Journal of Multivariate Analysis 136 (2015) 95–107

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Shift outliers in linear inference

D.R. Jensen [a], D.E. Ramirez [b,*]

[a] Department of Statistics, Virginia Tech, Blacksburg, VA 24061, United States

[b] University of Virginia, P. O. Box 400137, Charlottesville, VA 22904, United States

ARTICLE INFO

Article history:
Received 18 December 2013
Available online 22 January 2015

AMS subject classifications:
62J05
62J20

Keywords:
Vector outliers
Propagation of shifts
Anomalies in inference
Regression diagnostics

ABSTRACT

Shifts in responses typically are obscured from users, so that regression proceeds as if unshifted. At issue is the infusion of such shifts into classical analysis. On projecting outliers into the "Regressor" and "Error" spaces of a model, findings here are that shifts in responses may account for shifts in the *OLS* solutions, or for inflated residuals, or both. These in turn impact estimation, prediction, and hypothesis tests, all of vital interest to users, and all considered here. Tools for identifying shifts are given. Case studies illustrate effects of shifts on regression, to include a reexamination of studies from the literature.

1. Introduction

Classical linear inference begins with $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$ of full rank having n observations, p regressors, and uncorrelated errors with variance σ^2, giving $(\widehat{\boldsymbol{\beta}}, S^2)$ as Gauss–Markov (*OLS*) solutions and the Residual Mean Square (*RMS*). Such models long have been staples of theoretical and applied statistics; they serve as templates beyond linearity and *OLS*; and they carry a large body of supporting diagnostics in regard to model validation. Basic arrays include $\boldsymbol{H}_n = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$; its diagonal elements $\{h_{ii} \in (0,\ 1);\ 1 \leq i \leq n\}$ are *leverages* attributed to rows $\{\boldsymbol{x}_i';\ 1 \leq i \leq n\}$ of $\boldsymbol{X}$; and elements of $(\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{Y} = \boldsymbol{e}$ are observed residuals. In addition, conventional regression diagnostics seek to identify *outlying* data, and to label as *influential* those observations whose removal would alter essentials of the analysis. Traditional diagnostic procedures and references are surveyed in Appendix A.2; we return to these subsequently.

In addition to conventional diagnostics, the present study seeks to track effects on the regression exerted by a vector shift $\{\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{\omega}\}$ in the elements of $\boldsymbol{Y}$. Here $\boldsymbol{\omega} \in \mathbb{R}^n$ is a shift parameter taking fixed but unknown values in particular applications, as in any non-Bayesian setting. Specifically, $\boldsymbol{\omega}$ is decomposed into a "Regressor" component $\boldsymbol{\omega}_1$ and an "Error" component $\boldsymbol{\omega}_2$, accounting respectively for shifts in the *OLS* solutions and for inflated variation about the best-fitting line. Such shifts typically are obscured from the user, who then proceeds as for unshifted data, yet retains vital interest in whether such shifts may have occurred. Accordingly, critical effects on conventional inferences regarding $(\boldsymbol{\beta}, \sigma^2)$, as induced by shifts in $\boldsymbol{Y}$, are examined; and venues for securing evidence regarding $\boldsymbol{\omega}$ are given, to include estimation and hypothesis tests for $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$. This study fills a long-standing void in regression diagnostics, where the conventional *DFBETA*'s and *DIFFIT*s examine a succession of singleton shifts, together with effects of each on the estimated $\boldsymbol{\beta}$'s and predictors. Circumstances for the present approach are found in the sciences and engineering, where the recalibration of a calibrated device is often required, and in statistical process control. Details of the study are outlined next.

* Corresponding author.
E-mail addresses: djensen@vt.edu (D.R. Jensen), der@virginia.edu (D.E. Ramirez).

http://dx.doi.org/10.1016/j.jmva.2015.01.012

Supporting developments are given in Section 2, to include notes on g-inverses and the projection of shifts into the "Regressor" and "Error" spaces of a model, Sections 3 and 4 comprise the principal findings. Section 3 establishes effects of shifts on the outcomes of regression analysis, to include anomalies in estimation and tests for $(\boldsymbol{\beta}, \sigma^2)$. Estimating $\boldsymbol{\omega}$ is heretofore unavailable, requiring that $n+p+1$ parameters should be supported by n observation vectors. Nonetheless, Section 4 undertakes inferences regarding the unknown shift $\boldsymbol{\omega}$ and its components $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ on utilizing additional observations. Section 5 reports case studies, first for an elementary and transparent example, proceeding then to a reexamination of comprehensive data from the literature. Appendix A.1 derives the distribution of a ratio of correlated chi-squared variables required for tests regarding $\boldsymbol{\omega}_2$. Connections to other venues are noted, to include deletion diagnostics and robust regression. Outliers under deletions are special cases of those considered here, and the two approaches are revisited in Appendix A.2.

2. Preliminaries

2.1. Notation

Spaces here include $\mathbb{R}^n$ as Euclidean n-space, its positive orthant $\mathbb{R}^n_+$, and the real symmetric matrices $\mathbb{S}_n$ of order 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. Here $\boldsymbol{1}_n = [1, 1, \ldots, 1]' \in \mathbb{R}^n$ is the unit vector, and $\boldsymbol{0}$ a vector of zeros of dimension to be determined in context. If $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k]$ is of order $(n \times k)$ and rank $k < n$, then $S_p(\boldsymbol{B})$ designates the column span of $\boldsymbol{B}$, *i.e.*, the k-dimensional subspace of $\mathbb{R}^n$ spanned by $[\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k]$. The ordered eigenvalues of $\boldsymbol{A} \in \mathbb{S}_n$ are $\{\lambda_i(\boldsymbol{A}) = \alpha_i;\ 1 \le i \le n\}$ with $\{\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n\}$, and its spectral resolution is $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}_\alpha\boldsymbol{P}' = \sum_{i=1}^n \alpha_i \boldsymbol{p}_i\boldsymbol{p}_i'$, where $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n]$ is orthogonal and $\boldsymbol{D}_\alpha = \text{Diag}(\alpha_1, \ldots, \alpha_n)$. The range and null spaces of $\boldsymbol{A}$ are designated as $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{N}(\boldsymbol{A})$. Specifically, if $\{\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_r > \alpha_{r+1} = \cdots = \alpha_n = 0\}$ and if $\boldsymbol{P} = [\boldsymbol{P}_1, \boldsymbol{P}_2]$ with $\boldsymbol{P}_1 = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r]$ and $\boldsymbol{P}_2 = [\boldsymbol{p}_{r+1}, \ldots, \boldsymbol{p}_n]$, then $\mathcal{R}(\boldsymbol{A}) = S_p(\boldsymbol{P}_1)$ and $\mathcal{N}(\boldsymbol{A}) = S_p(\boldsymbol{P}_2)$.

Generalized inverses. Given $\boldsymbol{A}$ of order $(n \times m)$, g-inverses are pivotal in solving linear systems $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Consider $\boldsymbol{G}(m \times n)$ together with the properties: $A_1 : \boldsymbol{AGA} = \boldsymbol{A}$; $A_2 : \boldsymbol{GAG} = \boldsymbol{G}$; $A_3 : (\boldsymbol{AG})' = \boldsymbol{AG}$; $A_4 : (\boldsymbol{GA})' = \boldsymbol{GA}$. Any g-inverse $\boldsymbol{G}$ of $\boldsymbol{A}$ satisfies A_1; a *reflexive* g-inverse satisfies A_1 and A_2; and a *Moore–Penrose* inverse satisfies A_1–A_4, to be denoted by $\boldsymbol{A}^\dagger$. For reference see [23]. Some g-inverses of interest here are treated in the following.

Lemma 2.1. *Consider the model $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$ of order n and full rank $p < n$, and let $\boldsymbol{H}$ be $(n \times n)$ symmetric idempotent of rank $k < n$. Then*

(i) $\boldsymbol{X}^\dagger = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ *is the Moore–Penrose inverse of* $\boldsymbol{X}$*;*
(ii) $\boldsymbol{H}^\dagger = \boldsymbol{H}$ *is the Moore–Penrose inverse of* $\boldsymbol{H}$*;*
(iii) $\boldsymbol{H}_n = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{X}^\dagger$ *with self inverse* $(\boldsymbol{X}\boldsymbol{X}^\dagger)^\dagger = \boldsymbol{X}\boldsymbol{X}^\dagger = \boldsymbol{H}_n$.

Proof. Conclusions follow directly from the properties A_1–A_4. □

Distributions of note. Given a random $\boldsymbol{Y} \in \mathbb{R}^n$, its distribution, characteristic function (*chf*), mean vector, and dispersion matrix are denoted by $\mathcal{L}(\boldsymbol{Y})$, $\phi_Y(\boldsymbol{t})$, $\text{E}(\boldsymbol{Y})$, $\text{V}(\boldsymbol{Y}) = \boldsymbol{\Xi}$, say, with variance $\text{Var}(Y) = \sigma^2$ on $\mathbb{R}^1$. Specifically, $\mathcal{L}(\boldsymbol{Y}) = N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is Gaussian on $\mathbb{R}^n$ with $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as its mean and dispersion matrix. Distributions on $\mathbb{R}^1_+$ include $\chi^2(\nu, \lambda)$ as chi-squared having ν degrees of freedom, noncentrality parameter λ, *chf* $\phi(t) = (1-2it)^{-\nu/2} \exp[i\lambda t/(1-2it)]$, and with mean $(\nu + \lambda)$ and variance $2(\nu + 2\lambda)$; see [16, pp. 132–133]. In addition $F(\nu_1, \nu_2, \lambda_1, \lambda_2)$ is the doubly noncentral F-distribution with (ν_1, λ_1) and (ν_2, λ_2) as degrees of freedom and noncentralities in its numerator and denominator, and $t^2(\nu, \lambda_1, \lambda_2) = F(1, \nu, \lambda_1, \lambda_2)$ is the doubly noncentral Student's t^2. Identify $\{F > c_\alpha\}$ as the conventional α-level rejection rule based on $F(\nu_1, \nu_2, 0, 0)$.

The model. Take $\{Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \varepsilon_i;\ 1 \le i \le n\}$ to model response Y_i to regressors $\{X_{i1}, \ldots, X_{ik}\}$ through $p = k+1$ parameters $\boldsymbol{\beta}' = [\beta_0, \beta_1, \ldots, \beta_k]$. Arrayed as $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$, the entities $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, $\boldsymbol{e} = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$, and $S^2 = \boldsymbol{e}'\boldsymbol{e}/(n-p)$ are the *OLS* solutions, the residual vector, and the *RMS*, respectively, where *OLS* solutions are displayed as $\widehat{\boldsymbol{\beta}} = [\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k]'$. In this setting $\boldsymbol{H}_n$ is now $\boldsymbol{H}_n = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Conventional Gauss–Markov assumptions on error moments, then distributions, are as follow.

Assumptions A. A_1. $\text{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0} \in \mathbb{R}^n$, $\text{V}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}_n$; and
A_2. $\mathcal{L}(\boldsymbol{\varepsilon}) = N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$.

Outliers often are modeled as additive shifts, typically at designated observations to be deleted in deletion diagnostics. To the contrary, this study allows unfettered shifts $\{\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{\omega}\}$ in the collective data, to include single-case and subset deletions as special cases.

2.2. Classification of shifts

A critical issue, largely unexamined in the literature, is the manner in which a given shift $\{\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{\omega}\}$ is infused into outcomes of conventional regression analyses. To these ends recall $\boldsymbol{H}_n$ and $(\boldsymbol{I}_n - \boldsymbol{H}_n)$ as idempotent $(n \times n)$ matrices of ranks p and $(n-p)$, projecting into the "Regressor" space $\mathcal{R}(\boldsymbol{H}_n)$ and "Error" space $\mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$ generated by $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$. Critical insight is gained on decomposing any $\boldsymbol{\omega} \in \mathbb{R}^n$ as in the following.

Definition 1. A shift $\boldsymbol{\omega} \in \mathbb{R}^n$ is decomposed as $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$, where $\boldsymbol{\omega}_1 = \boldsymbol{H}_n\boldsymbol{\omega}$ and $\boldsymbol{\omega}_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$ are respective projections into the "Regressor" and "Error" spaces of $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$, where $\boldsymbol{H}_n = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. In addition, let θ_1 and θ_2 be respective angles between $(\boldsymbol{\omega}, \boldsymbol{\omega}_1)$ and $(\boldsymbol{\omega}, \boldsymbol{\omega}_2)$.

In consequence, shifts decompose into components lying in $\mathcal{R}(\boldsymbol{H}_n)$ and $\mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$. These in turn exert profound and differing impacts on the principal outcomes of regression analyses as shown subsequently. Some implications follow immediately:

Lemma 2.2. *Given the projection $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ in $\mathbb{R}^n$, it follows that*

(i) *$\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{H}_n)$ implies $\boldsymbol{\omega}_2 = \boldsymbol{0}$, and $\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$ implies $\boldsymbol{\omega}_1 = \boldsymbol{0}$;*
(ii) *$\boldsymbol{X}'\boldsymbol{\omega} = \boldsymbol{X}'\boldsymbol{\omega}_1$, and $\boldsymbol{X}^{\dagger}\boldsymbol{\omega} = \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_1$;*
(iii) *$\boldsymbol{X}'\boldsymbol{\omega}_2 = \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_2 = \boldsymbol{0} \in \mathbb{R}^p$; moreover, $\boldsymbol{H}_n\boldsymbol{\omega}_2 = \boldsymbol{0} \in \mathbb{R}^n$ and $(\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}_1 = \boldsymbol{0} \in \mathbb{R}^n$.*

Proof. These follow directly from Definition 1.

3. Propagation of shifts

3.1. Basics

The decomposition $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ in $\mathbb{R}^n$ is basic; it enables their effects to be tracked separately; and these are found to differ markedly. To fix ideas, consider outcomes $\{\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{e}_{\omega}, S^2_{\boldsymbol{\omega}}\}$ from the model $\{\boldsymbol{Y}_{\boldsymbol{\omega}} = \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{\omega}} + \boldsymbol{\varepsilon}_{\omega}\}$, modified under a fixed but unknown shift parameter $\{\boldsymbol{Y}_{\varnothing} \to \boldsymbol{Y}_{\boldsymbol{\omega}} = \boldsymbol{Y}_{\varnothing} + \boldsymbol{\omega}\}$ taking place during the course of an experiment. These stand in contrast to the conventional output $\{\widehat{\boldsymbol{\beta}}_{\varnothing} = \boldsymbol{X}^{\dagger}\boldsymbol{Y}_{\varnothing},\ \boldsymbol{e}_{\varnothing} = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{Y}_{\varnothing},\ S^2_{\varnothing} = \boldsymbol{e}'_{\varnothing}\boldsymbol{e}_{\varnothing}/(n-p)\}$ under the intended model $\{\boldsymbol{Y}_{\varnothing} = \boldsymbol{X}\boldsymbol{\beta}_{\varnothing} + \boldsymbol{\varepsilon}_{\varnothing}\}$ had no shifts occurred. The subscript $(Symbol_{\varnothing})$ identifies the quantities sought by experiment, but typically not recoverable under the shifted model.

Remark 1. Specifically, the parameter space is the Cartesian product $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\omega}) \in \mathbb{R}^p \times \mathbb{R}^1_+ \times \mathbb{R}^n$, a misspecified model as noted by a Referee. Even for $\{\boldsymbol{Y} = \beta_0\boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$ with intercept, attempts to reconfigure the shift $\{\boldsymbol{Y}_{\boldsymbol{\omega}} = \boldsymbol{Y} + \boldsymbol{\omega}\}$ as $\{\boldsymbol{Y} = (\beta_0\boldsymbol{1}_n - \boldsymbol{\omega}) + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$ fail under conventional *OLS* in having $n + p + 1$ parameters, some outside the span of the regressors, to be supported by n observations.

Instead we follow a different approach in order to examine effects exerted by a given shift on regression outcomes. Details follow.

3.2. Effects on regression

At issue is the manner in which a given shift $\boldsymbol{\omega} \in \mathbb{R}^n$ is infused into properties of $\{\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{e}_{\boldsymbol{\omega}}, S^2_{\boldsymbol{\omega}}, F_{\boldsymbol{\omega}}\}$. It is seen for fixed $\boldsymbol{\omega}$ that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}$ has expectation $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) \stackrel{\text{def}}{=} \boldsymbol{\beta}_{\boldsymbol{\omega}} = \boldsymbol{\beta} + \boldsymbol{\kappa}$ as a shifted version of $\boldsymbol{\beta}$ with $\boldsymbol{\kappa} = \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_1$. Conversely, write $\{\boldsymbol{Y}_{\varnothing} + \boldsymbol{\omega} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\omega}_1 + (\boldsymbol{\varepsilon} + \boldsymbol{\omega}_2)\}$; fix $\boldsymbol{\kappa}^* \in \mathbb{R}^p$ and lift this to $\boldsymbol{\omega}^*_1 = \boldsymbol{X}\boldsymbol{\kappa}^* \in \mathbb{R}^n$, taken to be a component of $\boldsymbol{\omega}^*$. Then the model $\{\boldsymbol{Y}_{\varnothing} + \boldsymbol{\omega}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\omega}^*_1 + (\boldsymbol{\varepsilon} + \boldsymbol{\omega}^*_2)\}$ emerges as $\{\boldsymbol{Y}_{\boldsymbol{\omega}*} = \boldsymbol{X}\boldsymbol{\beta}_{\boldsymbol{\omega}*} + (\boldsymbol{\varepsilon} + \boldsymbol{\omega}^*_2)\}$, since $\boldsymbol{\omega}^*_1 = \boldsymbol{X}\boldsymbol{\kappa}^*$. In short, $\boldsymbol{\beta}_{\boldsymbol{\omega}} \in \mathbb{R}^p \iff \boldsymbol{\omega}_1 \in \mathbb{R}^n$. Here and elsewhere we take $\boldsymbol{M} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$. A first look under moment assumptions follows.

Theorem 1. *Consider $\{\boldsymbol{Y}_{\boldsymbol{\omega}} = \boldsymbol{Y}_{\varnothing} + \boldsymbol{\omega}\}$ with fixed $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$, yielding $\{\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{e}_{\boldsymbol{\omega}}, S^2_{\boldsymbol{\omega}}\}$; let $\boldsymbol{\kappa} = \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_1 \in \mathbb{R}^p$ and $\nu = n - p$; and define $\boldsymbol{\beta}_{\boldsymbol{\omega}} = \boldsymbol{\beta} + \boldsymbol{\kappa}$.*

(i) *The model elements $\boldsymbol{\beta}_{\boldsymbol{\omega}} \in \mathbb{R}^p \iff \boldsymbol{\omega}_1 \in \mathbb{R}^n$ are in correspondence.*
Under Assumptions A_1 we have for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}$
(ii) *$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} = \widehat{\boldsymbol{\beta}}_{\varnothing} + \boldsymbol{\kappa}$; $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) = \boldsymbol{\beta}_{\boldsymbol{\omega}} = \boldsymbol{\beta} + \boldsymbol{\kappa}$; and $\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) = \sigma^2\boldsymbol{M}$;*
(iii) *Specifically, if $\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$, then $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} \equiv \widehat{\boldsymbol{\beta}}_{\varnothing}$ is observable;*
(iv) *The MSE efficiency ratio $E_{ff}(\widehat{\boldsymbol{\beta}}_{\varnothing} : \widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) = 1 + (\boldsymbol{\kappa}'\boldsymbol{\kappa}/\sigma^2 \mathrm{tr}\,\boldsymbol{M})$ quantifies the loss in efficiency due to $\boldsymbol{\omega}$ in estimating $\boldsymbol{\beta}$.*
Similarly for the residuals $\boldsymbol{e}_{\boldsymbol{\omega}}$, $\boldsymbol{e}'_{\boldsymbol{\omega}}\boldsymbol{e}_{\boldsymbol{\omega}}$, and $S^2_{\boldsymbol{\omega}}$ we have
(v) *$\boldsymbol{e}_{\boldsymbol{\omega}} = \boldsymbol{e}_{\varnothing} + \boldsymbol{\omega}_2$; $\mathrm{E}(\boldsymbol{e}_{\boldsymbol{\omega}}) = \boldsymbol{\omega}_2$; $\mathrm{V}(\boldsymbol{e}_{\boldsymbol{\omega}}) = \sigma^2(\boldsymbol{I}_n - \boldsymbol{H}_n)$; and, if $\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{H}_n)$, then $\boldsymbol{e}_{\boldsymbol{\omega}} \equiv \boldsymbol{e}_{\varnothing}$ is observable;*
(vi) *$\mathrm{E}(\boldsymbol{e}'_{\boldsymbol{\omega}}\boldsymbol{e}_{\boldsymbol{\omega}}) = \nu\sigma^2 + \boldsymbol{\omega}'_2\boldsymbol{\omega}_2$;*
(vii) *$S^2_{\boldsymbol{\omega}} = \boldsymbol{e}'_{\boldsymbol{\omega}}\boldsymbol{e}_{\boldsymbol{\omega}}/\nu$ and, if $\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{H}_n)$, then $S^2_{\boldsymbol{\omega}} \equiv S^2_{\varnothing}$ is observable;*
(viii) *$E(S^2_{\boldsymbol{\omega}}) = \sigma^2 + \lambda_2$ with $\lambda_2 = \boldsymbol{\omega}'_2\boldsymbol{\omega}_2/\nu$ and, if $\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{H}_n)$, then $E(S^2_{\boldsymbol{\omega}}) = \sigma^2$.*

Proof. Conclusion (i) was demonstrated in the paragraph preceding. For (ii) observe that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} = \boldsymbol{X}^{\dagger}\boldsymbol{Y}_{\boldsymbol{\omega}} = \boldsymbol{X}^{\dagger}(\boldsymbol{Y}_{\varnothing} + \boldsymbol{\omega}) = \widehat{\boldsymbol{\beta}}_{\varnothing} + \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_1$ using $\boldsymbol{X}^{\dagger}\boldsymbol{\omega} = \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_1$ from Lemma 2.2, and since $\boldsymbol{\omega}$ is fixed, if unknown, first and second moments follow as in conclusion (ii) under Assumptions A_1 since $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\varnothing}) = \boldsymbol{\beta}$ and $\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\varnothing}) = \sigma^2\boldsymbol{M}$. To continue, $\boldsymbol{\omega} \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$ in (iii) implies $\boldsymbol{\omega}_1 = \boldsymbol{0}$. In (iv) the loss in efficiency is the M_{SE} efficiency ratio $E_{ff}(\widehat{\boldsymbol{\beta}}_{\varnothing} : \widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) = M_{SE}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})/M_{SE}(\widehat{\boldsymbol{\beta}}_{\varnothing})$, where

$M_{SE}(\widehat{\boldsymbol\beta}_{\varnothing}) = \sigma^2 \mathrm{tr}\,\boldsymbol M$ and $M_{SE}(\widehat{\boldsymbol\beta}_{\boldsymbol\omega}) = \mathrm{tr}(\sigma^2\boldsymbol M + \boldsymbol\kappa\boldsymbol\kappa') = \sigma^2\mathrm{tr}\,\boldsymbol M + \boldsymbol\kappa'\boldsymbol\kappa$, to give conclusion (iv). In regard to residuals, observe that $\boldsymbol e_{\boldsymbol\omega} = (\boldsymbol I_n - \boldsymbol H_n)(\boldsymbol Y_{\varnothing} + \boldsymbol\omega) = \boldsymbol e_{\varnothing} + \boldsymbol\omega_2$ and $\boldsymbol\omega_2$ is fixed, so that $\mathrm{E}(\boldsymbol e_{\boldsymbol\omega}) = \boldsymbol 0 + \boldsymbol\omega_2$, $\mathrm{V}(\boldsymbol e_{\boldsymbol\omega}) = \mathrm{V}(\boldsymbol e_{\varnothing}) = \sigma^2(\boldsymbol I_n - \boldsymbol H_n)$; moreover, if $\boldsymbol\omega \in \mathcal{R}(\boldsymbol H_n)$, then $\boldsymbol\omega_2 = \boldsymbol 0$ and $\boldsymbol e_{\boldsymbol\omega} \equiv \boldsymbol e_{\varnothing}$ is observable, giving (v). Conclusion (vi) follows as the expectation of a noncentral quadratic form as in [21, p. 51]. Conclusion (vii) follows from (v) and (viii) from (vii). □

We draw the following conclusions:

C_1. The components $(\boldsymbol\omega_1,\ \boldsymbol\omega_2)$ induce exclusive shifts in $(\widehat{\boldsymbol\beta},\ \boldsymbol e)$, respectively.

C_2. The extremal case $\boldsymbol\omega = \boldsymbol\omega_1$ renders shifts in $\widehat{\boldsymbol\beta}$ that cannot be discerned through altered residuals. The other extremity, $\boldsymbol\omega = \boldsymbol\omega_2$, leaves $\widehat{\boldsymbol\beta}$ unscathed as from the intended model, while inflating variability about the intended best-fitting line.

C_3. Otherwise the angles (θ_1, θ_2) of Definition 1 quantify the extent to which $\boldsymbol\omega$ projects into $\mathcal{R}(\boldsymbol H_n)$ and $\mathcal{R}(\boldsymbol I_n - \boldsymbol H_n)$, respectively.

C_4. If $\boldsymbol\omega$ consists of a single outlier $\{Y_i+\delta_i\}$ at $\boldsymbol x_i'$, then $\mathrm{E}(\widehat{\boldsymbol\beta}_{\boldsymbol\omega}) = \boldsymbol\beta + \boldsymbol w_i\delta_i$ with $\boldsymbol w_i = (\boldsymbol X'\boldsymbol X)^{-1}\boldsymbol x_i$ from Theorem 1(ii). This exhibits the manner in which δ_i is distributed as bias across elements of $\widehat{\boldsymbol\beta}_{\boldsymbol\omega}$ as estimators for $\boldsymbol\beta$, as in Section 3.2 of Jensen [13] under single-case deletions.

Effects that shifts exert on fundamental distributions may be summarized as follows under the normality Assumption A_2. Here $F_{\boldsymbol\omega} = (\widehat{\boldsymbol\beta}_{\boldsymbol\omega} - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\widehat{\boldsymbol\beta}_{\boldsymbol\omega} - \boldsymbol\beta_o)/pS^2_{\boldsymbol\omega}$ is the conventional statistic, but applied under the shift $\{\boldsymbol Y_{\boldsymbol\omega} = \boldsymbol Y_{\varnothing} + \boldsymbol\omega\}$, for testing the intended $H_0^{\beta} : \boldsymbol\beta = \boldsymbol\beta_o$ against $H_1^{\beta} : \boldsymbol\beta \neq \boldsymbol\beta_o$.

Theorem 2. *Consider $\{\boldsymbol Y_{\boldsymbol\omega} = \boldsymbol Y_{\varnothing} + \boldsymbol\omega\}$ with fixed $\boldsymbol\omega = \boldsymbol\omega_1 + \boldsymbol\omega_2$, yielding $\{\widehat{\boldsymbol\beta}_{\boldsymbol\omega},\ \boldsymbol e_{\boldsymbol\omega}, S^2_{\boldsymbol\omega},\ F_{\boldsymbol\omega}\}$, and let $\nu = n - p$. Then under Assumption A we have*

(i) $\mathcal{L}(\widehat{\boldsymbol\beta}_{\boldsymbol\omega}) = N_p(\boldsymbol\beta + \boldsymbol\kappa,\ \sigma^2\boldsymbol M)$;
(ii) $\mathcal{L}(\boldsymbol e_{\boldsymbol\omega}) = N_n(\boldsymbol\omega_2, \sigma^2(\boldsymbol I_n - \boldsymbol H_n))$ *and* $\mathcal{L}(\boldsymbol e_{\boldsymbol\omega}'\boldsymbol e_{\boldsymbol\omega}/\sigma^2) = \chi^2(\nu, \lambda_2)$, *with* $\lambda_2 = \boldsymbol\omega_2'\boldsymbol\omega_2/\sigma^2$;
(iii) $\mathcal{L}(\nu S^2_{\boldsymbol\omega}/\sigma^2) = \chi^2(\nu, \lambda_2)$, *and if* $\boldsymbol\omega \in \mathcal{R}(\boldsymbol H_n)$, *then* $\mathcal{L}(\nu S^2_{\boldsymbol\omega}/\sigma^2) = \mathcal{L}(\nu S^2_{\varnothing}/\sigma^2) = \chi^2(\nu, 0)$.
(iv) *The M_{SE} efficiency ratio in estimating σ^2 is* $E_{ff}(S^2_{\varnothing} : S^2_{\boldsymbol\omega}) = 1 + [(\nu + \lambda_2)^2/2(\nu + 2\lambda_2)]$.
(v) $\mathcal{L}(F_{\boldsymbol\omega}) = F(p, \nu, \lambda_1, \lambda_2)$ *with* $\lambda_1 = (\boldsymbol\beta + \boldsymbol\kappa - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\boldsymbol\beta + \boldsymbol\kappa - \boldsymbol\beta_o)/\sigma^2$ *and* $\lambda_2 = \boldsymbol\omega_2'\boldsymbol\omega_2/\sigma^2$.

Proof. Conclusion (i) and the first part of conclusion (ii) are Gaussian versions of Theorem 1(ii), (v). To continue, for a random $\boldsymbol U \in \mathbb{R}^n$ having $\mathrm{E}(\boldsymbol U) = \boldsymbol\mu$, the noncentrality parameter for $\boldsymbol U'\boldsymbol A\boldsymbol U$ is the quadratic form $\boldsymbol\mu'\boldsymbol A\boldsymbol\mu$ in its expectation. Observe that $(\boldsymbol I_n - \boldsymbol H_n)\boldsymbol e_{\boldsymbol\omega} = (\boldsymbol I_n - \boldsymbol H_n)^2(\boldsymbol Y_{\varnothing} + \boldsymbol\omega) = \boldsymbol e_{\boldsymbol\omega}$ since $(\boldsymbol I_n - \boldsymbol H_n)$ is idempotent. Accordingly, $\boldsymbol e_{\boldsymbol\omega}'\boldsymbol e_{\boldsymbol\omega} = \boldsymbol e_{\boldsymbol\omega}'(\boldsymbol I_n - \boldsymbol H_n)\boldsymbol e_{\boldsymbol\omega}$ is a quadratic form of type $\boldsymbol U'\boldsymbol A\boldsymbol U$ with idempotent matrix $\boldsymbol A = (\boldsymbol I_n - \boldsymbol H_n)$ of rank $n - p$, and noncentrality $\boldsymbol\omega_2'(\boldsymbol I_n - \boldsymbol H_n)\boldsymbol\omega_2 = \boldsymbol\omega_2'\boldsymbol\omega_2$, to complete conclusion (ii). Conclusion (iii) follows from (ii). For the M_{SE} efficiency ratio it suffices to consider $E_{ff}(\boldsymbol e_{\boldsymbol\omega}'\boldsymbol e_{\boldsymbol\omega}/\sigma^2 : \boldsymbol e_o'\boldsymbol e_o/\sigma^2)$. From (iii) and moments of $\chi^2(\nu, \lambda_2)$ find $M_{SE}(\boldsymbol e_o'\boldsymbol e_o/\sigma^2) = \mathrm{Var}(\boldsymbol e_o'\boldsymbol e_o/\sigma^2) = 2(\nu + 2\lambda_2)$. Similarly, with $\mathrm{E}(\boldsymbol e_{\boldsymbol\omega}'\boldsymbol e_{\boldsymbol\omega}/\sigma^2) = (\nu + \lambda_2)$, we have $M_{SE}(\boldsymbol e_{\boldsymbol\omega}'\boldsymbol e_{\boldsymbol\omega}/\sigma^2) = 2(\nu + 2\lambda_2) + (\nu + \lambda_2)^2$. Combining gives conclusion (iv). Conclusion (v) follows along conventional lines since $(\widehat{\boldsymbol\beta}_{\boldsymbol\omega}, S^2_{\boldsymbol\omega})$ are mutually independent under Gaussian Assumptions, noting that the scale-invariance of the ratio $F_{\boldsymbol\omega}$ frees its distribution from dependence on σ^2. □

We next consider effects on conventional tests regarding $\boldsymbol\beta$ and σ^2, as exerted by shifts in the responses. Details follow.

3.3. Anomalies: tests regarding $(\boldsymbol\beta,\ \sigma^2)$

The test for $H_0^{\beta} : \boldsymbol\beta = \boldsymbol\beta_o$ against $H_1^{\beta} : \boldsymbol\beta \neq \boldsymbol\beta_o$ in unshifted data rejects for $\{F_{\varnothing} > c_\alpha\}$ with c_α from the upper tail of $F(p, n - p, 0, 0)$, where $F_{\varnothing}$ is the intended F-statistic. Under $\{\boldsymbol Y_{\boldsymbol\omega} = \boldsymbol Y_{\varnothing} + \boldsymbol\omega\}$, the observable statistic becomes $F_{\boldsymbol\omega} = (\widehat{\boldsymbol\beta}_{\boldsymbol\omega} - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\widehat{\boldsymbol\beta}_{\boldsymbol\omega} - \boldsymbol\beta_o)/S^2_{\boldsymbol\omega}$, as noted. Theorem 2(v) shows that $\mathcal{L}(F_{\boldsymbol\omega}) = F(p, n - p, \lambda_1, \lambda_2)$ with $\lambda_1 = (\boldsymbol\beta + \boldsymbol\kappa - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\boldsymbol\beta + \boldsymbol\kappa - \boldsymbol\beta_o)$ and $\lambda_2 = \boldsymbol\omega_2'\boldsymbol\omega_2$, where $\boldsymbol\kappa = \boldsymbol X^{\dagger}\boldsymbol\omega_1$ depends on $\boldsymbol\omega_1$.

Aberrations in testing $H_0^{\sigma} : \sigma^2 = \sigma_0^2$ against $H_1^{\sigma} : \sigma^2 \neq \sigma_0^2$ also are germane. Against one-sided upper alternatives $H_{1U}^{\sigma} : \sigma^2 > \sigma_0^2$, normal-theory tests reject at level α for $\{\nu S^2_{\varnothing}/\sigma_0^2 > c_\alpha\}$; against $H_{1L}^{\sigma} : \sigma^2 < \sigma_0^2$ the rejection rule is $\{\nu S^2_{\varnothing}/\sigma_0^2 < c_{1-\alpha}\}$ with $(c_{1-\alpha},\ c_\alpha)$ from lower and upper tails of $\chi^2(\nu, 0)$. These are as intended had there been no shifts; under shifts the altered statistic is $\nu S^2_{\boldsymbol\omega}/\sigma_0^2$. Effects of shifts in tests regarding both $\boldsymbol\beta$ and σ^2 are reported next.

Corollary 1. *Consider testing H_0^{β} vs. H_1^{β} using $F_{\boldsymbol\omega} = (\widehat{\boldsymbol\beta}_{\boldsymbol\omega} - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\widehat{\boldsymbol\beta}_{\boldsymbol\omega} - \boldsymbol\beta_o)/S^2_{\boldsymbol\omega}$ as altered under $\{\boldsymbol Y_{\boldsymbol\omega} = \boldsymbol Y_{\varnothing} + \boldsymbol\omega\}$; and let $\nu = n - p$. The test has the following properties.*

(i) *Suppose that $\boldsymbol\omega_2 = \boldsymbol 0$; then $\mathcal{L}(F_{\boldsymbol\omega}) = F(p, \nu, \lambda_1, 0)$ with $\lambda_1 = (\boldsymbol\beta + \boldsymbol\kappa - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\boldsymbol\beta + \boldsymbol\kappa - \boldsymbol\beta_o)$; and if H_0^{β} holds, then $\lambda_1 = \boldsymbol\kappa'\boldsymbol X'\boldsymbol X\boldsymbol\kappa = \boldsymbol\omega_1'\boldsymbol\omega_1$.*
(ii) *Suppose that $\boldsymbol\omega_1 = \boldsymbol 0$; then $\mathcal{L}(F_{\boldsymbol\omega}) = F(p, \nu, \lambda_1, \lambda_2)$ with $\lambda_1 = (\boldsymbol\beta - \boldsymbol\beta_o)'\boldsymbol X'\boldsymbol X(\boldsymbol\beta - \boldsymbol\beta_o)$ and $\lambda_2 = \boldsymbol\omega_2'\boldsymbol\omega_2$; and if H_0^{β} holds, then the null distribution is $\mathcal{L}(F_{\boldsymbol\omega} \mid H_0^{\beta}) = F(p, \nu, 0, \lambda_2)$.*

(iii) (a) *In consequence, for* $\boldsymbol{\omega}_1 = \mathbf{0}$*, the test is conservative in that* $P(F_{\omega} > c_{\alpha}) < \alpha$*; and* (b) *for* $\boldsymbol{\omega}_2 = \mathbf{0}$ *the test is anti-conservative in that* $P(F_{\omega} > c_{\alpha}) > \alpha$.

In testing H_0^{σ} *vs.* H_1^{σ} *using* $\{\nu S_{\omega}^2/\sigma_0^2\}$*, shifts* $\{\boldsymbol{Y}_{\omega} = \boldsymbol{Y}_{\varnothing} + \boldsymbol{\omega}\}$ *exert effects as follow.*

(iv) $\mathcal{L}(\nu S_{\omega}^2/\sigma_0^2 \mid H_0^{\sigma}) = \chi^2(\nu, \lambda_2)$ *with* $\lambda_2 = \boldsymbol{\omega}_2'\boldsymbol{\omega}_2/\sigma_0^2$.

(v) (a) *In consequence, for* $\boldsymbol{\omega}_2 \neq \mathbf{0}$*, the test for* H_{1L}^{σ} *using* $\{\nu S_{\omega}^2/\sigma_0^2 < c_{1-\alpha}\}$ *is conservative in that* $P(\nu S_{\omega}^2/\sigma_0^2 < c_{1-\alpha}) < \alpha$*; and* (b) *for* $\boldsymbol{\omega}_2 \neq \mathbf{0}$*, the test for* H_{1U}^{σ} *using* $\{\nu S_{\omega}^2/\sigma_0^2 > c_{\alpha}\}$ *is anti-conservative in that* $P(\nu S_{\omega}^2/\sigma_0^2 > c_{\alpha}) > \alpha$.

Proof. Conclusions (i), (ii) and (iv) are direct consequences of Theorem 2. Conclusion (iii) follows since $F(p, \nu, 0, \lambda_2)$ is stochastically smaller than $F(p, \nu, 0, 0)$ and $F(p, \nu, \lambda_1, 0)$ is stochastically larger than $F(p, \nu, 0, 0)$. Similarly, conclusion (v) follows since $\mathcal{L}(U) = \chi^2(\nu, \lambda_2)$ is stochastically larger than $\mathcal{L}(U) = \chi^2(\nu, 0)$. □

Remark 2. (a) Conclusions (iii(a)) and (v(a)) are akin to *Masking* in deletion diagnostics. That is, suppressing evidence in favor of H_1^{β} and H_{1L}^{σ}, respectively; specifically, suppressing that $\boldsymbol{\beta} \neq \boldsymbol{\beta}_o$, or that the actual variance σ^2 is smaller than the hypothetical σ_0^2.

(b) Conclusions (iii(b)) and (v(b)) are akin to *Swamping* in deletion diagnostics. Specifically, inferring through inflated statistics that $H_1^{\beta} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_o$ holds when in fact $\boldsymbol{\beta} = \boldsymbol{\beta}_o$; or that the actual variance σ^2 exceeds the hypothetical σ_0^2 when it does not. An assessment of the latter is to evaluate $P(\nu S_{\omega}^2/\sigma_0^2 > c_{\alpha})$ from the actual distribution $\mathcal{L}(\nu S_{\omega}^2/\sigma_0^2) = \chi^2(\nu, \lambda_2)$ under H_0^{σ}.

We next examine whether a given shift might exert differential effects on subsets of the betas.

3.4. Effects on subsets of betas

Consider a design $\{\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}\}$, semiorthogonal in that $\boldsymbol{X}_1 \perp \boldsymbol{X}_2$, *i.e.,* $\boldsymbol{X}_1'\boldsymbol{X}_2 = \mathbf{0}$. Then $\boldsymbol{X}'\boldsymbol{X} = \mathrm{Diag}(\boldsymbol{X}_1'\boldsymbol{X}_1, \boldsymbol{X}_2'\boldsymbol{X}_2)$ and thus

$$\boldsymbol{X}^{\dagger} = \begin{bmatrix} (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1' \\ (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2' \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1^{\dagger} \\ \boldsymbol{X}_2^{\dagger} \end{bmatrix}. \tag{3.1}$$

It is instructive to examine whether shifts may be induced in some estimators but not others. Shifting $\{\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{\omega}\}$ and taking $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ as before, it follows that $\widehat{\boldsymbol{\beta}}_{\omega} = \widehat{\boldsymbol{\beta}}_{\varnothing} + \boldsymbol{X}^{\dagger}\boldsymbol{\omega}_1$. Recalling that $\boldsymbol{\omega}_1 \in \mathcal{R}(\boldsymbol{H}_n) = \mathcal{R}(\boldsymbol{X})$, now partition $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_{11} + \boldsymbol{\omega}_{12}$ with $\boldsymbol{\omega}_{11} \in \mathcal{R}(\boldsymbol{X}_1)$ and $\boldsymbol{\omega}_{12} \in \mathcal{R}(\boldsymbol{X}_2)$. Then $\boldsymbol{\omega}_{11} \in \mathcal{N}(\boldsymbol{X}_2)$ and $\boldsymbol{\omega}_{12} \in \mathcal{N}(\boldsymbol{X}_1)$ since $\boldsymbol{X}_1 \perp \boldsymbol{X}_2$, where for convenience we use the notation $\mathcal{N}(\boldsymbol{X}_i)$ for $\{\mathcal{R}(\boldsymbol{I}_n - \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'); i = 1, 2\}$. It follows that

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\omega 1} \\ \widehat{\boldsymbol{\beta}}_{\omega 2} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\varnothing 1} \\ \widehat{\boldsymbol{\beta}}_{\varnothing 2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_1^{\dagger}\boldsymbol{\omega}_{11} \\ \boldsymbol{X}_2^{\dagger}\boldsymbol{\omega}_{12} \end{bmatrix}. \tag{3.2}$$

In short, it is seen that shifts in $\boldsymbol{Y}$ may induce shifts in $(\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \boldsymbol{e}, S^2)$.

Theorem 3. *Consider* $\{\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}\}$ *such that* $\boldsymbol{X}_1 \perp \boldsymbol{X}_2$*. Suppose that* $\{\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{\omega}\}$ *with* $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_{11} \in \mathcal{R}(\boldsymbol{X}_1)$.

(i) *This induces a shift in the component* $\widehat{\boldsymbol{\beta}}_{\omega 1}$ *only, as*

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\omega 1} \\ \widehat{\boldsymbol{\beta}}_{\omega 2} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{\varnothing 1} \\ \widehat{\boldsymbol{\beta}}_{\varnothing 2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_1^{\dagger}\boldsymbol{\omega}_1 \\ \mathbf{0} \end{bmatrix}. \tag{3.3}$$

(ii) *In consequence,* $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\omega 1}) = \boldsymbol{\beta}_1 + \boldsymbol{X}_1^{\dagger}\boldsymbol{\omega}_1$ *and* $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\omega 2}) = \boldsymbol{\beta}_2$.

(iii) *The residuals* $\boldsymbol{e}$, M_{SE} *and* S^2 *retain properties given in Theorems 1 and 2.*

Proof. Conclusion (i) follows from (3.2) since $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_{11}$ implies $\boldsymbol{\omega}_{12} = \mathbf{0}$ and thus (ii). Conclusion (iii) holds since $\boldsymbol{\omega} = \boldsymbol{\omega}_{11} + \boldsymbol{\omega}_2$, and that developments in Theorems 1 and 2 continue to apply regarding $\boldsymbol{\omega}_2$. □

4. Inferences regarding shifts

Evidence is sought regarding an unknown shift $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 \in \mathbb{R}^n$. To motivate, one of the authors was asked to consult with a manufacturer of automotive drive line subassemblies. A model relating a critical response to designated regressors had been established as a benchmark free of extraneous aberrations through a carefully controlled pilot study. It remained to see whether the benchmark model continues to apply under the vagaries of a full production line, in the framework of statistical process control as introduced in [14]. Moreover, an examination of elements of $\boldsymbol{\omega}$, if assessed empirically, would aid in identifying regressor settings under full production that are prone to shifts. Another venue in practice is to determine whether a calibrated instrument requires recalibration.

In this setting we postulate separate experiments, namely $\{\boldsymbol{Y}_\omega = \boldsymbol{X}\boldsymbol{\beta}_\omega + \boldsymbol{\varepsilon}_\omega\}$ having shifted responses as in Section 3, together with $\{\boldsymbol{Y}_o = \boldsymbol{X}_o\boldsymbol{\beta} + \boldsymbol{\varepsilon}_o\}$ taken to be free of shifts. Both $\boldsymbol{Y}_\omega$ and $\boldsymbol{Y}_o$ are observable. The two experiments may differ in design and size; they are commensurate in having the same $\boldsymbol{\beta}$'s; and they are to be carried out independently. In our experience researchers often know that aberrations may have occurred, in retrospect through careful notes taken during the course of an experiment.

Accordingly, we carry over from Section 3 the notation and findings for $\{\boldsymbol{Y}_\omega = \boldsymbol{X}\boldsymbol{\beta}_\omega + \boldsymbol{\varepsilon}_\omega\}$ as in the following display, together with corresponding items from the unshifted array $\{\boldsymbol{Y}_o = \boldsymbol{X}_o\boldsymbol{\beta} + \boldsymbol{\varepsilon}_o\}$.

$$\{n, \boldsymbol{X}, \boldsymbol{M}, \boldsymbol{X}^\dagger, \boldsymbol{H}_n, \widehat{\boldsymbol{\beta}}_\omega, \boldsymbol{e}_\omega, S_\omega^2, \boldsymbol{\kappa}\} \quad \text{under } \boldsymbol{Y} + \boldsymbol{\omega} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\{m, \boldsymbol{X}_o, \boldsymbol{M}_o, \boldsymbol{X}_o^\dagger, \boldsymbol{H}_o, \widehat{\boldsymbol{\beta}}_o, \boldsymbol{e}_o, S_o^2\} \quad \text{under } \boldsymbol{Y}_o = \boldsymbol{X}_o\boldsymbol{\beta} + \boldsymbol{\varepsilon}_o.$$

Specifics are $\boldsymbol{X}_o^\dagger = (\boldsymbol{X}_o'\boldsymbol{X}_o)^{-1}\boldsymbol{X}_o'$, $\widehat{\boldsymbol{\beta}}_o = \boldsymbol{X}_o^\dagger\boldsymbol{Y}_o$, $\boldsymbol{M}_o = (\boldsymbol{X}_o'\boldsymbol{X}_o)^{-1}$, $\boldsymbol{H}_o = \boldsymbol{X}_o(\boldsymbol{X}_o'\boldsymbol{X}_o)^{-1}\boldsymbol{X}_o'$, $\boldsymbol{e}_o = (\boldsymbol{Y}_o - \boldsymbol{X}_o\widehat{\boldsymbol{\beta}}_o)$, and $S_o^2 = \boldsymbol{e}_o'\boldsymbol{e}_o/(m-p)$, where $(\widehat{\boldsymbol{\beta}}_o, S_o^2)$ are unbiased in the conventional Gauss–Markov setting, namely Assumption A_o for the reference experiment on replacing n by m in Assumption A.

To continue, since $\boldsymbol{\omega} \in \mathbb{R}^n$, we seek evidence regarding $\boldsymbol{\omega}_1 \in \mathcal{R}(\boldsymbol{H}_n)$ and $\boldsymbol{\omega}_2 \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$. First consider $\widehat{\boldsymbol{\kappa}} = (\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o)$ as a prospective estimator for $\boldsymbol{\kappa} = \boldsymbol{X}^\dagger\boldsymbol{\omega}_1 \in \mathbb{R}^p$. This turns out to be unbiased with further properties to be cited. However, the challenge is to lift this from $\boldsymbol{\kappa} \in \mathbb{R}^p$ to $\boldsymbol{\omega}_1 \in \mathbb{R}^n$. As a tentative step note that both $\boldsymbol{\kappa} \in \mathbb{R}^p$ and $\boldsymbol{\omega}_1 \in \mathcal{R}(\boldsymbol{H}_n)$ in effect are p-dimensional. Accordingly, we next apply $\boldsymbol{X}$ to $\boldsymbol{\kappa} = \boldsymbol{X}^\dagger\boldsymbol{\omega}_1$ as $\boldsymbol{X}\boldsymbol{X}^\dagger\boldsymbol{\omega}_1 = \boldsymbol{H}_n\boldsymbol{\omega}_1 = \boldsymbol{\omega}_1$ to get the prospective estimator $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o) \in \mathbb{R}^n$. Details are given subsequently. Similarly, for the case that $m = n$, consider $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o)$ for estimating $\boldsymbol{\omega}_2 \in \mathbb{R}^n$. Looking ahead, the n-dimensional joint distributions $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1)$ and $\mathcal{L}(\widehat{\boldsymbol{\omega}}_2)$ on $\mathbb{R}^n$ necessarily will be singular of ranks p and $n-p$, respectively, as will their $(n \times n)$ dispersion matrices, since $\widehat{\boldsymbol{\omega}}_1 \in \mathcal{R}(\boldsymbol{H}_n)$ and $\widehat{\boldsymbol{\omega}}_2 \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$. Essential properties are collected in the following where, on occasion, considerable simplification accrues on allowing $n = m$ and the designs $\boldsymbol{X}$ and $\boldsymbol{X}_o$ to coincide.

Theorem 4. *Consider $\widehat{\boldsymbol{\kappa}} = (\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o) \in \mathbb{R}^p$ and $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o) \in \mathbb{R}^n$ as prospective estimators for $\boldsymbol{\kappa} = \boldsymbol{X}^\dagger\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_1 \in \mathcal{R}(\boldsymbol{H}_n)$. Further taking $m = n$, consider $\widehat{\boldsymbol{\omega}} = (\boldsymbol{Y}_\omega - \boldsymbol{Y}_o)$ and $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o) \in \mathbb{R}^n$ for estimating $\boldsymbol{\omega} \in \mathbb{R}^n$ and $\boldsymbol{\omega}_2 \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$. Under Assumption A and A_o we have*

(i) $\mathrm{E}(\widehat{\boldsymbol{\kappa}}) = \boldsymbol{\kappa}$, $\mathrm{V}(\widehat{\boldsymbol{\kappa}}) = \sigma^2(\boldsymbol{M} + \boldsymbol{M}_o)$ *and, if* $\boldsymbol{X} = \boldsymbol{X}_o$, *then* $\mathrm{V}(\widehat{\boldsymbol{\kappa}}) = 2\boldsymbol{M}$;
(ii) $\mathrm{E}(\widehat{\boldsymbol{\omega}}_1) = \boldsymbol{\omega}_1$, $\mathrm{V}(\widehat{\boldsymbol{\omega}}_1) = \sigma^2\boldsymbol{V}$ *with* $\boldsymbol{V} = \boldsymbol{X}(\boldsymbol{M} + \boldsymbol{M}_o)\boldsymbol{X}'$ *and, if* $\boldsymbol{X} = \boldsymbol{X}_o$, *then* $\boldsymbol{V} = 2\boldsymbol{H}_n$;
(iii) $\mathrm{E}(\widehat{\boldsymbol{\omega}}_2) = \boldsymbol{\omega}_2$ *and* $\mathrm{V}(\widehat{\boldsymbol{\omega}}_2) = 2(\boldsymbol{I}_n - \boldsymbol{H}_n)$;
(iv) $\mathrm{E}(\widehat{\boldsymbol{\omega}}) = \boldsymbol{\omega}$ *and* $\mathrm{V}(\widehat{\boldsymbol{\omega}}) = 2\sigma^2\boldsymbol{I}_n$.
Moreover, taking $m = n$, for $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_2$ we have
(v) $\mathcal{L}(\widehat{\boldsymbol{\kappa}}) = N_p(\boldsymbol{\kappa}, \sigma^2(\boldsymbol{M} + \boldsymbol{M}_o))$ *and, if* $\boldsymbol{X} = \boldsymbol{X}_o$, *then* $\mathcal{L}(\widehat{\boldsymbol{\kappa}}) = N_p(\boldsymbol{\kappa}, 2\sigma^2\boldsymbol{M})$;
(vi) $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1) = N_n(\boldsymbol{\omega}_1, \sigma^2\boldsymbol{V})$ *and, if* $\boldsymbol{X} = \boldsymbol{X}_o$, *then* $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1) = N_n(\boldsymbol{\omega}_1, 2\sigma^2\boldsymbol{H}_n)$;
(vii) $\mathcal{L}(\widehat{\boldsymbol{\omega}}_2) = N_n(\boldsymbol{\omega}_2, 2\sigma^2(\boldsymbol{I}_n - \boldsymbol{H}_n))$.
(viii) $\mathcal{L}(\widehat{\boldsymbol{\omega}}) = N_n(\boldsymbol{\omega}, 2\sigma^2\boldsymbol{I}_n)$ *and* $\mathcal{L}(\widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\omega}}/2\sigma^2) = \chi^2(n, \delta)$ *with* $\delta = \boldsymbol{\omega}'\boldsymbol{\omega}/2\sigma^2$.
(ix) $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1'\boldsymbol{V}^\dagger\widehat{\boldsymbol{\omega}}_1/\sigma^2) = \chi^2(p, \delta_1)$ *with* $\boldsymbol{V}^\dagger$ *as the Moore–Penrose inverse and* $\delta_1 = \boldsymbol{\omega}_1'\boldsymbol{V}^\dagger\boldsymbol{\omega}_1/\sigma^2$ *and, if* $\boldsymbol{X} = \boldsymbol{X}_o$, *then* $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2\sigma^2) = \chi^2(p, \delta_1)$ *with* $\delta_1 = \boldsymbol{\omega}_1'\boldsymbol{\omega}_1/2\sigma^2$.
(x) $\mathcal{L}(\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2\sigma^2) = \chi^2(n-p, \delta_2)$ *with* $\delta_2 = \boldsymbol{\omega}_2'\boldsymbol{\omega}_2/2\sigma^2$.
(xi) *For the case* $\boldsymbol{X} = \boldsymbol{X}_o$ *it follows that* $\widehat{\boldsymbol{\omega}} = \widehat{\boldsymbol{\omega}}_1 + \widehat{\boldsymbol{\omega}}_2$, *and the noncentrality parameters satisfy* $\delta = \delta_1 + \delta_2$.

Proof. Conclusions (i)–(iii) follow on combining properties of $(\widehat{\boldsymbol{\beta}}_\omega, \boldsymbol{e}_\omega)$ from Theorem 1(ii), (v) under Assumptions A_1, with those of $(\widehat{\boldsymbol{\beta}}_o, \boldsymbol{e}_o)$ under A_{01}, together with independence of the two experiments. Since $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o)$, we have $\mathrm{V}(\widehat{\boldsymbol{\omega}}_2) = \sigma^2[(\boldsymbol{I}_n - \boldsymbol{H}_n) + (\boldsymbol{I}_n - \boldsymbol{H}_o)]$. Because $S_p(\boldsymbol{X}) = S_p(\boldsymbol{X}_o)$, it follows that $\boldsymbol{H}_n$ and $\boldsymbol{H}_o$ are interchangeable in projecting to this common subspace. Accordingly, $\mathrm{V}(\widehat{\boldsymbol{\omega}}_2) = 2\sigma^2(\boldsymbol{I}_n - \boldsymbol{H}_n)$ as asserted in conclusion (iii). Conclusion (iv) follows since $\mathrm{V}(\boldsymbol{Y}_\omega - \boldsymbol{Y}_o) = 2\sigma^2\boldsymbol{I}_n$. Assertions (v)–(viii) follow directly from (i)–(iv) and linearity under Gaussian errors. Conclusion (ix) follows initially from Theorem 9.2.3 of Rao and Mitra [23, p. 173] since $\boldsymbol{V}^\dagger$ is also a reflexive g-inverse. The second part of (ix) follows since $\boldsymbol{V} = 2\boldsymbol{H}_n$ and $\boldsymbol{V}^\dagger = \frac{1}{2}\boldsymbol{H}_n$ from Lemma 2.1(iii), so that $\widehat{\boldsymbol{\omega}}_1'\boldsymbol{V}^\dagger\widehat{\boldsymbol{\omega}}_1 = \frac{1}{2}\widehat{\boldsymbol{\omega}}_1'\boldsymbol{H}_n\widehat{\boldsymbol{\omega}}_1 = \frac{1}{2}\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1$, and similarly $\boldsymbol{\omega}_1'\boldsymbol{V}^\dagger\boldsymbol{\omega}_1 = \frac{1}{2}\boldsymbol{\omega}_1'\boldsymbol{H}_n\boldsymbol{\omega}_1 = \frac{1}{2}\boldsymbol{\omega}_1'\boldsymbol{\omega}_1$ for its noncentral parameter. Conclusion (x) follows directly from (iii) and (vii). Conclusion (xi) follows from conclusion (viii) and the special case in the second part of (ix). □

Remark 3. For the general case that $\mathrm{V}(\widehat{\boldsymbol{\omega}}_1) = \boldsymbol{V} = \boldsymbol{X}(\boldsymbol{M} + \boldsymbol{M}_o)\boldsymbol{X}'$ of order $(n \times n)$ and rank p, its Moore–Penrose inverse $\boldsymbol{V}^\dagger$ may be found through its spectral decomposition $\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}'$ with $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{D}_1, \boldsymbol{0})$ and $\boldsymbol{D}_1 = \mathrm{Diag}(d_1, \ldots, d_p)$. Then $\boldsymbol{V}^\dagger = \boldsymbol{Q}\boldsymbol{D}^\dagger\boldsymbol{Q}'$ is the Moore–Penrose inverse of $\boldsymbol{V}$, where $\boldsymbol{D}^\dagger = \mathrm{Diag}(\boldsymbol{D}_1^{-1}, \boldsymbol{0})$.

We turn next to hypothesis tests regarding $\boldsymbol{\omega} = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$.

Table 1
Values $\{\boldsymbol{\kappa}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2\}$ from $\boldsymbol{X}$ and $\boldsymbol{\omega}' = [4, 0, 0, 0]$; observations $\boldsymbol{Y}$ of $\{Y_i = 13 + 2X_i + \varepsilon_i\}$; $\boldsymbol{Y}_\omega = \boldsymbol{Y} + \boldsymbol{\omega}$; together with *OLS* solutions $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_\omega)$ and residual vectors $(\boldsymbol{e}, \boldsymbol{e}_\omega)$.

$\boldsymbol{\kappa}$	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\boldsymbol{Y}$	$\widehat{\boldsymbol{\beta}}$	$\boldsymbol{e}$	$\boldsymbol{Y}+\boldsymbol{\omega}$	$\widehat{\boldsymbol{\beta}}_\omega$	$\boldsymbol{e}_\omega$
1.0	2.8	1.2	7.9935	12.8525	1.01570	11.9935	13.8525	2.21570
−0.6	1.6	−1.6	9.6834	1.9582	−1.21088	9.6834	1.3582	−2.81088
	0.4	−0.4	14.1854		−0.62534	14.1854		−1.02634
	−0.8	0.8	19.5477		0.82052	19.5477		1.62052

Theorem 5. *Consider* $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o) \in \mathbb{R}^n$, *and for* $m = n$, $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o) \in \mathbb{R}^n$ *as estimators for* $\boldsymbol{\omega}_1 \in \mathcal{R}(\boldsymbol{H}_n)$ *and* $\boldsymbol{\omega}_2 \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$, *respectively, with properties as in Theorem 4. Under Assumption A and* A_o *we have*

(i) *A normal-theory test for* $H_0 : \boldsymbol{\omega}_1 = \mathbf{0}$ *against* $H_1 : \boldsymbol{\omega}_1 \neq \mathbf{0}$ *at level* α *utilizes the statistic* $F = (\widehat{\boldsymbol{\omega}}_1' \boldsymbol{V}^\dagger \widehat{\boldsymbol{\omega}}_1 / pS_o^2)$ *together with the rejection rule* $F > c_\alpha$ *from* $F(p, m-p, 0)$; *specifically, if* $n = m$ *and* $\boldsymbol{X} = \boldsymbol{X}_o$, *then* $F = (\widehat{\boldsymbol{\omega}}_1' \widehat{\boldsymbol{\omega}}_1 / 2pS_o^2)$ *with* c_α *from* $F(p, n-p, 0)$.

(ii) *For the case* $n = m$ *and* $\boldsymbol{X} = \boldsymbol{X}_o$, *a normal-theory test for* $H_0 : \boldsymbol{\omega}_2 = \mathbf{0}$ *against* $H_1 : \boldsymbol{\omega}_2 \neq \mathbf{0}$ *at level* α *utilizes the statistic* $G = (\widehat{\boldsymbol{\omega}}_2' \widehat{\boldsymbol{\omega}}_2 / 2\boldsymbol{e}_o' \boldsymbol{e}_o)^{1/2}$ *together with the rejection rule* $G > c_\alpha^*$ *from the distribution* $G_\rho(n-p, n-p, 0)$ *of correlated ratios as in Lemma A.1*(iii).

Proof. Conclusions (i) and (ii) are complicated by prospective dependences between $(\widehat{\boldsymbol{\omega}}_1, \boldsymbol{e}_o)$ and $(\widehat{\boldsymbol{\omega}}_2, \boldsymbol{e}_o)$. Observe that $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\boldsymbol{X}^\dagger \boldsymbol{Y}_\omega - \boldsymbol{X}_o^\dagger \boldsymbol{Y}_o) = \boldsymbol{H}_n \boldsymbol{Y}_\omega - \boldsymbol{X}\boldsymbol{X}_o^\dagger \boldsymbol{Y}_o$. The cross-covariance in (i) is $\mathrm{Cov}(\widehat{\boldsymbol{\omega}}_1, \boldsymbol{e}_o) = \mathrm{Cov}((\boldsymbol{H}_n \boldsymbol{Y}_\omega - \boldsymbol{X}\boldsymbol{X}_o^\dagger \boldsymbol{Y}_o),\ (\boldsymbol{I}_n - \boldsymbol{H}_o)\boldsymbol{Y}_o) = \mathrm{Cov}(\boldsymbol{H}_n \boldsymbol{Y}_\omega,\ (\boldsymbol{I}_n - \boldsymbol{H}_o)\boldsymbol{Y}_o) - \mathrm{Cov}(\boldsymbol{X}\boldsymbol{X}_o^\dagger \boldsymbol{Y}_o,\ (\boldsymbol{I}_n - \boldsymbol{H}_o)\boldsymbol{Y}_o) = \mathbf{0} - \mathrm{Cov}(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_o,\ \boldsymbol{e}_o) = \mathbf{0}$. Conclusion (i) follows from Theorem 4(ix) together with the central distribution $\mathcal{L}((m-p)S_o^2/\sigma^2) = \chi^2(m-p, 0)$ and independence of $(\widehat{\boldsymbol{\omega}}_1, S_o^2)$ under Gaussian errors. The given statistic for testing $H_0 : \boldsymbol{\omega}_2 = \mathbf{0}$ against $H_1 : \boldsymbol{\omega}_2 \neq \mathbf{0}$ is complicated by the fact that $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o)$ and $\boldsymbol{e}_o$ are dependent. Details accounting for this are supplied in Lemma A.1, giving in conclusion (iii) of that lemma an expression for the *pdf* of $G_\rho(n-p, n-p, 0)$, its upper cutoff value c_α^* giving the rejection rule of conclusion (ii). □

5. Case studies

We begin with an elementary example in monitoring linear profiles, and then proceed to more expansive studies from the literature. Recalling that single-case and subset deletions are special cases restricting $\boldsymbol{\omega}$ to one or a few nonzero entries, further connections to deletion diagnostics are outlined in Appendix A.2.

5.1. Case study 1

An example with $(n = 4, p = 2)$ is taken from Kang and Albin [18] and Kim et al. [20], who carried out extensive simulation studies in regard to monitoring linear profiles in statistical process control. In the centered form of Kim et al. [20], observations $\{Y_i = 13 + 2X_i + \varepsilon_i;\ 1 \le i \le 4\}$ are generated as reported in Table 1 for $X_i \in [-3, -1,\ 1,\ 3]$ in order, where the disturbances are $N(0, 1)$ deviates. The shift is $\boldsymbol{\omega}' = [4, 0, 0, 0]$, a single outlier in the parlance of deletion diagnostics. Also listed in Table 1 are $\boldsymbol{\omega}_1 = \boldsymbol{H}_4 \boldsymbol{\omega}$ and $\boldsymbol{\omega}_2 = (\boldsymbol{I}_4 - \boldsymbol{H}_4)\boldsymbol{\omega}$ and $\boldsymbol{\kappa} = \boldsymbol{X}^\dagger \boldsymbol{\omega}_1$. Computations utilize the MINITAB package. As in Definition 1 the angles (θ_1, θ_2) between $(\boldsymbol{\omega}, \boldsymbol{\omega}_1)$ and $(\boldsymbol{\omega}, \boldsymbol{\omega}_2)$ are $\theta_1 = 33.2°$ and $\theta_2 = 56.8°$. These gauge the proximity of the shift $\boldsymbol{\omega}' = [4, 0, 0, 0]$ to the "Regressor" and "Error" spaces, favoring the latter, where $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ serve to perturb the *OLS* solutions and the residuals, respectively. In contrast to these, the shift $\boldsymbol{\omega} = [1, -2, 1, 0] \in \mathcal{R}(\boldsymbol{H}_n)$ yields $\theta_2 = 0°$, perturbing $\widehat{\boldsymbol{\beta}}$ but not $\boldsymbol{e}$. Alternatively, $\boldsymbol{\omega}' = [3, 2, 1, 0] \in \mathcal{R}(\boldsymbol{I}_n - \boldsymbol{H}_n)$ yields $\theta_1 = 0°$, perturbing the residuals $\boldsymbol{e}$ but not $\widehat{\boldsymbol{\beta}}$. Our analysis continues in regard to the shift $\boldsymbol{\omega}' = [4, 0, 0, 0]$.

For completeness we list the transpose of $\boldsymbol{X}^\dagger$ and the (4×4) matrix $\boldsymbol{H}_4$ as follows.

$$\boldsymbol{X}^{\dagger\prime} = \begin{bmatrix} -0.15 & 0.25 \\ -0.05 & 0.25 \\ 0.05 & 0.25 \\ 0.15 & 0.25 \end{bmatrix}; \qquad \boldsymbol{H}_4 = \begin{bmatrix} 0.7 & 0.4 & 0.1 & -0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.3 & 0.4 \\ -0.2 & 0.1 & 0.4 & 0.7 \end{bmatrix}. \tag{5.1}$$

Observe in Table 1 that $\boldsymbol{\kappa} = \boldsymbol{X}^\dagger \boldsymbol{\omega}_1$, $\widehat{\boldsymbol{\beta}}_\omega = \widehat{\boldsymbol{\beta}} + \boldsymbol{\kappa}$ and $\boldsymbol{e}_\omega = \boldsymbol{e} + \boldsymbol{\omega}_2$ are verified numerically.

To illustrate Remark 3, suppose instead that $X_i \in [-3, 0,\ 0,\ 3]$ in $\boldsymbol{X}_o$. Then the eigenvalues of $\boldsymbol{V} = \boldsymbol{X}(\boldsymbol{M} + \boldsymbol{M}_o)\boldsymbol{X}'$ are $[2.11111, 2, 0, 0]$. The Moore–Penrose inverse of $\boldsymbol{V}$ is

$$\boldsymbol{V}^\dagger = \begin{bmatrix} 0.33816 & 0.19605 & 0.05395 & -0.08816 \\ 0.19605 & 0.14868 & 0.10132 & 0.05395 \\ 0.05395 & 0.10132 & 0.14868 & 0.19605 \\ -0.08816 & 0.05395 & 0.19605 & 0.33816 \end{bmatrix}$$

to be compared with $(2\boldsymbol{H}_4)^\dagger = \frac{1}{2}\boldsymbol{H}_4$ from (5.1) for the case $\boldsymbol{M}_o = \boldsymbol{M}$. To illustrate Comment C_4 following Theorem 1, we compute $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_1\delta_1 = \begin{bmatrix} 4 & 0 \\ 0 & 20 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -3 \end{bmatrix}(4) = \begin{bmatrix} 1.0 \\ -0.6 \end{bmatrix}$ as the bias owing to the shift $\{Y_1 + 4\}$, given also as κ in Table 1.

Table 2
Responses for analyses, to include $\boldsymbol{Y}_\varnothing = f(\boldsymbol{X}) + \boldsymbol{\varepsilon}$, $\boldsymbol{Y}_\omega = (\boldsymbol{Y}_\varnothing + \boldsymbol{\omega}), (\boldsymbol{Y}_\varnothing + \boldsymbol{\omega}_{11}), (\boldsymbol{Y}_\omega + \boldsymbol{\omega}_{11})$, and $(\boldsymbol{Y}_\omega - \boldsymbol{\omega}_{12})$, with $\boldsymbol{\omega}_{11} = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o)$ and $\boldsymbol{\omega}_{12} = \boldsymbol{X}\boldsymbol{\beta}_o$; to include *OLS* solutions and S^2 for each; partial elements of residuals $\boldsymbol{e}_\varnothing$ and $\boldsymbol{e}_\omega$; and the bias $B(S^2)$ for S^2 taking values $S^2_\varnothing$ and S^2_ω.

Item	$\boldsymbol{Y}_\varnothing$	$\boldsymbol{Y}_\omega$	$\boldsymbol{Y}_\varnothing + \boldsymbol{\omega}_{11}$	$\boldsymbol{Y}_\omega + \boldsymbol{\omega}_{11}$	$\boldsymbol{Y}_\omega - \boldsymbol{\omega}_{12}$	$\boldsymbol{e}^*_\varnothing$	$\boldsymbol{e}^*_\omega$
$\widehat{\beta}_0$	−0.40989	−1.26129	−1.67118	−2.52258	−1.26129	0.918	3.309
$\widehat{\beta}_1$	0.99809	1.09743	1.09552	1.19486	0.09743	−0.508	1.891
$\widehat{\beta}_2$	1.04608	1.11081	1.15689	1.22162	0.11081	−1.489	0.918
$\boldsymbol{e}'\boldsymbol{e}$	17.16630	37.31680	17.16630	37.31680	37.31680	1.359	−0.266
S^2	0.78029	1.30230	0.78029	1.30230	1.30230	0.052	0.351
$B(S^2)$	0.00000	1.30830	0.00000	1.30830	1.30830	−1.389	−1.886

5.2. Case study 2: Hadi and Simonoff data

5.2.1. Background

Hadi and Simonoff [10] presented an artificial data set with two predictor variables $\{\boldsymbol{X}_1, \boldsymbol{X}_2\}$ having response $\boldsymbol{Y} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{\varepsilon}$, sample size $n = 25$, and outliers embedded in rows $\{1, 2, 3\}$ designed to be difficult to find. Their errors were generated from $N(0, 1)$ for rows $\{4, \dots, 25\}$ and zeros for rows $\{1, 2, 3\}$. Outliers are fixed at $\boldsymbol{\omega} = [4, 4, 4, 0, \dots, 0]' \in \mathbb{R}^n$. Since their responses were generated with a constant term of zero, we write $\{\boldsymbol{Y} = \beta_0 \mathbf{1}_n + \beta_1 \boldsymbol{X}_1 + \beta_2 \boldsymbol{X}_2 + \boldsymbol{\varepsilon}\}$, where the actual parameters generating the data are $\boldsymbol{\beta}_o' = [\beta_0, \beta_1, \beta_2] = [0, 1, 1]$. For our case studies all rows, including $\{1, 2, 3\}$, are disturbed by $N(0, 1)$ errors so as to be amenable to Gauss–Markov theory. The design matrix $\boldsymbol{X} = [\mathbf{1}_n, \boldsymbol{X}_1, \boldsymbol{X}_2]$ is retained throughout; components $\boldsymbol{\omega}_1 = \boldsymbol{H}_n \boldsymbol{\omega}$ and $\boldsymbol{\omega}_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$ in $\mathbb{R}^n$ are reported subsequently in Table 3, where we determine that $\boldsymbol{\omega}_1' \boldsymbol{\omega}_1 = 19.2173$ and $\boldsymbol{\omega}_2' \boldsymbol{\omega}_2 = 28.7827$. Accordingly, the angle θ_1 between $(\boldsymbol{\omega}, \boldsymbol{\omega}_1)$ is $\theta_1 = 50.8°$, indicating that $\boldsymbol{\omega}$ has only a slight propensity towards the "Error" space. Hadi and Simonoff [10] and others offer often intricate subset deletion algorithms for identifying outlying subsets. Our work supports the view that shifts in responses are tantamount to shifts in the *OLS* solutions, and to inflated variation about the best-fitting line, these being the principal deleterious consequences of outlying data. Accordingly, methods offered here would seem appropriate for first determining whether such shifts are apparent.

5.2.2. Case 2.1

We first illustrate basics of Theorems 1 and 2. Given the origins of *these* data, shifts otherwise unknown in practice, the unobservable underlying model $\{\boldsymbol{Y}_\varnothing = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$ of Section 3 now can be recovered. Specifically, outcomes $\{\boldsymbol{Y}_\varnothing, \widehat{\boldsymbol{\beta}}_\varnothing, \boldsymbol{e}_\varnothing, S^2_\varnothing\}$ all are now observable. In the Case 2.1 studies we retain the data of Hadi and Simonoff [10] as reported in their Table 1, to include their simulated $N(0, 1)$ disturbances. In addition we attach standard normal disturbances $\{Y_1 + 1.17062, Y_2 - 0.25768, Y_3 - 1.24093\}$ giving the shifted model $\{\boldsymbol{Y}_\omega = \boldsymbol{X}\boldsymbol{\beta}_\omega + \boldsymbol{\varepsilon}_\omega\}$ of Section 3. Subtracting $\boldsymbol{\omega} = [4, 4, 4, 0, \dots, 0]'$ from these responses gives the actual unshifted model $\{\boldsymbol{Y}_\varnothing = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$, retaining the same random disturbances in keeping with our model for shifted data. Accordingly, elements of $\{\widehat{\boldsymbol{\beta}}_\varnothing = \boldsymbol{X}^\dagger \boldsymbol{Y}_\varnothing, \ \boldsymbol{e}_\varnothing = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{Y}_\varnothing, \ S^2_\varnothing = \boldsymbol{e}_\varnothing' \boldsymbol{e}_\varnothing/(n-p)\}$ of Section 3 are now observable; these are listed in column 2 of Table 2, where in the interests of brevity $\boldsymbol{e}_\varnothing$ is given in part in column 7. For reference the *OLS* solutions in Table 2 are identified as $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}^*)$, with $\boldsymbol{Y}^*$ taking successive values heading columns 2–6. Specifically, $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}_\varnothing) = \widehat{\boldsymbol{\beta}}_\varnothing$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}_\omega) = \widehat{\boldsymbol{\beta}}_\omega$ in notation set earlier. Accordingly, elements of $\{\widehat{\boldsymbol{\beta}}_\omega = \boldsymbol{X}^\dagger \boldsymbol{Y}_\omega, \ \boldsymbol{e}_\omega = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{Y}_\omega, \ S^2_\omega = \boldsymbol{e}_\omega' \boldsymbol{e}_\omega/(n - p)\}$ are listed in column 3 of Table 2, with $\boldsymbol{e}_\omega$ appearing in part in column 8 with only the first six entries displayed, as in column 7.

The matrix $\boldsymbol{X}^\dagger = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ next is applied to $\boldsymbol{\omega} = [4, 4, 4, 0, \dots, 0]'$ to give $\boldsymbol{\kappa} = \boldsymbol{X}^\dagger \boldsymbol{\omega} = \boldsymbol{X}^\dagger \boldsymbol{\omega}_1 = [-0.85140, 0.09934, 0.06473]'$ as in Theorem 1. This illustrates from Table 2 the Theorem 1(ii) dictum that $\widehat{\boldsymbol{\beta}}_\omega = \widehat{\boldsymbol{\beta}}_\varnothing + \boldsymbol{\kappa}$, specifically,

$$\begin{bmatrix} -1.26129 \\ 1.09743 \\ 1.11081 \end{bmatrix} = \begin{bmatrix} -0.40989 \\ 0.99809 \\ 1.04608 \end{bmatrix} + \begin{bmatrix} -0.85140 \\ 0.09934 \\ 0.06473 \end{bmatrix}.$$

Moreover, Theorem 1(v) assertion that $\boldsymbol{e}_\omega = \boldsymbol{e}_\varnothing + \boldsymbol{\omega}_2$ is illustrated numerically, as seen for the subsets $(\boldsymbol{e}^*_\varnothing, \boldsymbol{e}^*_\omega)$ in columns 7 and 8 together with corresponding elements of $\boldsymbol{\omega}_2$ from Table 3. In particular, elements in the first row of columns 7 and 8 are related by $0.918 + 2.391 = 3.309$.

Further options are seen for manipulating the *OLS* solutions and their residuals on shifting $\boldsymbol{Y}_\varnothing$ and $\boldsymbol{Y}_\omega$ as in Theorem 1(i). Specifically, replace $\boldsymbol{\kappa}'$ by

$$(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o)' = [-1.26129, 0.09743, 0.11081]$$

with $\widehat{\boldsymbol{\beta}}_\omega$ from the third column of Table 2 and $\boldsymbol{\beta}_o = [0, 1, 1]'$ as the actual values giving the simulated data. From this we recover the corresponding $\boldsymbol{\omega}_{11}$ as $\boldsymbol{\omega}_{11} = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o)$, then apply this to $\boldsymbol{Y}_\varnothing$ as $\{\boldsymbol{Y}_\varnothing \to \boldsymbol{Y}_\varnothing + \boldsymbol{\omega}_{11}\}$ as in Table 2. It is verified numerically that $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}_\varnothing + \boldsymbol{\omega}_{11})$ in column 4 of Table 2 satisfies $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}_\varnothing + \boldsymbol{\omega}_{11}) = \widehat{\boldsymbol{\beta}}_\varnothing + \widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o$. Similarly, when applied to $\boldsymbol{Y}_\omega$ as $\{\boldsymbol{Y}_\omega \to \boldsymbol{Y}_\omega + \boldsymbol{\omega}_{11}\}$, it follows that $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}_\omega + \boldsymbol{\omega}_{11}) = \widehat{\boldsymbol{\beta}}_\omega + \widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o$ with values as reported in column 5 of Table 2.

Table 3
Data $\boldsymbol{Y}_o = \boldsymbol{X}_o\boldsymbol{\beta}_o + \boldsymbol{\varepsilon}_o$ and $\boldsymbol{Y}_\omega = \boldsymbol{X}\boldsymbol{\beta}_\omega + \boldsymbol{\varepsilon}_\omega$ as in Section 4 having $N(0, 1)$ deviates generated independently within and across samples, to include residuals $\boldsymbol{e}_o$ and $\boldsymbol{e}_\omega$ and values for $\boldsymbol{\omega}_1 = \boldsymbol{H}_n\boldsymbol{\omega}$, $\boldsymbol{\omega}_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$, $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o)$, and $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o)$.

$\boldsymbol{Y}_o$	$\boldsymbol{e}_o$	$\boldsymbol{Y}_\omega$	$\boldsymbol{e}_\omega$	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\widehat{\boldsymbol{\omega}}_1$	$\widehat{\boldsymbol{\omega}}_2$
30.263	0.635	34.414	3.526	1.610	2.391	1.260	2.891
29.169	−0.359	33.488	2.704	1.601	2.399	1.257	3.063
31.826	2.397	32.598	1.917	1.593	2.407	1.253	−0.480
29.609	−0.887	28.896	−2.595	1.625	−1.625	0.995	−1.708
5.965	−0.369	7.838	1.015	−0.299	0.299	0.489	1.384
15.658	−1.408	16.308	−1.159	0.497	−0.497	0.401	0.249
15.839	1.918	14.488	0.219	0.249	−0.249	0.348	−1.699
12.801	−0.573	13.593	−0.493	0.276	−0.276	0.710	0.081
14.802	−0.034	14.553	−1.195	0.424	−0.424	0.912	−1.161
15.164	−1.096	16.495	−0.755	0.547	−0.547	0.990	0.341
10.600	−1.109	12.536	−0.264	0.221	−0.221	1.091	0.845
2.042	0.086	1.596	−0.859	−0.628	0.628	0.499	−0.945
25.196	0.255	25.844	−0.497	1.281	−1.281	1.399	−0.751
15.447	2.364	15.374	1.367	0.294	−0.294	0.925	−0.997
11.920	0.515	13.260	0.604	0.228	−0.228	1.250	0.089
6.515	−1.035	7.731	−0.371	−0.195	0.195	0.552	0.663
27.092	−1.093	26.351	−2.777	1.441	−1.441	0.943	−1.684
21.759	0.388	21.063	−0.749	0.830	−0.830	0.442	−1.137
9.069	0.802	8.496	0.085	−0.218	0.218	0.143	−0.717
17.883	−0.331	18.181	−1.531	0.791	−0.791	1.498	−1.200
6.398	−0.531	5.462	−1.934	−0.258	0.258	0.467	−1.404
0.810	0.582	1.677	1.228	−0.811	−0.811	0.222	0.645
14.5854	−0.491	15.868	0.726	0.283	−0.283	0.065	1.217
21.962	−0.710	23.719	−0.266	1.093	−1.093	1.313	0.444
5.045	0.086	7.102	2.055	−0.478	0.478	0.088	1.969

Special features emerge if instead we lift $\boldsymbol{\beta}_o = [0, 1, 1]'$ in $\mathbb{R}^p$ as $\boldsymbol{\omega}_{12} = \boldsymbol{X}\boldsymbol{\beta}_o$ in $\mathbb{R}^n$ and apply as the shift $\{\boldsymbol{Y}_\omega \to \boldsymbol{Y}_\omega - \boldsymbol{\omega}_{12}\}$. Then the beta values for $(\boldsymbol{Y}_\omega - \boldsymbol{\omega}_{12})$ in column 6 are seen to be $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}_\omega - \boldsymbol{\omega}_{12}) = \widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o$, generating the discrepancies between $\widehat{\boldsymbol{\beta}}_\omega$ and $\boldsymbol{\beta}_o$. In particular, these values give $Q = (\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o)'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \boldsymbol{\beta}_o) = 26.5775$ as the numerator for the normal-theory F-statistic in testing $H_0^\beta : \boldsymbol{\beta}_\omega = \boldsymbol{\beta}_o$ against $H_1^\beta : \boldsymbol{\beta}_\omega \neq \boldsymbol{\beta}_o$. But since $\sigma^2 = 1.0$ for these data and $\mathcal{L}(Q \mid H_0) = \chi^2(3, 0)$, the p-value in testing at level $\alpha = 0.05$ is $P(Q > 26.5775) = 0.722 \times 10^{-5}$. This in turn offers overwhelming evidence not only that responses have shifted in the Table 1 data of Hadi and Simonoff [10], but that these shifts are accompanied by riveting standardized changes in the betas.

Further evidence resides in the residuals. Note first that shifts in $\boldsymbol{Y}_\varnothing$ where $\boldsymbol{\omega}_2 \neq \boldsymbol{0}$ generate shifted residuals as in columns 3, 5 and 6. Residuals remain unshifted when $\boldsymbol{\omega}_2 = \boldsymbol{0}$ as in columns 2 and 4. In consequence, $\mathrm{E}(S_\omega^2) = \sigma^2 + \boldsymbol{\omega}_2'\boldsymbol{\omega}_2/(n-p)$ from Theorem 1(viii), with $\boldsymbol{\omega}_2'\boldsymbol{\omega}_2/(n-p) = 28.7827/22 = 1.3083$ as the bias $B(S_\omega^2)$ shown in Table 2. Moreover, the normal-theory statistic in testing $H_0^\sigma : \sigma^2 = \sigma_0^2$ against $H_1^\sigma : \sigma^2 > \sigma_0^2$ is $W = (n-p)S^2/\sigma_0^2$ having distribution $\mathcal{L}((n-p)S^2/\sigma_0^2 \mid H_0^\sigma) = \chi^2(n-p, 0)$ with $(n-p) = 22$. The critical value for a test at level $\alpha = 0.05$ is $c_\alpha = 33.924$. Accordingly, the p-values for $S_\varnothing^2$ and S_ω^2 in testing $H_0^\sigma : \sigma^2 = 1.0$ against upper alternatives are $P((n-p)S_\varnothing^2 > 17.16630) = 0.7540$ and $P((n-p)S_\omega^2 > 37.31680) = 0.0218$ from Table 2, as evaluated from $\chi^2(22, 0)$. Consequently, evidence points towards a significant increase in variability in concert with the shift $\{\boldsymbol{Y}_\omega = \boldsymbol{Y}_\varnothing + \boldsymbol{\omega}\}$ with $\boldsymbol{\omega}' = [4, 4, 4, 0, \ldots, 0]'$. In reality, under Gaussian errors we have $\mathcal{L}((n-p)S_\omega^2/\sigma^2) = \chi^2(n-p, \lambda_2)$ with $\sigma^2\lambda_2 = \boldsymbol{\omega}_2'\boldsymbol{\omega}_2 = 28.7827$ from Theorem 2(iii), under which $P((n-p)S_\omega^2 > 37.31680) = 0.8620$.

5.2.3. Case 2.2

Our Case 2.1 analyses exploited the known structure of the Hadi and Simonoff [10] data in recovering $\boldsymbol{Y}_\varnothing = \boldsymbol{Y}_\omega - \boldsymbol{\omega}$. This enabled us to illustrate essentials of Theorems 1 and 2. Since such structure typically is unavailable, we turn next to essentials of Section 4.

To these ends, we continue as in Section 4 in the context of the (1993) study. Specifically, we take $\boldsymbol{X}_o = \boldsymbol{X} = [\boldsymbol{1}_n, \boldsymbol{X}_1, \boldsymbol{X}_2]$ as before, with $n = m = 25$; $p = 3$; $\boldsymbol{Y}_o = \boldsymbol{X}\boldsymbol{\beta}_o + \boldsymbol{\varepsilon}_o$ which is free of shifts; whereas $\boldsymbol{Y}_\omega = \boldsymbol{X}\boldsymbol{\beta}_\omega + \boldsymbol{\varepsilon}_\omega$ has its response vector shifted by $\boldsymbol{\omega} = [4, 4, 4, 0, \ldots, 0]'$ as before. In addition, the error vectors $\{\boldsymbol{\varepsilon}_o, \boldsymbol{\varepsilon}_\omega\}$ are $N(0, 1)$ deviates generated independently within and across samples in keeping with the tenets of Section 4. Values $\boldsymbol{Y}_o$ so generated are listed in column 1 of Table 3; similarly, values $\boldsymbol{Y}_\omega$ are listed in column 3 of Table 3; each is disturbed by $N(0, 1)$ deviates generated anew within and across samples. Moreover, $S_o^2 = \boldsymbol{e}_o'\boldsymbol{e}_o/(n-p)$ and $S_\omega^2 = \boldsymbol{e}_\omega'\boldsymbol{e}_\omega/(n-p)$; $\boldsymbol{H}_o = \boldsymbol{H}_n = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$; $\boldsymbol{\omega}_1 = \boldsymbol{H}_n\boldsymbol{\omega}$; and $\boldsymbol{\omega}_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$ are as reported in Table 3.

From Theorem 4 we carry forward $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o) \in \mathbb{R}^n$ and $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o) \in \mathbb{R}^n$ with values as reported in Table 3. Estimated parameters are determined to be

$$\widehat{\boldsymbol{\beta}}_o = \begin{bmatrix} -0.25596 \\ 1.00180 \\ 0.99046 \end{bmatrix}; \qquad \widehat{\boldsymbol{\beta}}_\omega = \begin{bmatrix} -0.05315 \\ 1.12669 \\ 0.93606 \end{bmatrix}; \qquad \begin{bmatrix} S_o^2 \\ S_\omega^2 \end{bmatrix} = \begin{bmatrix} 1.18595 \\ 2.67888 \end{bmatrix}.$$

The vectors $\{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2\}$ are orthogonal, with $\boldsymbol{\omega}_1'\boldsymbol{\omega}_1 = 19.2173$ and $\boldsymbol{\omega}_2'\boldsymbol{\omega}_2 = 28.7827$ such that $\boldsymbol{\omega}'\boldsymbol{\omega} = 48$. Similarly $\{\widehat{\boldsymbol{\omega}}_1, \widehat{\boldsymbol{\omega}}_2\}$ are orthogonal, with $\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1 = 20.0377$ and $\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2 = 44.8657$ such that $\widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\omega}} = 64.9034$. Table 3 records the response vectors $\boldsymbol{Y}_o$ and $\boldsymbol{Y}_\omega$; residual vectors $\boldsymbol{e}_o$ and $\boldsymbol{e}_\omega$; the shift vectors $\{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2\}$ and their estimates $\{\widehat{\boldsymbol{\omega}}_1, \widehat{\boldsymbol{\omega}}_2\}$. Computations utilized the software package MINITAB.

The distributions $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2\sigma^2) = \chi^2(p,\ \boldsymbol{\omega}_1'\boldsymbol{\omega}_1/2\sigma^2)$ and $\mathcal{L}(\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2\sigma^2) = \chi^2(n-p,\ \boldsymbol{\omega}_2'\boldsymbol{\omega}_2/2\sigma^2)$ are as in Theorem 4. For our case study with $\sigma^2 = 1.0$, the p-values are

$$P(\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2 > 20.0377/2) = 0.602 \quad \text{from } \mathcal{L}(\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2) = \chi^2(3, 19.2173/2)$$
$$P(\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2 > 44.8657/2) = 0.934 \quad \text{from } \mathcal{L}(\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2) = \chi^2(22, 28.7827/2)$$

showing that both quantities $\{\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1, \widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2\}$ are within the range of their respective 95% confidence intervals.

To test $H_0 : \boldsymbol{\omega} = \mathbf{0}$ against $H_1 : \boldsymbol{\omega} \neq \mathbf{0}$, we test both $H_{01} : \boldsymbol{\omega}_1 = \mathbf{0}$ and $H_{02} : \boldsymbol{\omega}_2 = \mathbf{0}$ against $H_{11} : \boldsymbol{\omega}_1 \neq \mathbf{0}$ and $H_{12} : \boldsymbol{\omega}_2 \neq \mathbf{0}$. One application in statistical process control is to assess whether a process has remained in control ($\boldsymbol{\omega} = \mathbf{0}$) or has shifted ($\boldsymbol{\omega} \neq \mathbf{0}$); and whether $\boldsymbol{\omega}_1 \neq \mathbf{0}$ has shifted the *OLS* solutions $\widehat{\boldsymbol{\beta}}$, or whether $\boldsymbol{\omega}_2 \neq \mathbf{0}$ has shifted the residuals and thus inflated the variation about the best-fitting line.

For testing $H_{01} : \boldsymbol{\omega}_1 = \mathbf{0}$, not assuming σ^2 to be known but estimated unbiasedly by S_o^2, Theorem 5(i) shows that $F = \widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2pS_o^2$ has under H_{01} the distribution $\mathcal{L}(F) = F(p, n-p, 0)$, with critical value $c_\alpha = 3.04912$ from $F(3, 22, 0)$ at $\alpha = 0.05$. With the values from our case study, we find $P(F > 20.0377/2(3)(1.18595) = 2.81598) = 0.06280$ as borderline evidence in favor of $H_{11} : \boldsymbol{\omega}_1 \neq \mathbf{0}$. Equivalently, as in the Case 2.1 study, we infer not only that responses have shifted in the $\boldsymbol{Y}_\omega$ data, but that these shifts are accompanied by standardized changes in the betas.

For the test $H_0 : \boldsymbol{\omega}_2 = \mathbf{0}$, Theorem 5(ii) gives $R^2 = \widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2\boldsymbol{e}_o'\boldsymbol{e}_o = \widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2(n-p)S_o^2$ as a ratio of dependent chi-squared variates with parameter $\rho^2 = 1/2$ and with degrees of freedom $(n-p, n-p)$. Its positive square root has the distribution $\mathcal{L}(R) = G_\rho(n-p, n-p, 0)$. Values from our case study give $R^2 = \widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2(n-p)S_o^2 = 44.8657/2(22)(1.18595) = 0.85980$ and $R = 0.92725$ with p-value $P(R > 0.92725) = 0.59776$ from $\mathcal{L}(R) = G_\rho(22, 22, 0)$, which in turn supports the null hypothesis $H_{02} : \boldsymbol{\omega}_2 = \mathbf{0}$.

The experiment was repeated 250 times to recover estimates for the expected values of $F = \widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2pS_o^2$ and $R = [\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2(n-p)S_o^2]^{\frac{1}{2}}$ as 2.680 and 0.7250, respectively, with associated p-values $P(F > 2.680) = 0.0718$ from $F(3, 22, 0)$, and $P(R > 0.7250) = 0.8520$ from $\mathcal{L}(R) = G_\rho(22, 22, 0)$, indicating borderline evidence in favor of $H_{11} : \boldsymbol{\omega}_1 \neq \mathbf{0}$, while supporting the null hypothesis $H_{02} : \boldsymbol{\omega}_2 = \mathbf{0}$.

5.2.4. Shifts: moment estimation

The Hadi and Simonoff data in Section 5.2 were shifted by $\boldsymbol{\omega} \neq \mathbf{0}$ units. In this section we show how to compute a moment estimator $\widetilde{\boldsymbol{\omega}}$ for the shift vector $\boldsymbol{\omega}$. To conform with the notation of Jensen and Ramirez [15] as in Appendix A.2, these follow on eliminating s rows $[\boldsymbol{Y}_I, \boldsymbol{Z}, \boldsymbol{\varepsilon}_I]$ from $[\boldsymbol{Y}_o, \boldsymbol{X}_o, \boldsymbol{\varepsilon}_o]$, leaving $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\}$ of full rank with $r = n - s > p$ rows, giving observed residuals $\boldsymbol{e}_o' = [\boldsymbol{e}', \boldsymbol{e}_I']$ and $(\widehat{\boldsymbol{\beta}}_I, S_I^2)$ from the reduced data. The case $s = 1$ has $\{\boldsymbol{Z} = \boldsymbol{z}_i', \widehat{\boldsymbol{\beta}}_I = \widehat{\boldsymbol{\beta}}_{(i)}, S_I^2 = S_{(i)}^2\}$. Taking $\{\boldsymbol{Y}_o \to \boldsymbol{Y}_o + \boldsymbol{\omega}\}$ with $\boldsymbol{\omega}' = [\boldsymbol{\gamma}', \boldsymbol{\delta}'] \in \mathbb{R}^n$ and $\boldsymbol{\gamma}$ fixed, Lemma A.3 of Jensen and Ramirez [15] under Assumption A gives

$$\widetilde{\boldsymbol{\delta}}(\boldsymbol{\gamma}) = (\boldsymbol{Y}_I - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}_I) + \boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\gamma} : \quad s > 1; \tag{5.2}$$
$$\widetilde{\delta}_i(\boldsymbol{\gamma}) = (Y_i - \boldsymbol{z}_i'\widehat{\boldsymbol{\beta}}_{(i)}) + \boldsymbol{z}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\gamma} : \quad s = 1; \tag{5.3}$$

as unbiased for $(\boldsymbol{\delta}, \delta_i)$ with $\boldsymbol{\gamma}$ fixed, having dispersion matrix $\mathrm{V}(\widetilde{\boldsymbol{\delta}}) = \sigma^2[\boldsymbol{I}_s + \boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}']$ not depending on $\boldsymbol{\gamma}$. A seminal tool in outlier detection is the ratio $\lambda = \boldsymbol{e}_o'\boldsymbol{e}_o/\boldsymbol{e}_I'\boldsymbol{e}_I$, with $\lambda >> 1$ supporting the conjecture that outlying shifts $\{\boldsymbol{Y}_I \to \boldsymbol{Y}_I + \boldsymbol{\delta}\}$ have occurred in the subset $\boldsymbol{Y}_I$.

The data of Hadi and Simonoff [10] are intended to hide the shifts in rows $\{1, 2, 3\}$. We next demonstrate that moment equations of type (5.3) provide good estimates for the hidden shifts. To these ends we analyze responses from Table 3, where $\boldsymbol{Y}_\omega$ by construction has shifts of 4.0 in rows $\{1, 2, 3\}$ in keeping with Hadi and Simonoff [10]. To illustrate the methodology, we suppose that strict experimental control may have lapsed during the first six runs and, accordingly, that the researcher is concerned about prospective shifts $\{Y_i + \delta_i; 1 \le i \le 6\}$, to be denoted as $\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6\}$, with no shifts having occurred subsequently. The statistic $\lambda = 58.94/18.82 = 3.13$ supports the claim that there are outlier shifts in $\boldsymbol{Y}_I$.

The moment equation for δ_1 is given by

$$\delta_1 = (Y_1 - \boldsymbol{z}_1'\widehat{\boldsymbol{\beta}}_{(1)}) + \boldsymbol{z}_1'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'[\delta_2, \delta_3, \delta_4, \delta_5, \delta_6, 0, \ldots, 0]'$$

where elements $[\delta_2, \delta_3, \delta_4, \delta_5, \delta_6, 0, \ldots, 0]'$ replace $\boldsymbol{\gamma}$, of order (24×1), in (5.3), with similar equations for $\{\delta_2, \ldots, \delta_6\}$. For $\boldsymbol{Y}_\omega$ the six consistent moment equations in the six unknowns are

$$\delta_1 = +4.0752 + 0.1550\delta_2 + 0.1542\delta_3 + 0.1574\delta_4 - 0.0293\delta_5 + 0.0479\delta_6$$
$$\delta_2 = +3.1204 + 0.1548\delta_1 + 0.1532\delta_3 + 0.1563\delta_4 - 0.0287\delta_5 + 0.0478\delta_6$$
$$\delta_3 = +2.2083 + 0.1538\delta_1 + 0.1530\delta_2 + 0.1553\delta_4 - 0.0281\delta_5 + 0.4772\delta_6$$

$$\delta_4 = -3.1216 + 0.1637\delta_1 + 0.1629\delta_2 + 0.1621\delta_3 - 0.0370\delta_5 + 0.1030\delta_6$$
$$\delta_5 = +1.1100 - 0.0277\delta_1 - 0.0272\delta_2 - 0.0267\delta_3 - 0.0336\delta_4 + 0.0357\delta_6$$
$$\delta_6 = -1.2916 + 0.0462\delta_1 + 0.0463\delta_2 + 0.0462\delta_3 + 0.0954\delta_4 + 0.0364\delta_5$$

with solutions $\widetilde{\delta}_1 = 4.99$, $\widetilde{\delta}_2 = 4.16$, $\widetilde{\delta}_3 = 3.37$, $\widetilde{\delta}_4 = -1.19$, $\widetilde{\delta}_5 = 0.78$, $\widetilde{\delta}_6 = -0.80$. The estimated standard deviation from the reduced model eliminating $\{Y_i;\ 1 \le i \le 6\}$ is $S_I = 1.0845$, so that standardized shifts are given by $\widetilde{\delta}_i/S_I$ with values $\{4.60,\ 3.84,\ 3.10,\ -1.10, 0.72,\ -0.74\}$, namely, the first three shifts near 4.0, and remaining shifts near 0.0. This provides evidence that the only essential shifts are in rows $\{1, 2, 3\}$. We next analyze these rows separately.

Given shifts only in rows $I = \{1, 2, 3\}$, Eq. (5.3) gives

$$\delta_1 = 4.0752 + 0.1550\delta_2 + 0.1542\delta_3$$
$$\delta_2 = 3.1204 + 0.1548\delta_1 + 0.1532\delta_3$$
$$\delta_3 = 2.2083 + 0.1538\delta_1 + 0.1530\delta_2$$

with solutions $\widetilde{\delta}_1 = 5.35$, $\widetilde{\delta}_2 = 4.52$, $\widetilde{\delta}_3 = 3.72$. The estimated standard deviation from the reduced model eliminating $\{Y_1, Y_2, Y_3\}$ is $S_I = 1.0443$. The standardized shifts are given by $\widetilde{\delta}_i/S_I$ with values $\{5.12, 4.33, 3.56\}$, empirically supporting the claim that the shifts in rows $\{1, 2, 3\}$ are around 4.0.

5.2.5. A caveat

For completeness it should be recorded that the analysis of Hadi and Simonoff [10] in regard to their Table 1 is strictly inadmissible. Specifically, they applied *OLS* when their first three observations are deterministic since devoid of random disturbances. This is not countenanced in Gauss–Markov theory; indeed, weighted regression would entail infinite weights; and otherwise the equations $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*\}$, with $\boldsymbol{\varepsilon}^* = [0, 0, 0, \boldsymbol{\varepsilon}']' \in \mathbb{R}^n$, are inconsistent. We have preempted this pitfall here on assigning random $N(0, 1)$ disturbances to all observations.

6. Conclusions

Traditional deletion diagnostics focus on shifts in one or a proper subset of elements of $\boldsymbol{Y}$. The present study allows a fixed but unknown vector $\boldsymbol{\omega}$ to perturb all elements of $\boldsymbol{Y}$. On technical grounds $\boldsymbol{\omega}$ is decomposed into orthogonal components, the "Regressor" component $\boldsymbol{\omega}_1$ accounting for shifts in the *OLS* solutions according to the rule $\widehat{\boldsymbol{\beta}}_\omega = \widehat{\boldsymbol{\beta}} + \boldsymbol{\kappa}$ with $\boldsymbol{\kappa} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\omega}_1$, and the "Error" component $\boldsymbol{\omega}_2$ accounting for inflated variation about the best-fitting line according to the rule $\boldsymbol{e}_\omega = \boldsymbol{e} + \boldsymbol{\omega}_2$. The distributions of $(\widehat{\boldsymbol{\beta}}_\omega, \boldsymbol{e}_\omega)$ are given in Theorem 2 under Gaussian errors, and anomalies in conventional tests regarding $(\boldsymbol{\beta}, \sigma^2)$ are given in Corollary 1.

Specifically, when $\boldsymbol{\omega} = \boldsymbol{\omega}_2 \ne \boldsymbol{0}$, the test for $H_0^\beta : \boldsymbol{\beta} = \boldsymbol{\beta}_o$ is conservative, akin to the concept of *Masking* in deletion diagnostics. In contrast, for the case $\boldsymbol{\omega} = \boldsymbol{\omega}_1 \ne \boldsymbol{0}$, the test is anti-conservative, akin to *Swamping* in deletion diagnostics. Under Gauss–Markov error moments, Theorem 4 shows that $\widehat{\boldsymbol{\omega}}_1 = \boldsymbol{X}(\widehat{\boldsymbol{\beta}}_\omega - \widehat{\boldsymbol{\beta}}_o)$ is unbiased for $\boldsymbol{\omega}_1 = \boldsymbol{H}_n\boldsymbol{\omega}$; for the case $n = m$, $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o)$ is unbiased for $\boldsymbol{\omega}_2 = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{\omega}$. Moreover, both $\widehat{\boldsymbol{\omega}}_1$ and $\widehat{\boldsymbol{\omega}}_2$ have normal distributions; for the case $\boldsymbol{X} = \boldsymbol{X}_o$, their dispersion matrices are $\mathrm{V}(\widehat{\boldsymbol{\omega}}_1) = 2\sigma^2\boldsymbol{H}_n$ and $\mathrm{V}(\widehat{\boldsymbol{\omega}}_2) = 2\sigma^2(\boldsymbol{I}_n - \boldsymbol{H}_n)$, respectively. The associated quadratic forms have noncentral chi-squared distributions, namely, $\mathcal{L}(\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2\sigma^2) = \chi^2(p,\ \boldsymbol{\omega}_1'\boldsymbol{\omega}_1/2\sigma^2)$ and $\mathcal{L}(\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2\sigma^2) = \chi^2(n-p,\ \boldsymbol{\omega}_2'\boldsymbol{\omega}_2/2\sigma^2)$.

Connections to other venues deserve note, to include deletion diagnostics as in Appendix A.2. On visualizing shifts instead as contaminants in robust regression, even here $\boldsymbol{\omega}$ would be limited in scope, as the most resistant algorithms would tolerate at most 50% nonzero contaminants.

Our case studies arise in monitoring a linear profile as in [20], together with a reexamination of data from Hadi and Simonoff [10]. To test that there are no shifts in the model, that is for $H_0 : \boldsymbol{\omega} = \boldsymbol{0}$, we have given the distributions under the null hypothesis for both $H_{01} : \boldsymbol{\omega}_1 = \boldsymbol{0}$ and $H_{02} : \boldsymbol{\omega}_2 = \boldsymbol{0}$. The critical value for the former is from a central F-distribution; for the latter its critical value is from the distribution of the ratio of correlated chi-squared variables as derived afresh in Appendix A.1 to follow. Moreover, moment estimators for shifts are given in the context of Hadi and Simonoff [10] based on single-case deletion diagnostics.

Appendix

A.1. Distribution: ratio of correlated variables

For the case that $\boldsymbol{X} = \boldsymbol{X}_o$ and $\boldsymbol{H}_n = \boldsymbol{H}_o$, recall that $\mathcal{L}(\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2)$ is noncentral and $\mathcal{L}(\boldsymbol{e}_o'\boldsymbol{e}_o)$ is central under Assumption A and A_o. Their ratio $\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\omega}}_2/2\boldsymbol{e}_o'\boldsymbol{e}_o$ would be scale-invariant to the unknown σ^2, as done in Theorem 5(i) in testing $H_0 : \boldsymbol{\omega}_1 = \boldsymbol{0}$ against $H_1 : \boldsymbol{\omega}_1 \ne \boldsymbol{0}$ at level α using $F = (\widehat{\boldsymbol{\omega}}_1'\widehat{\boldsymbol{\omega}}_1/2pS_o^2)$. However, $(\widehat{\boldsymbol{\omega}}_2, \boldsymbol{e}_o) = [(\boldsymbol{e}_\omega - \boldsymbol{e}_o), \boldsymbol{e}_o]$ are dependent; this in turn generates dependent quadratic forms having bivariate χ^2-distributions, as well as a nonstandard F-distribution of their ratio. These have been studied in the literature, but not for the case of singular joint and marginal distributions of $(\widehat{\boldsymbol{\omega}}_2, \boldsymbol{e}_o)$ as encountered here. For completeness we proceed to undertake the required modifications.

To these ends consider $\boldsymbol{Z}' = [\boldsymbol{e}'_\omega, \boldsymbol{e}'_o] \in \mathbb{R}^{2n}$ such that $\mathrm{V}(\boldsymbol{Z}) = \sigma^2 \mathrm{Diag}(\boldsymbol{A}, \boldsymbol{B})$. To adjust for scale let $\sqrt{2}\boldsymbol{R} = (\boldsymbol{e}_\omega - \boldsymbol{e}_o) = \widehat{\boldsymbol{\omega}}_2$ and $\boldsymbol{S} = \boldsymbol{e}_o$, and eventually $\mathrm{Q}(\boldsymbol{R}) = \widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2$ and $\mathrm{Q}(\boldsymbol{S}) = \boldsymbol{e}'_o\boldsymbol{e}_o$. Then their dispersion matrix is given by the following, together with its value when $\boldsymbol{A} = \boldsymbol{B} = (\boldsymbol{I}_n - \boldsymbol{H}_n)$, namely

$$\mathrm{V}\left(\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{S}\end{bmatrix}\right) = \begin{bmatrix} c^2(\boldsymbol{A}+\boldsymbol{B}) & -c\boldsymbol{B}\\ -c\boldsymbol{B} & \boldsymbol{B}\end{bmatrix} = \begin{bmatrix} (\boldsymbol{I}_n - \boldsymbol{H}_n) & -c(\boldsymbol{I}_n - \boldsymbol{H}_n)\\ -c(\boldsymbol{I}_n - \boldsymbol{H}_n) & (\boldsymbol{I}_n - \boldsymbol{H}_n)\end{bmatrix} \tag{A.1}$$

with $c^2 = 1/2$. Their joint distribution is singular since $(\boldsymbol{I}_n - \boldsymbol{H}_n)$ is idempotent of order $(n \times n)$ and rank $(n-p)$. Accordingly, its spectral decomposition is $(\boldsymbol{I}_n - \boldsymbol{H}_n) = \boldsymbol{QDQ}'$ with $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{I}_\nu, \boldsymbol{0})$, so let $\boldsymbol{T} = \boldsymbol{Q}'\boldsymbol{R}$ and $\boldsymbol{U} = \boldsymbol{Q}'\boldsymbol{S}$. Their singular joint dispersion matrix is

$$\mathrm{V}\left(\begin{bmatrix}\boldsymbol{T}\\ \boldsymbol{U}\end{bmatrix}\right) = \begin{bmatrix} \mathrm{Diag}(\boldsymbol{I}_\nu, \boldsymbol{0}) & -c\mathrm{Diag}(\boldsymbol{I}_\nu, \boldsymbol{0})\\ -c\mathrm{Diag}(\boldsymbol{I}_\nu, \boldsymbol{0}) & \mathrm{Diag}(\boldsymbol{I}_\nu, \boldsymbol{0})\end{bmatrix}. \tag{A.2}$$

To proceed, let $\psi(x; g) = x^{g-1}e^{-x}/\Gamma(g)$ and identify $\{L_h^{(g-1)}(x);\ h \in \{0, 1, 2, \ldots\}\}$ as the system of Laguerre polynomials of degree h and orthogonal with respect to $\psi(x; g)$. Take $\rho = 1/\sqrt{2}$ and $\nu^* = (n-p)/2$. Essentials are reported in the following.

Lemma A.1. *Let $\widehat{\boldsymbol{\omega}}_2 = (\boldsymbol{e}_\omega - \boldsymbol{e}_o)$ under Assumption A and A_0, for experiments having $\boldsymbol{X} = \boldsymbol{X}_o$ and $\boldsymbol{H}_n = \boldsymbol{H}_o$. Let $f(q_1, q_2)$ be the joint density of $Q_1 = \widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2$ and $Q_2 = \boldsymbol{e}'_o\boldsymbol{e}_o$, and identify $\mathcal{L}(\widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2\boldsymbol{e}'_o\boldsymbol{e}_o) = F_\rho(n-p, n-p, \lambda)$ as the distribution of their ratio.*

(i) *Then the joint central pdf with arguments (q_1, q_2), $\rho = 1/\sqrt{2}$ and $\nu^* = (n-p)/2$, is*

$$f(q_1, q_2) = \psi(q_1; \nu^*)\psi(q_2; \nu^*)\sum_{k=0}^{\infty} \rho^i L_k^{(\nu^*-1)}(q_1)L_k^{(\nu^*-1)}(q_2). \tag{A.3}$$

(ii) *The pdf of $F_\rho(n-p, n-p, 0)$, as the null distribution of $Z = (\widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2\boldsymbol{e}'_o\boldsymbol{e}_o)$, is given by*

$$f(z) = \frac{(1-\rho^2)^{\frac{\nu}{2}} z^{\frac{\nu}{2}-1}}{B\left(\frac{\nu}{2}, \frac{\nu}{2}\right)(1+z)^\nu}\left[1 - \frac{4\rho^2 z}{(1+z)^2}\right]^{-\frac{\nu+1}{2}} \tag{A.4}$$

with $\nu = n - p$ and $B(\cdot, \cdot)$ as the beta function.

(iii) *The pdf of $G_\rho(n-p, n-p, 0)$, as the null distribution of $W = (\widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2\boldsymbol{e}'_o\boldsymbol{e}_o)^{\frac{1}{2}}$, is given by*

$$g(w) = \frac{2(1-\rho^2)^{\frac{\nu}{2}} w^{\nu-1}}{B\left(\frac{\nu}{2}, \frac{\nu}{2}\right)(1+w^2)^\nu}\left(1 - \frac{4\rho^2 w^2}{(1+w^2)^2}\right)^{-\frac{\nu+1}{2}}. \tag{A.5}$$

Proof. Partition $\boldsymbol{T}' = [\boldsymbol{T}'_1, \boldsymbol{T}'_2]$ and $\boldsymbol{U}' = [\boldsymbol{U}'_1, \boldsymbol{U}'_2]$ in (A.2), with $(\boldsymbol{T}_1, \boldsymbol{U}_1) \in \mathbb{R}^{n-p}$. Their nonsingular joint dispersion matrix is $\mathrm{V}\left(\begin{bmatrix}\boldsymbol{T}_1\\ \boldsymbol{U}_1\end{bmatrix}\right) = \begin{bmatrix}1 & -c\\ -c & 1\end{bmatrix} \otimes \boldsymbol{I}_\nu$ with $\boldsymbol{A} \otimes \boldsymbol{B}$ as the Kronecker product. Clearly $\mathrm{Q}(\boldsymbol{R}) = \widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2 = \boldsymbol{T}'_1\boldsymbol{T}_1 + \boldsymbol{T}'_2\boldsymbol{T}_2$ and $\mathrm{Q}(\boldsymbol{S}) = \boldsymbol{e}'_o\boldsymbol{e}_o = \boldsymbol{U}'_1\boldsymbol{U}_1 + \boldsymbol{U}'_2\boldsymbol{U}_2$ but, since the distribution $\mathcal{L}(\boldsymbol{T}_2, \boldsymbol{U}_2)$ is degenerate at $(\boldsymbol{0}, \boldsymbol{0})$, stochastic properties of $(\widehat{\boldsymbol{\omega}}'_2\widehat{\boldsymbol{\omega}}_2/2, \boldsymbol{e}'_o\boldsymbol{e}_o)$ are determined completely by $Q_0(\boldsymbol{R}) = \boldsymbol{T}'_1\boldsymbol{T}_1$ and $Q_0(\boldsymbol{S}) = \boldsymbol{U}'_1\boldsymbol{U}_1$. Here $\mathcal{L}(\boldsymbol{T}_1, \boldsymbol{U}_1)$ is nonsingular such that elements of $\boldsymbol{T}_1$ are mutually independent, as are elements of $\boldsymbol{U}_1$, whereas their pair-wise correlation is $\rho = -1/\sqrt{2}$. In this setting the expansion (A.3) in Laguerre polynomials traces back to Kibble [19]; the expression (A.3) follows on specializing Eq. (3.14) of Jensen [12] with $\rho = 1/\sqrt{2}$ as the common Hotelling [11] canonical correlation between elements of $(\boldsymbol{T}_1, \boldsymbol{U}_1)$. Expression (A.4), reported as Eq. (13) in [17, p. 222] (correcting their beta function $B(\frac{1}{2}, \frac{\nu}{2})$), derives from this distribution. Expression (A.5) derives in turn from (A.4) through the change of variables $w^2 = z$, but was given earlier in [4,7] using different methods. □

A.2. Deletion diagnostics: a critique

Deletion diagnostics follow on eliminating one or a subset of rows from $\{\boldsymbol{Y}_o = \boldsymbol{X}_o\boldsymbol{\beta} + \boldsymbol{\varepsilon}_o\}$. The remaining data give solutions $(\widehat{\boldsymbol{\beta}}_{(i)}, S^2_{(i)})$ on deleting $[Y_i, \boldsymbol{x}'_i, \varepsilon_i]$, and $(\widehat{\boldsymbol{\beta}}_I, S^2_I)$ on deleting $[\boldsymbol{Y}_I, \boldsymbol{Z}, \boldsymbol{\varepsilon}_I]$ comprising s rows indexed by I. Of interest is either a single shift $\{Y_i \to Y_i + \delta\}$, or a vector shift $\{\boldsymbol{Y}_I \to \boldsymbol{Y}_I + \boldsymbol{\delta}\}$. The observed residuals $\boldsymbol{e}_o = (\boldsymbol{I}_n - \boldsymbol{H}_n)\boldsymbol{Y}_o$ are partitioned as $\boldsymbol{e}'_o = [\boldsymbol{e}', e_i]$ and $\boldsymbol{e}'_o = [\boldsymbol{e}', \boldsymbol{e}'_I]$ for these cases, respectively, where $\boldsymbol{H}_n = \boldsymbol{X}_o(\boldsymbol{X}'_o\boldsymbol{X}_o)^{-1}\boldsymbol{X}'_o$. Many deletion diagnostics, long deemed to be staples of regression, are studied in [3,6,2,1,24,5,22,8], and others. Designs fully estimable after deletions are studied in [9].

Remark 4. These shifts specialize those of Section 3.1, namely $\boldsymbol{\omega}' = [\boldsymbol{0}', \delta]$ for single-case, and $\boldsymbol{\omega}' = [\boldsymbol{0}', \boldsymbol{\delta}']$ for subset deletions. Connections to the present study follow.

Various *influence* diagnostics seek to track changes in the regression output owing to deleted observations. The quantity $\{DFB_{ij} = (\widehat{\beta}_j - \widehat{\beta}_{j(i)})/(S_{(i)}\sqrt{c_{jj}})\}$, also called *DFBETA*, is a scaled divergence between $(\widehat{\beta}_j, \widehat{\beta}_{j(i)})$ as elements of $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_{(i)})$ with and without Y_i, with c_{jj} from the diagonal of $(\boldsymbol{X}_o'\boldsymbol{X}_o)^{-1}$. Similarly $\{DFT_i = (\widehat{Y}_i - \widehat{Y}_{i(i)})/(S_{(i)}\sqrt{h_{ii}})\}$, known also as *DIFFIT*, is a scaled divergence between predictors at $\boldsymbol{x}_i$ with and without Y_i. Thus Y_i is deemed to be *influential* for *estimating* β_j, or *for predicting* at $\boldsymbol{x}_i'$, according as DFB_{ij} or DFT_i relate to designated cutoff values. See especially Belsley et al. [3]. Unfortunately, these concepts fail to grasp the actual changes, as demonstrated in the following.

Remark 5. In Case Study 1 of Section 5.1, the *asserted* difference in DFB_{ij} is $(\widehat{\beta}_j - \widehat{\beta}_{j(i)})$. Instead the *actual* difference induced by the shift $\{Y_1 \to Y_1 + 4\}$ is $(\widehat{\beta}_{\omega j} - \widehat{\beta}_j) = \kappa_j$ as elements of $(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} - \widehat{\boldsymbol{\beta}}) = \boldsymbol{\kappa}$ from Table 1. On deleting the outlying Y_1 we compute $\widehat{\boldsymbol{\beta}}_{(1)} = [12.0061, 2.4661]'$ so that $(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} - \widehat{\boldsymbol{\beta}}_{(1)}) = [1.8464, -1.1079]$ as components of $[DFB_{11}, DFB_{12}]$. These clearly miss the mark in excess, if intended to gage the change in $\widehat{\boldsymbol{\beta}}$ owing to deleting the shifted $\{Y_1 \to Y_1 + 4\}$. The *actual* difference in $\widehat{\boldsymbol{\beta}}$ induced by $\{Y_1 \to Y_1 + 4\}$ is $\boldsymbol{\kappa} = [1.0, -0.6]'$.

Remark 6. Similarly, the *asserted* difference in DFT_1 is $(\widehat{Y}_1 - \widehat{Y}_{1(1)}) = \boldsymbol{x}_1'(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}} - \widehat{\boldsymbol{\beta}}_{(1)}) = 12.5269$ in Case Study 1, with $\boldsymbol{x}_1' = [13, -3]$. This again misses the mark if intended to gage the change in the predictor at $\boldsymbol{x}_1$ owing to $\{Y_1 \to Y_1 + 4\}$. Instead the *actual* difference induced by $\{Y_1 \to Y_1 + 4\}$ is $\boldsymbol{x}_1'\boldsymbol{\kappa} = 14.8000$ from Table 1.

References

[1] A.C. Atkinson, Plots, Transformations, and Regression, Oxford University Press, Oxford, UK, 1985.
[2] V. Barnett, T. Lewis, Outliers in Statistical Data, third ed., Wiley, New York, 1994.
[3] D.A. Belsley, E. Kuh, R.E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, Wiley, New York, 1980.
[4] S.S. Bose, P.C. Mahalanobis, On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population, Sankhyā 2 (1935) 65–72.
[5] S. Chatterjee, A.S. Hadi, Sensitivity Analysis in Linear Regression, Wiley, New York, 1988.
[6] R.D. Cook, S. Weisberg, Residuals and Influence in Regression, Chapman and Hall, London, 1982.
[7] D.J. Finney, The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population, Biometrika 30 (1938) 190–192.
[8] J. Fox, Regression Diagnostics, Sage, Newbury Park, CA, 1991.
[9] S. Ghosh, On robustness of designs against incomplete data, Sankhyā B 40 (1978) 204–208.
[10] A.S. Hadi, J.S. Simonoff, Procedures for the identification of multiple outliers in linear models, J. Amer. Statist. Assoc. 88 (1993) 1264–1272.
[11] H. Hotelling, Relations between two sets of variates, Biometrika 28 (1936) 321–377.
[12] D.R. Jensen, The joint distribution of quadratic forms and related distributions, Austral. J. Statist. 12 (1970) 13–22.
[13] D.R. Jensen, The use of studentized diagnostics in regression, Metrika 52 (2000) 213–223.
[14] D.R. Jensen, Y.V. Hui, P.M. Ghare, Monitoring an input–output model for production. I. The control charts, Manage. Sci. 30 (1984) 1197–1206.
[15] D.R. Jensen, D.E. Ramirez, Noncentralities induced in regression diagnostics, J. Stat. Theory Pract. 8 (2014) 141–165.
[16] N.L. Johnson, S. Kotz, Distributions in Statistics: Continuous Univariate Distributions-2, Houghton Mifflin, Boston, 1970.
[17] N.L. Johnson, S. Kotz, Distributions in Statistics: Continuous Multivariate Distributions, Wiley, New York, 1972.
[18] L. Kang, S.L. Albin, On-line monitoring when the process yields a linear profile, J. Qual. Technol. 32 (2000) 418–426.
[19] W.F. Kibble, A two-variate gamma type distribution, Sankhyā 5 (1941) 137–150.
[20] K. Kim, M.A. Mahmoud, W.H. Woodall, On the monitoring of linear profiles, J. Qual. Technol. 35 (2003) 317–328.
[21] A.M. Mathai, S.B. Provost, Quadratic Forms in Random Variables, Marcel Dekker, New York, 1992.
[22] R.H. Myers, Classical and Modern Regression with Applications, second ed., PWS-KENT, Boston, MA, 1990.
[23] C.R. Rao, S.K. Mitra, Generalized Inverse of Matrices and its Applications, Wiley, New York, 1971.
[24] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.