theorem 2(vi) such that $L(T^2_i) = T^2_k(u; r - 1, 0)$. The row being evaluated for a potential shift is set to $i = 10$. Computations yielding $T^2_i$ were undertaken for each repetition, with results as summarized in Table 3.

Table 3: Tabulated and empirical critical values for $\{T^2_i \geq c_{\alpha}\}$, $N = 40,000$ runs, for correlated $Z_0$ with varying $\rho$ such that $V(Z_0) = \Omega(\theta) \otimes \Sigma$ with $\theta = \rho/(1 - \rho)$, $n = 10$, and $k = 2$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tabulated $c_{\alpha}$</td>
<td>7.45</td>
<td>10.83</td>
<td>14.95</td>
<td>21.82</td>
</tr>
<tr>
<td>$\rho = 0.0$</td>
<td>7.40</td>
<td>10.77</td>
<td>14.65</td>
<td>21.84</td>
</tr>
<tr>
<td>$\rho = 0.2$</td>
<td>7.40</td>
<td>10.77</td>
<td>14.65</td>
<td>21.84</td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td>7.40</td>
<td>10.77</td>
<td>14.65</td>
<td>21.84</td>
</tr>
<tr>
<td>$\rho = 0.8$</td>
<td>7.40</td>
<td>10.77</td>
<td>15.65</td>
<td>21.84</td>
</tr>
</tbody>
</table>

Table 4 reports the empirical power for $T^2_i$ with noncentrality parameter $\lambda_i = \frac{\delta^T_i \Sigma^{-1} \delta_i}{2}$ as in Theorem 2(vi). The corresponding tabulated power as in Theorem 2(vi) is from $L(T^2_i) = T^2_k(u; r - 1, \lambda_i)$, where the shifts were added to the residual for row $i = 10$.

Table 4 demonstrates that the powers for Hotelling’s $T^2_i$ diagnostics under equicorrelated data are equivalent to those tabulated for independent data. Recalling that $L(T^2_i) = F(u; k, r - k, \lambda_i, 0)$, the noncentral $F$ probabilities were computed using the Keisan Online Calculator provided by the Casio Computer Co., Ltd.

(ii) Mixture Experiments.
To demonstrate the validity of $T_\alpha^2$ in mixture distributions as in Theorem 3, $N = 40,000$ random samples $Y_0 \in F_{n \times k}$ of size $n = 10$ and $k = 2$ were generated from $N_n\times k(\mathbf{0}, I_n \otimes \Sigma)$ having zero means and dispersion matrix $\Sigma = \begin{bmatrix} 0 & 0.8 \\ 0.8 & 0 \end{bmatrix}$. As before, correlations among the rows were induced through $Y_0 \rightarrow Z_i = [\Omega(\theta_i)]^\frac{1}{2} Y_0$ so that $V(Z_i) = \Omega(\theta_i) \otimes \Sigma$ using Remark A.1. This was repeated with $Y_0 \rightarrow Z_2 = [\Omega(\theta_2)]^\frac{1}{2} Y_0$ to form another correlated data set.

For a 50%–50% mixture, $n = 5$ observations were randomly chosen from each of $Z_1$ and $Z_2$ and stacked in random order to form a data set of order $(10 \times 2)$. Table 5 reports the empirical critical values for $T_\alpha^2$ from Theorem 2(v), with the corresponding tabulated critical values $c_\alpha$ from Theorem 2(vi) where $R(T_\alpha^2) = T_\alpha^2(u; r-1, 0)$. Values $\rho \in [0, 0.2, 0.5]$ were used with $\theta = \rho/(1-\rho)$. The row being evaluated for a potential shift was set to be $i = 10$.

Table 5 demonstrates empirically the invariance of $T_\alpha^2$ in mixture experiments. Observe that the second and third cases comprise contamination of $I_n \otimes \Sigma$, the classical model, with 50% contamination using $\Omega(\theta) \otimes \Sigma$.

### 4.3 Running Times Example.

Woodward (1970) studied the running times for $n = 22$ baseball players who ran three different paths rounding first base. These data, as used by Morrison (2005) to test for outliers, are reported in Appendix Table 9. The times that appear to be abnormal are those for Player 14 and Player 22. Using $T_\alpha^2$ we see in Table 6 that the running times for Player 22 are indeed outlying with $p$-value 0.0138. Times for Player 14 are not flagged as outliers. However, the last four rows of Table 6 give $T_\alpha^2$ for Player 14 assuming improvements in his running times in units of $\delta \in [0.1, 0.2, 0.3, 0.4]$.

Beckett (1977) has identified that the $n = 22$ data points consist 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. Observe that the rank order of the times $[Y_1, Y_2, Y_3]$ for Players 1, 6, and 18 differ from the rank order for other players. In consequence, Morrison’s (2005) search for outliers using $T_\alpha^2$ is in dispute when based on the validating model of the day.

### Table 4: Tabulated and empirical powers for $\{T_\alpha^2 \geq c_\alpha\}$. $N = 40,000$, with varying shifts $\delta' = [\delta_1, \delta_2]$ for correlated $Z_0$ such that $V(Z_0) = \Omega(\theta) \otimes \Sigma$ with $\theta = 1$, $n = 10$, and $\lambda_i = \frac{\xi_0}{n}\delta_i [\Sigma^{-1}]_{ii}$.

<table>
<thead>
<tr>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$k$</th>
<th>$r-k$</th>
<th>$\lambda_i$</th>
<th>Tabulated Power</th>
<th>Empirical Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.0496</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>1.000</td>
<td>0.1020</td>
<td>0.1036</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>4.000</td>
<td>0.2848</td>
<td>0.2822</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>9.000</td>
<td>0.5670</td>
<td>0.5647</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>16.000</td>
<td>0.8188</td>
<td>0.8177</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>25.000</td>
<td>0.9504</td>
<td>0.9499</td>
</tr>
</tbody>
</table>
namely $V(Y) = I_n \otimes \Sigma$, instead of the apparent mixing over clusters.

A fortunate conclusion from Theorems 2 and 3 is that searches for outliers using Hotelling’s $T^2$ now has been validated for data sets from cohorts which are equicorrelated. And, in addition, even for data sets arising as mixture distributions, in a belated verification of Morrison’s (2005) analysis.

4.4 Adjusting to a Common Standard.

Following the notation of Section 4.1, we consider a Case II example in which the data $Y_0 = [Y_1, \ldots, Y_n]'$ are adjusted against a standard $w = [w_1, \ldots, w_k]$. Here $N = 40,000$ random samples $Y_0 \in F_{n,k}$ were generated from $N_{n,k}(0, I_n \otimes \Sigma(\rho))$ of size $n = 20$ and $k = 4$ with rows as independent Gaussian vectors having zero means and dispersion matrix $\Sigma(\rho) = (1 - \rho)I_k + \rho I_k I_k'$ with $\rho = 0.8$. The standardizing vector $w$ was random from $N_k(0, \Sigma)$, with $\{Z_j = Y_j - w; 1 \leq j \leq n\}$ and with $Z_0 = AY_0^\prime$ as equicorrelated data with $V(Z_0) = \Omega(\theta) \otimes \Sigma$, $\theta = 1.0$ and corresponding $\rho = \theta/(1 + \theta) = 0.5$.

Lemma 4.1(iii) notes that $T^2 = (r - 1)\text{tr}(Q_1 Q_2^{-1})$ remains exact in level and power for the equicorrelated data. Table 7 reports the tabulated and empirical critical values for this study, affirming that the critical values remain the same for data equicorrelated by adjustment to a common standard.

Table 8 reports the tabulated and empirical powers for $\{T^2 \geq c_\alpha\}$ with $c_{0.05} = 14.67$ and with non-centrality parameter $\lambda_i = \frac{r}{k} \delta_i \Sigma^{-1} \delta_i$. The shifts $\delta_i = [\delta_1, \ldots, \delta_k]$ used a common value $\delta_i$ and were added to the residuals for row $i = 20$.

5 Conclusions

Given Gaussian vectors $Y_0 = [Y_1, \ldots, Y_n]'$ having a common mean $\mu$ and dispersion matrix $\Sigma$, normal-theory diagnostics for outlying vector observations are reexamined using the $R$–Hotelling statistics $\{T^2_i; 1 \leq i \leq n\}$ when successive vector observations are correlated. The latter is attained on replacing the classical structure $V(Y_0) = I_n \otimes \Sigma$ with $V(Y_0) = \Omega \otimes \Sigma$ such that $\Omega \in \{\Omega(\theta), \Omega(\rho), \Omega(\xi)\}$ as given in Section 2.4, and with mixtures over these.

Table 5: Tabulated and empirical critical values for $\{T^2_i \geq c_\alpha\}, N = 40,000$, with varying $\rho$, for 50%-50% mixtures with correlated $Z$ having $V(Z) = \Omega(\theta) \otimes \Sigma$, and with $Z$ having $V(Z) = \Omega(\theta_2) \otimes \Sigma$, where $\theta = \rho/(1 - \rho)$, $m = 5$, $k = 2$, and $n = 10$.

<table>
<thead>
<tr>
<th>$\Omega(\theta_1)$</th>
<th>$\Omega(\theta_2)$</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1=0.0$</td>
<td>$\rho_2=0.0$</td>
<td>7.40</td>
<td>10.77</td>
<td>14.65</td>
<td>21.84</td>
</tr>
<tr>
<td>$\rho_1=0.0$</td>
<td>$\rho_2=0.2$</td>
<td>7.50</td>
<td>10.83</td>
<td>15.07</td>
<td>21.90</td>
</tr>
<tr>
<td>$\rho_1=0.0$</td>
<td>$\rho_2=0.5$</td>
<td>7.31</td>
<td>10.51</td>
<td>14.55</td>
<td>21.41</td>
</tr>
<tr>
<td>$\rho_1=0.2$</td>
<td>$\rho_2=0.5$</td>
<td>7.21</td>
<td>10.29</td>
<td>14.27</td>
<td>20.90</td>
</tr>
<tr>
<td>Tabulated $c_\alpha$</td>
<td></td>
<td>7.45</td>
<td>10.83</td>
<td>14.95</td>
<td>21.82</td>
</tr>
</tbody>
</table>
Table 6: Values of $T_2^i$ for Player 22 and Player 14 with varying improvements $\delta' = [\delta_1, \delta_2, \delta_3]$ for Player 14.

<table>
<thead>
<tr>
<th>Player</th>
<th>$\delta = [\delta_1, \delta_2, \delta_3]'$</th>
<th>$T_2^i$</th>
<th>$n$</th>
<th>$k$</th>
<th>$r - 1$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>15.589</td>
<td>22</td>
<td>3</td>
<td>20</td>
<td>0.0138</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4.886</td>
<td>22</td>
<td>3</td>
<td>20</td>
<td>0.2572</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>(-0.1, -0.1, -0.1)</td>
<td>6.835</td>
<td>22</td>
<td>3</td>
<td>0.1432</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>(-0.2, -0.2, -0.2)</td>
<td>9.094</td>
<td>22</td>
<td>3</td>
<td>0.0744</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>(-0.3, -0.3, -0.3)</td>
<td>11.694</td>
<td>22</td>
<td>3</td>
<td>0.0367</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>(-0.4, -0.4, -0.4)</td>
<td>14.624</td>
<td>22</td>
<td>3</td>
<td>0.0175</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Tabulated and empirical critical values for $\{T_2^i \geq c_{\alpha}\}$, $N = 40,000$ runs, for correlated $\{Z_j = Y_j - \bar{W}\}$ as $Y_j$ adjusted to the standard $\bar{W}$, with $V(Z_0) = \Omega(\theta) \otimes \Sigma$ and $\theta = 1.0$, $\rho = 0.5$, $n = 20$ and $k = 4$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Tabulated $c_{\alpha}$</th>
<th>Empirical $c_{\alpha, \rho = 0.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>11.33</td>
<td>11.32</td>
</tr>
<tr>
<td>0.05</td>
<td>14.67</td>
<td>14.70</td>
</tr>
<tr>
<td>0.025</td>
<td>18.26</td>
<td>18.24</td>
</tr>
<tr>
<td>0.01</td>
<td>23.49</td>
<td>23.38</td>
</tr>
</tbody>
</table>

In particular, the invariance of distributions of $T_2^i$ for $V(Y_0) = \Omega \otimes \Sigma$ is established in Theorem 2 for $\Omega \in \{\Omega(\theta), \Omega(\rho), \Omega(\xi)\}$, and in Theorem 3 for the cited mixtures. Theorems 2 and 3 are multivariate extensions of the univariate results of Jensen and Ramirez (2015), made tractable through a fundamental characterization of Wishart from $\chi^2$ distributions as in Appendix A.1.

Case studies, to include simulations, are given to demonstrate Theorems 2 and 3. In addition, we have reexamined the Running Times for Baseball Players of Woodward (1975), containing observations from two clusters of players. With mild assumptions, Theorem 3 allows for the use of $T_2^i$ to detect outliers in this case. In other venues in practice, researchers often standardize data using, for example, the transformation $\{Z_j = Y_j + \bar{W} \in \mathbb{R}^k; 1 \leq j \leq n\}$, inducing correlations $V(Z_0) = \Omega(\theta) \otimes \Sigma$ with $\theta = 1$. Lemma 4.1 gives criteria assuring that the critical values for independent Gaussian data can also be used in such circumstances.

In short, Hotelling’s $T_2^i$ diagnostics continue to remain exact in level and power for classes of distributions and their mixtures, updated substantially beyond the classical theory in keeping with needs of some contemporary experiments. Thus diagnostics using $T_2^i$ are genuinely nonparametric in being distribution–free over nonstandard classes of distributions and their mixtures.
Marginals:
(ii) Wishart Distributions

We collect basics for matrix distributions essential to the present study. First partition $\mathbf{Y}_0$ by columns as $\mathbf{Y}_0 = [\mathbf{Y}_1, \ldots, \mathbf{Y}_k] \in \mathbb{L}^n$.

<table>
<thead>
<tr>
<th>$\delta_i$</th>
<th>$k$</th>
<th>$r - k$</th>
<th>$\lambda_i$</th>
<th>Tabulated Power</th>
<th>Empirical Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>15</td>
<td>0.000</td>
<td>0.5000</td>
<td>0.0503</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>15</td>
<td>1.118</td>
<td>0.0956</td>
<td>0.0933</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>15</td>
<td>4.471</td>
<td>0.2683</td>
<td>0.2681</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>15</td>
<td>10.059</td>
<td>0.5673</td>
<td>0.5694</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>15</td>
<td>17.882</td>
<td>0.8391</td>
<td>0.8388</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>15</td>
<td>27.941</td>
<td>0.9661</td>
<td>0.9661</td>
</tr>
</tbody>
</table>

A Appendix
A.1 Matrix Distributions

We collect basics for matrix distributions essential to the present study. First partition $\mathbf{Y}_0$ by columns as $\mathbf{Y}_0 = [\mathbf{Y}_1, \ldots, \mathbf{Y}_k] \in \mathbb{L}^n$. Alternatively, if instead we row–partition as $\mathbf{Y}_0 = [\mathbf{Z}_1, \ldots, \mathbf{Z}_m]$, then we adopt the convention that $\mathbb{V}(\mathbf{Y}_0) = \mathbf{\Omega} \otimes \mathbf{S}$, so that $\{\mathbb{V}(\mathbf{Z}_i) = \omega_i \mathbf{S}, 1 \leq i \leq n\}$. Moreover, if $\mathbb{V}(\mathbf{Y}) = \mathbf{\Omega} \otimes \mathbf{S}$ of order $(nk \times nk)$, then for fixed $(\mathbf{A}, \mathbf{B})$ and for $\mathbf{U} = \mathbf{A} \mathbf{Y} \mathbf{B}'$, the corresponding moment arrays are as follow.

\[ \mathbb{E}(\mathbf{Z}) = \mathbf{A} \mathbf{B}' \] and $\mathbb{V}(\mathbf{Z}) = \mathbf{A} \mathbf{\Omega} \mathbf{A}' \otimes \mathbf{B} \mathbf{S} \mathbf{B}'$.

(i) Gaussian Distributions

Designate the distribution of $\mathbf{Y} \in \mathbb{L}^n$ as $\mathcal{L}(\mathbf{Y}) = N_{n \times k}(\mathbf{M}, \mathbf{\Omega} \otimes \mathbf{S})$, the matrix Gaussian distribution in $\mathbb{L}^n$ having $\mathbb{E}(\mathbf{Y}) = \mathbf{M}$ and $\mathbb{V}(\mathbf{Y}) = \mathbf{\Omega} \otimes \mathbf{S}$. Its pdf is

\[ f_{nk}(\mathbf{Y}) = (2\pi)^{-\frac{n^2 k}{2}} |\mathbf{\Omega}|^{-\frac{1}{2}} |\mathbf{S}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \text{tr} \mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{M})\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{M})' \right]. \] (A.1)

Next partition $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2], \mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2]$, and $\mathbf{\Sigma} = [\mathbf{\Sigma}_{ij}; i, j = 1, 2]$. Here $(\mathbf{Y}_1, \mathbf{M}_1)$ are of order $(n \times r)$; $(\mathbf{Y}_2, \mathbf{M}_2)$ are of order $(s \times s)$; and elements of $\mathbf{\Sigma}$ are $\{\mathbf{\Sigma}_{11}(r \times r), \mathbf{\Sigma}_{12}(r \times s), \mathbf{\Sigma}_{21}(s \times r), \mathbf{\Sigma}_{22}(s \times s)\}$. Then

- Marginals: $\mathcal{L}(\mathbf{Y}_1) = N_{n \times r}(\mathbf{M}_1, \mathbf{\Omega} \otimes \mathbf{\Sigma}_{11}); \mathcal{L}(\mathbf{Y}_2) = N_{n \times s}(\mathbf{M}_2, \mathbf{\Omega} \otimes \mathbf{\Sigma}_{22})$.

- Conditional: $\mathcal{L}(\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2) = N_{n \times r}(\mathbf{M}_{1,2}, \mathbf{\Omega} \otimes \mathbf{\Sigma}_{11,2})$ with $\mathbf{M}_{1,2} = \mathbf{M}_1 + (\mathbf{y}_2 - \mathbf{M}_2)\mathbf{R}'$ and $\mathbf{R} = \mathbf{\Sigma}_{12}(s \times r)\mathbf{\Sigma}_{22}^{-1}$.

(ii) Wishart Distributions

Take $\mathcal{L}(\mathbf{Y}) = N_{nk}(\mathbf{M}, \mathbf{I}_n \otimes \mathbf{S});$ let $\mathbf{W} = \mathbf{Y} \mathbf{Y}'$; and partition $\mathbf{W} = [\mathbf{W}_{ij}; i, j = 1, 2]$ and $\mathbf{S} = [\mathbf{\Sigma}_{ij}; i, j = 1, 2]$, conformably. Here elements of $\mathbf{W}$ are $[(\mathbf{W}_{11} \times r \times s), \mathbf{W}_{21}(s \times r), \mathbf{W}_{22}(s \times s)]$. Define $\mathbf{W}_{11} = \mathbf{W}_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}$, and $\mathbf{\Sigma}_{11,2}$ as before.
- $W$ is said to have the Wishart distribution $\mathcal{L}(W) = \mathcal{W}_k(n; \Sigma, \Theta)$ with $n$ degrees of freedom, the scale parameters $\Sigma$, and the noncentrality matrix $\Theta = MM^T$.
- Marginals: $\mathcal{L}(W_{11}) = \mathcal{W}_r(n; \Sigma_{11}, M_1^T M_1)$ and $\mathcal{L}(W_{22}) = \mathcal{W}_s(n; \Sigma_{22}, M_2^T M_2)$.
- Conditional: $\mathcal{L}(W_{11}|Y_2 = y_2) = \mathcal{W}_r(n - s; \Sigma_{11|2}, \Theta(y_2))$ with $\Theta(y_2) = M_1^T M_{12}$ and $M_{12} = M_1 + R(y_2 - M_2)$ with $R = \sigma_{12} \Sigma_{22}^{-1}$.
- A standard result is that if $A$ is idempotent of rank $\nu$, and if $W = Y'AY$, then $\mathcal{L}(W) = \mathcal{W}_k(\nu, \Sigma, \Theta)$ with $\Theta = MM^T$.

(iii) A characterization

There are fundamental connections between noncentral chi–squared and noncentral Wishart distributions. To wit: The noncentral Chi–squared and Wishart $\chi^2$'s are

$$
\phi_Z(t) = (1-2it\sigma^2)^{-\frac{n}{2}} \exp[it\theta^2/(1-2it\sigma^2)]
$$

$$
\phi_W(T) = |I_k - 2iT\Sigma|^{-\frac{n}{2}} \exp[i\text{tr}(T\Theta(I_k - 2iT\Sigma)^{-1})],
$$

respectively. The following is germane.

**Theorem 5.** Let $\mathcal{L}(Y) = N_{nk}(M, I_n \otimes \Sigma)$, and take $W = Y'AY$ such that $A$ is idempotent of rank $\nu$.

(i) If $\mathcal{L}(W) = \mathcal{W}_k(\nu, \Sigma, \Theta)$, then $\mathcal{L}(uWu) = \chi^2(\nu, \sigma^2, \lambda(u))$ for every $u \in \mathbb{R}^k$, where $\sigma^2 = u^T \Sigma u$ and $\lambda(u) = u^T \Theta u$.

(ii) Conversely, if $\mathcal{L}(uWu) = \chi^2(\nu, \sigma^2, \lambda(u))$ for every $u \in \mathbb{R}^k$, then $\mathcal{L}(W) = \mathcal{W}_k(\nu, \Sigma, \Theta)$.

(iii) Define $W_1 = Y'A_1Y$ and $W_2 = Y'A_2Y$ with $A_1 \neq A_2$. Then $(W_1, W_2)$ are mutually independent Wishart matrices if and only $(uW_1u, uW_2u)$ are mutually independent $\chi^2$ variates for every $u \in \mathbb{R}^k$.

**Proof.** Conclusion (i) follows on substituting $t uu'$ for $T$ in the $\chi^2$ for $W$, together with the fact that the nonvanishing eigenvalue of $uu'$ is $u^T u$. Conclusion (ii) follows on lifting from $\mathbb{R}^+_k$ to $\mathbb{S}^+_k$; this may be done using the characterization of Cramér and Wold (1936), as carried out in Jensen (1982). Conclusion (iii) follows from (i) and (ii) on verifying that the joint $\chi^2$'s of $(W_1, W_2)$ and of $(uW_1u, uW_2u)$ factor into the product of their marginal $\chi^2$'s. □

**Remark A.2.** The central version of conclusions (i) and (ii) was given in Result (ii) of Rao (1973; p.535).

### A.2 Regression Diagnostics

Single–case univariate deletion diagnostics in continuing vogue include

\{t_i, DFFIT, DFBETA, COVRATIO, FVARATIO, OUT_i, AP_i, C_i, WK_i, W_i, D_i\},
the second through fifth from Belsley et al. (1980); $\text{OUT}_i$, $C_i$ as $d_i$, and $WK_i$ as in Section 2.3; $AP_i$ from Andrews and Pregibon (1978); $W_i$ of Welsch (1982); and the diagnostic $D_i$ of Jensen (2001a). For detailed properties see Table 1 of Jensen (2000).

These diagnostics are known to be excessive, their conventional usage as inconsistent and contradictory, “based on ad hoc reasoning” accounting “for the diversity of recommendations;” Cook (1986). See also Chatterjee and Hadi (1986). In addition, “advocates often disclaim that diagnostics and their benchmarks are not to test hypotheses, but rather to ‘nominate’ points for ‘special attention’;” LaMotte (1999). A resolution rests on the one–to–one correspondence of the listed diagnostics with $t_i$ or $t_i^2$ and, accordingly, that the critical $c_{\alpha}$ value for $t_i^2$ serves to translate into consistent external benchmarks for all. See LaMotte (1999), Jensen (2000) and Naik (2003). The relevance here is that these properties carry forward to illuminate the Table 1 diagnostics for vector deletions, as in Section 2.3 and Theorem 3(iii).

A.3 Running Times Data

The data employed in Section 4.3 are listed here as reported in Morrison (2005; p.102).
Table 9: Running times around first base for $k = 3$ paths and $n = 22$ players.

<table>
<thead>
<tr>
<th>Player</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>Player</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.40</td>
<td>5.50</td>
<td>5.55</td>
<td>12</td>
<td>5.65</td>
<td>5.55</td>
<td>5.45</td>
</tr>
<tr>
<td>2</td>
<td>5.85</td>
<td>5.70</td>
<td>5.75</td>
<td>13</td>
<td>5.60</td>
<td>5.35</td>
<td>5.45</td>
</tr>
<tr>
<td>3</td>
<td>5.20</td>
<td>5.60</td>
<td>5.50</td>
<td>14</td>
<td>5.05</td>
<td>5.00</td>
<td>4.95</td>
</tr>
<tr>
<td>4</td>
<td>5.55</td>
<td>5.50</td>
<td>5.40</td>
<td>15</td>
<td>5.50</td>
<td>5.50</td>
<td>5.40</td>
</tr>
<tr>
<td>5</td>
<td>5.90</td>
<td>5.85</td>
<td>5.70</td>
<td>16</td>
<td>5.45</td>
<td>5.55</td>
<td>5.50</td>
</tr>
<tr>
<td>6</td>
<td>5.45</td>
<td>5.55</td>
<td>5.60</td>
<td>17</td>
<td>5.55</td>
<td>5.55</td>
<td>5.35</td>
</tr>
<tr>
<td>7</td>
<td>5.40</td>
<td>5.40</td>
<td>5.35</td>
<td>18</td>
<td>5.45</td>
<td>5.50</td>
<td>5.55</td>
</tr>
<tr>
<td>8</td>
<td>5.45</td>
<td>5.50</td>
<td>5.35</td>
<td>19</td>
<td>5.50</td>
<td>5.45</td>
<td>5.25</td>
</tr>
<tr>
<td>9</td>
<td>5.25</td>
<td>5.05</td>
<td>5.00</td>
<td>20</td>
<td>5.65</td>
<td>5.60</td>
<td>5.40</td>
</tr>
<tr>
<td>10</td>
<td>5.85</td>
<td>5.80</td>
<td>5.70</td>
<td>21</td>
<td>5.70</td>
<td>5.65</td>
<td>5.55</td>
</tr>
<tr>
<td>11</td>
<td>5.25</td>
<td>5.20</td>
<td>5.10</td>
<td>22</td>
<td>6.30</td>
<td>6.30</td>
<td>6.25</td>
</tr>
</tbody>
</table>

Acknowledgement.

The authors encounter on occasion the skeptic’s view that dependencies need not be accounted for in view of available marginal procedures. The present authors disagree vigorously, to include the use of $T^2_i$, and we have supplied Sections 1.2 and 4.1 in support of our view.

Bibliography.


