Received: 23 February 2021 | Revised: 31 August 2021 | Accepted: 1 September 2021

DOI: 10.1002/sam.11550

WILEY

RESEARCH ARTICLE

Residual's influence index (*RINFIN*), bad leverage and unmasking in high dimensional L_2-regression

Yannis G. Yatracos[1,2]

[1]Yau Mathematical Sciences Center, Tsinghua University, Beijing, China

[2]Beijing Institute of Mathematical Sciences and Applications, Beijing, China

Correspondence
Yannis G. Yatracos, Yau Mathematical Sciences Center, Tsinghua University, Beijing, China.
Email: yatracos@tsinghua.edu.cn; yannis.yatracos@gmail.com

Abstract

In linear regression of Y on $\mathbf{X}(\in R^p)$ with parameters $\beta(\in R^{p+1})$, statistical inference is unreliable when observations are obtained from gross-error model, $F_{\epsilon,G} = (1-\epsilon)F + \epsilon G$, instead of the assumed probability F;G is gross-error probability, $0 < \epsilon < 1$. Residual's influence index (*RINFIN*) at $(\mathbf{x}, y)$ is introduced, *with components measuring also the local influence of* $\mathbf{x}$ *in the residual* and large value *flagging* a bad leverage case (from G), thus causing unmasking. Large sample properties of *RINFIN* are presented to confirm significance of the findings, but often the large difference in the *RINFIN* scores of the data is indicative. *RINFIN* is successful with microarray data, simulated, high dimensional data and classic regression data sets. *RINFIN*'s performance improves as p increases and can be used in multiple response linear regression.

KEYWORDS

big data, influence function, leverage, local influence, masking, post-*P*-value era, residual's influence index

1 | INTRODUCTION

Tukey [29, p. 60] wrote: "Procedures of *diagnosis*, and procedures to *extract indications* rather than extract conclusions, will have to play a large part in the future of data analyses and graphical techniques offer great possibilities in both areas." This philosophy is widely adopted nowadays in Data Science, realizes ASA's hope for a "post-*P*-value era" [31] and motivates this work.

Data cleaning should precede statistical analysis. In linear regression of Y on $\mathbf{X}$ and parameters β, it is often erroneously assumed that the data follow probability F instead of the gross-error model $F_{\epsilon,G} = (1-\epsilon)F + \epsilon G$ [19]; G is gross-error probability, $0 < \epsilon < 1$, $Y \in R$, $\mathbf{X} \in R^p$, $\beta \in R^{p+1}$. A case $(\mathbf{x}, y)$ with $\mathbf{x}$ far away from the bulk of F's factor space is called "leverage" case [25] and influences the statistical analysis. In particular, a "bad" leverage case $(\mathbf{x}, y)$ from G forces the regression hyperplane determined by F (the F-regression) and *the associated F-residuals* to change drastically when $\mathbf{x}$ becomes more remote. The goal of this work is to provide *a simple and easy to implement procedure* extracting indications for *flagging* bad leverage cases (from G) in least squares (L_2) regression, without the use of *P*-values and hypothesis testing that lead to a hard "yes–no"-decision.

A tool indicating the influence of $(\mathbf{x}, y)$ in the value of a statistical functional, $T(F)$, is T's influence function, $IF(\mathbf{x}, y; T, F)$ [13]. In linear L_2-regression, the empirical influence function of a non-robustified estimator, $\widehat{\beta}_i$, of β_i measures the change of $\widehat{\beta}_i$ when $(\mathbf{x}, y)$ is added in the sample, but suffers from the *masking effect* that is due to neighboring cases of $(\mathbf{x}, y)$ in the sample. For example, from (10),

in simple, linear L_2-regression with sample $(x_1, y_1), \ldots, (x_n, y_n)$, the empirical influence function of the slope at (x, y) has form $C \cdot r \cdot (x - \bar{x}_n)$; r is the residual of (x, y), C is independent of x, y. If (x, y) is one of few, neighboring bad leverage cases in the sample, the difference $(x - \bar{x}_n)$ will have large absolute value whereas r may be near zero, thus $|r \cdot (x - \bar{x}_n)|$ may take a moderate value *masking* (x, y). To the contrary, from (12), the x-derivative of the slope's influence function measures *local influence* of (x, y) (see (7)) and separates the factors of the influence function, obtaining instead the sum of $C \cdot \widehat{\beta}(x - \bar{x}_n)$ and $C \cdot r$, which has large absolute value even when x is masked and r is near 0; $\widehat{\beta}$ is the L_2-estimate of the slope. The index, *RINFIN*, introduced to flag bad leverage cases in multiple, linear L_2-regression, depends on the $\mathbf{x}$-partial derivatives of the influence functions of regression coefficients and shares the advantage of the factors' separation in $\widehat{\beta}_i$'s influence function, $i = 0, \ldots, p$. The same holds for L_2-regression with diagonal matrix of weights, W, independent of $\mathbf{x}, r$.

The $\mathbf{x}$-partial derivatives of regression coefficients' influence functions appear naturally in changes of regression residuals for small $\mathbf{x}$-perturbations under models F and $F_{\epsilon,\mathbf{x},y}$; see Section 4.2. The so-obtained (population) L_2-*RINFIN* is[1]

$$RINFIN(\mathbf{x}, y; \epsilon, \beta) = \epsilon \cdot \sum_{i=1}^{p} \left(IF_i + \frac{\partial IF_0}{\partial x_i} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i} \right)^2; \tag{1}$$

IF_j denotes the influence function of the jth regression coefficient, $j = 0, 1, \ldots, p$. Note that *RINFIN* uses also the information in the influence functions. For simple linear regression,

$$RINFIN(x, y; \epsilon, \beta) = \epsilon \cdot \left\{ \frac{2r_2(x,y)(x - EX) - \beta[(x - EX)^2 + Var(X)]}{Var(X)} \right\}^2, \tag{2}$$

with L_2-residual (r_2), slope (β), mean (EX), and variance $(Var X)$ all under F.

RINFINABS is obtained by replacing in the sum in (1) the squares by absolute values. Using G's averages, $(\bar{\mathbf{x}}, \bar{y})$, $RINFIN(\bar{\mathbf{x}}, \bar{y}; \epsilon, \beta)$ is calculated. Asymptotic properties of $RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n)$ are also presented; $\widehat{\beta}_n$ is β's L_2-estimate.

In practice, sample *RINFIN* score, $RINFIN(\mathbf{x}, y; 1/n, \widehat{\beta}_n)$, is obtained for every $(\mathbf{x}, y)$ in the sample. Large *RINFIN* scores provide indications for bad leverage cases. Since the percentage of G-cases in $F_{\epsilon,\mathbf{x},y}$ is expected to be 10% or less, potential bad leverage cases in the sample are those $(\mathbf{x}, y)$ with the 10% larger *RINFIN* scores, and especially those with the same ordering in *RINFINABS* scores. The spacings of ordered *RINFIN* scores are also informative. The 10% threshold value can be replaced by another value. Once a case is flagged, it can be grouped with neighboring cases and using their average, $(\bar{\mathbf{x}}, \bar{y})$, and the group's proportion in the sample, ϵ, *RINFIN* scores can be re-calculated as described in Appendix C. Comparison of the findings with the ordering of the scores obtained with other indices, or with results of methods that extract conclusions using statistical significance are informative. Proposition 5 can also be used to determine significance of the *RINFIN* scores.

When n is smaller than p, sample *RINFIN* values are calculated sequentially, for the y-response and sub-vectors of $\mathbf{x}$-covariates with dimension $q < n$;p is multiple of q. For each case, *the total* of its $\frac{p}{q}$ sample *RINFIN* values is its *RINFIN* score. *RINFIN* can also be used with multiple response linear regression, adding for $(\mathbf{x}, y)$ the sample *RINFIN* scores for each response. The *RINFIN* approach can be used for *LASSO* and ridge regression with the influence functions derived in Ollerer et al. [24].

RINFIN provides satisfactory results with high dimensional data. It is successful with the microarray data used in Zhao et al. [37, 38] for which $n = 120$ and $p = 1500$. In simulations with gross-error normal mixtures F, G and fixed sample size n, the misclassification proportion of G-cases using $RINFIN(\mathbf{x}, y; 1/n, \widehat{\beta}_n)$ decreases to zero as p increases, $p < n$. The *blessing of high dimensionality* is due to the "separation" of the mixtures' components measured, for example, by their Hellinger's distance, as p increases ([34, 35], Section 8, Proposition 8.1). *RINFIN* also identifies bad leverage cases in classic regression data sets.

Due to the flood of Big Data, new influence measures have been recently introduced. In Genton and Ruiz-Gazen [11], an *observation* is influential "whenever a change in its value leads to a radical change in the estimate" and the *hair-plot* is used for *visual identification*. Local and global influence measures are proposed using partial derivative of the *estimate*. She and Owen [28] have as goals outlier identification and robust coefficient estimation, both achieved using a non-convex sparsity criterion. Zhao et al. [37] propose a high dimensional influence measure (*HIM*) based on marginal correlations between the response and the individual covariates and the leave-one-out observation idea [32]. For a particular regression model, Zhao et al. [38] propose a novel procedure, for *multiple* influential point detection (*MIP*).

Robustness tools have been used extensively in outlier detection. The influence function has been used by Campbell [4] in discriminant analysis and subsequent results appeared, among others, in Boente et al. [3]. Rousseeuw and van Zomeren [26] used standardized least

[1] Alternatively, the sum next to ϵ in (1) is divided by p.

trimmed squares residuals against robust distance to classify observations in regression. Hubert et al. [21] present a survey of high breakdown robust methods to detect outlying observations. The influence of observations and of model assumptions in estimates' values and identification of outliers have been also studied by several authors, among others by Cook [7], Belsley et al. [2], Cook and Weisberg [9], Hawkins [17], Huber [20], Velleman and Welsch [30], Welsch [33], Hawkins et al. [18], Carroll and Ruppert [5], Hampel [15], Cook [8], Hampel et al. [16], Lawrance [22], and Hadi and Simonoff [12].

In Section 2, *RINFIN* applications are presented. In Section 3, the derivative of the influence function is introduced for measuring *local influence* of $(\mathbf{x}, y)$; it is the main tool to obtain *RINFIN* scores. In Section 4, local influence of $(\mathbf{x}, y)$ in L_2-regression residual is studied, *RINFIN* and *RINFINABS* are defined and asymptotic properties of *RINFIN*'s estimate are obtained. In Appendix A, $\mathcal{E}$-matrix is introduced to obtain a simple and insightful form for *RINFIN* when the $\mathbf{X}$-covariates are uncorrelated. Proofs follow in Appendix B. Directions for the use of *RINFIN* with data are in Appendix C.

2 | *RINFIN* IN ACTION

2.1 | *RINFIN* and simulations, $p < n$

Data $(\mathbf{X}, Y)$ from probability F follow a linear regression model with p parameters, $\beta = (1.5, 0.5, 0, 1, 0, 0, 1.5, 0, 0, 0, 1, 0, \ldots, 0)$; when $p < 11$, β's first p coordinates are used. $\mathbf{X}$ is obtained from p-dimensional normal distribution, $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, with $\mathbf{\Sigma}$'s entries $\Sigma_{i,j} = 0.5^{|j-i|}$, $1 \le i, j \le p$, as in Alfons et al. [1, p. 11]. For gross-error model, $F_{\epsilon,G}$, the proportion ϵ is 10%. For each contaminated $\mathbf{X}$ (from G) the first $[\gamma \cdot p]$ coordinates are *independent*, normal with mean μ and variance 1; $0 < \gamma \le 1$, $[x]$ denotes the integer part of x. Various values for γ, p, and μ are used and p is smaller than the sample size n. The regression errors are independent, standard normal random variables. Each of the $N = 100$ simulated samples has size $n = 100$. Cases 1–10 are contaminated and compared with those having the 10 larger sample *RINFIN* scores for calculating the misclassification proportion.

In Table 1, the misclassification proportion decreases as p increases except for an anomaly when $p = 90$ due to its proximity to $p = n = 100$, for which L_2-regression breaks down. By increasing n to 150 cases this anomaly disappears, for example, for $\mu = 1$ the misclassification proportion is 0.105.

In Table 2, for *fixed* contamination proportion in the first $[\gamma \cdot p]$ $\mathbf{x}$-coordinates, the *RINFIN* misclassification proportion decreases as p increases. The anomaly is still observed when $p = 90$. The blessing of high dimensionality is observed in both Tables 1 and 2.

TABLE 1 Average misclassification proportion with *RINFIN*'s orderings

Complete contamination ($\gamma = 1$)				
p	$\mu = 0.5$	$\mu = 1$	$\mu = 1.5$	$\mu = 2$
10	0.857	0.624	0.320	0.117
30	0.802	0.394	0.079	0.003
50	0.775	0.254	0.016	0.000
70	0.728	0.162	0.000	0.000
90	0.740	0.208	0.009	0.000

2.2 | *RINFIN* and real, high dimensional data, $p > n$

RINFIN is used for the microarray data in Zhao et al. [38], obtained from Chiang et al. [6] and previously analyzed by Zhao et al. [37]: 120 twelve-week-old male offspring were selected for tissue harvesting from the eyes. The microarray contains over 30,000 different probe sets. Probe gene *TR32* is used as the response and the covariates are 1500 genes mostly correlated with it.

Since $n = 120 < p = 1500$, *RINFIN* values are calculated for the response *TR32* and each group of 100 $\mathbf{x}$-covariates partitioning the microarray data, with coordinates $100(j-1)+1, \ldots, 100j$, $1 \le j \le 15$. For each of the 120 cases, the *total* of its 15 *RINFIN* values is its score. In Table 3, cases with the higher 16 *RINFIN* scores are provided; more than 10% of the cases are presented in order to get an idea of the *spacings* in the successive scores.

Indications for leverage cases from G in the gross-error model are given for cases 80, 95, 32, 120, and 59, after which the spacings' in the *RINFIN* scores are reduced. In Table 4, the highest 16 *RINFINABS*-scores are provided. Cases 80, 95, 32, 120, and 59 have still the same order as in Table 3, but the order of the remaining cases changes.

Cases 80, 95, 32, 120, and 59 are also supported by diagnostics *HIM* and *MIP* which are based, respectively, on analyses of correlations and covariances, whereas *RINFIN* is based on perturbations of residuals. According to Leng [23], diagnostic *HIM* [37] identifies 15 influential cases,

80, 95, 120, 32, 75, 70, 107, 28, 59, 38, 67, 27, 17, 51, 98

and diagnostic *MIP* [38] identifies 7 influential cases,

80, 95, 120, 32, 75, 28, 59.

TABLE 2 Average misclassification proportion with *RINFIN*'s ordering

Partially contaminated data in the first $\gamma \cdot p$ X-coordinates						
p	$\mu = 1$, $\gamma = 0.2$	$\mu = 1$, $\gamma = 0.4$	$\mu = 1$, $\gamma = 0.6$	$\mu = 1.5$, $\gamma = 0.2$	$\mu = 1.5$, $\gamma = 0.4$	$\mu = 1.5$, $\gamma = 0.6$
10	0.859	0.834	0.747	0.811	0.695	0.550
30	0.822	0.753	0.599	0.719	0.516	0.296
50	0.804	0.676	0.506	0.663	0.364	0.164
70	0.787	0.612	0.416	0.598	0.250	0.089
90	0.784	0.605	0.435	0.611	0.294	0.116

TABLE 3 Cases with the higher *RINFIN* scores

Microarray data								
Case	80	95	32	120	59	64	85	112
RINFIN	824,471	146,639	40,295	24,749	14,802	12,849	12,582	11,683
Case	38	40	24	117	27	28	84	90
RINFIN	11,680	10,973	10,476	8478	7516	6214	5689	5536

TABLE 4 Most influential cases with *RINFINABS*

Microarray data								
Case	80	95	32	120	59	85	38	112
RINFINABS	1744.5	797.4	488.1	379.4	319.3	285.6	282.1	273.6
Case	64	24	40	27	117	6	84	28
RINFINABS	261.8	259.4	254.4	228.7	226	193.5	191.9	191.4

RINFIN and *RINFINABS* both identify case 28 in Tables 3 and 4, at the top 12% of the 120 *RINFIN* scores. Case 75 is identified by both *RINFINs* as the case with the 21st larger *RINFIN* score, at the top 17.5% of the *RINFIN* scores. *RINFIN* values of case 75 were not in the top 10 *RINFIN* values in any of the 15 groups partitioning the microarray data. *RINFIN* values of case 28 were ranked 9th in the group of 1–100 coordinates and 10th in the group of 101–200 coordinates. The total square distances of cases 75 and 28 from the means of the $p = 1500$ coordinates were at the 80th quantile of all the distances. However, their maximum square distances over all coordinates were between the 40th and 50th quantiles of all cases. *RINFIN* targets bad leverage cases and 28 and 75 do not fall in this category.

2.3 | *RINFINABS* and classic, regression data sets

The top 6–8 *RINFINABS* scores are obtained using (35), with sum of absolute values instead of *RINFIN*'s sum of squares, for 6 known data sets; those without references are in Rousseeuw and Leroy [25].

TABLE 5 Data: Kootenay River ($p = 1$, $n = 13$)

Case	4	7	2	12	6	1
RINFINABS	8.906	0.106	0.052	0.044	0.030	0.015

In the Kootenay River data, case 4 is remote and has large *RINFINABS* score compared with the rest (see Table 5).

In the Hertzsprung–Russel star data, cases 11, 20, 30, 34 correspond to giant stars, that is, remote, x-neighboring cases. *RINFINABS* scores for each case, as well as for groups $G_1 = \{11, 20, 30, 34\}$, $G_2 = \{7, 14\}$ of x-neighboring cases, support that {11, 20, 30, 34} are bad leverage cases (see Table 6).

In the Hadi and Simonoff [12] data, cases 1–3, 4, 17 are more distant from the origin than the remaining cases. *RINFINABS* scores obtained for each case and after grouping indicate cases 1–3 are bad leverage cases. Hadi and Simonoff [12] identify these as *true outliers* (see Table 7).

In the education data, case 50 (Alaska) is far from the origin and its *RINFINABS* score compared with those of the rest indicates bad leverage (see Table 8).

TABLE 6 Data: Hertzsprung–Russel stars ($p = 1$, $n = 47$)

Case	*RINFINABS*	Group	*RINFINABS*	Group	*RINFINABS*
34	0.545	11, 20, 30,3 4	26.555	11, 20, 30, 34	39.654
30	0.387	14	0.276	7, 14	0.447
20	0.272	36	0.131	17	0.159
14	0.198	4	0.131	36	0.149
7	0.191	2	0.131	4	0.143
11	0.162	17	0.125	2	0.143

TABLE 7 Data: Hadi–Simonoff ($p = 2$, $n = 25$)

Case	*RINFINABS*	Group	*RINFINABS*
22	0.677	1, 2, 3	1.074
4	0.572	4	0.645
17	0.527	17	0.620
12	0.527	22	0.607
25	0.374	12	0.464
1	0.351	13	0.399
2	0.346	25	0.328
3	0.340	24	0.298

TABLE 8 Data: education ($p = 3$, $n = 50$)

Case	50	33	7	44	29	5
RINFINABS	1.20	0.486	0.476	0.402	0.336	0.304

TABLE 9 Data: salinity ($p = 3$, $n = 28$)

Case	16	15	5	3	9	4
RINFINABS	2.216	0.418	0.327	0.307	0.293	0.288

In the salinity data, Carroll and Ruppert [5] indicate that remote case 16 and the other influential case 3 are masking case 5. *RINFINABS* scores support the findings for case 16 but not case's 5 masking (see Table 9).

In the modified wood data, cases 4, 6, 8, 19 are neighboring and remote in *each* **x**-coordinate. *RINFINABS* scores obtained for each case and with cases 4, 6, 8, 19 as group support these are bad leverage cases (see Table 10).

TABLE 10 Data: modified wood ($p = 5$, $n = 20$)

Case	*RINFINABS*	Group	*RINFINABS*
19	1.579	4, 6, 8, 19	34.729
8	1.532	11	1.710
6	1.452	7	1.460
4	1.332	12	1.390
12	1.324	10	1.084
11	1.161	16	0.785
7	1.158	1	0.779
10	1.075	17	0.738

3 | LOCAL INFLUENCE AND THE DERIVATIVE OF THE INFLUENCE FUNCTION

To study local perturbation of the residual, r, which provides *RINFIN*, the derivative of r's influence function is used. Hampel [13] introduced the influence function, $IF(\mathbf{x}; T, F)$, of a functional T with real values, to measure the influence of $\mathbf{x}$ in the value of T for the gross-error model, $(1-\epsilon)F + \epsilon\Delta_{\mathbf{x}}$:

$$IF(\mathbf{x}; T, F) = \lim_{\epsilon \to 0} \frac{T[(1-\epsilon)F + \epsilon\Delta_{\mathbf{x}}] - T(F)}{\epsilon}, \tag{3}$$

when this limit exists; $\mathbf{x}(\in R^p)$, F is a probability, $\Delta_{\mathbf{x}}$ is the probability with all its mass at $\mathbf{x}$, $0 < \epsilon < 1$. $T(F)$ is usually a parameter of model F, for example, the expected value of a random variable X from F with $T(F) = E_F(X)$.

$IF(\mathbf{x}; T, F)$ determines the "bias" in the value of T at F when using instead $(1-\epsilon)F + \epsilon\Delta_{\mathbf{x}}$:

$$\begin{aligned} T[(1-\epsilon)F + \epsilon\Delta_{\mathbf{x}}] - T(F) &= \epsilon IF(\mathbf{x}; T, F) + o(\epsilon) \\ &\approx \epsilon IF(\mathbf{x}; T, F), \end{aligned} \tag{4}$$

"$\approx$" is used since

$$\lim_{\epsilon \to 0} \frac{T[(1-\epsilon)F + \epsilon\Delta_{\mathbf{x}}] - T(F)}{\epsilon IF(\mathbf{x}; T, F)} = 1.$$

Hampel [14, p. 389] introduced also *local-shift-sensitivity*,

$$\lambda^* = \sup_{\mathbf{x} \neq \mathbf{y}} \frac{|IF(\mathbf{x}; T, F) - IF(\mathbf{y}; T, F)|}{\|\mathbf{x} - \mathbf{y}\|}, \tag{5}$$

as "a measure for the *worst* (approximate) effect of wiggling the observations"; $\|\cdot\|$ is a Euclidean distance in R^p.

Local-shift-sensitivity was never fully exploited. One reason is that, in reality, it is a "global" measure as supremum over all $\mathbf{x}$, $\mathbf{y}$. Thus, λ^* cannot be used to study the bias in the value of T at $(1-\epsilon)F+\epsilon\Delta_{\mathbf{x}}$ for $\mathbf{x}$'s small perturbation, from $\mathbf{x}$ to $\mathbf{x}+\mathbf{h}$, and $\|\mathbf{h}\|$ small, that is

$$T[(1-\epsilon)F+\epsilon\Delta_{\mathbf{x}+\mathbf{h}}]-T[(1-\epsilon)F+\epsilon\Delta_{\mathbf{x}}]. \tag{6}$$

Local influence (6) of $\mathbf{x}$ in $T[(1-\epsilon)F+\epsilon\Delta_x]$ is measured by the partial derivatives of the influence function, as the next lemma indicates for $x\in R$ and is confirmed for the residuals in Proposition 2 with $\mathbf{x}\in R^p$.

Lemma 1. *Assume that F is defined on the real line and that* (6) *is evaluated at neighboring points x, $x+h$, $x\in R$, $h\in R$, $|h|$ small. Then,*

$$\lim_{h\to 0}\lim_{\epsilon\to 0}\frac{T[(1-\epsilon)F+\epsilon\Delta_{x+h}]-T[(1-\epsilon)F+\epsilon\Delta_x]}{\epsilon h} = \frac{dIF(x;T,F)}{dx}=IF'(x;T,F), \tag{7}$$

when the limit exists.

Remark 1. Under mild conditions, the limits in (7) can be interchanged without affecting the limit, for example, for any function g for which the derivative g' exists and

$$T(F)=\int g(y)dF(y).$$

$IF'(x;\,T,\,F)$ is used to approximate local influence (6) for small ϵ, $|h|$:

$$T[(1-\epsilon)F+\epsilon\Delta_{x+h}]-T[(1-\epsilon)F+\epsilon\Delta_x]\approx\epsilon hIF'(x;T,F). \tag{8}$$

(8) is the *main tool* used to approximate L_2-residuals of gross-error models and determine *RINFIN*. When (8) is used, the influence function's derivative is always evaluated at F.

Example 1. Consider a simple, linear regression model, $Y=\beta_0+\beta_1X+e$, with error e having mean zero and finite second moment, F is the joint distribution of $(X,\,Y)$.

The influence functions for the L_2-parameters $\beta_0(F)$, $\beta_1(F)$, obtained at F are

$$IF(x,y;\beta_0(F),F)=[y-\beta_0(F)-\beta_1(F)x]\frac{EX^2-xEX}{Var(X)} = r(x,y;F)\frac{EX^2-xEX}{Var(X)}, \tag{9}$$

$$IF(x,y;\beta_1(F),F)=[y-\beta_0(F)-\beta_1(F)x]\frac{x-EX}{Var(X)} = r(x,y;F)\frac{x-EX}{Var(X)}; \tag{10}$$

EU and $Var(U)$ denote, respectively, U's mean and variance. The x(-partial)-derivatives of (9), (10) are

$$IF'_{x,0}=\frac{\partial IF(x,y;\beta_0(F),F)}{\partial x}=-\beta_1(F)\frac{EX^2-xEX}{Var(X)} - r(x,y;F)\frac{EX}{Var(X)}, \tag{11}$$

$$IF'_{x,1}=\frac{\partial IF(x,y;\beta_1(F),F)}{\partial x}=-\beta_1(F)\frac{x-EX}{Var(X)} - r(x,y;F)\frac{1}{Var(X)}. \tag{12}$$

Observe in (9) and (10) the *multiplicative* effects of r with $(x-EX)$ and EX^2-xEX and their conversions to *additive* effects in (11) and (12).

Remark 2. The y-derivatives of L_2-influence functions (9) and (10) are, respectively, $(EX^2-xEX)/Var(X)$ and $(x-EX)/Var(X)$, do not provide information for $r(x,y;F)$ and their sample versions are maximized at the extreme x-values in the sample.

4 | RESIDUALS, INFLUENCE, LEVERAGE CASES AND *RINFIN*

4.1 | Influence functions in regression

Let $(\mathbf{X},\,Y)$ follow probability model F,

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_pX_p+e; \tag{13}$$

$\mathbf{X}=(X_1,\,\ldots,\,X_p)^T$ is the covariates' vector, Y is the response, $\beta=(\beta_0,\,\ldots,\,\beta_p)^T=(\beta_0(F),\,\ldots,\,\beta_p(F))^T$.

4.1.1 | The assumptions

($\mathcal{A}1$) The error, e, has mean zero and finite second moment.

($\mathcal{A}2$) Case $(\mathbf{x},y)$ is mixed with cases from model F with probability ϵ (model $F_{\epsilon,\mathbf{x},y}$).

The L_2-regression coefficients β are obtained minimizing Ee^2; E denotes expected value.

RINFIN has a simple form providing insight when an additional assumption is used:

($\mathcal{A}3$) $X_1,\,\ldots,X_p$ are uncorrelated random variables.

($\mathcal{A}3$) is not necessary to use *RINFIN* in practice; see (34) and Remark 5.

4.1.2 | Notation

The jth regression coefficient obtained by L_2-minimization at model $F_{\epsilon,\mathbf{u},v}$ is denoted by $\beta_j(F_{\epsilon,\mathbf{u},v}), j = 0, 1, \ldots, p$, and their vector by $\beta(F_{\epsilon,\mathbf{u},v})$.

Denote the L_2- residuals for model $F_{\epsilon,\mathbf{u},v}$ at $(\mathbf{x}, y)$ by

$$r(\mathbf{x}, y; F_{\epsilon,\mathbf{u},v}) = y - \beta_0(F_{\epsilon,\mathbf{u},v}) - \sum_{j=1}^{p} \beta_j(F_{\epsilon,\mathbf{u},v}) x_j; \qquad (14)$$

r is also used to denote $r(\mathbf{x}, y; F)$.

Add h in the ith component of $\mathbf{x}$, with $h \in R$, $|h|$ small, to obtain

$$\mathbf{x}_{i,h} = \mathbf{x} + (0, \ldots, h, \ldots, 0), \qquad (15)$$

such that $(\mathbf{x}_{i,h}, y)$, $(\mathbf{x}, y+h)$ are small perturbations of $(\mathbf{x}, y)$.

The influence function of β_j is evaluated at $(\mathbf{x}, y)$ for F, thus use

$$IF_j = IF(\mathbf{x}, y; \beta_j, F), \quad IF'_{v,j} = \frac{\partial IF(\mathbf{x}, y; \beta_j, F)}{\partial v}, \quad v = y, x_i, \ i = 1, \ldots, p, \qquad (16)$$

that is, in words, $IF'_{v,j}$ is the derivative of IF_j with respect to $v, j = 0, 1, \ldots, p$.

4.1.3 | Influence functions

Influence functions of L_2-regression coefficients at F are solutions of the system:

$$IF_0 + IF_1 EX_1 + \cdots + IF_p EX_p = r(\mathbf{x}, y; F), \qquad (17)$$

$$IF_0 EX_i + \cdots + IF_j EX_i X_j + \cdots + IF_p EX_i X_p = x_i r(\mathbf{x}, y; F), \quad i = 1, \ldots, p. \qquad (18)$$

Equations (17) and (18) are obtained by interchanging in the normal equations,

$$\frac{\partial E_H (Y - \beta_0 - \beta_1 X_1 - \cdots - \beta_p X_p)^2}{\partial \beta_i} = 0, \quad i = 0, 1, \ldots, p, \qquad (19)$$

the expected value with the partial derivatives and, after evaluation at the models $H = F$ and $H = (1-\epsilon)F + \epsilon\Delta_{(\mathbf{x},y)}$, subtracting the equations for the ith partial derivative for both models, dividing by ϵ and taking the limit as ϵ converges to zero.

The influence functions in (17) and (18) are now provided when, in addition, $(\mathcal{A}3)$ holds. With an additional assumption on the error, e, influence functions of L_1-regression coefficients have also been obtained ([36], Proposition 3.2).

Proposition 1. *For regression model* (13) *with assumptions* $(\mathcal{A}1)$–$(\mathcal{A}3)$ *and notation* (16), *the influence functions of L_2-regression coefficients at* $(\mathbf{x}, y)$ *for model F are*:

$$IF_0 = r\left[1 - p + \sum_{j=1}^{p} \frac{EX_j^2 - x_j EX_j}{\sigma_j^2}\right], \quad IF_j = r\frac{x_j - EX_j}{\sigma_j^2}, \quad j = 1, \ldots, p; \qquad (20)$$

$r = r(\mathbf{x}, y; F)$, σ_j^2 *is the variance of* $X_j, j = 1, \ldots, p$.

4.2 | Perturbations of L_2-residuals for models F and $F_{\epsilon,\mathbf{x},y}$

The goal is to compare small $(\mathbf{x}, y)$-residual changes in L_2 regressions for $F_{\epsilon,\mathbf{x},y}$ and F:

- **i.** when $(\mathbf{x}_{i,h}, y)$ replaces $(\mathbf{x}, y)$ in the ϵ-mixture, that is, under $F_{\epsilon,\mathbf{x},y}$ and $F_{\epsilon,\mathbf{x}_{i,h},y}$, $r(\mathbf{x}_{i,h}, y; F_{\epsilon,\mathbf{x}_{i,h},y}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y})$, and
- **ii.** when $(\mathbf{x}, y+h)$ replaces $(\mathbf{x}, y)$ in the ϵ-mixture, that is, under $F_{\epsilon,\mathbf{x},y}$ and $F_{\epsilon,\mathbf{x},y+h}$, $r(\mathbf{x}, y+h; F_{\epsilon,\mathbf{x},y+h}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y})$.

A lemma follows that is used repeatedly to calculate residuals' differences (i) and (ii).

Lemma 2. *For regression model* (13), $(\mathcal{A}1)$, $(\mathcal{A}2)$ *and* ϵ, $|h|$ *both small it holds*:

$$\beta_j(F_{\epsilon,\mathbf{x},y}) \approx \beta_j(F) + \epsilon IF_j, \quad \beta_j(F_{\epsilon,\mathbf{x}_{i,h},y}) \approx \beta_j(F_{\epsilon,\mathbf{x},y}) + \epsilon h \frac{\partial IF(\mathbf{x}, y; \beta_j, F)}{\partial x_i}, \quad 0 \le j \le p. \qquad (21)$$

Proposition 2. *For regression model* (13) *with* $(\mathcal{A}1)$, $(\mathcal{A}2)$, $\mathbf{x}_{i,h}$ *the perturbation of* $\mathbf{x}$ (*see* (15)) *and for* ϵ *and* $|h|$ *both small*:

- **a.** *the difference of* $(\mathbf{x}, y)$*-residuals at* $F_{\epsilon,\mathbf{x}_{i,h},y}$ *and* $F_{\epsilon,\mathbf{x},y}$ *is*:

$$r(\mathbf{x}_{i,h}, y; F_{\epsilon,\mathbf{x}_{i,h},y}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) + \beta_i h \approx -\epsilon h \left[IF_i + \frac{\partial IF_0}{\partial x_i} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i}\right], \quad i = 1, \ldots, p, \qquad (22)$$

- **b.** *the difference of* $(\mathbf{x}, y)$*-residuals at* $F_{\epsilon,\mathbf{x},y+h}$ *and* $F_{\epsilon,\mathbf{x},y}$ *is*:

$$r(\mathbf{x}, y+h; F_{\epsilon,\mathbf{x},y+h}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) - h \approx -\epsilon h \left[\frac{\partial IF_0}{\partial y} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial y}\right]. \qquad (23)$$

Remark 3. The right side of (22) involves influence functions and their derivatives. An index using it to detect bad leverage is less affected by masking than diagnostics based solely on influence functions, as explained in the Introduction.

Corollary 1. *Under the assumptions of Proposition* 2 *and* $(\mathcal{A}3)$, *with* $r = r(\mathbf{x}, y; F)$:

(a_1)

$$r(\mathbf{x}_{i,h}, y; F_{\epsilon,\mathbf{x}_{i,h},y}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) + \beta_i h \approx -\epsilon h \left\{ 2\frac{r(x_i - EX_i)}{\sigma_i^2} - \beta_i \left[1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2} \right] \right\}. \quad (24)$$

(a_2) *If*, *in addition*, $|x_i|$ *is large*,

$$r(\mathbf{x}_{i,h}, y; F_{\epsilon,\mathbf{x}_{i,h},y}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) \approx \epsilon h \cdot 3\beta_i \frac{(x_i - EX_i)^2}{\sigma_i^2}, \quad (25)$$

(b)

$$r(\mathbf{x}, y + h; F_{\epsilon,\mathbf{x},y+h}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) - h \approx -\epsilon h \left[1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2} \right]. \quad (26)$$

4.3 | Residual's influence index, *RINFIN*$(\mathbf{x}, y; \epsilon, \beta)$

Local influence of $(\mathbf{x}, y)$ is determined using the distance of residuals' partial derivatives at $(\mathbf{x}, y)$ for model F and gross-error model $F_{\epsilon,\mathbf{x},y}$. The larger this distance is, the larger the local influence of $(\mathbf{x}, y)$ is.

4.3.1 | Local $\mathbf{x}$-influence on L_2-residuals

For $(\mathbf{x}_{i,h}, y)$ and $(\mathbf{x}, y)$ both under model F,

$$\frac{r(\mathbf{x}_{i,h}, y; F) - r(\mathbf{x}, y; F)}{h} + \beta_i = 0, \quad i = 1, \ldots, p. \quad (27)$$

For gross-error models $F_{\epsilon,\mathbf{x},y}$, $F_{\epsilon,\mathbf{x}_{i,h},y}$, the difference in partial derivatives of residuals is obtained from (22) for small ϵ,

$$\lim_{h\to 0} \frac{r(\mathbf{x}_{i,h}, y; F_{\epsilon,\mathbf{x}_{i,h},y}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y})}{h} + \beta_i \approx -\epsilon \left[IF_i + \frac{\partial IF_0}{\partial x_i} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i} \right], \quad i = 1, \ldots, p. \quad (28)$$

From (27) and (28), the right side of (28) measures influence of $\mathbf{x}$'s ith coordinate in the residual's derivative and provides the motivation for defining influence.

Definition 1. For gross-error model $F_{\epsilon,\mathbf{x},y}$,

a. the influence of $\mathbf{x}$'s ith coordinate in the L_2-residual is

$$INF(i) = \epsilon \cdot \left| IF_i + \frac{\partial IF_0}{\partial x_i} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i} \right|, \quad i = 1, \ldots, p. \quad (29)$$

b. The L_2-*RINFIN* is

$$RINFIN(\mathbf{x}, y; \epsilon, \beta) = \epsilon \cdot \sum_{i=1}^{p} \left(IF_i + \frac{\partial IF_0}{\partial x_i} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i} \right)^2. \quad (30)$$

Remark 4. Replacing in (30) the squares by absolute values, *RINFINABS*$(\mathbf{x}, y; \epsilon, \beta)$ is obtained, which is used to complement *RINFIN*'s ordering as described in Section 2.2. The sum next to ϵ in (30) can be divided by p, both for *RINFIN* and *RINFINABS*.

The equations' system (17) and (18) can be written in matrix notation

$$\widetilde{\mathcal{E}} \cdot \mathbf{IF} = \mathbf{q}(\mathbf{x}, y; \beta); \quad (31)$$

$\widetilde{\mathcal{E}}$ is the symmetric matrix of EX_i, EX_iX_j and 1, $1 \le i, j \le p$, **IF** is the vector of β-influence functions and

$$\mathbf{q} = (r(\mathbf{x}, y; F), x_1 r(\mathbf{x}, y; F), \ldots, x_p r(\mathbf{x}, y; F))^T.$$

Using the notation,

$$\mathcal{E}^* = (e_{ij}^*) = \widetilde{\mathcal{E}}^{-1}, \quad 0 \le i, j \le p, \quad (32)$$

it is shown in Lemma B.2 for regression model (13) under $(\mathcal{A}1)$, $(\mathcal{A}2)$, $(\mathbf{x}, y) \in R^{p+1}$, that

$$INF[i] = \epsilon \cdot \left| 2r \left[e_{i0}^* + \sum_{k=1}^{p} e_{ik}^* x_k \right] - \beta_i \left(e_{00}^* + 2\sum_{j=1}^{p} x_j e_{j0}^* + \mathbf{x}^T \mathcal{E}^* \mathbf{x} \right) \right|, \quad i = 1, \ldots, p, \quad (33)$$

and (30) is written

$$RINFIN(\mathbf{x}, y; \epsilon, \beta) = \epsilon \cdot \sum_{i=1}^{p} \left\{ 2r \left[e_{i0}^* + \sum_{k=1}^{p} e_{ik}^* x_k \right] - \beta_i \left(e_{00}^* + 2\sum_{j=1}^{p} x_j e_{j0}^* + \mathbf{x}^T \mathcal{E}^* \mathbf{x} \right) \right\}^2. \quad (34)$$

Remark 5. To calculate the *RINFIN* value (34) of $(\mathbf{x}_i, y_i)$, the rest of the sample is used to estimate β, $\widetilde{\mathcal{E}}$, $\mathcal{E}^*$, $i = 1, \ldots, n$. If one $|\widehat{\beta}_i|$ is near zero, the effect of "bad leverage" remains in another term in (34).

Assuming in addition $(\mathcal{A}3)$ and using (B2), a more accessible and insightful form of *RINFIN* is obtained.

$$RINFIN(\mathbf{x}, y; \epsilon, \beta) = \epsilon \cdot \sum_{i=1}^{p} \left\{ 2\frac{r(\mathbf{x}, y; F)(x_i - EX_i)}{\sigma_i^2} - \beta_i \left[1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2} \right] \right\}^2. \quad (35)$$

Proposition 3. *Under* $(\mathcal{A}1)$–$(\mathcal{A}3)$, *with G unit mass at* $(\mathbf{x}, y)$, $\epsilon = 1/n$,

$$\lim_{|x_i| \to \infty} RINFIN(\mathbf{x}, y; \epsilon, \beta) = \infty. \quad (36)$$

4.3.2 | Local y-influence on L_2-residuals

For $(\mathbf{x}, y + h)$ and $(\mathbf{x}, y)$ both under model F,

$$\frac{r(\mathbf{x}, y + h; F) - r(\mathbf{x}, y; F)}{h} = 1, \quad i = 1, \ldots, p. \quad (37)$$

Proposition 4. *For models* $F, F_{\epsilon,\mathbf{x},y}, F_{\epsilon,\mathbf{x},y+h}$, ϵ *small and* L_2 *regression under* $(\mathcal{A}1) - (\mathcal{A}3)$:

$$\lim_{h \to 0} \frac{r(\mathbf{x}, y + h; F_{\epsilon,\mathbf{x},y+h}) - r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y})}{h} - 1 \approx -\epsilon \left[1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2} \right]. \quad (38)$$

Remark 6. From (38), the y-influence index is

$$\sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2}; \quad (39)$$

it is maximized for cases in the extremes of the $\mathbf{x}$-coordinates. Thus, *RINFIN* is restricted to the influence of factor space cases.

4.4 | Large sample properties of $RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n)$

Consistency of $RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n)$ and its asymptotic distribution follow from properties of the least squares estimate $\widehat{\beta}_n$. For Proposition 5 the notation is changed: $\mathbf{X}(\in R^{p+1})$ has 1 as first coordinate, $\widetilde{\mathcal{E}}$ is $E\mathbf{X}\mathbf{X}^T$ and $\mathbf{x}(\in R^p)$ still denotes a factor space vector.

Proposition 5. [2] *Let* $(\mathbf{X}, Y), (\mathbf{X_1}, Y_1), \ldots, (\mathbf{X_n}, Y_n)$ *be independent, identically distributed random vectors with form* $\mathbf{X}^T = (1, X_1, \ldots, X_p) \in R^{p+1}$, $Y \in R$,

$$Y = \mathbf{X}^T\beta + \epsilon. \quad (40)$$

Let $\widehat{\beta}_n$ *be the least squares estimate of* β.

a. *Assume that* (i) *Rank* $\widetilde{\mathcal{E}} = Rank\ E\mathbf{X}\mathbf{X}^T = p+1$, (ii) $E\mathbf{X}\epsilon = \mathbf{0}$, (iii) $E\epsilon^2 < \infty$.
Then, for every $(\mathbf{x}, y) \in R^{p+1}$, $RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n)$ *is consistent estimate for* $RINFIN(\mathbf{x}, y; \epsilon, \beta)$, $\epsilon > 0$.

b. *Assume in addition to* (i) *and* (ii) *in* (a): (iv) $E\epsilon^4 < \infty$ *and* $E\|\mathbf{X}\|_2^4 < \infty$; $\|\mathbf{u}\|_2$ *denotes the Euclidean* L_2 *norm of vector* $\mathbf{u}$. (v) *For at least one* β*-coordinate, for example, the ith*:

$$g_i = \frac{\partial RINFIN(\mathbf{x}, y; \epsilon, \beta)}{\partial \beta_i} \neq 0. \quad (41)$$

Then, $RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n)$ *is asymptotically normal*:

$$\sqrt{n}[RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n) - RINFIN(\mathbf{x}, y; \epsilon, \beta)] \xrightarrow{\mathcal{D}} N(0, \mathbf{g}^T V \mathbf{g}); \quad (42)$$

$V = \widetilde{\mathcal{E}}^{-1} E(\mathbf{X}_i \mathbf{X}_i^T \epsilon_i^2) \widetilde{\mathcal{E}}^{-1}$ *is the covariance matrix of the asymptotic normal distribution of* $\widehat{\beta}_n$ *and* $\mathbf{g}$ *has coordinates* g_i *in* (41), $i = 0, 1, \ldots, p$.

Remark 7. *RINFIN*'s advantage, that is, making additive the effects of $\mathbf{x}$ and r, remains for L_2-regression with diagonal weight matrix, W, independent of $\mathbf{x}$, r; Proposition 5 still holds with known V (W) in (42). When W depends on $\mathbf{x}$, r, the decomposition of the influence function in Dollinger and Staudte [10, Theorem 3, Equation (2)] indicates that *RINFIN*'s advantage may not hold, depending on the form of the weights.

ACKNOWLEDGMENTS

Many thanks are due to the Professor Bertrand Clarke, Editor, the AE, and referees for the comments that improved the presentation of the paper. Many thanks are also due to Professor Douglas Hawkins, for the useful comments and the encouragement about this work, and to Professor Chenlei Leng, who has kindly communicated to us earlier the results in Zhao et al. [38] and provided the microarray data used in applications. Thanks are due to Miss Eleni Yatracos, for confirming in 2017 properties of the $\mathcal{E}$-matrix and an erroneous referee's comment about *RINFIN*.

[2] Provided for completeness and for interested readers, even though *P*-values are not used herein.

DATA AVAILABILITY STATEMENT

Data source appears in the references.

ORCID

Yannis G. Yatracos https://orcid.org/0000-0001-9972-3337

REFERENCES

1. A. Alfons, C. Croux, and S. Gelper, *Sparse least trimmed squares regression for analyzing high-dimensional large data sets*, Ann. Appl. Stat. 7 (2013), 226–248.
2. D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression diagnostics: Identifying influential data and sources of collinearity*, Wiley, New York, 1980.
3. G. Boente, A. M. Pires, and I. M. Rodrigues, *Influence functions and outlier detection under the common principal components model: A robust approach*, Biometrika 89 (2002), 861–875.
4. N. A. Campbell, *The influence function as an aid in outlier detection in discriminant analysis*, Appl. Statist. 27 (1978), 251–258.
5. R. J. Carroll and D. Ruppert, *Transformations in regression: A robust Analysis*, Technometrics 27 (1985), 1–12.
6. A. P. Chiang, J. S. Beck, H. J. Yen, M. K. Tayeh, T. E. Scheetz, R. E. Swiderski, D. Y. Nishimura, T. A. Braun, K. Y. Kim, J. Huang, K. Elbedour, R. Carmi, D. C. Slusarski, T. L. Casavant, E. M. Stone, and V. C. Sheffield, *Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11)*, Proc. Natl. Acad. Sci. U. S. A. 103 (2006), 6287–6292.
7. R. D. Cook, *Detection of influential observations in linear regression*, J. Amer. Statist. Assoc. 74 (1977), 169–174.
8. R. D. Cook, *Assessment of local influence (with discussion)*, J. R. Stat. Soc. Ser. B 48 (1986), 133–169.
9. R. D. Cook and S. Weisberg, *Characterizations of an empirical influence function for detecting influential cases in regression*, Technometrics 22 (1980), 495–508.
10. M. B. Dollinger and R. G. Staudte, *Influence functions of iteratively reweighted least squares estimators*, J. Amer. Statist. Assoc. 86 (1991), 709–716.
11. M. G. Genton and A. Ruiz-Gazen, *Visualizing influential observations in dependent data*, J. Comput. Graph. Statist. 19 (2010), 808–825.
12. A. S. Hadi and J. S. Simonoff, *Procedures for the identification of multiple outliers in linear models*, J. Amer. Statist. Assoc. 88 (1993), 1264–1272.
13. F. R. Hampel, *A general qualitative definition of robustness*, Ann. Math. Stat. 42 (1971), 1887–1896.
14. F. R. Hampel, *The influence curve and its role in robust estimation*, J. Amer. Statist. Assoc. 69 (1974), 383–394.
15. F. R. Hampel, *The breakdown point of the mean combined with some rejection rules*, Technometrics 27 (1985), 95–107.
16. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust statistics: The approach based on influence functions*, Wiley, New York, 1986.
17. D. M. Hawkins, *The identification of outliers*, Chapman and Hall, London, 1980.
18. D. M. Hawkins, D. Bradu, and G. V. Kass, *Location of several outliers in multiple regression data using elemental sets*, Technometrics 26 (1984), 197–208.
19. P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Stat. 35 (1964), 73–101.
20. P. J. Huber, *Robust statistics*, Wiley, New York, 1981.
21. M. Hubert, P. J. Rousseeuw, and S. Van Aelst, *High-breakdown robust multivariate methods*, Stat. Sci. 23 (2008), 92–119.
22. A. J. Lawrance, *Regression transformation diagnostics using local influence*, J. Amer. Statist. Assoc. 83 (1988), 1067–1072.
23. C. Leng, Private communication, 2017.
24. V. Ollerer, C. Croux, and A. Alfons, *The influence function of penalized regression estimators*, Statistics 49 (2015), 741–765.
25. P. J. Rousseeuw and A. M. Leroy, *Robust regression & outlier detection*, Wiley, New York, 1987.
26. P. J. Rousseeuw and B. C. van Zomeren, *Unmasking multivariate outliers and leverage points*, J. Amer. Statist. Assoc. 85 (1990), 633–639.
27. R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, New York, 1980.
28. Y. She and A. B. Owen, *Outlier detection using nonconvex penalized regression*, J. Amer. Statist. Assoc. 106 (2011), 626–639.
29. J. W. Tukey, *The future of data analysis*, Ann. Math. Stat. 33 (1962), 1–68.
30. P. F. Velleman and R. E. Welsch, *Efficient computing of regression diagnostics*, Amer. Statist. 35 (1981), 234–242.
31. R. Wasserstein and N. Lazar, *The ASA's statement on* p*-values: Context, process, and purpose*, Amer. Statist. 70 (2016), 129–133.
32. S. S. Weisberg, *Applied linear regression*, 2nd ed., Wiley, New York, 1985.
33. R. E. Welsch, *Influence functions and regression diagnostics*, in *Modern Data Analysis*, Academic Press, New York, 1982.
34. Y. G. Yatracos, *Detecting clusters in the data from variance decompositions of its projections*, J. Classif. 30 (2013), 30–55.
35. Y. G. Yatracos, *Discussion of "Random-projection ensemble classification" by T. I. Cannings and R. J. Samworth*, J. R. Stat. Soc. Ser. B 79 (2017), 1026–1027.
36. Y. G. Yatracos, *The derivative of influence function, location breakdown point, group influence and regression residuals' plots*. https://arxiv.org/pdf/1607.04384v3.pdf, 2018
37. J. Zhao, C. Leng, L. Li, and H. Wang, *High-dimensional influence measure*, Ann. Stat. 41 (2013), 2639–2667.
38. J. Zhao, C. Liu, L. Niu, and C. Leng, *Multiple influential point detection in high dimensional regression spaces*, J. R. Stat. Soc. Ser. B 81 (2019), 385–408. https://arxiv.org/abs/1609.03320

How to cite this article: Y. G. Yatracos, *Residual's influence index (*RINFIN*), bad leverage and unmasking in high dimensional* L_2*-regression*, Stat. Anal. Data Min.: ASA Data Sci. J. (2021), 1–14. https://doi.org/10.1002/sam.11550

APPENDIX A. $\mathcal{E}$-MATRIX

To prove Proposition 1, the general form of a symmetric, $(n+1)$ by $(n+1)$ matrix $\mathcal{E}_n$ is introduced. $\mathcal{E}_n$'s entries are motivated by the expected values in the equations' system (17) and (18) when the n covariates are uncorrelated.

$\mathcal{E}_n$'s cofactors are obtained and used to determine in *closed form* the influence functions of L_2-regression coefficients.

Under assumption ($\mathcal{A}3$), the coefficients in the system of Equations (17) and (18) form $\mathcal{E}_n$-matrix; n is the covariates' dimension. As an illustration, for real numbers a, b, c, A, B, C,

$$\mathcal{E}_3 = \begin{pmatrix} 1 & a & b & c \\ a & A & ab & ac \\ b & ba & B & bc \\ c & ca & cb & C \end{pmatrix}.$$

For $\mathcal{E}_3$, the corresponding linear regression model with uncorrelated covariates X_1, X_2, X_3 provides $a = EX_1$, $b = EX_2$, $c = EX_3$ and $A = EX_1^2$, $B = EX_2^2$, $C = EX_3^2$.

Definition A.1. The $\mathcal{E}_n$-matrix[3] with real entries has form:

$$\mathcal{E}_n = \begin{pmatrix} 1 & a_1 & a_2 \ldots & a_n \\ a_1 & A_1 & a_1a_2 \ldots & a_1a_n \\ a_2 & a_2a_1 & A_2 \ldots & a_2a_n \\ \ldots & & & \\ a_n & a_na_1 & a_na_2 \ldots & A_n \end{pmatrix}. \tag{A1}$$

Notation: $\mathcal{E}_{n,-k}$ denotes the matrix obtained from $\mathcal{E}_n$ by deleting its kth column and kth row, $2 \le k \le n+1$.

Property of $\mathcal{E}_n$-matrix: Deleting the kth row and the kth column of $\mathcal{E}_n$-matrix, the obtained matrix $\mathcal{E}_{n,-k}$ is $\mathcal{E}_{n-1}$ matrix formed by $\{1, a_1, \ldots, a_n\} - \{a_{k-1}\}$, $2 \le k \le n+1$.

The cofactors of $\mathcal{E}_n$-matrix are needed to solve the system of Equations (17) and (18).

Proposition A.1.

a. *The determinant of $\mathcal{E}_n$-matrix* (A1) *is*

$$|\mathcal{E}_n| = \Pi_{m=1}^{n}(A_m - a_m^2). \tag{A2}$$

b. *Let $C_{i+1,j+1}$ be the cofactor of element $(i+1, j+1)$ in $\mathcal{E}_n$, then*:

$$\begin{aligned} C_{i+1,j+1} &= 0, \ \text{if } i > 0, \ j > 0, \ i \ne j, \\ C_{1,j+1} &= -a_j \Pi_{k \ne j}(A_k - a_k^2). \\ C_{i+1,1} &= -a_i \Pi_{j \ne i}(A_j - a_j^2), \ \text{if } i > 0, \\ C_{1,1} &= |\mathcal{E}_n| + \sum_{k=1}^{n} a_k^2 |\mathcal{E}_{n,-k}|. \end{aligned}$$

Proof for Proposition A.1.

a. Induction is used.

For $n = 1$, the determinant is $A_1 - a_1^2$.

For $n = 2$, the determinant is

$$\begin{aligned} &(A_1A_2 - a_1^2a_2^2) - a_1 \cdot (a_1A_2 - a_1a_2^2) + a_2 \cdot (a_1^2a_2 - A_1a_2) \\ &\quad = A_1A_2 - a_1^2A_2 + a_1^2a_2^2 - A_1a_2^2 \\ &\quad = A_2(A_1 - a_1^2) - a_2^2(A_1 - a_1^2) = (A_1 - a_1^2)(A_2 - a_2^2). \end{aligned}$$

Assume that (A2) holds also for $\mathcal{E}_n$. It is enough to show (A2) holds for

$$\mathcal{E}_{n+1} = \begin{pmatrix} 1 & a_1 & a_2 \ldots & a_n & a_{n+1} \\ a_1 & A_1 & a_1a_2 \ldots & a_1a_n & a_1a_{n+1} \\ a_2 & a_2a_1 & A_2 \ldots & a_2a_n & a_2a_{n+1} \\ \ldots & & & & \\ a_n & a_na_1 & a_na_2 \ldots & A_n & a_na_{n+1} \\ a_{n+1} & a_{n+1}a_1 & a_{n+1}a_2 \ldots & a_{n+1}a_n & A_{n+1} \end{pmatrix}.$$

$|\mathcal{E}_{n+1}|$ is obtained using line $(n+1)$ and its cofactors $C_{n+1,1}, \ldots, C_{n+1,n+1}$:

$$\begin{aligned} |\mathcal{E}_{n+1}| = a_{n+1}C_{n+1,1} + a_{n+1}a_1C_{n+1,2} + \cdots \\ + a_{n+1}a_nC_{n+1,n} + A_{n+1}C_{n+1,n+1}. \end{aligned} \tag{A3}$$

Observe that for $2 \le j \le n$, cofactor $C_{n+1,j}$ is obtained from a matrix where the last column is a multiple of its first column by a_{n+1}, thus,

$$C_{n+1,j} = 0, \ \ j = 2, \ldots, n. \tag{A4}$$

For the matrix in cofactor $C_{n+1,1}$, observe that in its last column a_{n+1} is common factor and if taken out of the determinant the remaining column is the vector generating $\mathcal{E}_n$, that is, $\{1, a_1, \ldots, a_n\}$. With $n-1$ successive interchanges to the left, this column becomes first and $\mathcal{E}_n$ appears. Thus,

$$C_{n+1,1} = (-1)^{n+2}(-1)^{n-1} \cdot a_{n+1}|\mathcal{E}_n| = -a_{n+1}|\mathcal{E}_n|. \tag{A5}$$

In cofactor $C_{n+1,n+1}$, the determinant is that of $\mathcal{E}_n$,

$$C_{n+1,n+1} = (-1)^{2(n+1)}|\mathcal{E}_n| = |\mathcal{E}_n|. \tag{A6}$$

From (A3)–(A6) it follows that

$$|\mathcal{E}_{n+1}| = -a_{n+1}^2|\mathcal{E}_n| + A_{n+1}|\mathcal{E}_n| = \Pi_{m=1}^{n+1}(A_m - a_m^2).$$

b. We now work with $\mathcal{E}_n$. For $i > 0$, $j > 0$, $i \ne j$, after deleting row $(j+1)$ the remaining of column $(j+1)$ in the cofactor is a multiple of column 1, thus $|C_{i+1,j+1}|$ vanishes.

[3] $\mathcal{E}$ for Eleni.

For $C_{1,j+1}$, using column $j+1$ to calculate $\mathcal{E}_n$, it holds:

$$a_jC_{1,j+1} + A_jC_{j+1,j+1} = |\mathcal{E}_n| \Rightarrow a_jC_{1,j+1} = -a_j^2\Pi_{k\neq j}(A_k - a_k^2) \Rightarrow C_{1,j+1} = -a_j\Pi_{k\neq j}(A_k - a_k^2).$$

For $C_{i+1,1}$, $i>0$, after deletion of row $(i+1)$ in $\mathcal{E}_n$ the remaining of column $(i+1)$ in the cofactor's matrix is multiple of a_i and the basic vector creating $\mathcal{E}_{n,-i}$. Column 1 of $\mathcal{E}_n$ is also deleted and for column $(i+1)$ in the cofactor's matrix to become first column $(i-1)$ exchanges of columns are needed. Thus,

$$C_{i+1,1} = (-1)^{i+2}\cdot a_i\cdot(-1)^{i-1}\Pi_{k\neq i}(A_k - a_k^2) = -a_i\cdot\Pi_{k\neq i}(A_k - a_k^2).$$

For $C_{1,1}$ we express $|\mathcal{E}_n|$ as sum of cofactors along the first row of $\mathcal{E}_n$,

$$C_{1,1} + a_1C_{1,2} + \cdots + a_nC_{1,n} = |\mathcal{E}_n| \Rightarrow C_{1,1} = \Pi_{k=1}^n(A_k - a_k^2) + a_1^2\Pi_{k\neq 1}(A_k - a_k^2) + \cdots + a_n^2\Pi_{k\neq n}(A_k - a_k^2).$$

APPENDIX B. PROOFS

Proof of Lemma 1. Equality (7) is obtained by adding and subtracting $T(F)$ in the numerator of its left side and by taking first the limit with respect to ϵ.

Proof of Proposition 1. For system of Equations (17), (18) and matrix $\mathcal{E}_p$ with $a_j = EX_j$, $A_j = EX_j^2$, $j = 1, \ldots, p$, from Proposition A.1 with $r = r(\mathbf{x}, y; F)$,

$$IF_j = \frac{C_{1,j+1}r + C_{j+1,j+1}rx_j}{|\mathcal{E}_p|} = r\frac{-EX_j\Pi_{k\neq j}\sigma_k^2 + x_j\Pi_{k\neq j}\sigma_k^2}{\Pi_{k=1}^p\sigma_k^2} = r\frac{x_j - EX_j}{\sigma_j^2},\quad j = 1, \ldots, p.$$

$$IF_0 = \frac{C_{1,1}r + \sum_{j=1}^p C_{1,j+1}rx_j}{|\mathcal{E}_p|} = r\frac{\Pi_{k=1}^p\sigma_j^2 + \sum_{j=1}^p(EX_j)^2\Pi_{k\neq j}\sigma_k^2 - \sum_{j=1}^p x_jEX_j\Pi_{k\neq j}\sigma_k^2}{\Pi_{k=1}^p\sigma_k^2} = r\left[1 + \sum_{j=1}^p\frac{EX_j^2 - \sigma_j^2 - x_jEX_j}{\sigma_j^2}\right] = r\left[1 - p + \sum_{j=1}^p\frac{EX_j^2 - x_jEX_j}{\sigma_j^2}\right].$$

Lemma B.1. *For the influence functions* (20) *with* $r = r(\mathbf{x}, y; F)$ *it holds*:

(a)
$$IF_0 + \sum_{j=1}^p x_jIF_j = r\left[1 + \sum_{j=1}^p\frac{(x_j - EX_j)^2}{\sigma_j^2}\right], \tag{B1}$$

(b)
$$IF_i + IF'_{x_i,0} + \sum_{j=1}^p x_jIF'_{x_i,j} = 2\frac{r\cdot(x_i - EX_i)}{\sigma_i^2} - \beta_i\left[1 + \sum_{j=1}^p\frac{(x_j - EX_j)^2}{\sigma_j^2}\right]. \tag{B2}$$

$$\approx -3\beta_i\frac{(x_i - EX_i)^2}{\sigma_i^2},\ \textit{if } |x_i - EX_i| \textit{ is very large}, \tag{B3}$$

(c)
$$IF'_{y,0} + \sum_{j=1}^p x_jIF'_{y,j} = 1 + \sum_{j=1}^p\frac{(x_j - EX_j)^2}{\sigma_j^2}. \tag{B4}$$

Proof of Lemma B.1.

a. From (20),

$$IF_0 + \sum_{j=1}^p x_jIF_j = r\left[1 - p + \sum_{j=1}^p\frac{EX_j^2 - x_jEX_j}{\sigma_j^2}\right] + \sum_{j=1}^p x_j\frac{r(x_j - EX_j)}{\sigma_j^2} = r\left[1 - p + \sum_{j=1}^p\frac{EX_j^2 - 2x_jEX_j + x_j^2}{\sigma_j^2}\right] = r\left[1 + \sum_{j=1}^p\frac{(x_j - EX_j)^2}{\sigma_j^2}\right].$$

b. Proof is provided for $i = 1$. Since

$$IF_0 = r\left[1 - p + \sum_{j=1}^p\frac{EX_j^2 - x_jEX_j}{\sigma_j^2}\right],$$

$$IF_j = r\frac{x_j - EX_j}{\sigma_j^2},\quad j = 1, \ldots, p,$$

$$IF'_{x_1,0} = -\beta_1\left[1 - p + \sum_{j=1}^p\frac{EX_j^2 - x_jEX_j}{\sigma_j^2}\right] - r\frac{EX_1}{\sigma_1^2}$$

$$IF'_{x_1,1} = -\beta_1\frac{x_1 - EX_1}{\sigma_1^2} + \frac{r}{\sigma_1^2} \Rightarrow x_1IF'_{x_1,1} = -\beta_1\frac{x_1^2 - x_1EX_1}{\sigma_1^2} + r\frac{x_1}{\sigma_1^2}$$

$$IF'_{x_1,j} = -\beta_1\frac{x_j - EX_j}{\sigma_j^2} \Rightarrow x_jIF'_{x_1,j} = -\beta_1\frac{x_j^2 - x_jEX_j}{\sigma_j^2},\quad j \neq 1.$$

Thus,

$$
\begin{aligned}
&x_1 IF'_{x_1,1} + x_2 IF'_{x_1,2} + \cdots + x_p IF'_{x_1,p} \\
&\quad = r\frac{x_1}{\sigma_1^2} - \beta_1 \sum_{j=1}^{p} \frac{x_j^2 - x_j EX_j}{\sigma_j^2} \\
&\quad \Rightarrow IF_1 + IF'_{x_1,0} + x_1 IF'_{x_1,1} + x_2 IF'_{x_1,2} + \cdots + x_p IF'_{x_1,p} \\
&\quad = 2\frac{r(x_1 - EX_1)}{\sigma_1^2} - \beta_1 \left[1 - p + \sum_{j=1}^{p} \frac{x_j^2 - 2x_j EX_j + EX_j^2}{\sigma_j^2} \right] \\
&\quad = 2\frac{r(x_1 - EX_1)}{\sigma_1^2} - \beta_1 \left[1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2} \right].
\end{aligned}
$$

Since

$$
\begin{aligned}
r(x_1 - EX_1) &= y(x_1 - EX_1) - \beta_1 x_1 (x_1 - EX_1) \\
&\quad - (x_1 - EX_1) \sum_{j=2}^{p} \beta_j x_j \\
&= y(x_1 - EX_1) - \beta_1 (x_1 - EX_1)^2 - \beta_1 (x_1 - EX_1) EX_1 \\
&\quad - (x_1 - EX_1) \sum_{j=2}^{p} \beta_j x_j,
\end{aligned}
$$

if $|x_1 - EX_1|$ is very large dominating all the other terms, then

$$
\begin{aligned}
&IF_1 + IF'_{x_1,0} + x_1 IF'_{x_1,1} + x_2 IF'_{x_1,2} + \cdots + x_p IF'_{x_1,p} \\
&\quad \approx -3\beta_1 \frac{(x_1 - EX_1)^2}{\sigma_1^2}.
\end{aligned}
$$

c. From (20),

$$
IF'_{y,0} = 1 - p + \sum_{j=1}^{p} \frac{EX_j^2 - x_j EX_j}{\sigma_j^2}, \quad IF'_{y,j} = \frac{x_j - EX_j}{\sigma_j^2}, \quad j = 1, \ldots, p.
$$

Thus,

$$
\begin{aligned}
IF'_{y,0} + \sum_{j=1}^{p} x_j IF'_{y,j} &= 1 - p \\
&\quad + \sum_{j=1}^{p} \frac{EX_j^2 - x_j EX_j + x_j^2 - x_j EX_j}{\sigma_j^2} \\
&= 1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2}.
\end{aligned}
$$

Proof of Lemma 2. Use approximations (4) and (8).

Proof of Proposition 2.

a. Is provided for $i = 1$ using repeatedly Lemma 2:

$$
\begin{aligned}
r(\mathbf{x}_{1,h}, y; F_{\epsilon,\mathbf{x}_{1,h},y}) &= y - \beta_0(F_{\epsilon,\mathbf{x}_{1,h},y}) \\
&\quad - \beta_1(F_{\epsilon,\mathbf{x}_{1,h},y})(x_1 + h) - \cdots - \beta_p(F_{\epsilon,\mathbf{x}_{1,h},y}) x_p \\
&\approx y - \{\beta_0(F_{\epsilon,\mathbf{x},y}) + \epsilon h IF'_{x_1,0}\} \\
&\quad - \{\beta_1(F_{\epsilon,\mathbf{x},y}) + \epsilon h IF'_{x_1,1}\}(x_1 + h) \\
&\quad - \cdots - \{\beta_p(F_{\epsilon,\mathbf{x},y}) + \epsilon h IF'_{x_1,p}\} x_p \\
&= r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) - \beta_1(F_{\epsilon,\mathbf{x},y}) h \\
&\quad - \epsilon h [IF'_{x_1,0} + x_1 IF'_{x_1,1} + x_2 IF'_{x_1,2} + \cdots + x_p IF'_{x_1,p}] \\
&\quad - \epsilon h^2 IF'_{x_1,1} \\
&= r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) - \beta_1 h \\
&\quad - \epsilon h [IF_1 + IF'_{x_1,0} + x_1 IF'_{x_1,1} + x_2 IF'_{x_1,2} + \cdots + x_p IF'_{x_1,p}] \\
&\quad - \epsilon h^2 IF'_{x_1,1}.
\end{aligned}
$$

b. Lemma 2 is also used.

$$
\begin{aligned}
r(\mathbf{x}, y + h; F_{\epsilon,\mathbf{x},y+h}) &= y + h - \beta_0(F_{\epsilon,\mathbf{x},y+h}) \\
&\quad - \beta_1(F_{\epsilon,\mathbf{x},y+h}) x_1 - \cdots - \beta_p(F_{\epsilon,\mathbf{x},y+h}) x_p \\
&\approx y + h - \{\beta_0(F_{\epsilon,\mathbf{x},y}) + \epsilon h IF'_{y,0}\} \\
&\quad - \{\beta_1(F_{\epsilon,\mathbf{x},y}) + \epsilon h IF'_{y,1}\} x_1 \\
&\quad - \cdots - \{\beta_p(F_{\epsilon,\mathbf{x},y}) + \epsilon h IF'_{y,p}\} x_p \\
&= r(\mathbf{x}, y; F_{\epsilon,\mathbf{x},y}) + h - \epsilon h \left[IF'_{y,0} + \sum_{j=1}^{p} x_j IF'_{y,j} \right].
\end{aligned}
$$

Proof of Corollary 1.

- (a_1) The right side of (24) follows from (B2).
- (a_2) If $|x_i|$ is large and $|h|$ is small, $\beta_i h$ and $\epsilon h^2 IF'_{x_i,i}$ are of smaller order than the remaining terms and (B3) implies (25).
- (b) The right side of (26) follows from (B4).

Proof of Proposition 3.

$$
\begin{aligned}
&\lim_{|x_i| \to \infty} RINFIN(\mathbf{x}, y; \epsilon, L_2) \geq \epsilon \cdot \\
&\lim_{|x_i| \to \infty} \left\{ 2\frac{r(\mathbf{x}, y)(x_i - EX_i)}{\sigma_i^2} - \beta_i \left[1 + \sum_{j=1}^{p} \frac{(x_j - EX_j)^2}{\sigma_j^2} \right] \right\}^2 \\
&\approx \epsilon \cdot \lim_{|x_i| \to \infty} 3^2 \beta_i^2 \frac{(x_i - EX_i)^4}{\sigma_i^4} = \infty;
\end{aligned}
$$

the last approximation follows from (B3).

Proof of Proposition 4. Follows from (26) dividing both its sides by h and taking the limit with h converging to zero.

Lemma B.2. *For regression model* (13) *under* ($\mathcal{A}1$), ($\mathcal{A}2$), $(\mathbf{x}, y) \in R^{p+1}$,

$$
\begin{aligned}
INF[i] = \epsilon \cdot \Bigg| & 2r \left[e_{i0}^* + \sum_{k=1}^{p} e_{ik}^* x_k \right] \\
& - \beta_i \left(e_{00}^* + 2\sum_{j=1}^{p} x_j e_{j0}^* + \mathbf{x}^T \mathcal{E}^* \mathbf{x} \right) \Bigg|, \quad i = 1, \ldots, p.
\end{aligned} \tag{B5}
$$

Proof of Lemma B.2. From (31),

$$\mathbf{IF} = \mathcal{E}^* \cdot \mathbf{q}, \quad \mathcal{E}^* = (e^*_{ij}) = \widetilde{\mathcal{E}}^{-1}, 0 \le i, j \le p. \tag{B6}$$

The influence function of β_j has form

$$IF_j = \sum_{k=0}^{p} e^*_{jk} q_k(\mathbf{x}, y; \beta) = r e^*_{j0} + r \sum_{k=1}^{p} e^*_{jk} x_k, \quad j = 0, 1, \ \dots \ , p. \tag{B7}$$

For $j = 0, 1, \ \dots \ , p, i = 1, \ \dots \ , p$

$$\frac{\partial IF_j}{\partial x_i} = e^*_{j0} \frac{\partial r}{\partial x_i} + \sum_{k=1}^{p} e^*_{jk} \frac{\partial \ (x_k \cdot r)}{\partial x_i}$$

$$= -\beta_i \left(e^*_{j0} + \sum_{k=1}^{p} e^*_{jk} x_k \right) + r e^*_{ji},$$

$$\sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i} = -\beta_i \sum_{j=1}^{p} x_j e^*_{j0} - \beta_i \sum_{j=1}^{p} x_j \sum_{k=1}^{p} e^*_{jk} x_k + r \sum_{j=1}^{p} x_j e^*_{ji}$$

$$\epsilon \cdot \left[IF_i + \frac{\partial IF_0}{\partial x_i} + \sum_{j=1}^{p} x_j \frac{\partial IF_j}{\partial x_i} \right]$$

$$= \epsilon \cdot \left[r e^*_{i0} + r \sum_{k=1}^{p} e^*_{ik} x_k - \beta_i \left(e^*_{00} + \sum_{k=1}^{p} e^*_{0k} x_k \right) \right.$$

$$+ r e^*_{0i} - \beta_i \sum_{j=1}^{p} x_j e^*_{j0}$$

$$\left. - \beta_i \sum_{j=1}^{p} x_j \sum_{k=1}^{p} e^*_{jk} x_k + r \sum_{j=1}^{p} x_j e^*_{ji} \right]$$

$$= \epsilon \cdot \left\{ 2r \left[e^*_{i0} + \sum_{k=1}^{p} e^*_{ik} x_k \right] \right.$$

$$\left. - \beta_i \left(e^*_{00} + 2 \sum_{j=1}^{p} x_j e^*_{j0} + \sum_{j=1}^{p} \sum_{k=1}^{p} x_j e^*_{jk} x_k \right) \right\}.$$

Proof of Proposition 5. (a) Conditions (i)–(iii) imply that the least squares estimate $\widehat{\beta}_n$ is consistent estimate of β. From (30) and (B5), $RINFIN(\mathbf{x}, y; \epsilon, \beta)$ is continuous function of β and therefore $RINFIN(\mathbf{x}, y; \epsilon, \widehat{\beta}_n)$ is consistent estimate of $RINFIN(\mathbf{x}, y; \epsilon, \beta)$. (b) Conditions (i), (ii), and (iv) imply that $\widehat{\beta}_n$ has asymptotically multivariate normal distribution with covariance matrix $\widetilde{\mathcal{E}}^{-1} E(\mathbf{X}_i \mathbf{X}_i^T \epsilon_i^2) \widetilde{\mathcal{E}}^{-1}$. From (30) and (B5), $RINFIN(\mathbf{x}, y; \epsilon, \beta)$ has continuous first partial derivatives at β which are not all zero from (v). Thus, $RINFIN(\mathbf{x}, y; \epsilon, \beta)$ has nonzero differential at β. The result follows from Serfling [27, Corollary in section 3.3, p. 124].

APPENDIX C. USING *RINFIN* WITH DATA

The data

$$D_n = \{(\mathbf{x}_1, y_1), \ \dots \ , (\mathbf{x}_n, y_n)\}, \quad D_{n,-m} = D_n - \{(\mathbf{x}_m, y_m)\},$$
$$m = 1, \ \dots \ , n.$$

We describe first how to calculate the *RINFIN* score of $(\mathbf{x}_m, y_m)$ with (35) and $\epsilon = 1/n$:

a. Use $D_{n,-m}$ to obtain L_2-estimates $\widehat{\beta}$ and $\widehat{r}_2(\mathbf{x}_m, y_m)$.

b. Estimate EX_i and σ_i^2, respectively, by the sample average and sample variance $\mathbf{x}$-data's ith coordinate in $D_{n,-m}$, $i = 1, \ \dots \ , p$.

c. Use $\widehat{\beta}$'s ith coordinate and replace x_i with $\mathbf{x}_m$'s ith coordinate, $i = 1, \ \dots \ , p$.

If a group G of k remote $\mathbf{x}$-neighboring cases exists,

$$G = \{(\mathbf{x}_{i_1}, y_{i_1}), \ \dots \ , (\mathbf{x}_{i_k}, y_{i_k})\} \subset D_n,$$

D_n may follow a gross-error model. Let $\overline{g}$ be the average of the elements in G and use, instead of D_n, new data

$$(D_n - G) \cup \{\overline{g}\}.$$

Calculate *RINFIN* scores following (a)–(c). For the *RINFIN* score of $\overline{g}$ use $\epsilon = k/n$; in the remaining $(n-k)$ cases weights are $1/n$.

With J groups, $G_1, \ \dots \ , G_J$, of remote $\mathbf{x}$-neighboring cases, $G_k \cap G_l = \emptyset$, $k \neq l$, obtain averages $\overline{g}_1, \ \dots \ , \overline{g}_J$, and use data set

$$\left(D_n - \bigcup_{j=1}^{J} G_j \right) \cup \{\overline{g}_1, \ \dots \ , \overline{g}_J\}.$$

Proceed with (a)–(c). For the *RINFIN* score of $\overline{g}_j$ use $\epsilon_j = k_j/n$, k_j is the cardinality of G_j, $j = 1, \ \dots \ , J$; in the remaining cases the weights are $1/n$.

To calculate the *RINFIN* score of $(\mathbf{x}_m, y_m)$ with (34), $D_{n,-m}$ is used to estimate $\widetilde{\mathcal{E}}$ with the corresponding sample averages and inverting it to obtain the corresponding estimate of $\mathcal{E}^*$ (see (32)) without the mth case, $m = 1, \ \dots \ , n$. The β's are also estimated. The approach can be repeated with J groups, $G_1, \dots, G_J$, of remote $\mathbf{x}$-neighboring cases, as described above; $J \ge 1$.