

Depth induced regression medians and uniqueness

YIJUN ZUO

Department of Statistics and Probability, Michigan State University
East Lansing, MI 48824, USA
zuo@msu.edu

March 20, 2020

Abstract

Notion of median in one dimension is a foundational element in nonparametric statistics. It has been extended to multi-dimensional cases both in location and in regression via notions of data depth.

Regression depth (RD) and projection regression depth (PRD) represent the two most promising notions in regression. Carrizosa depth D_C is another depth notion in regression. Depth induced regression medians (maximum depth estimators) serve as robust alternatives to the classical least squares estimator.

The uniqueness of regression medians is indispensable in the discussion of their properties and the asymptotics (consistency and limiting distribution) of sample regression medians. Are the regression medians induced from RD, PRD, and D_C unique? Answering this question is the main goal of this article. It is found that only the regression median induced from PRD possesses the desired uniqueness property. The conventional remedy measure for non-uniqueness, taking average of all medians, might yield an estimator that no longer possesses the maximum depth in both RD and D_C cases.

These and other findings indicate that the PRD and its induced median are highly favorable among their leading competitors.

AMS 2000 Classification: Primary 62G08, 62G35; Secondary 62J05, 62J99.

Key words and phrase: uniqueness, regression depth, maximum depth estimator, regression median, robustness.

Running title: Uniqueness of regression medians.

1 Introduction

Regular univariate sample median defined as the innermost (deepest) point of a data set is unique*. The population median defined as the $\frac{1}{2}$ -th quantile[†] of the underlying distribution (there are other versions of definition) is also unique. The most outstanding feature of the univariate median is its robustness. In fact, among all translation equivariant location estimators, it has the best possible breakdown point (Donoho (1982)[2]) (and the minimum maximum bias if underlying distribution has a unimodal symmetric density (Huber(1964)[4]). Besides serving as a promising robust location estimator, the univariate median also provides a base for a center-outward ordering (in terms of the deviations from the median), an alternative to the traditional left-to-right ordering.

To extend the univariate median to multidimensional settings and to share its outstanding robustness property and an alternative ordering scheme is desirable for multidimensional data. One approach, among others, is via notions of data depth. General notions of data depth have been increasingly pursued and studied (Liu, et al (1999)[6], Zuo and Serfling (2000) (ZS00)[26]) since the pioneer proposal of Tukey (1975)[15] (see Donoho and Gasko (1992)[3]). Besides Tukey depth, another prevailing depth, among others, is the projection depth (PD) (Liu (1992)[5], [26], and Zuo (2013)[21]).

Depth notions in location have also been extended to regression. Regression depth (RD) of Rousseeuw and Hubert (1999) (RH99)[10], the most famous, exemplifies a direct extension of Tukey location depth to regression. Projection regression depth (PRD) of Zuo (2018a) (Z18a)[22] is another example of the extension of prominent PD in location to regression. The RD and PRD represent the two leading notions of depth in regression ([22]) which satisfy desirable axiomatic properties. Carrizosa depth D_C (Carrizosa (1996) (C96))[1] (defined in Section 2.2) is one of the other notions of depth in regression ([22]).

One of the outstanding advantages of depth notions is that they can be directly employed to introduce median-type deepest estimating functionals (or estimators in the empirical case) for the location or regression parameters in a multi-dimensional setting based on a general min-max stratagem. The maximum (deepest) regression depth estimator (also called regression median) serves as a *robust* alternative to the classical least squares or least absolute deviations estimator for the unknown parameters in a general linear regression model:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + e_i, \text{ for } i = 1, \dots, n, \quad (1)$$

where \top denotes the transpose of a vector, and vector $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i(p-1)})^\top$ and parameter vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ are in \mathbb{R}^p ($p \geq 2$), and e_i is a random variable in \mathbb{R} . One can regard the observations (y_i, \mathbf{x}_i^\top) as a sample from random vector $(y, \mathbf{x}^\top) \in \mathbb{R}^{p+1}$.

*If the sample median is defined to be the point θ that minimizes the sum of its distances to sample points (i.e. $\theta = \arg \min_{\theta \in \mathbb{R}^1} \sum_{i=1}^n |\theta - x_i|$, where $x_i, i = 1, \dots, n$ are the given n sample points in \mathbb{R}^1), then it is not unique. However, to overcome this drawback, conventionally it is defined as $\theta = \text{Median}\{x_i\} := x_{(\lfloor \frac{n+1}{2} \rfloor)} + x_{(\lfloor \frac{n+2}{2} \rfloor)} / 2$, where $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ are ordered values of x_i 's and $\lfloor \cdot \rfloor$ is the floor function. Namely, it is the innermost point (from both left and right direction) or the average of two deepest sample points. Hence, it is unique.

[†]Recall, for any univariate distribution function F , and for $0 < p < 1$, the quantity $F^{-1}(p) := \inf\{x : F(x) \geq p\}$ is called the p th *quantile* or *fractile* of F (see page 3 of Serfling (1980)[14]).

Robustness of the median induced from RD and PRD have been investigated in Van Aelst and Rousseeuw (2000) (VAR00)[16] and Zuo (2018b)[23], respectively. These medians, just like their location or univariate counterpart, indeed possess high breakdown point robustness.

Regression median, as the deepest regression hyperplane, just like their location or univariate counterpart, is expected to be unique because non-uniqueness would result in vagueness in the inference (prediction and estimation) via regression median. Uniqueness is the indispensable feature and axiomatic property when one (i) investigates the population median, or (ii) deals with the convergence in probability or in distribution of the sample regression median to its inevitably unique population version (iii) it is also an essential property in the computation of the sample regression medians for the convergence of approximate algorithms. The uniqueness issue of multidimensional location medians has been addressed in [21].

Are the medians induced from regression depth notions via the min-max scheme generally unique? Answering this question is the goal of this article. It turns out that the regression depth induced medians are not necessarily unique. The conventional remedy measure for this issue is taking average of all. It, however, might not work (in the sense that the resulting estimator might no longer possess the maximum depth) for both RD of RH99 and D_C of C96. On the other hand, PRD induced regression medians are unique.

The rest of article is organized as follows. Section 2 introduces leading regression depth notions and induced medians and show these medians indeed recover the regular univariate sample median in the special univariate case. Empirical examples of regression depth and medians and their behavior are illustrated in Section 3. Section 4 establishes general results on uniqueness of regression medians. Brief concluding remarks in Section 5 end the article.

2 Maximum depth functionals (regression medians)

Let $D(\beta; P)$ be a generic non-negative functional on $\mathbb{R}^p \times \mathcal{P}$, where $\beta \in \mathbb{R}^p$ and \mathcal{P} is a collection of distributions F_Z of $Z = (y, \mathbf{x}^\top) \in \mathbb{R}^{p+1}$ (F_Z and P are used interchangeably).

If $D(\beta; P)$ satisfies four axiomatic properties: **(P1)** (regression, scale and affine) invariance; **(P2)** maximality at center; **(P3)** monotonicity relative to any deepest point and **(P4)** vanishing at infinity, then it is called a regression depth functional (see [22] for details). The maximum regression depth functional, or the regression median, can be defined as

$$\beta^*(F_Z) := \arg \max_{\beta \in \mathbb{R}^p} D(\beta; F_Z) \quad (2)$$

Note that β^* might not be unique, and a conventional remedy measure is to take the average of all maximum depth points. Unfortunately, this could lead to a scenario where the resulting functional (or estimator) might not have the maximum depth any more. For detailed discussions on $D(\beta; F_Z)$ and $\beta^*(F_Z)$, see [22]. In the following we elaborate three examples.

2.1 Median induced from regression depth of RH99

Definition 2.1 For any $\beta \in \mathbb{R}^p$ and the joint distribution P of (y, \mathbf{x}^\top) in (1), [10] defined the regression depth of β , denoted hence by $\text{RD}_{RH}(\beta; P)$, to be the minimum probability mass

that needs to be passed when tilting (the hyperplane induced from) β in any way until it is vertical. The maximum regression depth functional $\beta_{RD_{RH}}^*$ (regression median) is defined as

$$\beta_{RD_{RH}}^*(P) = \arg\max_{\beta \in \mathbb{R}^p} RD_{RH}(\beta; P) \quad (3)$$

The $RD_{RH}(\beta; P)$ definition above is rather abstract and not easy to comprehension, many characterizations of it, or equivalent definitions, have been given in the literature though, see, e.g., [22] and references cited therein.

2.2 Median induced from Carrizosa depth of C96

Among regression depth notions investigated in [22], Carrizosa depth $D_C(\beta; P)$

$$D_C(\beta; P) = \inf_{\alpha \in \mathbb{R}^p} P(|r(\beta)| \leq |r(\alpha)|), \quad (4)$$

for any $\beta \in \mathbb{R}^p$ and underlying probability measure P associated with (y, \mathbf{x}^\top) , was a pioneer regression depth notion introduced in [1] and thoroughly investigated in [22], where $r(\gamma) = y - \mathbf{x}^\top \gamma$. As characterized in [22] (see Proposition 2.2 there), it turns out that ($p \geq 2$)

$$D_C(\beta; P) = P(r(\beta) = 0). \quad (5)$$

The maximum regression depth functional (or regression median) was then defined as

$$\beta_{D_C}^*(P) = \arg\max_{\beta \in \mathbb{R}^p} D_C(\beta; P). \quad (6)$$

As shown in [22], $\beta_{D_C}^*$ always exists if the assumption: **(A)** $P(H_v) = 0$, for any vertical hyperplane H_v , holds. Unfortunately, as D_C violates **(P3)** generally (see [22]), we will not focus on it in the sequel. On the other hand, under **(A)** RD_{RH} above satisfies **(P1)**-**(P4)**.

2.3 Median induced from projection regression depth of Z18a

Hereafter, assume that R is a univariate regression estimating functional which satisfies

(A1) regression, scale and affine equivariant. That is, respectively,

$$R(F_{(y+xb, x)}) = R(F_{(y, x)}) + b, \quad \forall b \in \mathbb{R}^1, \text{ and}$$

$$R(F_{(sy, x)}) = sR(F_{(y, x)}), \quad \forall s \in \mathbb{R}^1, \text{ and}$$

$$R(F_{(y, ax)}) = a^{-1}R(F_{(y, x)}), \quad \forall a(\neq 0) \in \mathbb{R}^1.$$

where $x, y \in \mathbb{R}^1$ are random variables. Throughout, the lower case x stands for a variable in \mathbb{R}^1 while the bold \mathbf{x} for a vector in \mathbb{R}^p ($p > 1$). $F_{(x_1, x_2)}$ is the distribution of vector (x_1, x_2) .

(A2) $\sup_{\mathbf{v} \in \mathbb{S}^{p-1}} |R(F_{(y, \mathbf{x}^\top \mathbf{v})})| < \infty$, where $\mathbb{S}^{p-1} := \{\mathbf{u} \in \mathbb{R}^p, \|\mathbf{u}\| = 1\}$

(A3) $R(F_{(y-\mathbf{x}^\top \beta, \mathbf{x}^\top \mathbf{v})})$ is continuous in β and \mathbf{v} , and quasi-convex in β , for $\beta \in \mathbb{R}^p$, $\mathbf{v} \in \mathbb{S}^{p-1}$.

Let S be a positive scale estimating functional that is *scale equivariant* and *location invariant*.

R will be restricted to the form $R(F_{(y-\mathbf{x}^\top \beta, \mathbf{x}^\top \mathbf{v})}) = T(F_{(y-\mathbf{x}^\top \beta)/(\mathbf{x}^\top \mathbf{v})})$ and T will be a univariate location functional that is location, scale and affine equivariant (see pages 158-159 of Rousseeuw and Leroy (1987) (RL87)[11] for definitions). Hereafter we assume that

(A0) $P(\mathbf{x}^\top \mathbf{v} = 0) = 0$ for any $\mathbf{v} \in \mathbb{S}^{p-1}$ (see (I) of Remarks 4.1 for the explanations). Examples of T include, among others, the mean, weighted mean, and quantile functionals. A particular example of $R(F_{(y-\mathbf{x}^\top \boldsymbol{\beta}, \mathbf{x}^\top \mathbf{v})})$ is $\text{Med}_{\mathbf{x}^\top \mathbf{v} \neq 0}(F_{(y-\mathbf{x}^\top \boldsymbol{\beta})/\mathbf{x}^\top \mathbf{v}})$, where Med stands for the median functional. Typical examples of S include the standard deviation and weighted deviation functionals (Wu and Zuo (2008)[18]) and the median of absolute deviations (MAD) functional.

Equipped with a pair of T and S , we can introduce a corresponding projection based regression estimating functional. By modifying a functional in Marrona and Yohai (1993)[9] to achieve scale equivariance, [22] defined

$$\text{UF}_{\mathbf{v}}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T) := |T(F_{(y-\mathbf{x}^\top \boldsymbol{\beta})/\mathbf{x}^\top \mathbf{v}})|/S(F_y), \quad (7)$$

which represents unfitness of $\boldsymbol{\beta}$ at $F_{(y, \mathbf{x}^\top)}$ w.r.t. T along the $\mathbf{v} \in \mathbb{S}^{p-1}$. If R is a *Fisher consistent* regression estimating functional, then $T(F_{(y-\mathbf{x}^\top \boldsymbol{\beta}_0)/\mathbf{x}^\top \mathbf{v}}) = 0$ for some $\boldsymbol{\beta}_0$ (the true parameter of the model) and $\forall \mathbf{v} \in \mathbb{S}^{p-1}$. Thus overall, one expects $|T|$ to be small and close to zero for a candidate $\boldsymbol{\beta}$, independent of the choice of \mathbf{v} and $\mathbf{x}^\top \mathbf{v}$. The magnitude of $|T|$ measures the unfitness of $\boldsymbol{\beta}$ along the \mathbf{v} . Taking the supremum over all $\mathbf{v} \in \mathbb{S}^{p-1}$ yields

$$\text{UF}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T) = \sup_{\|\mathbf{v}\|=1} \text{UF}_{\mathbf{v}}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T), \quad (8)$$

the *unfitness* of $\boldsymbol{\beta}$ at $F_{(y, \mathbf{x}^\top)}$ w.r.t. T . Now applying the min-max scheme, [22] obtained the *projection regression estimating functional* (also denoted by $\boldsymbol{\beta}_{\text{PRD}}^*$) w.r.t. the pair (T, S)

$$\begin{aligned} \boldsymbol{\beta}^*(F_{(y, \mathbf{x}^\top)}, T) &= \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \text{UF}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T) \\ &= \arg \max_{\boldsymbol{\beta} \in \mathbb{R}^p} \text{PRD}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T), \end{aligned} \quad (9)$$

where the *projection regression depth* (PRD) functional was defined in [22] as

$$\text{PRD}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T) = \left(1 + \text{UF}(\boldsymbol{\beta}; F_{(y, \mathbf{x}^\top)}, T)\right)^{-1}, \quad (10)$$

Just like S (which is for achieving scale invariance and is nominal), T sometimes is also suppressed in above functionals for simplicity. [22] showed that PRD satisfies **(P1)**-(**P4**).

For robustness consideration, in the sequel, (T, S) is the fixed pair (Med, MAD), unless otherwise stated. Hereafter, we write $\text{Med}(Z)$ rather than $\text{Med}(F_Z)$. For this special choice of T and S , we have that

$$\begin{aligned} T(F_{(y-\mathbf{x}^\top \boldsymbol{\beta})/\mathbf{x}^\top \mathbf{v}}) &= \text{Med}_{\mathbf{x}^\top \mathbf{v} \neq 0}\left(\frac{y - \mathbf{x}^\top \boldsymbol{\beta}}{\mathbf{x}^\top \mathbf{v}}\right), \\ S(F_y) &= \text{MAD}(F_y). \end{aligned}$$

To end this section, we show that the three maximum depth estimators above indeed deserve to be called regression median since they recover the regular univariate sample median in the special univariate case. (The result below also holds true for the population case).

Proposition 2.1 For univariate data, the $\boldsymbol{\beta}_{RD_{RH}}^*$, $\boldsymbol{\beta}_{DC}^*$ and $\boldsymbol{\beta}_{PRD}^*$ all recover the univariate sample median.

Proof: (i) For $\beta_{RD_{RH}}^*$, this has already been discussed and claimed in [10] (page 390). So we only need to focus on the other two.

(ii) For $\beta_{D_C}^*$, we no longer can use (5) and have to invoke (4). Note that $r(\beta) = y - \beta$ in this case (no slope term any more).

$$\begin{aligned} D_C(\beta, P) &= \inf_{\alpha \in \mathbb{R}^1} P(|y - \beta| \leq |y - \alpha|) \\ &= \min \left\{ \inf_{\alpha > \beta} P(|y - \beta| \leq |y - \alpha|), \inf_{\alpha \leq \beta} P(|y - \beta| \leq |y - \alpha|) \right\} \\ &= \min \{P(y \leq \beta), P(y \geq \beta)\}. \end{aligned}$$

That is, $D_C(\beta, P_n) = \min \{ \sum_{i=1}^n I(y_i \leq \beta)/n, \sum_{i=1}^n I(y_i \geq \beta)/n \}$. The latter immediately leads to $\beta_{D_C}^* = \text{Med}_i\{y_i\}$ (the average of all solutions).

(iii) For β_{PRD}^* , first we note that (without loss of generality, assume that $S(F_y) = 1$)

$$\beta_{PRD}^* = \arg \min_{\beta \in \mathbb{R}^p} \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} \left| \text{Med}_i \left\{ \frac{y_i - \mathbf{x}_i^\top \beta}{\mathbf{x}_i^\top \mathbf{v}} \right\} \right|. \quad (11)$$

When $p = 1$, it reduces to the following

$$\beta_{PRD}^* = \arg \min_{\beta \in \mathbb{R}} \sup_{v = \pm 1} \left| \text{Med}_i \left\{ \frac{y_i - \beta}{v} \right\} \right|. \quad (12)$$

It is readily seen that

$$\beta_{PRD}^* = \arg \min_{\beta \in \mathbb{R}} |\text{Med}_i\{y_i - \beta\}| = \arg \min_{\beta \in \mathbb{R}} |\text{Med}_i\{y_i\} - \beta|, \quad (13)$$

where the first equality follows from (12) and the oddness of median operator, the second one follows from the translation equivalence (see page 249 of [11] for definition) of the median as a location estimator. The last display means that β_{PRD}^* recovers the sample median. ■

3 Examples of regression depths and regression medians

For a better comprehension of depth notions and depth induced medians in the last section, we present empirical examples below. We will confine attention to RD and PRD only since $D_C(\beta, P)$ is just the probability mass carried by the hyperplane determined by $y = \mathbf{x}^\top \beta$.

Example 3.1 (Empirical RD_{RH} and PRD). How do empirical RD_{RH} and PRD look like? To answer the question, 30 random bivariate standard normal points are generated (plotted in Figure 1) and RD_{RH} and PRD are computed w.r.t. these points.

We select 961 equal-spaced grid points from the square of $[x, y]$ with range of $|x| \leq 3$ and $|y| \leq 3$, then treat each point (x, y) as a $\beta^\top = (\beta_1, \beta_2)$ and compute its regression depth (RD_{RH} and PRD) w.r.t. the 30 bivariate normal points. The depths of these 961 points are plotted in Figure 2.

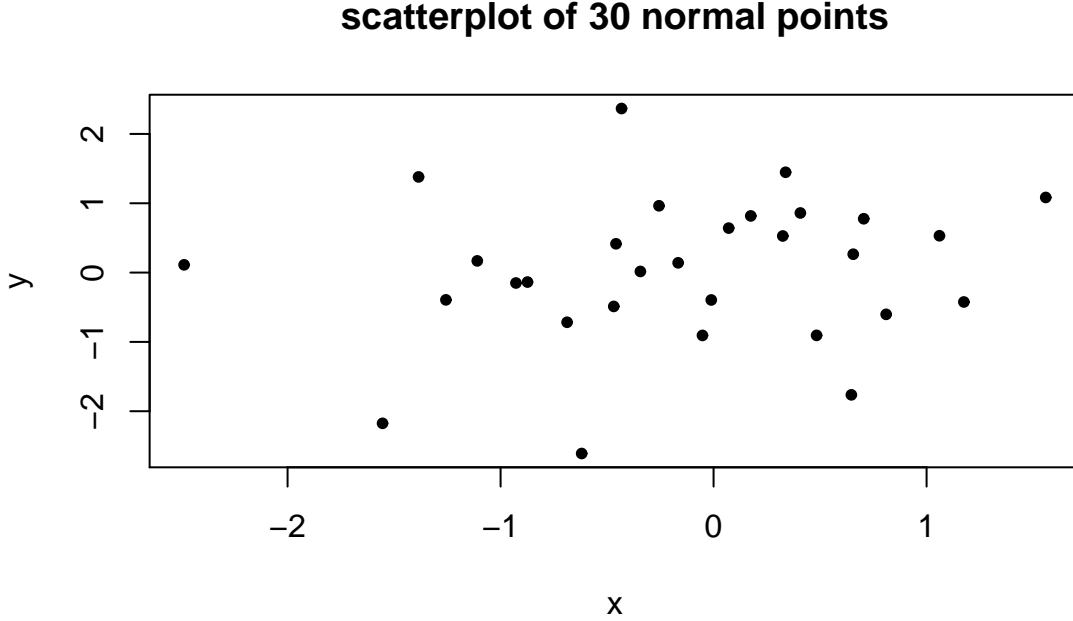


Figure 1: 30 bivariate standard normal points.

Inspecting the Figure reveals that (i) sample RD_{RH} function is a step-wise increasing function (each step in this case is $1/30$). For this roughly symmetric data case, it can attain maximum depth around the center of symmetry (the origin), while (ii) on the other hand, PRD is a strictly monotonically increasing function and attains its maximum value at the center of symmetry, sharply contrasting the behavior of RD_{RH} around the center (one has a unique maximum depth point and the other is opposite (multiple maximum depth points)).

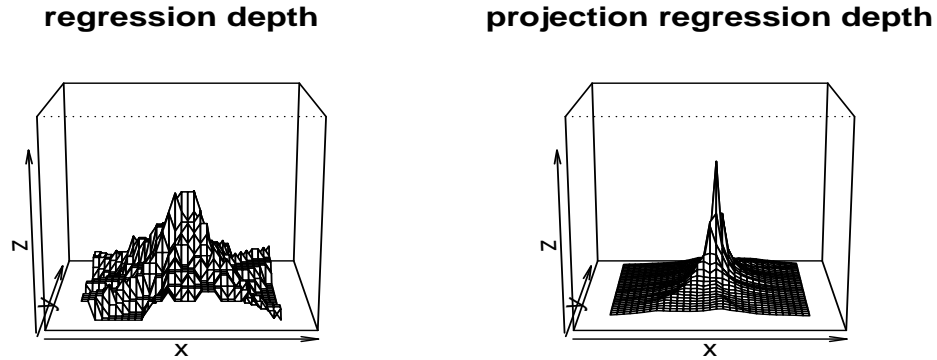


Figure 2: Regression depth (RD_{RH}) (left) and projection regression depth (PRD) (right) of 961 candidate parameter β^\top 's w.r.t. 30 bivariate standard normal points.

Example 3.2 Uniqueness of medians induced from empirical RD_{RH} and PRD
This example illustrates the uniqueness behavior of the regression depth (RD_{RH} and PRD) induced medians in the empirical distribution case via a concrete example on the real data from the Hertzsprung-Russell diagram of the star cluster CYG OB1 (see Table 3 in chapter

Permormance of regression depth medains on uniqueness

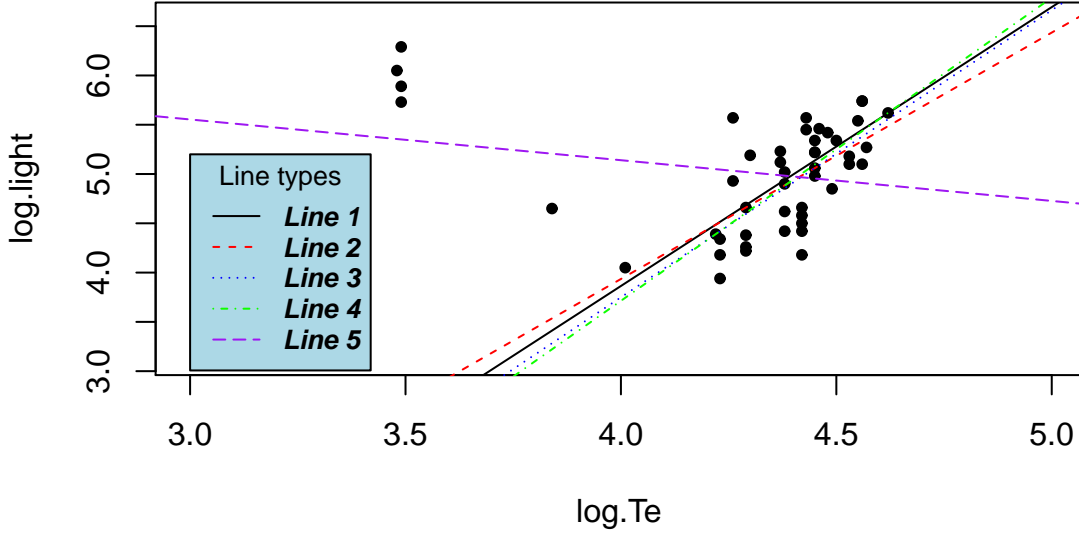


Figure 3: Five regression depth median lines based on the data from Hertzsprung-Russell diagram of the star cluster CYG OB1 (solid black for β_{PRD}^* ; dashed red, dotted blue, and dotdash green all for β_{RD}^* ; longdash purple for LS).

2 of [11]), which contains 47 stars in the direction of Cygnus. Here x is the logarithm of the effective temperature at the surface of the star (T_e), and y is the logarithm of its light intensity (L/L_0); see Figure 3 for the plot of the data set.

Five regression lines are plotted in Figure 3. Among them, three (dashed red, dotted blue, and dotdash green) are regression medians from RD_{RH} , one (solid black) from PRD, and the other (longdash purple) is the least squares line. Note that the classical least squares regression estimator (as well as many traditional regression estimators) could be regarded as a depth induced median under the general “objective depth” D_{Obj} framework (see [22]). Thus, for the benchmark purpose, the least squares line is also plotted in Figure 3 alongside the four other median lines.

The LS line also justifies the legitimacy of the existence of RD_{RH} and PRD induced medians (as robust alternatives) since the LS line fails to capture the main-sequence/pattern of the data cloud (stars) and is heavily affected by four giant stars whereas the other four depth medians resist the four leverage points (outliers) and catch the main trend/cluster.

It turns out that there exist three maximum depth lines (medians) induced from RD_{RH} . Each of the three lines goes through exactly two data points. In terms of (intercept, slope) form, they are $(-6.065000, 2.500000)$, $(-8.586500, 3.075000)$, and $(-7.903043, 2.913043)$. These lines are plotted by dash red, dotted blue, and dotdash green in Figure 3. All three possess regression depth 21/47. Note that the average of the three deepest lines is $(-7.518181, 2.829348)$ which possesses RD 20/47. *That is, it no longer possesses the maximum regression depth.*

On the other hand, there exists only one maximum regression line (median), $(-7.453665, 2.829416)$, induced from PRD, plotted in solid black in Figure 3, with PRD value 0.8585901. Incidentally, the LS line is $(6.7934673, -0.4133039)$, plotted in longdash purple.

The computation issues of RD_{RH} have been discussed in RH99, Rousseeuw and Struyf (1998)[12], and Liu and Zuo (2014)[7]. For the discussion on the computation of the PRD and induced regression medians, see Zuo (2019b) (Z19b)[25].

After obtaining $(\hat{\beta}_1, \hat{\beta}_2)$, one can immediately get the fitted line $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ (which has actually already been plotted in Figure 3), and the predicted values: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$, and hence the residuals: $r_i := y_i - \hat{y}_i$. All these involve the uniqueness issue, we first need to have a unique fitted line for each method. Here due to the non-uniqueness of the deepest RD lines, we select the first deepest line $(-6.065000, 2.500000)$ among the three as a representative. Then we construct a table with nine columns: 1st is the id's of observations, 2nd is the explanatory variable x_i values, 3rd is dependent variable y_i values, 4th-6th are the predicted \hat{y}_i values for LS, RD, PRD methods, respectively, 7th-9th are the residuals r_i for LS, RD, PRD methods, respectively.

Residuals analysis for three regression methods								
i	x	y	ls	\hat{y}		ls	r	
				rd	prd		rd	prd
1	4.37	5.23	4.987329	4.860	4.910883	0.2426707	0.370	0.3191171
2	4.56	5.74	4.908802	5.335	5.448472	0.8311985	0.405	0.2915280
3	4.26	4.93	5.032793	4.585	4.599647	-0.1027927	0.345	0.3303528
4	4.56	5.74	4.908802	5.335	5.448472	0.8311985	0.405	0.2915280
5	4.30	5.19	5.016261	4.685	4.712824	0.1737395	0.505	0.4771762
6	4.46	5.46	4.950132	5.085	5.165530	0.5098681	0.375	0.2944696
7	3.84	4.65	5.206380	3.535	3.411292	-0.5563803	1.115	1.2387076
8	4.57	5.27	4.904668	5.360	5.476766	0.3653315	-0.090	-0.2067661
9	4.26	5.57	5.032793	4.585	4.599647	0.5372073	0.985	0.9703528
10	4.37	5.12	4.987329	4.860	4.910883	0.1326707	0.260	0.2091171
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
45	4.55	5.54	4.912935	5.310	5.420178	0.62706545	0.230	0.1198222
46	4.45	4.98	4.954265	5.060	5.137236	0.02573506	-0.080	-0.1572362
47	4.42	4.50	4.966664	4.985	5.052354	-0.46666406	-0.485	-0.5523537

Table 1: Residuals of three regression methods

Next the residuals of three methods are plotted below in the Figure 4.

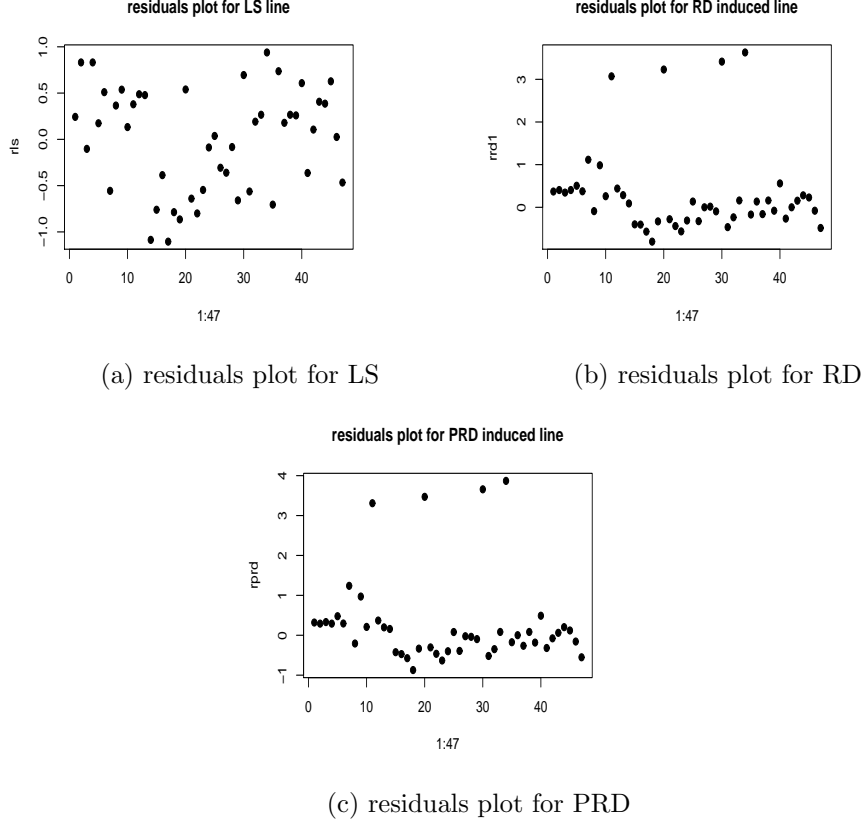


Figure 4: Residuals plots for three types of regression methods for the star cluster data (a) Residuals plot for the LS method. (b) Residuals plot for one of the deepest line induced from RD. (c) Residuals plot for the unique deepest line induced from PRD.

Inspecting the residuals plot immediately reveals that in this case, the residuals of LS method are rather deceptively homogeneous, its plot fails to identify any outliers whereas the robust regression median lines all can easily spot the four obvious outliers and two groups of stars. Based on the residual plot, one can make some conclusions. For example, the four outliers are not necessarily errors but might be exceptional observations (they come from a different group of stars), and LS line does not provide a good fit (only explained 4.4% of total variation in observations of y). ■

In the empirical distribution case, one can always take the average over all regression medians to take care of the non-uniqueness issue. Nevertheless, challenges arise computationally if there exist infinitely many medians in higher dimensions. Furthermore, the average sometimes will no longer be a deepest line/hyperplane (as seen in this example and more in Section 4).

The non-uniqueness issue is more vital with the population case since without the uniqueness, there will be no uniquely defined median and it is impossible to discuss the convergence (or consistency) and the limiting distribution of the unique empirical regression median.

4 Uniqueness of regression medians

From the empirical example in the previous section, we see that there can exist multiple empirical regression medians induced from RD_{RH} while in the case of PRD there exists a unique one. These results are just empirical special examples and not for general cases. In the following, we address general cases and draw general conclusions.

4.1 Non-uniqueness of $\beta_{RD_{RH}}^*$ and $\beta_{D_C}^*$

Under certain symmetry assumption (e.g. regression symmetry of Rousseeuw and Struyf (2004) (RS04)) [13] and other conditions, the regression median induced from RD_{RH} can be unique (see Theorem 3 and Corollary 3 of [13]). However, generally speaking, we have

Proposition 4.1 $\beta_{RD_{RH}}^*(F_{(y, \mathbf{x})})$ is not unique in general. The average of all $\beta_{RD_{RH}}^*(F_{(y, \mathbf{x})})$ might not possess the maximum depth any more.

PROOF: A counterexample suffices.

In fact, the real data example 3.2 could serve as one counterexample, where one has three maximum depth lines and the average line no longer possesses the maximum RD_{RH} value.

An even simpler counterexample could be constructed. Assume that there are three sample points $A = (-1, 0)$, $B = (0, 1)$ and $C = (1, 0)$. Then it is readily seen that three lines each of which formed by two sample points are $(1, -1)$; $(1, 1)$ and $(0, 0)$ in terms of (intercept, slope) form and each line has the maximum RD_{RH} $2/3$ whereas the average of all maximum depth lines is $(2/3, 0)$ which has RD_{RH} only $1/3$. ■

For special distributions, the median induced from Carrizosa depth can also be unique. But generally speaking, it is not.

Proposition 4.2 $\beta_{D_C}^*(F_{(y, \mathbf{x})})$ is not unique in general. The average of all $\beta_{D_C}^*(F_{(y, \mathbf{x})})$ might not possess the maximum depth value any more.

PROOF: A counterexample suffices.

Denote by H_{β} the hyperplane determined by $y = \mathbf{x} \top \beta$ for any $\beta \in \mathbb{R}^p$ and by θ_{β} the acute angle formed between the hyperplane H_{β} and the horizontal hyperplane H_h ($y=0$).

Assume that $\beta_i \in \mathbb{R}^p$, ($i = 1, 2$), $\beta_1 \neq \beta_2$, and H_{β_i} each contains $1/2$ probability mass; any hyperline in H_{β_i} contains no probability mass ($i = 1$ or 2); $\theta_{\beta_1} = \theta_{\beta_2}$, and H_{β_1} intersects with H_{β_2} at a hyperline in the horizontal hyperplane H_h .

Now in light of characterization (5) of D_C , it is readily seen that at each β_i , $D_C(\beta_i; P)$ attains the maximum depth value $1/2$.

Let $\gamma = (\beta_1 + \beta_2)/2$, then it is readily seen that the $D_C(\gamma; P) = 0$, and H_γ is no longer a hyperplane with the maximum depth value. \blacksquare

4.2 Uniqueness of β_{PRD}^*

For two univariate random variables X, Y defined on the sample space Ω , $X < Y$ stands for $X(\omega) < Y(\omega)$, $\forall \omega \in \Omega$. We say that $T(F_{(y-\mathbf{x}^\top \beta)/\mathbf{x}^\top \mathbf{v}})$ is *strictly monotonic* at point β_0 iff $T(F_{(y-\mathbf{x}^\top \beta_0)/\mathbf{x}^\top \mathbf{v}}) > T(F_{(y-\mathbf{x}^\top \beta_1)/\mathbf{x}^\top \mathbf{v}})$ whenever $-\mathbf{x}^\top \beta_0 > -\mathbf{x}^\top \beta_1 \quad \forall \beta_1 \in \mathbb{R}^p$, for any $\mathbf{v} \in \mathbb{S}^{p-1}$.

Proposition 4.3 If **(A0)** holds and $T(F_{(y-\mathbf{x}^\top \beta)/\mathbf{x}^\top \mathbf{v}})$ (i) is strictly monotonic at $\mathbf{0}$ and (ii) satisfies **(A1)**, **(A2)**, and **(A3)**, then $\beta_{PRD}^*(F_{(y, \mathbf{x})})$ exists uniquely.

PROOF: To prove the proposition, we first invoke the following result.

Lemma 4.1 [22] The PRD and β_{PRD}^* satisfy the following propoerties.

- (i) The $\beta_{PRD}^*(F_{(y, \mathbf{x}^\top)})$ is regression, scale and affine equivariant in the sense that

$$\beta^*(F_{(y+\mathbf{x}^\top \mathbf{b}, \mathbf{x}^\top)}) = \beta^*(F_{(y, \mathbf{x}^\top)}) + \mathbf{b}, \quad \forall \mathbf{b} \in \mathbb{R}^p;$$

$$\beta^*(F_{(sy, \mathbf{x}^\top)}) = s\beta^*(F_{(y, \mathbf{x}^\top)}), \quad \forall \text{ scalar } s(\neq 0) \in \mathbb{R};$$

$$\beta^*(F_{(y, A^\top \mathbf{x})}) = A^{-1}\beta^*(F_{(y, \mathbf{x}^\top)}), \quad \forall \text{ nonsingular } A \in \mathbb{R}^{p \times p},$$

respectively.

- (ii) The maximum of $\text{PRD}(\beta; F_{(y, \mathbf{x}^\top)})$ exists and is attained at a $\beta_0 \in \mathbb{R}^p$ with $\|\beta_0\| < \infty$.
 (iii) The $\text{PRD}(\beta; F_{(y, \mathbf{x}^\top)})$ monotonically decreases along any ray stemming from a deepest point in the sense that for any $\beta \in \mathbb{R}^p$ and $\lambda \in [0, 1]$,

$$\text{PRD}(\lambda\beta^* + (1-\lambda)\beta; F_{(y, \mathbf{x}^\top)}) \geq \text{PRD}(\beta; F_{(y, \mathbf{x}^\top)}),$$

where β^* is a maximum depth point of $\text{PRD}(\beta; F_{(y, \mathbf{x}^\top)})$ for any $\beta \in \mathbb{R}^p$. \blacksquare

Now we are in a position to prove the proposition.

Assume, w.l.o.g., that $S(F_y) = 1$ (since it does not involve \mathbf{v} and β , it has nothing to do with the maximum depth point β_{PRD}^*). The existence of the maximum depth point (the regression median) is guaranteed in light of Lemma 4.1 above. We thus focus on the uniqueness. Assume that there are two maximum depth points $\beta_1^* \neq \beta_2^*$. We seek a contradiction.

Let $\beta_0^* = (\beta_1^* + \beta_2^*)/2$. By virtue of Lemma 4.1 above, β_0^* is also a maximum depth point. By the invariance of the projection regression depth functional (see [22]) and Lemma 4.1 above, assume (w.l.o.g.) that $\beta_0^* = \mathbf{0}$.

For a given $\beta \in \mathbb{R}^p$, write $g(\beta, \mathbf{v}) := T(F_{(y-\mathbf{x}^\top \beta)/\mathbf{x}^\top \mathbf{v}})$. In light of the continuity of T in \mathbf{v} , the generalized extreme value theorem on a compact set, and **(A1)**, there exists a $\mathbf{v}_\beta \in \mathbb{S}^{p-1}$ such that

$$g(\beta, \mathbf{v}_\beta) = \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} |T(F_{(y-\mathbf{x}^\top \beta)/\mathbf{x}^\top \mathbf{v}})| \quad (14)$$

For simplicity, denote by \mathbf{v}_0 for $\mathbf{v}_{\beta_0^*}$. Then we have

$$g(\beta_0^*, \mathbf{v}_0) = T(F_{(y-\mathbf{x}\top\beta_0^*)/\mathbf{x}\top\mathbf{v}_0}) = T(F_{y/\mathbf{x}\top\mathbf{v}_0}) = \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} |T(F_{(y-\mathbf{x}\top\beta_0^*)/\mathbf{x}\top\mathbf{v}})| := \alpha^*. \quad (15)$$

Denote by $l(\beta_1^*, \beta_2^*)$ the hyperline that connects β_1^* and β_2^* in the parameter space of $\beta \in \mathbb{R}^p$. Consider two cases.

Case I \mathbf{x} does not concentrate on any single hyperplane. In light of this assumption, there exists at least one $\gamma \in \mathbb{R}^p$ on $l(\beta_1^*, \beta_2^*)$ in the parameter space \mathbb{R}^p such that $-\mathbf{x}\top\gamma \neq 0$. Assume (w.l.o.g.) that $-\mathbf{x}\top\gamma < 0 = -\mathbf{x}\top\beta_0^*$. By (15) and the strictly monotonicity of T , one has that for the \mathbf{v}_γ defined in (14)

$$\begin{aligned} \alpha^* &= \inf_{\beta \in \mathbb{R}^p} \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} |T(F_{(y-\mathbf{x}\top\beta)/\mathbf{x}\top\mathbf{v}})| \leq \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} |T(F_{(y-\mathbf{x}\top\gamma)/\mathbf{x}\top\mathbf{v}})| = T(F_{(y-\mathbf{x}\top\gamma)/\mathbf{x}\top\mathbf{v}_\gamma}) \\ &< T(F_{(y-\mathbf{x}\top\beta_0^*)/\mathbf{x}\top\mathbf{v}_\gamma}) \leq \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} |T(F_{y/\mathbf{x}\top\mathbf{v}})| \\ &= T(F_{y/\mathbf{x}\top\mathbf{v}_0}) = \alpha^* \end{aligned} \quad (16)$$

which is a contradiction. This completes the proof of the Case I.

Case II \mathbf{x} concentrates on a single hyperplane. This implies that there is a $\mathbf{v} \in \mathbb{S}^{p-1}$ such that $\mathbf{x}\top(\omega)\mathbf{v} = 0$ for any $\omega \in \Omega$. This contradicts **(A0)**, however. This completes the proof of Case II and thus the proposition. \blacksquare

Remarks 4.1

(I) **(A0)** automatically holds if \mathbf{x} has density or if \mathbf{x} is not degenerated to concentrate on a single $(p-1)$ dimensional hyperplane. The latter means all \mathbf{x} lie on the same point for $p=1$, and they lie on a single line for $p=2$, and lie on a plane for $p=3$, and so on.

(II) **(A1)**, **(A2)** and **(A3)** hold for a large class of T , such as the mean, weighted mean (Wu and Zuo (2009)[19]), and quantile functionals.

(III) There also exists a large class of T that is strictly monotonic. For example (i) If $T(F_{(y-\mathbf{x}\top\beta)/\mathbf{x}\top\mathbf{v}}) = E((y-\mathbf{x}\top\beta)/\mathbf{x}\top\mathbf{v})$, then T is strictly monotonic at any β as long as the related expectations exist and $E(\mathbf{x}\top\alpha/\mathbf{x}\top\mathbf{v}) > 0$ whenever $\mathbf{x}\top\alpha > 0$ for any $\alpha \in \mathbb{R}^p$ and $\mathbf{v} \in \mathbb{S}^{p-1}$. (ii) When $T(F_{(y-\mathbf{x}\top\beta)/\mathbf{x}\top\mathbf{v}}) = Q_q((y-\mathbf{x}\top\beta)/\mathbf{x}\top\mathbf{v})$, $q \in (0, 1)$, where $Q_q(Z)$ is the q th quantile associated with the random variable Z (i.e. $Q_q(Z) = \inf\{z : P(Z \leq z) \geq q\}$), then T is strictly monotonic at any β as long as the CDF of $Z(\beta; \mathbf{v}, y, \mathbf{x}) := (y-\mathbf{x}\top\beta)/\mathbf{x}\top\mathbf{v}$ is not flat at β for a given $\mathbf{v} \in \mathbb{S}^{p-1}$.

(IV) The proposition covers the sample case. That is, when $F_{(y, \mathbf{x}\top)}$ is replaced by its sample version in the proposition, we have the uniqueness of the sample regression median induced from PRD, which is very helpful in the practical computation of the median and consistent with the finding in Figure 2. \blacksquare

5 Concluding remarks

Why do we care about the non-uniqueness of regression medians

Uniqueness is actually implicitly assumed when we discuss the property (such as the Fisher consistency, regression, scale and affine equivariance, or asymptotic breakdown point) of regression medians. Without the uniqueness, (i) the sample regression median can never converge in probability or in distribution to its population version, (ii) deepest regression will yield more than one response and residual for a given \mathbf{x} , (iii) algorithms for the approximate computation of sample medians can never converge.

Uniqueness is so essential in our discussion of medians that there is a conventional remedy measure for non-uniqueness: to take average of all medians. This works in many scenarios, but not for $\beta_{RD_{RH}}^*$ and β_{DC}^* . This phenomenon for $\beta_{RD_{RH}}^*$ was first noticed by Mizera and Volaufo (2002)[8] and Van Aelst et al (2002)[17]. Concrete examples such as the real data example 3.2 and the artificially constructed one in the proof of Proposition 4.1 are presented here though.

Why do we just treat three regression medians

D_C ([1]) and RD_{RH} ([10]) are two pioneer notions of regression depth. PRD is recently introduced in [22]. The latter systematically studied the three regression depth notions w.r.t. four axiomatic properties, that is, **(P1)**, **(P2)**, **(P3)** and **(P4)** (see Section 2). It is found out that both regression depth RD_{RH} and projection regression depth PRD are real depth notions in regression since both satisfy **(P1)**-**(P4)**. While the former needs an extra assumption **(A)** (see Section 2.2), the latter does not need any extra assumptions. On the other hand, Carrizosa depth D_C violates **(P3)** in general, hence is not a real regression depth notion w.r.t. the definition in [22]. That motivates us to just focus on RD_{RH} and PRD throughout.

Summary and Conclusion

In terms of robustness, both depth induced medians are indeed robust. In fact, the median $\beta_{RD_{RH}}^*$ can asymptotically resist up to 33% ([16]) contamination, whereas β_{PRD}^* can resist up to 50% ([23]) contamination without breakdown, sharply contrasting to the 0% of the classical LS estimator.

In terms of efficiency, sample β_{PRD}^* could possess a higher relative efficiency when compared with sample $\beta_{RD_{RH}}^*$ (see [25]).

Now in terms of uniqueness, β_{PRD}^* again distinguishes itself from the leading depth median $\beta_{RD_{RH}}^*$ by generally possessing the desirable uniqueness property.

From computational point of view, RD (and $\beta_{RD_{RH}}^*$) has an edge over PRD (and β_{PRD}^*). The former is relatively easier to compute than the latter (see [25]).

By virtue of the performance criteria above, we conclude that PRD and β_{PRD}^* are promising options among the leading regression depths and their induced medians.

Acknowledgments

The author thanks Hanshi Zuo for his careful proofreading of the manuscript and two anonymous referees for their helpful and constructive comments and suggestions, all of which have

led to improvements in the manuscript. Special thanks go to the Managing Editor, Ms. Yuanyuan Yang for her enthusiastic invitation, encourage, and full support.

References

- [1] Carrizosa, E. A characterization of halfspace depth. *J. Multivariate Anal.* (1996), 58, 21-26.
- [2] Donoho, D. L. Breakdown properties of multivariate location estimators. PhD Qualifying paper, Harvard Univ. (1982).
- [3] Dohono, D. L. and Gasko, M. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* (1992), 20, 1803-1827.
- [4] Huber, P. J. Robust estimation of a location parameter. *Ann. Math. Statist.* (1964), 35, 73-101.
- [5] Liu, R. Y. Data depth and multivariate rank tests. In *Y. Dodge (ed.), L1-Statistical Analysis and Related Methods*, North-Holland, Amsterdam, (1992), 279-294.
- [6] Liu, R. Y., Parelius, J. M. and Singh, K. Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics* (1999), 27, 783-858, (With discussion).
- [7] Liu, X. and Zuo, Y. Computing halfspace depth and regression depth. *Communications in Statistics - Simulation and Computation* (2014), 43(5), 969-985.
- [8] Mizera, I. and Volau, M. Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *J. Multivariate Anal.* (2002), 83, 365-388.
- [9] Maronna, R. A., and Yohai, V. J. Bias-Robust Estimates of Regression Based on Projections. *Ann. Statist.* (1993), 21(2), 965-990.
- [10] Rousseeuw, P. J., and Hubert, M. Regression depth (with discussion). *J. Amer. Statist. Assoc.* (1999), 94, 388-433.
- [11] Rousseeuw, P.J., and Leroy, A. *Robust regression and outlier detection*. Wiley New York. (1987).
- [12] Rousseeuw, P. J., and Struyf, A. Computing location depth and regression depth in higher dimensions. *Statistics and Computing* (1998), 8, 193-203.
- [13] Rousseeuw, P. J., and Struyf, A. Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference* (2004), 122, 161-173.
- [14] Serfling, R. J. *Approximation Theorems for mathematical statistics*. John Wiley & Sons Inc. New York. (1980).
- [15] Tukey, J. W. Mathematics and the picturing of data. In: *James, R.D. (ed.), Proceeding of the International Congress of Mathematicians*, Vancouver 1974 (Volume 2), Canadian Mathematical Congress, Montreal, 1975, 523-531.
- [16] Van Aelst, S., and Rousseeuw, P. J. Robustness of Deepest Regression. *J. Multivariate Anal.* (2000), 73, 82-106.
- [17] Van Aelst S., Rousseeuw P.J., Hubert M., Struyf A. The deepest regression method. *J. Multivariate Anal.* (2002), 81, 138-166.
- [18] Wu, M., and Zuo, Y. Trimmed and Winsorized Standard Deviations based on a scaled deviation. *Journal of Nonparametric Statistics* (2008), 20(4), 319-335.
- [19] Wu, M., and Zuo, Y. Trimmed and Winsorized means based on a scaled deviation. *J. Statist. Plann. Inference* (2009), 139(2), 350-365.

- [20] Zuo, Y. Projection-based depth functions and associated medians. *Ann. Statist.* (2003), 31, 1460-1490.
- [21] Zuo, Y. Multidimensional medians and uniqueness. *Computational Statistics and Data Analysis* (2013), 66, 82-88.
- [22] Zuo, Y. On general notions of depth in regression. *Statistical Science* (in press) (2018a), arXiv:1805.02046.
- [23] Zuo, Y. Robustness of deepest projection regression depth functional. *Statistical Papers* (2018b), <https://doi.org/10.1007/s00362-019-01129-4>, arXiv:1806.09611.
- [24] Zuo, Y. Asymptotics for the maximum regression depth estimator. (2019a), arXiv:1809.09896.
- [25] Zuo, Y. Computation of projection regression depth and its induced median. (2019b), arXiv:1905.11846.
- [26] Zuo, Y., Serfling, R. General notions of statistical depth function. *Ann. Statist.* (2000), 28, 461-482.