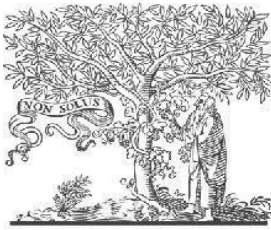


COPY RIGHT



ELSEVIER
SSRN

2026 IJEMR. Personal use of this material is permitted. Permission from IJEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper; all copy right is authenticated to Paper Authors

IJEMR Transactions, online available on 7th April 2026. Link

<https://ijiemr.org/downloads.php?vol=Volume-15&issue=issue04>

DOI: 10.48047/IJEMR/V15/ISSUE 04/19

Title New Diagnostic Measures for Identifying Influential Observations in Linear Statistical Models

Volume 15, Issue 04, Pages: 234 – 253

Paper Authors

Shaik Hussan Saheb, B Sarojamma



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per **UGC Guidelines** We Are Providing A Electronic Bar code

New Diagnostic Measures for Identifying Influential Observations in Linear Statistical Models

Shaik Hussan Saheb¹, B Sarojamma²

¹Research scholar, Department of Statistics, S.V University, Tirupati.

²Professor and HoD, Department of Statistics, S.V University, Tirupati.

Corresponding Author : saroja14397@gmail.com

Abstract

The reliability of a fitted linear regression model depends on more than the overall goodness of fit. A single observation, or a small group of observations, can exert a disproportionate effect on the estimated coefficients, the fitted values, the residual variance and the conclusions drawn from hypothesis tests. Classical case-deletion diagnostics such as Cook's distance, DFFITS, DFBETAS and the covariance ratio remain the standard tools for detecting such observations, yet each of them is built from only two quantities, a residual and a leverage value, and each is interpreted against a fixed rule-of-thumb cut-off that ignores the sample at hand. This paper examines the classical diagnostics from a unified algebraic perspective, shows that they differ mainly in how they weight the same two ingredients, and identifies the limitations that follow from this structure. A new diagnostic statistic, the Standardized Composite Influence Index, is then developed. The index combines four separately standardized components, describing residual outlyingness, leverage potential, sensitivity of the fitted values and sensitivity of the coefficient vector, through a robust location-scale standardization computed from the sample itself. The coefficient component introduces a direction-dependent term that is absent from Cook's distance and DFFITS, and the robust standardization makes the index comparable across data sets and less susceptible to masking. Algebraic relationships between the new index and the existing measures are derived, invariance properties are established, an inequality linking the coefficient and fitted-value components is proved, and threshold rules together with a step-by-step diagnostic algorithm are proposed. Numerical evaluation is left for future work.

Keywords: Regression diagnostics; Influential observations; Outliers; Leverage; Cook's distance; DFFITS; DFBETAS; Masking.

Introduction

Ordinary least squares is the most widely used estimation method in applied regression analysis, and it is also one of the most sensitive to individual data points. Because every observation enters the normal equations with equal weight, a single point located far from the remaining data, either in the space of the predictors or in the direction of the response, can pull the fitted hyperplane towards itself and thereby distort the coefficient estimates, the fitted values, the residual variance and the associated inferences. Observations of this kind are called influential. The detection of influential observations is therefore an indispensable part of any careful regression analysis, and it has been studied intensively since the work of Cook (1977) and Belsley, Kuh and Welsch (1980).

Three ideas are central to the subject and must be kept distinct. An observation is an outlier in the response direction when its residual is unusually large relative to the residuals of the other cases. An observation is a high-leverage point when its vector of predictor values lies far from the centre of the predictor cloud. An observation is influential when its removal from the sample produces a substantial change in some feature of the fitted model. The three properties are related, but none of them implies the others. A response outlier with ordinary leverage may change the residual variance without changing the coefficients appreciably; a high-leverage point that lies close to the regression surface fitted to the other cases, sometimes called a good leverage point, may improve the precision of the estimates rather than damage

them; and a high-leverage point with a large deviation from that surface, a bad leverage point, typically changes everything at once. Chatterjee and Hadi (1986) gave a systematic account of these distinctions and of the many statistics proposed to measure them.

The classical case-deletion diagnostics were designed to quantify influence one observation at a time. Cook's distance measures the change in the whole coefficient vector in the metric of its estimated covariance matrix; DFFITS measures the change in the fitted value of the deleted case; DFBETAS measures the change in each individual coefficient; and the covariance ratio measures the change in the generalized variance of the estimates. Despite their different definitions, all of these statistics are exact functions of just two quantities for each case, a suitably scaled residual and the diagonal element of the hat matrix. This algebraic fact, which is derived in Section 3, explains why the measures usually agree and also why they share the same weaknesses. Each of them is judged against a cut-off value that was proposed as a convenient rule of thumb rather than derived from a sampling distribution, each of them can be defeated by groups of similar observations that mask one another, and none of them provides a single interpretable index that reflects the several directions in which an observation can be influential.

The purpose of this paper is to develop a new diagnostic measure that responds to these limitations while remaining simple to compute and to interpret. The measure, called the Standardized Composite Influence Index, is constructed in three stages. First, four

component diagnostics are chosen so that together they describe residual outlyingness, leverage, sensitivity of the fitted values and sensitivity of the coefficient vector. Second, each component is standardized using a robust location and scale computed from the sample of n component values, so that a component is measured in units that reflect the variability of that diagnostic in the data at hand rather than a fixed constant. Third, the standardized components are combined through a weighted Euclidean norm. The resulting index has a clear geometric interpretation as the distance of an observation from the bulk of the data in a standardized diagnostic space.

The remainder of the paper is organized as follows. Section 2 reviews the literature on regression diagnostics, influence functions, robust estimation and multiple-outlier detection. Section 3 sets out the notation, the linear model, the deletion identities and the classical measures, and formalizes the research gap. Section 4 develops the proposed index, derives its relationship to the classical measures, establishes its theoretical properties, discusses threshold selection and presents the diagnostic algorithm. Section 5 concludes. No simulation or real-data analysis is reported; the contribution of this paper is methodological, and empirical evaluation is identified as the natural next step.

2. Review of Literature

2.1 Residuals and the hat matrix

The modern theory of regression diagnostics rests on the projection matrix that maps the observed response vector onto the fitted

values. Hoaglin and Welsch (1978) gave the first comprehensive treatment of this matrix, which they named the hat matrix, and showed that its diagonal elements measure the distance of each case from the centroid of the predictor space and bound the extent to which a case can determine its own fitted value. They proposed the rule that a diagonal element larger than twice its average value indicates a high-leverage case, and Velleman and Welsch (1981) later suggested a stricter multiple of three for situations in which the number of predictors is small relative to the sample size. The same authors described computationally efficient methods for obtaining leverage values and deletion statistics from a single fit without refitting the model.

Residuals had been used for model checking long before the hat matrix was introduced, but the recognition that raw residuals have unequal variances led to the definition of studentized residuals. Two versions are now standard. The internally studentized residual divides each raw residual by an estimate of its standard deviation computed from the full sample. The externally studentized residual uses an estimate of the error variance obtained after deleting the case in question; Belsley, Kuh and Welsch (1980) called this quantity RSTUDENT and showed that it follows a Student t distribution under the null hypothesis that the case is not an outlier, which provides an exact test. Atkinson (1981, 1985) gave a detailed treatment of the graphical use of these residuals, including half-normal plots with simulation envelopes. The prediction residual, obtained by

predicting a case from a fit that excludes it, was introduced by Allen (1974) in connection with variable selection and later became a standard ingredient of deletion diagnostics.

2.2 Case-deletion measures of influence

The systematic study of influence began with Cook (1977), who proposed measuring the effect of deleting a case by the distance between the full-sample coefficient vector and the deleted-sample coefficient vector in the metric defined by the estimated covariance matrix of the coefficients. Cook (1979) extended the idea and related the statistic to confidence ellipsoids, suggesting that its value be compared with percentage points of the F distribution as a rough guide to interpretation. Cook and Weisberg (1980) showed that the statistic can be viewed as a sample version of the influence curve of Hampel (1974), and their monograph (Cook and Weisberg, 1982) placed residuals and influence within a single coherent framework.

Belsley, Kuh and Welsch (1980) developed a parallel set of tools, building on the technical report of Welsch and Kuh (1977). Their DFFITS statistic measures the scaled change in the fitted value of a case when that case is deleted; their DFBETAS statistics measure the scaled change in each regression coefficient; and their covariance ratio measures the change in the determinant of the estimated covariance matrix of the coefficients. They proposed size-adjusted cut-offs for each statistic, chosen so that the cut-off declines as the sample size grows, and were careful to describe these as approximations rather than critical values.

Andrews and Pregibon (1978) approached the problem through the determinant of the augmented cross-product matrix, obtaining a statistic that factorizes into a leverage part and a residual part. Draper and John (1981) compared several of these proposals and pointed out that they can disagree on which observations matter, and Pregibon (1981) extended the deletion approach to logistic regression.

Chatterjee and Hadi (1986) surveyed the rapidly growing collection of influence measures and showed that many of them can be written as quadratic forms in the change of the coefficient vector, differing only in the weight matrix and the scaling used. Their monograph (Chatterjee and Hadi, 1988) developed this unification in detail and argued that a small number of measures, supplemented by graphical displays, suffices in practice. Atkinson (1982, 1985) proposed a modification of Cook's distance that uses the externally studentized residual and a rescaling intended to place the values on a scale comparable to that of the residuals. Hadi (1992a) introduced a measure of overall potential influence that adds a leverage term to a term based on the normalized residual and possesses attractive invariance properties, and recommended a two-dimensional plot of the two terms in preference to a single cut-off. Reviews of these developments were given by Hocking (1983) in a broad survey of linear regression methodology and by Ullah and Pasha (2009).

Two further lines of development concern the form of the perturbation. Cook (1986) introduced the local influence approach, in which the model is perturbed continuously

rather than by deletion, and the normal curvature of the likelihood displacement is examined in the direction of maximum change. This approach recovers the deletion diagnostics as a special case of case-weight perturbation and extends naturally to perturbations of the response, the predictors and the variance. Peña (2005) reversed the usual viewpoint by asking how the forecast of a given case changes when each of the other cases is deleted, and showed that the resulting sensitivity statistic can reveal groups of similar high-leverage outliers that Cook's distance fails to detect.

2.3 Influence functions and robust estimation

The influence curve of Hampel (1974) provides the theoretical foundation for case-deletion diagnostics. It measures the effect on an estimator of an infinitesimal contamination at a single point and thus describes local robustness. Huber (1981) and Hampel, Ronchetti, Rousseeuw and Stahel (1986) developed the theory of robust estimation on this basis, and Maronna, Martin and Yohai (2006) gave a modern account. A different aspect of robustness, the breakdown point, concerns the largest fraction of contamination an estimator can tolerate before its value becomes arbitrary. Rousseeuw (1984) introduced least median of squares regression, which attains the maximal breakdown point of one half, and Rousseeuw and Leroy (1987) described the least trimmed squares estimator and its use in outlier detection. Yohai (1987) proposed MM-estimators, which combine a high breakdown point with high efficiency under normal errors.

Because the hat matrix is computed from the very data that may be contaminated, leverage values can fail to reveal groups of remote points. Rousseeuw and van Zomeren (1990) proposed replacing the classical Mahalanobis distance by a robust distance based on the minimum volume ellipsoid estimator of location and scatter, and introduced the diagnostic plot of robust residuals against robust distances that separates vertical outliers, good leverage points and bad leverage points. Hubert, Rousseeuw and Van Aelst (2008) reviewed high-breakdown multivariate methods, including the minimum covariance determinant estimator, and Rousseeuw and Hubert (2011) summarized the use of robust statistics for outlier detection. The earlier multivariate outlier literature, including Wilks (1963) and the distance of Mahalanobis (1936), remains relevant for the assessment of leverage.

2.4 Multiple outliers, masking and swamping

Single-case deletion diagnostics examine each observation on the assumption that the remaining observations are clean. When several unusual cases are present this assumption fails, and two related phenomena arise. Masking occurs when the presence of one unusual case hides another, so that neither is detected when cases are examined one at a time. Swamping occurs when unusual cases distort the fit sufficiently to make ordinary cases appear unusual. Hawkins (1980) and Barnett and Lewis (1994) discussed both phenomena in the general outlier literature, and Davies and Gather (1993) gave a formal treatment of the

identification of multiple outliers. Lawrance (1995) analysed the algebra of deletion influence under masking and showed how conditional and joint deletion effects differ. Beckman and Cook (1983) gave an extensive survey of the outlier problem, with discussion.

Several strategies for overcoming masking were proposed before 2020. Hadi (1992b) and Hadi and Simonoff (1993) developed forward procedures that start from a small clean subset and add observations sequentially, testing each candidate against the current fit. Atkinson (1994) combined a fast robust starting point with a forward search, and Atkinson and Riani (2000) developed the forward search into a general framework for robust diagnostic regression analysis. Peña and Yohai (1995) introduced an influence matrix whose eigenvectors reveal influential subsets, and Peña and Yohai (1999) gave a fast procedure suitable for large problems. Billor, Hadi and Velleman (2000) proposed the BACON algorithm for nominating outliers in large multivariate data sets. Imon (2005) proposed generalized versions of the studentized residual and DFFITS based on group deletion, Habshah, Norazan and Imon (2009) developed the diagnostic-robust generalized potential for identifying multiple high-leverage points, and Nurunnabi, Hadi and Imon (2014) and Nurunnabi, Nasser and Imon (2016) extended group-deletion diagnostics to the identification and classification of multiple outliers, high-leverage points and influential observations. Resampling ideas from the jackknife and bootstrap literature (Efron, 1979; Wu, 1986) provide a further route to data-dependent

thresholds, and the false discovery rate of Benjamini and Hochberg (1995) offers a principled way to control the number of false identifications when many cases are tested simultaneously.

2.5 Textbook treatments and applied practice

The diagnostics reviewed above have entered standard textbooks on regression, including Weisberg (2005), Chatterjee and Price (1991), Montgomery, Peck and Vining (2012) and the introductory monograph of Fox (1991), and are implemented in all major statistical software systems. In applied work the conventional practice is to compute several diagnostics, compare each with its rule-of-thumb cut-off, and examine graphical displays of residuals against leverage. This practice is serviceable when there is at most one unusual case, but it leaves the analyst to reconcile conflicting signals from different measures and offers no formal guidance when the cut-offs are inappropriate for the sample size and dimension at hand.

2.6 Summary of the review

Three conclusions emerge from the literature. First, the classical single-case measures are variations on a common theme: each is a function of a residual and a leverage value, and the choice among them is largely a choice of weighting. Second, the cut-offs in common use are approximate rules that do not adapt to the sample and can be seriously misleading under masking. Third, the methods that do address masking, such as the forward search and group-deletion diagnostics, are procedures rather than single statistics, and they require choices of initial

subsets or group sizes. There is therefore room for a single, interpretable, self-calibrating index that combines the complementary information in the classical measures while borrowing the robustness of sample-based standardization. The construction of such an index is the subject of Section 4.

3. Methodology

3.1 The linear model and notation

Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{y}$ is an $n \times 1$ vector of responses, $\mathbf{X}$ is an $n \times p$ design matrix of full column rank whose first column is a column of ones, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$. Throughout, p counts the intercept, so that all cut-offs quoted below are consistent with this convention. The i th row of $\mathbf{X}$ is written $\mathbf{x}_i'$.

The ordinary least squares estimator and the associated quantities are

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, & \hat{\mathbf{y}} &= \mathbf{H}\mathbf{y}, & \mathbf{e} \\ &= (\mathbf{I}_n - \mathbf{H})\mathbf{y}, & & & (2) \end{aligned}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the hat matrix. Its diagonal elements $h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$ are the leverage values. Because $\mathbf{H}$ is symmetric and idempotent,

$$\frac{1}{n} \leq h_{ii} \leq 1, \quad \sum_{i=1}^n h_{ii} = \text{tr}(\mathbf{H}) = p, \quad \sum_{j=1}^n h_{ij}^2 = h_{ii}. \quad (3)$$

The residual mean square is $s^2 = \mathbf{e}'\mathbf{e}/(n - p)$. A subscript (i) in parentheses denotes a quantity computed after deleting the i th case;

thus $\hat{\boldsymbol{\beta}}_{(i)}$ is the coefficient vector estimated from the remaining $n - 1$ cases and $s_{(i)}^2$ is the corresponding residual mean square. It is convenient to write $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$, with elements c_{jk} and j th row $\mathbf{c}_j'$.

3.2 Deletion identities

All single-case diagnostics can be computed from a single fit through the following identities, which follow from the Sherman–Morrison formula (Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982):

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)} = \frac{\mathbf{C}\mathbf{x}_i\mathbf{e}_i}{1 - h_{ii}}, \quad (4)$$

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)} = \mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}) = \frac{\mathbf{h}_i\mathbf{e}_i}{1 - h_{ii}}, \quad (5)$$

$$\begin{aligned} &(n - p - 1) s_{(i)}^2 \\ &= (n - p) s^2 \\ &\quad - \frac{e_i^2}{1 - h_{ii}}, \end{aligned} \quad (6)$$

where $\mathbf{h}_i$ is the i th column of $\mathbf{H}$ and $\hat{\mathbf{y}}_{(i)}$ denotes the vector of fitted values at all n design points computed from $\hat{\boldsymbol{\beta}}_{(i)}$. Identity (5) shows that the change in the fitted value at the deleted point itself is $h_{ii}\mathbf{e}_i/(1 - h_{ii})$, and identity (6) shows that the deleted residual mean square depends on the full-sample residual only through $e_i^2/(1 - h_{ii})$.

3.3 Studentized residuals

The internally and externally studentized residuals are

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}}, \quad t_i = \frac{e_i}{s_{(i)}\sqrt{1 - h_{ii}}}, \quad (7)$$

and by (6) they are related through

$$t_i = r_i \sqrt{\frac{n-p-1}{n-p-r_i^2}}. \quad (8)$$

Under normal errors and the null hypothesis that case i follows the model, t_i has a Student t distribution with $n-p-1$ degrees of freedom. The relationship (8) is monotone, so the two residuals order the cases identically; the external version is preferred for diagnostics because a single large residual does not inflate its own scale.

3.4 The classical influence measures

Cook's distance (Cook, 1977) is

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' \mathbf{X}' \mathbf{X} (\hat{\beta} - \hat{\beta}_{(i)})}{p s^2} \\ = \frac{r_i^2}{p} \cdot \frac{h_{ii}}{1 - h_{ii}}, \quad (9)$$

where the second form follows by substituting (4) and using $\mathbf{x}_i' \mathbf{C} \mathbf{x}_i = h_{ii}$. DFFITS (Belsley, Kuh and Welsch, 1980) is the scaled change in the fitted value of the deleted case,

$$\text{DFFITS}_i = \frac{\hat{y}_i - \hat{y}_{i(i)}}{s_{(i)} \sqrt{h_{ii}}} = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}}, \quad (10)$$

and DFBETAS is the scaled change in the j th coefficient,

$$\text{DFBETAS}_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{s_{(i)} \sqrt{c_{jj}}} = \frac{\mathbf{c}_j' \mathbf{x}_i e_i}{(1 - h_{ii}) s_{(i)} \sqrt{c_{jj}}}, \quad j \\ = 1, \dots, p. \quad (11)$$

The covariance ratio compares the generalized variances of the coefficient estimates with and without case i ,

$$\text{COVRATIO}_i = \frac{\det(s_{(i)}^2 \mathbf{C}_{(i)})}{\det(s^2 \mathbf{C})} \\ = \frac{1}{(1 - h_{ii})} \left(\frac{n-p}{n-p-1+t_i^2} \right)^p, \quad (12)$$

and the modified Cook statistic of Atkinson (1982) is

$$C_i = |t_i| \sqrt{\frac{n-p}{p} \cdot \frac{h_{ii}}{1 - h_{ii}}}. \quad (13)$$

Hadi's (1992a) measure of overall potential influence is the sum of a leverage term and a residual term,

$$H_i = \frac{h_{ii}}{1 - h_{ii}} + \frac{p}{1 - h_{ii}} \cdot \frac{d_i^2}{1 - d_i^2}, \quad d_i \\ = \frac{e_i}{\sqrt{\mathbf{e}' \mathbf{e}}}. \quad (14)$$

The commonly cited cut-offs are $h_{ii} > 2p/n$ or $3p/n$ for leverage, $|t_i| > 2$ for residuals, $D_i > 4/n$ or $D_i > 1$ for Cook's distance, $|\text{DFFITS}_i| > 2\sqrt{p/n}$, $|\text{DFBETAS}_{ij}| > 2/\sqrt{n}$ and $|\text{COVRATIO}_i - 1| > 3p/n$ (Belsley, Kuh and Welsch, 1980; Cook and Weisberg, 1982).

3.5 The common structure of the classical measures

Comparison of (9), (10), (12) and (13) shows that Cook's distance, DFFITS, the covariance ratio and the modified Cook statistic are all functions of the pair (t_i, h_{ii}) only; indeed, using (8) and (6),

$$\text{DFFITS}_i^2 = \frac{s^2}{s_{(i)}^2} p D_i, \quad (15)$$

so that DFFITS and Cook's distance differ only through the ratio of the two variance estimates. The leverage value carries all the information about the position of the case in

predictor space that these measures use; the direction of $\mathbf{x}_i$ relative to the coordinate axes plays no role. The DFBETAS statistics in (11) are the exception: they depend on the direction of $\mathbf{x}_i$ through the products $\mathbf{c}'_j \mathbf{x}_i$, but they form a vector of p values per case rather than a single index, and in practice they are inspected coefficient by coefficient.

3.6 Research gap

The review in Section 2 and the algebra in Section 3.5 lead to the following statement of the gap. Existing single-case measures reduce to weightings of a residual and a leverage value; they discard the direction of the case in coefficient space; they are compared with fixed constants whose adequacy depends on n , p and the error distribution; and their values are not comparable from one data set to another. A diagnostic statistic is required that (i) incorporates the direction-dependent information contained in the coefficient changes, (ii) is standardized with respect to the sample itself so that it is comparable across data sets and less vulnerable to masking, (iii) retains a transparent connection to the classical measures, and (iv) admits a clear geometric interpretation and an explicit decision rule. The next section develops such a statistic.

4. Main Research: The Standardized Composite Influence Index

4.1 Component diagnostics

Four nonnegative components are defined for each case. The residual component measures outlyingness in the response direction,

$$A_i = t_i^2. \quad (16)$$

The leverage component is the potential of Hadi (1992a),

$$L_i = \frac{h_{ii}}{1 - h_{ii}}, \quad (17)$$

which is a monotone transformation of the leverage that increases without bound as h_{ii} approaches one. The fitted-value component measures the sensitivity of the fitted values to the deletion of case i . By (5) and (3), the squared norm of the change in the entire fitted vector is

$$\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)}\|^2 = \frac{\mathbf{h}'_i \mathbf{h}_i e_i^2}{(1 - h_{ii})^2} = \frac{h_{ii} e_i^2}{(1 - h_{ii})^2}, \quad (18)$$

and dividing by $s_{(i)}^2$ gives

$$F_i = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(i)}\|^2}{s_{(i)}^2} = t_i^2 \frac{h_{ii}}{1 - h_{ii}} = \text{DFFITS}_i^2. \quad (19)$$

Thus the squared DFFITS statistic, usually described as the change in a single fitted value, is also the scaled change in the whole fitted vector, and by (15) it equals $p D_i$ up to the factor $s^2/s_{(i)}^2$.

The coefficient component is the mean of the squared DFBETAS values,

$$B_i = \frac{1}{p} \sum_{j=1}^p \text{DFBETAS}_{ij}^2. \quad (20)$$

Substituting (11) yields the closed form

$$\begin{aligned} B_i &= \frac{t_i^2}{1 - h_{ii}} \cdot \frac{g_i}{p}, \quad g_i = \sum_{j=1}^p \frac{(\mathbf{c}'_j \mathbf{x}_i)^2}{c_{jj}} \\ &= \mathbf{x}'_i \mathbf{C} \mathbf{D}^{-1} \mathbf{C} \mathbf{x}_i, \end{aligned} \quad (21)$$

where $\mathbf{D} = \text{diag}(c_{11}, \dots, c_{pp})$. The quantity g_i is a coordinate-weighted analogue of the leverage $h_{ii} = \mathbf{x}'_i \mathbf{C} \mathbf{x}_i$; the two coincide in weighting only when $\mathbf{C}$ is diagonal, and in

general g_i depends on the direction of $\mathbf{x}_i$ relative to the coefficient axes. It is this quantity that introduces information not contained in the pair (t_i, h_{ii}) .

4.2 Two properties of the coefficient component

Proposition 1. For every i , $g_i \leq p h_{ii}$, and consequently $B_i \leq F_i$.

Proof. Write $\mathbf{c}_j = \mathbf{C}\mathbf{u}_j$, where $\mathbf{u}_j$ is the j th unit vector, so that $\mathbf{c}_j'\mathbf{x}_i = \mathbf{u}_j'\mathbf{C}\mathbf{x}_i$. Since $\mathbf{C}$ is symmetric positive definite, the Cauchy–Schwarz inequality in the inner product $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}'\mathbf{C}\mathbf{b}$ gives

$$\begin{aligned} (\mathbf{u}_j'\mathbf{C}\mathbf{x}_i)^2 &\leq (\mathbf{u}_j'\mathbf{C}\mathbf{u}_j)(\mathbf{x}_i'\mathbf{C}\mathbf{x}_i) \\ &= c_{jj} h_{ii}. \end{aligned} \quad (22)$$

Dividing by c_{jj} and summing over j gives $g_i \leq p h_{ii}$. Substituting into (21) and comparing with (19) gives $B_i \leq t_i^2 h_{ii} / (1 - h_{ii}) = F_i$. ◻

Proposition 2. $\sum_{i=1}^n g_i = p$, so that the average of g_i over the sample equals the average leverage p/n .

Proof. For fixed j , $\sum_{i=1}^n (\mathbf{c}_j'\mathbf{x}_i)^2 = \mathbf{c}_j'\mathbf{X}'\mathbf{X}\mathbf{c}_j = \mathbf{u}_j'\mathbf{C}\mathbf{X}'\mathbf{X}\mathbf{C}\mathbf{u}_j = \mathbf{u}_j'\mathbf{C}\mathbf{u}_j = c_{jj}$.

Dividing by c_{jj} and summing over j gives the result. ◻

Proposition 1 shows that the mean squared coefficient change can never exceed the squared DFFITS statistic, with equality only when $\mathbf{C}\mathbf{x}_i$ is proportional to every unit vector simultaneously, which is impossible for $p > 1$ unless the design is orthogonal. The ratio

$$\rho_i = \frac{B_i}{F_i} = \frac{g_i}{p h_{ii}} \in (0,1] \quad (23)$$

is therefore a natural measure of how the influence of case i is distributed across the coefficients: values near one indicate that the case moves all coefficients in proportion to their standard errors, whereas values near zero indicate that a large fitted-value change is produced by a change concentrated in a few coefficients, or by compensating changes among correlated coefficients. Proposition 2 shows that g_i shares the average of the leverage values, so that the same multiple-of-average rules used for h_{ii} can be applied to g_i if a separate screening of coefficient-space leverage is desired.

4.3 Robust standardization of the components

Let c_{ki} , $k = 1, 2, 3, 4$, denote the four component values A_i, L_i, F_i and B_i for case i . Each component is nonnegative and, under a clean sample, right-skewed; its magnitude also depends on n and p . Rather than comparing each component with a fixed constant, the proposal is to standardize each component by a robust location and scale computed from its own n values:

$$\begin{aligned} m_k &= \text{med}_{1 \leq i \leq n} c_{ki}, & \tau_k \\ &= 1.4826 \\ &\times \text{med}_{1 \leq i \leq n} |c_{ki} - m_k|, \end{aligned} \quad (24)$$

and to define the standardized component as the positive part of the scaled deviation,

$$z_{ki} = \max\left\{0, \frac{c_{ki} - m_k}{\tau_k}\right\}. \quad (25)$$

The constant 1.4826 makes τ_k a consistent estimator of the standard deviation when the component values are normally distributed (Hampel et al., 1986; Rousseeuw and Leroy, 1987). The truncation at zero reflects the fact

that only large values of a component are diagnostic; a case whose leverage or residual is smaller than typical is of no interest. Two consequences of this standardization deserve emphasis. First, the median and the median absolute deviation have breakdown point one half, so that m_k and τ_k are not inflated by a minority of extreme component values; a group of similar influential cases therefore does not raise the scale against which each of them is judged, which is precisely the mechanism by which masking defeats fixed cut-offs based on the full-sample s^2 . Second, because z_{ki} is a dimensionless quantity relative to the bulk of the sample, values are comparable across data sets with different n , p and response scale. If $\tau_k = 0$, which can occur when many component values are identical, the median absolute deviation is replaced by the mean absolute deviation from the median.

4.4 Definition of the index

The Standardized Composite Influence Index is the weighted Euclidean norm of the standardized component vector,

$$D_i^* = \sqrt{\sum_{k=1}^4 w_k z_{ki}^2}, \quad w_k \geq 0, \quad \sum_{k=1}^4 w_k = 1. \quad (26)$$

With equal weights $w_k = 1/4$, which is the default proposed here, D_i^* is the root mean square of the four standardized components. Geometrically, $\mathbf{z}_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})'$ locates case i in a four-dimensional diagnostic space whose origin corresponds to a typical case, and D_i^* is its weighted distance from that origin. A large value of D_i^* arises

when a case is far from typical in one component or moderately far in several, and the vector $\mathbf{z}_i$ itself reveals which components are responsible, so that the index is accompanied by a decomposition that supports interpretation. The weights allow the analyst to emphasize a particular aspect of influence, for example coefficient sensitivity when the goal is inference about $\boldsymbol{\beta}$ or fitted-value sensitivity when the goal is prediction; the influence of a choice of weights on the ranking can be examined by computing the index under several weight vectors.

Alternative aggregations were considered. A maximum norm, $\max_k z_{ki}$, is simpler but discards the information that a case is moderately unusual in several respects; a sum of the standardized components is equivalent to a weighted L_1 norm and is less sensitive to a single extreme component. The Euclidean form is retained because it corresponds to the familiar quadratic-form structure of Cook's distance and because it admits the approximate reference distribution discussed in Section 4.7.

4.5 Relationship to existing measures

The index is related to the classical statistics through the components. By (9), (15) and (19),

$$F_i = \text{DFFITS}_i^2 = \frac{s^2}{s_{(i)}^2} p D_i, \quad (27)$$

so that the third component is, up to a factor that tends to one for a non-outlying case as n grows, Cook's distance scaled by p . By (21) and (23),

$$B_i = \rho_i F_i, \quad (28)$$

so that the fourth component is the third multiplied by the coefficient-concentration ratio. The first two components are the squared externally studentized residual and Hadi's potential term. Therefore, for fixed h_{ii} and fixed direction of $\mathbf{x}_i$, all four components are nondecreasing functions of t_i^2 , and for fixed t_i and fixed direction the components L_i , F_i and B_i are increasing functions of h_{ii} . These monotonicities are transmitted to D_i^* when the reference quantities m_k and τ_k are held fixed, since the truncation, squaring, weighted sum and square root in (25) and (26) are all nondecreasing. It follows that the index respects the ordering imposed by the classical measures whenever those measures agree, and departs from it only through the direction term ρ_i and through the sample-based standardization.

The index is not a monotone function of any single classical measure. Two cases with the same Cook's distance may have different D_i^* values because one has a large residual and small leverage while the other has a small residual and large leverage, so that their standardized residual and leverage components differ, or because they have different values of ρ_i . Conversely, the index differs from Hadi's measure (14), which adds a leverage term and a residual term without standardization and contains no coefficient-direction information, and from Peña's (2005) statistic, which measures the sensitivity of the forecast of case i to the deletion of every other case rather than combining the deletion effects of case i itself. To the best of the author's knowledge, after a search of the literature summarized in Section 2, no previously published influence statistic

combines a residual component, a leverage component, a fitted-value component and a coefficient-vector component through sample-based robust standardization into a single index.

4.6 Theoretical properties

Property 1 (invariance to location and scale of the response). If $\mathbf{y}$ is replaced by $a\mathbf{1} + b\mathbf{y}$ with $b \neq 0$, then e_i , s and $s_{(i)}$ are each multiplied by $|b|$ up to sign, while h_{ii} , $\mathbf{C}$ and g_i are unchanged. Hence t_i , and with it A_i , F_i and B_i , are unchanged, L_i is unchanged, and D_i^* is invariant.

Property 2 (invariance to rescaling of the predictors). Let $\mathbf{X}$ be replaced by $\mathbf{XS}$ with $\mathbf{S}$ diagonal and nonsingular. Then $\mathbf{H}$ and the residuals are unchanged, so A_i , L_i and F_i are unchanged. Further, $\mathbf{C}$ becomes $\mathbf{S}^{-1}\mathbf{CS}^{-1}$, $\mathbf{D}$ becomes $\mathbf{S}^{-2}\mathbf{D}$, $\mathbf{x}_i$ becomes $\mathbf{S}\mathbf{x}_i$, and

$$(\mathbf{S}\mathbf{x}_i)'(\mathbf{S}^{-1}\mathbf{CS}^{-1})(\mathbf{S}^{-2}\mathbf{D})^{-1}(\mathbf{S}^{-1}\mathbf{CS}^{-1})(\mathbf{S}\mathbf{x}_i) = \mathbf{x}_i'\mathbf{CD}^{-1}\mathbf{C}\mathbf{x}_i = g_i, \quad (29)$$

so B_i and hence D_i^* are unchanged. Under a general nonsingular transformation $\mathbf{XT}$, the components A_i , L_i and F_i remain invariant but B_i does not, because DFBETAS refers to individual coefficients whose meaning changes under such a transformation. This is a property of the coefficient component rather than a defect: an index that is fully invariant to arbitrary reparametrization cannot carry information about individual coefficients.

Property 3 (sensitivity to genuine influence). As shown in Section 4.5, D_i^* is nondecreasing in t_i^2 for fixed leverage and direction and nondecreasing in h_{ii} for fixed residual and direction, when the reference

location and scale are held fixed. A case that substantially changes the coefficient vector has large F_i and, unless ρ_i is small, large B_i ; the corresponding standardized components are large and the index increases.

Property 4 (robustness to non-influential outliers). A case with large $|t_i|$ and small h_{ii} has a large residual component but small leverage, fitted-value and coefficient components, because the last three all contain the factor $h_{ii}/(1 - h_{ii})$ or $g_i/(1 - h_{ii})$. With equal weights the index is then at most half the value it would take if all four components were equally extreme, and the decomposition $\mathbf{z}_i$ identifies the case as a residual outlier rather than an influential point. The index therefore does not equate a large residual with influence.

Property 5 (detection of leverage influence). A bad leverage point has large h_{ii} and, because its residual is computed from a fit that has been pulled toward it, a residual that may be only moderately large. Its leverage component is large, and by (19) and (21) the factor $h_{ii}/(1 - h_{ii})$ amplifies even a moderate residual in the fitted-value and coefficient components; the index increases accordingly. A good leverage point has large L_i but small t_i , so only its leverage component is large, and the decomposition classifies it as high leverage rather than influential.

Property 6 (behaviour under multiple influential cases). Suppose that a fraction $\gamma < 1/2$ of the cases are influential in a similar manner. The full-sample s^2 and the fixed cut-offs of the classical measures are affected by all of them jointly, which is the

mechanism of masking under single-case deletion. The medians and median absolute deviations in (24) are determined by the majority of the component values and are therefore unaffected by the minority in the sense of breakdown; each influential case is then standardized against a reference that describes the clean majority. This argument concerns the standardization only; the components themselves are still computed from the contaminated full-sample fit, and the index inherits from single-case deletion its inability to detect cases whose deletion effects are individually small but jointly large. For such configurations the index should be used as the starting point for group-deletion procedures of the kind reviewed in Section 2.4.

Property 7 (interpretability). The index is a distance in a standardized diagnostic space whose axes are the four classical concepts of residual size, leverage, fitted-value change and coefficient change. A value of $D_i^* = 3$ with equal weights, for example, means that the root mean square of the four standardized excesses is three robust standard deviations; the vector $\mathbf{z}_i$ shows how this total is distributed.

Property 8 (behaviour under the uncontaminated model). Under model (1) with normal errors, t_i has a t_{n-p-1} distribution, so A_i is distributed as the square of a t_{n-p-1} variable, or approximately χ_1^2 for large n . The leverage values are fixed by the design and, for a design whose rows are a random sample from a p -variate distribution with finite second moments, they are concentrated near p/n with a spread of the same order. The components F_i and B_i are

products of A_i with functions of the leverage and direction, and their marginal distributions depend on the design. An exact null distribution of D_i^* is therefore not available in closed form, and the reference behaviour of the index must be characterized by approximation or by resampling, as discussed next.

4.7 Threshold selection

Four decision rules are proposed and compared on theoretical grounds; their empirical calibration is a matter for simulation, which is not undertaken here.

Rule 1: approximate chi-square reference. If the untruncated standardized components $(c_{ki} - m_k)/\tau_k$ were independent standard normal variables, then $\sum_k z_{ki}^2$ would be a mixture of chi-square distributions with between zero and four degrees of freedom, and the equal-weight index would satisfy $4(D_i^*)^2 \leq \chi_4^2$ in distribution. A conservative cut-off is therefore

$$D_i^* > \sqrt{\frac{\chi_{4;1-\alpha}^2}{4}}, \quad (30)$$

which for $\alpha = 0.05$ gives approximately 1.54 and for $\alpha = 0.01$ approximately 1.82. The components are neither normal nor independent, so this rule is only a rough guide, and its actual false-positive rate will differ from α .

Rule 2: robust self-referential rule. In the spirit of the rules used by Hadi (1992a) and Peña (2005) for their own statistics, a case is flagged when

$$D_i^* > \text{med}_j D_j^* + c \times 1.4826 \times \text{med}_j |D_j^* - \text{med}_l D_l^*|, \quad (31)$$

with $c = 3$ as a default. This rule adapts automatically to the sample, and because the index has already been standardized it is expected to be less sensitive to the choice of c than the corresponding rules for unstandardized measures. Its weakness is that a substantial fraction of moderately influential cases can raise the median of the index and reduce power.

Rule 3: bootstrap threshold. Because the index is a function of the design, a case-resampling bootstrap that resamples rows of $(\mathbf{y}, \mathbf{X})$ would alter the leverage structure; a residual bootstrap is preferable. Residuals from a robust fit, or the ordinary residuals after removing the cases flagged by Rule 2, are centred, resampled with replacement and added to the fitted values from the same fit to produce bootstrap responses at the original design points. The index is computed for each bootstrap sample, and the empirical $1 - \alpha$ quantile of the maximum index over cases is taken as the threshold. Using the maximum controls the family-wise probability of flagging any clean case. The procedure is computationally cheap because each bootstrap replicate requires one least-squares fit.

Rule 4: false discovery rate control. When approximate per-case p-values are available, for example from the bootstrap distribution of the individual index values, the procedure of Benjamini and Hochberg (1995) can be applied to the n p-values to control the expected proportion of flagged cases that are not influential. This is the most appropriate rule when the analyst expects several influential cases and is prepared to tolerate a

small proportion of false identifications in exchange for greater power.

Rule 1 requires no computation beyond the index and is suitable for a first screening; Rule 2 is recommended as the default because it is self-calibrating and robust; Rule 3 is recommended when a formal error rate is required; and Rule 4 is recommended when many cases are expected to be flagged. Whatever rule is used, a flagged observation is a candidate for investigation, not a candidate for deletion.

4.8 Classification of observations

Once the index and its decomposition have been computed, each case is assigned to one of the following classes. The class boundaries use the standardized components with a common constant κ , for which $\kappa = 2$ is suggested, together with the chosen threshold δ for the index.

Non-influential: $D_i^* \leq \delta$ and $z_{ki} < \kappa$ for all k .

Residual outlier: $z_{1i} \geq \kappa$, $z_{2i} < \kappa$, and $D_i^* \leq \delta$.

High leverage (good leverage point): $z_{2i} \geq \kappa$, $z_{1i} < \kappa$, and $D_i^* \leq \delta$.

Influential: $D_i^* > \delta$; within this class, a case is described as a bad leverage point when $z_{2i} \geq \kappa$ and as a coefficient-directed influential case when ρ_i is small relative to the sample median of ρ_j .

Potentially masked: a case that is not flagged in the full sample but becomes flagged when the cases in the influential class are deleted and the index is recomputed.

Potentially swamped: a case flagged in the full sample that ceases to be flagged when the other flagged cases are deleted and the index is recomputed.

The last two classes are defined by a single round of joint deletion. They provide a practical safeguard against the two failure modes of single-case diagnostics without requiring the full machinery of a forward search, although a forward search remains preferable when the number of suspected cases is large.

4.9 The diagnostic algorithm

The proposed procedure is as follows.

Step 1. Fit model (1) by ordinary least squares and obtain $\hat{\beta}$, e , s^2 and C .

Step 2. Compute the leverage values h_{ii} from the hat matrix, and the deleted variance estimates $s_{(i)}^2$ from (6).

Step 3. Compute the externally studentized residuals t_i from (7).

Step 4. Compute Cook's distance, DFFITS, DFBETAS and the covariance ratio from (9) to (12) for comparison and reporting.

Step 5. Compute the components A_i , L_i , F_i and B_i from (16), (17), (19) and (21), and the ratio ρ_i from (23).

Step 6. Compute the robust location and scale of each component from (24), the standardized components from (25), and the index D_i^* from (26) with the chosen weights.

Step 7. Determine the threshold δ by one of Rules 1 to 4 of Section 4.7 and classify the cases according to Section 4.8.

Step 8. Delete the cases in the influential class jointly, recompute Steps 1 to 7 on the reduced sample, and identify potentially masked and potentially swamped cases by comparing the two classifications.

Step 9. Report, for the full sample and for the sample with the influential cases removed, the coefficient estimates, their standard errors, the coefficient of determination, the residual standard error and the confidence intervals of the coefficients of interest, and interpret any material differences in the light of the data-quality investigation and subject-matter knowledge.

The computational cost of Steps 1 to 7 is that of a single least-squares fit together with $O(np^2)$ operations for the leverage values and the quantities g_i , since g_i requires only the products $\mathbf{C}\mathbf{x}_i$, which are already needed for the leverage. The algorithm is therefore no more expensive than the computation of the classical measures themselves, and Step 8 doubles this cost. Rule 3 adds one fit per bootstrap replicate.

4.10 Extensions

The construction extends directly to weighted least squares by replacing $\mathbf{X}$ and $\mathbf{y}$ by their weighted versions. For generalized linear models the deletion identities (4) to (6) hold only approximately through the one-step approximation of Pregibon (1981), and the components would be defined from the corresponding one-step deletion quantities. A robust version of the index can be obtained by computing the components from a high-breakdown fit such as the MM-estimator of Yohai (1987), with the hat matrix replaced by a robust distance in the manner of Rousseeuw

and van Zomeren (1990); the standardization (24) to (26) and the aggregation (26) remain unchanged. These extensions are noted for completeness and are not developed here.

5. Conclusion

This paper has re-examined the classical single-case influence diagnostics for linear regression and shown, by direct algebra, that Cook's distance, DFFITS, the covariance ratio and the modified Cook statistic are all functions of the externally studentized residual and the leverage alone, so that they discard the direction of an observation in coefficient space and differ from one another only in weighting. This structural observation, together with the dependence of the conventional cut-offs on unstated assumptions about the sample size, the number of predictors and the absence of masking, defines the gap that the paper addresses.

The Standardized Composite Influence Index proposed here combines four components, the squared externally studentized residual, the leverage potential, the squared DFFITS statistic and the mean squared DFBETAS statistic, each standardized by a robust location and scale computed from the sample and then aggregated through a weighted Euclidean norm. The coefficient component introduces a direction-dependent term whose relationship to the fitted-value component is governed by a concentration ratio bounded between zero and one; this bound and the equality of the average of the new leverage-type quantity with the average leverage were proved. The index was shown to be invariant

to location and scale changes of the response and to rescaling of the predictors, to be monotone in the residual and the leverage in the appropriate sense, to distinguish residual outliers and good leverage points from genuinely influential cases through its decomposition, and to resist the inflation of its reference scale by a minority of influential cases. Four threshold rules were proposed, ranging from a simple chi-square approximation to bootstrap and false-discovery-rate procedures, and a nine-step algorithm was given that produces a classification of every observation and a one-round check for masking and swamping.

Several limitations should be stated plainly. The index is built from single-case deletion quantities and therefore shares the inability of all such quantities to detect groups of observations whose individual deletion effects are small; the robust standardization mitigates masking of the reference scale but does not remove it. The exact null distribution of the index is unknown, and the proposed thresholds are approximations or resampling procedures whose operating characteristics must be established empirically. The choice of equal weights is a default rather than an optimum. Finally, and most importantly, no numerical evidence has been presented: the performance of the index relative to the classical measures under vertical outliers, good and bad leverage points, multiple contamination, masking and swamping, for a range of sample sizes, predictor dimensions and error distributions, remains to be determined by Monte Carlo experiment and by application to real data. These

investigations are the natural continuation of the present work.

It bears repeating that an influential observation is not necessarily an erroneous one. The purpose of the index, as of every diagnostic reviewed in this paper, is to identify the observations on which the conclusions of an analysis depend, so that those observations can be examined with care. Any decision to remove or to down-weight an observation should rest on statistical evidence, on an investigation of data quality and on knowledge of the subject matter, and never on the value of a diagnostic statistic alone.

References

- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16(1), 125–127.
- Andrews, D. F., & Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society, Series B*, 40(1), 85–93.
- Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68(1), 13–20.
- Atkinson, A. C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *Journal of the Royal Statistical Society, Series B*, 44(1), 1–36.
- Atkinson, A. C. (1985). *Plots, transformations and regression: An introduction to graphical methods of diagnostic regression analysis*. Oxford: Clarendon Press.
- Atkinson, A. C. (1986). Comment: Aspects of diagnostic regression analysis. *Statistical Science*, 1(3), 397–402.

Atkinson, A. C. (1994). Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association*, 89(428), 1329–1339.

Atkinson, A. C., & Riani, M. (2000). *Robust diagnostic regression analysis*. New York: Springer.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Chichester: Wiley.

Beckman, R. J., & Cook, R. D. (1983). Outlier.....s (with discussion). *Technometrics*, 25(2), 119–163.

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57(1), 289–300.

Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, 34(3), 279–298.

Chatterjee, S., & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression (with discussion). *Statistical Science*, 1(3), 379–393.

Chatterjee, S., & Hadi, A. S. (1988). *Sensitivity analysis in linear regression*. New York: Wiley.

Chatterjee, S., & Price, B. (1991). *Regression analysis by example* (2nd ed.). New York: Wiley.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.

Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, 74(365), 169–174.

Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B*, 48(2), 133–169.

Cook, R. D., & Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, 22(4), 495–508.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Davies, L., & Gather, U. (1993). The identification of multiple outliers (with discussion). *Journal of the American Statistical Association*, 88(423), 782–792.

Draper, N. R., & John, J. A. (1981). Influential observations and outliers in regression. *Technometrics*, 23(1), 21–26.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.

Fox, J. (1991). *Regression diagnostics: An introduction*. Newbury Park, CA: Sage.

Habshah, M., Norazan, M. R., & Imon, A. H. M. R. (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics*, 36(5), 507–520.

Hadi, A. S. (1992a). A new measure of overall potential influence in linear regression. *Computational Statistics & Data Analysis*, 14(1), 1–27.

Hadi, A. S. (1992b). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society, Series B*, 54(3), 761–771.

Hadi, A. S., & Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*, 88(424), 1264–1272.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346), 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York: Wiley.

Hawkins, D. M. (1980). *Identification of outliers*. London: Chapman and Hall.

Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, 32(1), 17–22.

Hocking, R. R. (1983). Developments in linear regression methodology: 1959–1982 (with discussion). *Technometrics*, 25(3), 219–249.

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

Hubert, M., Rousseeuw, P. J., & Van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, 23(1), 92–119.

Imon, A. H. M. R. (2005). Identifying multiple influential observations in linear regression. *Journal of Applied Statistics*, 32(9), 929–946.

Lawrance, A. J. (1995). Deletion influence and masking in regression. *Journal of the Royal Statistical Society, Series B*, 57(1), 181–189.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49–55.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics: Theory and methods*. Chichester: Wiley.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Hoboken, NJ: Wiley.

Nurunnabi, A. A. M., Hadi, A. S., & Imon, A. H. M. R. (2014). Procedures for the identification of multiple influential observations in linear regression. *Journal of Applied Statistics*, 41(6), 1315–1331.

Nurunnabi, A. A. M., Nasser, M., & Imon, A. H. M. R. (2016). Identification and classification of multiple outliers, high

leverage points and influential observations in linear regression. *Journal of Applied Statistics*, 43(3), 509–525.

Peña, D. (2005). A new statistic for influence in linear regression. *Technometrics*, 47(1), 1–12.

Peña, D., & Yohai, V. J. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society, Series B*, 57(1), 145–156.

Peña, D., & Yohai, V. J. (1999). A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association*, 94(446), 434–445.

Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, 9(4), 705–724.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388), 871–880.

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *WIREs Data Mining and Knowledge Discovery*, 1(1), 73–79.

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

Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411), 633–639.

Ullah, M. A., & Pasha, G. R. (2009). The origin and developments of influence measures in regression. *Pakistan Journal of Statistics*, 25(3), 295–307.

Velleman, P. F., & Welsch, R. E. (1981). Efficient computing of regression diagnostics. *The American Statistician*, 35(4), 234–242.

Weisberg, S. (2005). *Applied linear regression* (3rd ed.). Hoboken, NJ: Wiley.

Welsch, R. E., & Kuh, E. (1977). *Linear regression diagnostics* (Technical Report No. 923-77). Cambridge, MA: Sloan School

of Management, Massachusetts Institute of Technology.

Wilks, S. S. (1963). Multivariate statistical outliers. *Sankhyā, Series A*, 25, 407–426.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics*, 14(4), 1261–1295.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2), 642–656.