ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2024, Volume 19, Issue 1

Robust Regression as a Sensible Alternative to the Weighted Ordinary Least Squares Regression in case of Heteroskedasticity. A Tutorial

Annalisa Orenti[(1)], *Anna Zolin*[(1)], *Ettore Marubini*[(*)], *Paolo Antonelli*[(2,#)], *Federico Ambrogi*[(1)], *Bruno Mario Cesana*[(1,#)]

(1) University of Milan, Department of Clinical Sciences and Community Health, Laboratory of Medical Statistics, Biometry and Epidemiology "G.A. Maccacaro", Milan, Italy
(2) Retired Professor of Calculus of Probabilities, Statistics and Operative Research at the State Industrial Technical Institute (ITIS) Benedetto Castelli, Brescia, Italy
(*) deceased
(#) retired

CORRESPONDING AUTHOR: Annalisa Orenti, Department of Clinical Sciences and Community Health, Laboratory of Medical Statistics, Biometry and Epidemiology "Maccacaro", University of Milan. Address: Via Celoria 22, 20133 Milan, Italy. E-mail: annalisa.orenti@unimi.it

SUMMARY

Background: The robust regression is rarely used in the statistical analyses in comparison with the Ordinary Least Squares regression and the Weighted Regression. In addition, in the frequent case of the heteroskedasticity of the residuals, a weighted regression carried out once is the main suggestion of the statistical books and the resulting reduced heteroscedasticity is usually considered sufficiently satisfactory.
Methods: We showed the OLS regression analysis on data simulated with a well evident heteroskedasticity and an ad hoc outlier, followed by a weighted regression iteratively carried out by using iteratively reweighted least squares, an estimation method used also in several procedures of the robust regression analysis. Therefore, the link between the iteratively performed weighted regression and the robust regression becomes immediate. Furthermore, the same data have been analysed using some robust regression procedures.
Results: It has been shown that in a simulated sample of heteroscedastic data with and without an obvious artificially created outlier the weighted regression performs worse with more biased parameter estimates than robust regression procedures (such as the robust MO procedure) as the presence of the outlier is not adequately neutralized.
Discussion: In presence of a heteroskedastic pattern of the residuals, the suggestion to use robust regression procedures which can also deal with the almost sure presence of outliers seems more sensible. Among the robust regression procedures carried out, the performance of the robust MO procedure appears particularly appealing since it allows biostatisticians a more reasoned management of the outliers shown in a very illustrative "ad hoc" plot. Robust regression procedures represent a sensible alternative to OLS regression taking into account that its assumptions are practically not always fulfilled and that outliers, which are almost certainly present, are not only difficult to handle in classical OLS regression but can also provide highly biased estimates.

Keywords: weighted regression; heteroscedasticity; iterative reweighted least squares, robust regression procedures (MO).

DOI: 10.54103/2282-0930/26484
Accepted: 23rd September 2024

INTRODUCTION

Robust regression does not appear adequately used by professional and not professional biostatisticians despite the theoretical advances on robust regression analysis done in the last thirty years as witnessed by excellent comprehensive textbooks from, among others, Atkinson and Riani[1], Rousseeuw and Leroy [2] Maronna et al. [3,4], Huber [5], and Huber and Ronchetti [6].

The aim of this tutorial is to show that, in case of a heteroskedastic pattern of the residuals, robust regression methods can replace much more effectively the weighted regression based on the Iterative ReWeighted Least Squares (IRWLS). In addition, robust regression methods allow optimal management of potential or real outliers.

This proposal can be considered an advance compared to the usual recommendation to resort to a single weighted regression usually suggested by many authors [7,8,9] to remove or at least reduce heteroskedasticity.

We therefore hope that biostatisticians, both professional and non-professional, will be more willing to adopt robust regression techniques also because the commercial and non-commercial software packages available today offer adequate data processing capabilities, capable of managing the computational requirements of these robust procedures.

Readers are assumed to be familiar with the statistical methodology of Ordinary Least Squares Regression (OLS-R) with a focus of methods for testing its assumptions and detecting outliers. In addition, at least a basic knowledge of Weighted Least Squares Regression (WLS-R, considered in more detail in this paper) is required. The relationships between the main equations used by OLS-R and WLS-R are shown in Table S4 of the supplementary material (s.m., thereafter).

Linear Regression - Statistical Theory

In this tutorial we will consider the second type of linear regression with both independent and dependent random variables. In fact, the first or "classic" type with the independent variable as a fixed variable is rather a theoretical model, useful however in the calibration of a new measurement method and the third type with both variables with a measurement error in addition to the biological variability belongs to the so-called measurement error models not considered in this paper. Furthermore, we will consider a simple linear regression (without loss of generality, assuming the number of the independent variables k equal to 1), so that the data points can be displayed in a scatter plot of X and Y to easily investigate the homogeneous or non-homogeneous pattern of the distribution of their sample values. However, our example is easily extendable to multiple regression.

For the Ordinary Least Squares (OLS) method, used to obtain the vector **b**, estimate of the parameter vector **β**, readers can refer to standard regression books such as Draper and Smith [7], Kuther et al. [8] or Chatterjee and Hadi [9] and the section "Linear regression – Statistical Theory" of the s.m. Here we would like to emphasize that the vector **b**, as given by equation (4) in Table S4 of the s.m., is the Best Linear Unbiased Estimator (BLUE) of the regression parameters (intercept and regression coefficients). This property follows from the Gauss-Markov theorem (see Kutner et al. [8, Chapter 1, page 18]) which states that among the unbiased linear estimators of **b**, the estimator with minimum variance is obtained by the Minimum Weighted Squares (MWS) method with the weight matrix (**W**) equal to the inverse of the error variance-covariance matrix: $\mathbf{W}=\mathbf{\Sigma}^{-1}$. Thus, OLS method is a special case of "Weighted Least Squares" (WLS) method when, due to homoscedasticity, equal residual variances can be collected to a common factor leaving all weights equal to 1. In turn, "Weighted Least Squares" method is a particular case of "Generalized Least Squares" (GLS) one when the error variance-covariance matrix is diagonal (i.e., the error terms are uncorrelated) with heteroskedasticity (the weights are different: $w_{i,i}= 1/\sigma_i^2$, being $w_{i,i}$ a generic element on the main diagonal of the weights matrix **W**; in fact, the inverse of a diagonal matrix is equal to a diagonal matrix with the reciprocal of its values on the diagonal).

Furthermore, the unbiasedness of the estimators does not require that the errors be normally distributed nor that they be independent and identically distributed as long as they are uncorrelated with zero mean and homoskedastic with finite variance. However, the requirement of the unbiasedness property has to be maintained since there also exist estimators with lower variance but biased.

Moreover, Carroll and Ruppert [12] do not consider the WLS but only the GLS that have the advantage of being applied without any distributional assumptions but specifying only the model for mean, variances and their relationship.

The OLS estimates are obtained by minimizing the sum of the squared residuals; namely,

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \alpha - \beta X_i\right)^2 \qquad (1)$$

Otherwise, in the case of the WLS the function to be minimized becomes:

$$\sum_{i=1}^{n} w_{i,i}\varepsilon_i^2 = \sum_{i=1}^{n} w_{i,i} \left(Y_i - \alpha - \beta X_i\right)^2 \qquad (2)$$

where $w_{i,i}$ is the i-th diagonal element of the (nxn) matrix **W**; the other terms are been defined in the paragraph "Linear Regression – Statistical Theory" of the s.m.. Thus, a weighted sum of the squared residuals is minimized, where each squared residual

is weighted by the reciprocal of its variance. In other words, when estimating **b**, less weight is given to the observations for which the linear relationship (to be estimated) is noisier and more weight to those for which it is less noisy.

Unfortunately, after an OLS-R, the recommended steps to test the statistical assumptions of the model (errors: identically, independently and normally distributed – the latter particularly relevant for the validity of the statistical tests on the estimates and the ANOVA table of the regression analysis) and also the adequacy of the model (straight line) are not systematically carried out even by professional biostatisticians.

Therefore, one may not observe a fan-shaped (or megaphone-like) pattern shown by the (externally studentized) residuals plotted against the fitted (predicted) values as an expression of a heteroskedastic distribution (higher variance for higher values of the fitted values) rather than the homoskedastic one required for the validity of the OLS analysis.

Methods: statistical analyses

We performed a simple OLS regression on the simulated dataset according to a heteroskedastic regression model as described in the s.m.. In particular, the mean and the standard deviation of the independent variable (X) were equal to 14 (μ_X) and to 6 (σ_X), respectively. The regression slope (β_1) and intercept (β_0) parameters are both 0.9.

For showing the advantage of performing a robust regression (see after), observation n.4 of the dataset was created as an outlier by increasing its ordinate to 15.09 from the original value of 2.33 and keeping the original simulated abscissa value of 1.59. Figure 1 shows the diagram plot of the simulated data (Panel A) and the diagram plot of the same data with observation n. 4 modified as an outlier (Panel B). In Figure 1, points whose OLS residuals were found outside some thresholds for outlier diagnostics are shown in red and marked with the observation number.

In order not to lose the thread of this tutorial from OLS heteroscedasticity to robust regression passing via weighted regression, the OLS regression results (ANOVA table, parameter estimates together with their standard error, t-statistics with their p-value, mean error squares, coefficient of determination (R^2) not adjusted and adjusted) are shown in the s.m. (Table S1 and Table S1.1, respectively). The paragraph "Considerations about the coefficient of determination" which deals with some theoretical aspects of the unadjusted and adjusted coefficient of determination of OLS and WLS regressions, the paragraphs "outlier diagnostics: theory" and "outlier diagnostics: data with the outlier", together with some plots obtained with the keyword "influence" from the OLS regression by SAS® Proc REG [26] are shown and commented in the s.m. to which interested readers are referred.

Figure 2 shows the plot of the "externally studentized residuals" or "jackknifed residuals" (see s.m.) called "Rstudent" in SAS®, more effective in detecting outlying Y observations than "internally studentized residuals", vs. the fitted values as a practical example of residual heteroskedasticity of the fifty simulated data without and with the artificially created outlier. The points are marked with the observation number to better understand why these observations exceeded some thresholds to be considered "suspected outliers" (see s.m.).

In Figure 2-Panel A, a fan-shaped pattern is quite evident, especially considering that these residuals

Figure 1. Scatter diagram of the simulated data (Panel A) and the same data with observation n.4 modified as an outlier (Panel B)

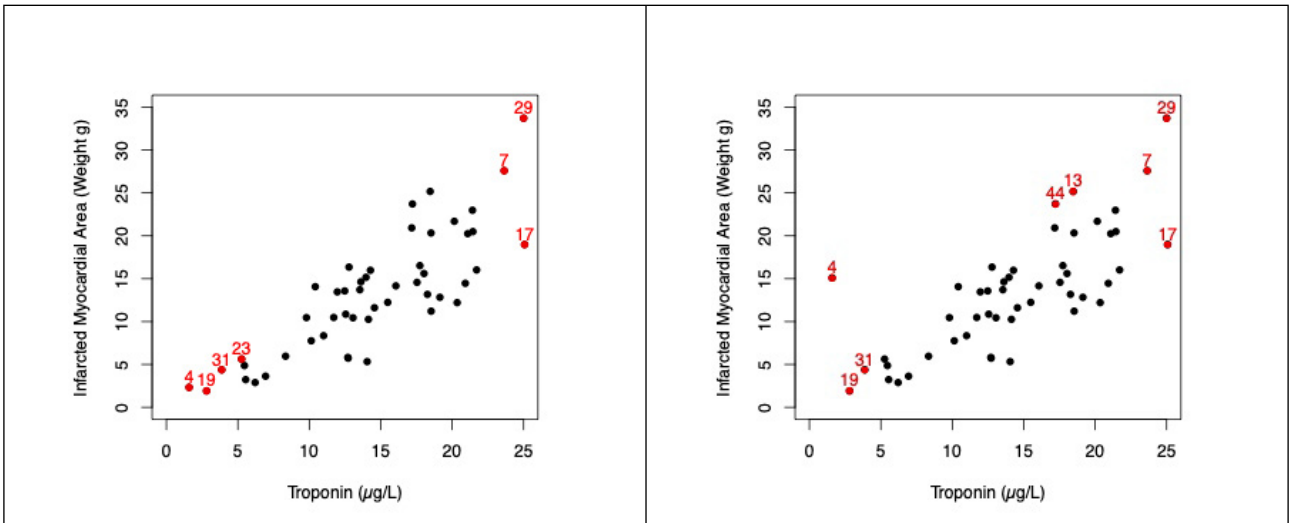


Figure 1- Panel A | *Figure 1 - Panel B*

Figure 2. Scatter diagram of the externally studentized residuals and fitted values for the simulated data (Panel A) and the same data with observation n.4 modified as an outlier (Panel B)

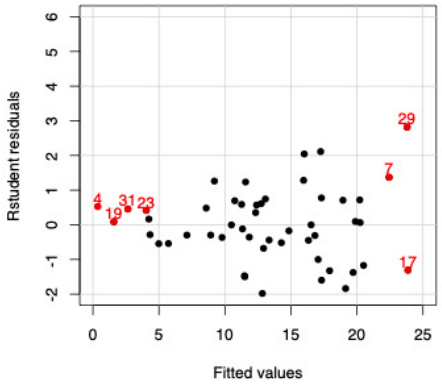


Figure 2-Panel A

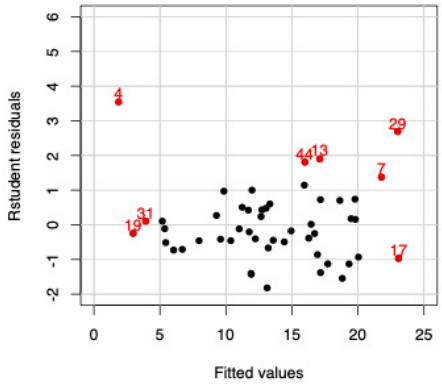


Figure 2-Panel B

come from simulated data with heteroskedastic errors. The same pattern is shown in the Figure 2-Panel B for the data with the artificially created Y-direction outlier, although the outlier (n.4 in the top left of the Figure 2-Panel B) is well above the fan-shaped pattern better evident in Figure 2-Panel A without the outlier.

It should be noted that, due to heteroscedasticity, OLS estimator still provides unbiased and consistent estimates but no longer of minimum variance (see Kutner et al. [8 Chapter 11]). Finally, according to Chatterjee and Hadi [9, 5th Ed. Page 193] it is possible to conclude that the coefficients "lack precision in a theoretical sense".

The presence of the heteroskedasticity can be formally tested by means of some formal statistical tests, as reported in Appendix A of the s.m.. However, since the statistical tests may provide inconsistent results, it is strongly recommended to base the judgement of heteroskedasticity on the pattern of the (externally studentized) residuals against the independent or the dependent or the fitted variables.

Outlier Diagnostics

Another relevant point in any type of statistical analysis, but particularly relevant in the context of the regression analysis, is the identification of outliers (usually defined as Outlier Diagnostics; see s.m.) in the sample dataset. In fact, the presence of just one outlier can dramatically modify the OLS estimates. Furthermore, it should be emphasized that the presence of a heteroskedastic pattern makes more difficult the identification of the outliers. For the outlier diagnostics (leverage, Mahalanobis distance or better the squared Mahalanobis distance -MDi^2-, the "standardized residuals", the "studentized residuals" or the "internally studentized residual", the DFFITS(i) statistics, the DFBETASj(i) statistics, and the "COVRATIO") together with the related diagnostic thresholds, readers are referred to the already cited paragraphs "Outlier diagnostics: theory" and "Outlier diagnostics: data with the outlier" of the s.m.

Statistical approaches to deal with the heteroskedasticity

Heteroskedasticity can be removed by a suitable transformation of the variables according to Cohen et al. [32] and Mosteller and Tukey [33] with their Bulging Rule (or Ladder of Powers) suggesting power transformations (of X or of Y or both) with exponents of 2, 1, 0.5, -0.5, -1 and -2 including the logarithmic transformation. However, this approach leads to the very relevant problem of attributing a meaning to the relationship between the transformed variables with respect to what the researcher wanted to evaluate between the original variables.

However, even non-professional biostatisticians are well aware that, in the case of heteroskedasticity, several statistical books dedicated to regression [7,8,9] suggest the Weighted Regression (WR) analysis as a more sensible alternative to transformations. WR is a procedure based on a generalization of the regression model, which is implemented by assigning different weights to each observation, instead of the weight equal to 1 given by the OLS method, being homoscedasticity. Therefore, it is necessary to calculate weighted least squares (WLS) estimates instead of OLS. Furthermore, the use of weights will (legitimately) impact the widths of the statistical intervals.

However, there is a practical difficulty in determining the weights to estimate the error variances (or standard deviations) that provide the **W** matrix to be used to obtain the WLS estimator ($\mathbf{b}_w$).

In some cases, weights values may be based on theory or previous research. For example, when the error variance is proportional to an independent variable, the natural weights are the reciprocal of the independent variable.

However, in the usual case where the structure of the matrix **W** is unknown, it is necessary to estimate the variance or the standard deviation function, accordingly.

In experiments designed with large numbers of replicates or in the case of some measurements of the dependent variable at the same or nearly the same value of the independent variable (replicates or nearly replicates), weights can be estimated directly from the sample variances of the response variable at each combination of the same or nearly the same values of the independent variable. An exemplification of this approach is shown in Draper and Smith's book [7, paragraph 9.2. Generalized Least Squares and Weighted Least Squares, page 221] where the variances of the Y values, computed at five means of equal or nearly equal X values, were regressed on the X means and the X means squared, due to a suggested quadratic relationship. Of course, this regression allows the variance pattern to be modelled and the regression coefficients to be used obtain the fitted variance values for each X. Then, the reciprocal of the fitted variance at each X value was used as the weight for the WR between the original Y and X dataset.

As a further example of this approach, consider the analysis of the dataset (https://online.stat.psu.edu/stat501/lesson/13/13.1) consisting of seven observations of the pea diameter (in inches) of the parent plant (X) and the mean diameter (in inches) of up to 10 plants grown from seeds of the parent plant (Y) made by Sir Francis Galton (16 February 1822 Birmingham England – 17 January 1911 Haslemere, Surrey, England). Therefore, it is possible to calculate the variance of the progeny plants and use its reciprocal as weights for the WR analysis.

Further issues are reported in Appendix B of the s.m. to which the interested reader is referred.

Usually, statistical books report an example of WR together with the plot of the WLS residuals against the variable for which a megaphone pattern has been highlighted and conclude that a satisfactory or rather satisfactory reduction in heteroskedasticity has been achieved. For example, Draper and Smith [7, page 229] report "The residuals plots in Figure 9.2 reveal that the vertical spread of residuals is now *roughly (our italic)* the same at the two main levels of the transformed response. At lower levels there are only two observations so that there is not much of an estimate of the spread there. The employment of weighted least squares here appears to be justified and useful".

WLS estimates of coefficients are generally close to the "ordinary" unweighted OLS estimates. However, Kutner et al. [8, 1974, page 426] and Chatterjee et al. [9] point out that if the estimated WLS coefficients differ substantially from the estimated OLS coefficients it is recommended to repeat the WLS regression until the estimated coefficients stabilize by using the revised weights obtained by the residual of the previous WLS regression to re-estimate the variance or the standard deviation function; this process gives the "*iteratively reweighted least squares (IRLS or IRWLS)*." Often the stabilization of the coefficients is achieved in no more than one or two iterations.

Similarly, the same procedure should be followed in the case of an unsatisfactory removal of the heteroskedasticity after the first WLS regression, as shown by the plots of the residuals towards the variable with which the megaphone pattern is highlighted at the first OLS-R. Only the studentized residuals take into account the weights that are used to model the different values of the variance and, consequently, these residuals must be used to draw the diagnostic plots.

However, the iterative steps of the weighed regression are a demanding procedure especially if they have to be performed without the availability of ad hoc software.

Indeed, after the externally (better) studentized OLS residuals plotted against a predictor (or fitted values) show a megaphone shape, a second OLS should be performed to estimate the variance function (or the standard deviation function) by regressing the squared residuals (or the absolute residuals) on the predictor or fitted values with which a megaphone pattern was evidenced. Indeed, if the first OLS-R model is correct the i-th squared residual is an estimate of σ_i^2 and the i-th absolute residual is an estimate of σ_i to be preferred in presence of outliers with expected largest residuals so as not to have a very low weight being $w_i=1/\sigma_i^2$ or $1/\sigma_i$.

Then the reciprocal of the fitted values by the estimated variance ($1/\sigma_i^2$) or standard deviation function ($1/\sigma_i$) are used to obtain the weights for the first WLS-R. Next, the procedure needs to be repeated using the set of externally studentized residuals from the WLS-R to re-model the variance (or standard deviation) with an OLS-R and the resulting residuals will be the weights to be used in the second WLS-R, and so on.

An "ad hoc" software that allows to perform the iteratively reweighted least squares procedure could be very useful and in fact a code in the open-source R language is available upon request to the corresponding author.

Table S2 s.m. shows the results of the iterative process using IRLS starting from the OLS regression with a heteroskedasticity pattern clearly evident from the plot of the OLS externally studentized residuals and the fitted $\hat{Y}$ (Y-hat) variable (Figure S1-Panel A equal to Figure 1-Panel A).

Finally, the OLS regression results are reported in s.m. together with those of the iterative steps of the weighted regression (Figure S1, Panel A, Panel B, Panel C, and Panel D).

Iteratively reweighted Least Squares (IRWLS)

The IRWLS estimator used above in iterative weighted regression is also used in some "robust regression" procedures, among other estimating approaches.

However, in weighted regression the parameter estimates are obtained by minimizing the weighted sum of the squared residuals; otherwise in the robust regression the parameter estimates are obtained by minimizing a particular function of the squared residuals.

Robust regression instead of the OLS regression allows to dampen the influence of outliers that inevitably exist in medical and biological datasets and that are difficult for the researcher to handle.

In fact, it is worth remembering the following clarification of Maronna et al. [3, page 51]: "… while in the classical approach to statistics one aims at estimates which have desirable properties at an exactly specified model, the aim of robust methods is loosely speaking to develop estimates which have a 'good' behaviour in a 'neighbourhood' of a model".

An online SAS® documentation reports the connection between robust regression and weighted least squares. [34] In fact, the use of the IRWLS estimator naturally leads to considering a robust regression whose main advantage consists in an adequate handling of outliers.

Other relevant references are the papers of Holland and Welsch [35], Street, Carroll and Ruppert [36] Heiberger and Becker [37], and Green [38].

Readers interested in a more in-depth clarification of the robust regression methodology and its estimating procedure may referred to Appendix B of s.m..

Since the weights change from one iteration to another, the *weighted* residual sum of squares could not decrease at each iteration. Indeed, for removing this restriction, it has to specify the keyword "NOHALVE" in the PROC NLIN of SAS® [26].

We would like to consider as an explanatory approach the robust Multiple Options (MO) procedure proposed by Orenti and Marubini [39] (see Appendix C of s.m. for an in-depth illustration) together with the robust regression results obtained from the Least Trimmed Squares (LTS) and the MM estimator.

The MO procedure has the particular characterization of iteratively checking for outliers after obtaining a first bulk of observations and has demonstrated satisfactorily behaviour for identifying outliers. The others two robust procedures were considered for their satisfactorily performance and frequent use.

Table 1. Results on data with the outlier

Estimates	OLS	OLS*	IRWLS	MO	MM	LTS
Intercept (b_0) SE P	-1.22785 1.49385 0.4152	0.42970 1.67672 0.7988	1.6778 1.4586 0.256	-0.65877 1.45311 0.652	-0.6064 1.6354 0.712	0.16334 1.47560 0.912
Slope (b_1) Se P	1.00084 0.09645 <.0001	0.90325 0.10825 <0.0001	0.8133 0.1056 <0.0001	0.95421 0.09482 <0.0001	0.9520 0.1303 <0.0001	0.85023 0.09634 <0.0001
MSE	3.94245	4.42505	4.495	3.535	4.072	3.576
R^2	0.6917	0.5919	0.5528	0.6784	0.6303	0.6338
Adj-R^2	0.6853	0.5834	0.5435	0.6717	0.6226	0.6257

OLS - OLS regression with the observation n.4 made as an outlier owing to the increase of the simulated value of its ordinate.*

OLS, WLS, and Robust regression procedure results

Table 1 shows the results of the fitted regression models. The estimates of the weighted regression performed with the iterative weighted estimators and of the MO, MM, and LST robust regressions are those computed at the final step of their iterative process.

Column 2 (OLS) and column 3 (OLS*) allow to capture the large difference between the OLS intercept estimates without and with the observation n. 4 created as an outlier. Indeed, it was expected that the outlier in the Y-direction located approximately at the beginning of the observations would shift the regression line clockwise leading to a positive intercept rather than a negative one. Furthermore, since there is only one outlier the OLS slope estimate is relatively little affected with a reduction of about 10%.

The IRWLS estimates of the weighted regression are very influenced by the presence of only one outlier with the greatest positive intercept and the lowest slope values leading us to conclude that, at least in this case, the weighted regression with IRWL was not a sensible choice.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2024, Volume 19, Issue 1

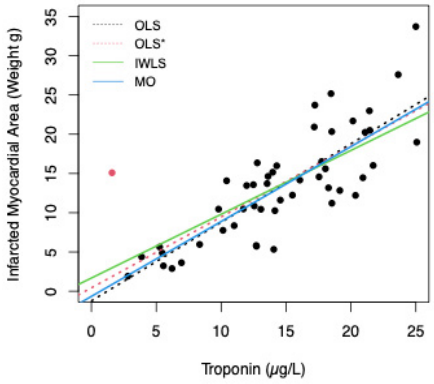


Figure 3-Panel A
OLS = OLS regression with the outlier*
OLS = OLS Regression without the outlier (observation n.4 with original data)

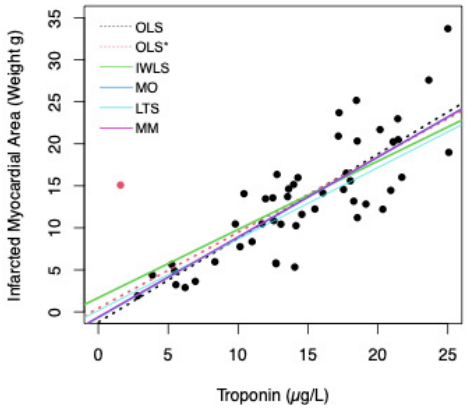


Figure 3-Panel B. The regression lines of the LTS and MM estimator of the robust regression are in addition shown

Furthermore, the estimates of the MO procedure are very close to those of the OLS without the outlier; finally, taking as a reference the OLS estimates of the dataset without the outlier, the MO estimates turned out to be a little less biased than the MM and LTS estimates.

Figure 3-Panel A shows the regression lines fitted with the OLS, the OLS*, IRWLS, and MO robust regression. Figure 3-Panel B shows also the regression lines fitted from the LST and MM robust regressions. This figure makes it easier to understand the comments reported about the Table 1 with the OLS* regression line deviating by the created outlier and the MO, LTS and MM regression lines close to the OLS line. Finally, it is evident that the weighted regression with the IRWLS estimator (line with the greatest intercept) is not able to overcome the influence of the outlier.

Table 2 shows the results of the regression models fitted without the outlier. The first (OLS) and the second (OLS*) column are equal to the corresponding columns in Table 1 and have been reported for easier comparison. The estimates of the weighted regression performed with the iterative weighted estimators and of the MO, MM, and LST robust regressions are those calculated in the final step of their iterative process.

Again, it is worth highlighting the poor performance of the IRWLS weighted regression with the and the lower bias of the robust MO procedure compared to the other two robust procedures.

Figure 4 shows the regression lines fitted with the OLS, IRWLS, MO, LTS, and MM robust regressions. In particularly, due to their very similar intercept and slope estimates the MO and MM regression lines overlap: specifically, the intercept is -0.78189 and -0.7723, and the slope is 0.96160 and 0.9607, respectively.

Table 2. Results from data without the outlier

Estimates	OLS	OLS*	IRWLS	MO	MM	LTS
Intercept (b_0) SE P	-1.22785 1.49385 0.4152	0.42970 1.67672 0.7988	0.19188 0.49302 0.699	-0.78189 1.33895 0.562	-0.7723 1.2122 0.527	0.2768 1.2859 0.831
Slope (b_1) SE P	1.00084 0.09645 <.0001	0.90325 0.10825 <0.0001	0.88750 0.05756 <0.0001	0.96160 0.08818 <0.0001	0.9607 0.1128 <0.0001	0.8433 0.0851 <0.0001
MSE	3.94245	4.42505	2.285	3.435	3.859	3.325
R^2	0.6917	0.5919	0.832	0.7124	0.6627	0.6858
Adj-R^2	0.6853	0.5834	0.8285	0.7064	0.6659	0.6788

OLS - OLS regression with the observation n.4 created as an outlier by increasing the simulated value of its ordinate*

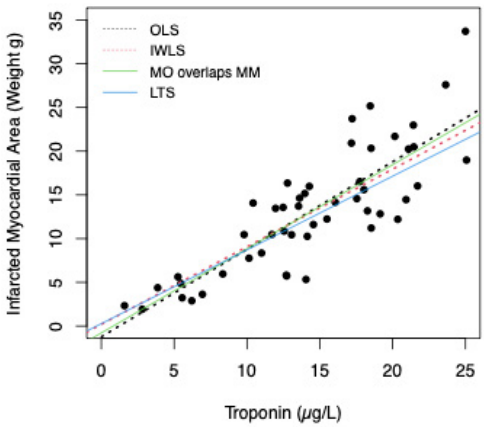


Figure 4

Figure 5, Panel A and Panel B shows the very informative plot obtained from the final iteration of the robust MO regression [39]. Indeed, Figure 5 shows four quadrants: two quadrants are above the horizontal line drawn at 0.5 which delimits the value of weight given to the observations (under or above) and the two remaining quadrants are to the right and left of the vertical line drawn at 0.89517, a value given by the Neperian logarithm of the square root of the 0.95 quantile of a χ^2 distribution with 2 degrees of freedom ($\ln\sqrt{5.99146} = \ln(2.44775) = 0.89517$) (see Appendix C of the s.m.). This line corresponds to the threshold of the Neperian logarithm of the robust distance ($\ln {}_{Z}RD$) that delimits the low (left) and high (right) leverage points. In particular, in the two quadrants below the line drawn at the weight value of 0.5 there are the observations considered as outliers and "bad leverage points" for the observations on the right of the vertical line drawn at the above reported value of 0.89517. In addition, the two quadrants over the horizontal line drawn at the weight value of 0.5 are the location of the "bulk" (left quadrant) and of the "good leverage points" (right quadrant) as opposed to the "bad leverage points" as they influence the regression fitting without providing biased estimates compared to those that would be obtained with the bulk data.

It is possible to see two observations in the dataset with the outlier (n.4 and n. 29 with weights of 0.16241 and 0.45452, respectively) under the line of the 0.5 threshold in the bottom right quadrant. Of course, in the dataset without observation n.4 created as an outlier, only observation n. 29 is considered an "outlier" with an attributed weight of 0.45155. According to MO, observation n. 7 is at a relevant distance (exceeding the above reported threshold of Cook's robust distance) from the bulk for both datasets as a "good leverage point".

Moreover, observations n. 13, and 44 are close to or above the threshold for the dataset with the outlier and show a further shift to the right for the dataset without the outlier. For the latter dataset, observation n.13 becomes a "good leverage point" while observation n.4 and observation n. 44 are close to but to the left of the threshold and just at the threshold, respectively.

The robust MO procedure assigns weights to the observations without the imputed outlier with a mean of 0.9110 (±0.1139, s.d.), median equal to 0.9674, first (Q_1) and third (Q_3) quartiles equal to 0.8647 and 0.9894, respectively, minimum and maximum values of 0.4516 (obs. n.29) and 1.0000, respectively. Only two observations (n. 19 and n. 28) have a weight of 1, contrary to what happens in the OLS regression in which all observations have a weight of 1. In addition, even in this case the negative skewness of the distribution is evident since the median is much greater than the arithmetic mean.

The weights given by MO to the observations with the imputed outlier have mean of 0.9010 (±0.1548, s.d.), median equal to 0.9654, first (Q_1) and third (Q_3) quartiles equal to 0.8681 and 0.9894, respectively, minimum and maximum values of 0.1624 (obs. n. 4) and 0.9999 (obs. n. 28), respectively. No observations have a weight of 1, contrary to what happens in the dataset without the observation n.4 created as an outlier and in the OLS regression. In addition, the negative skewness of the distribution is evident since the median is much greater than the arithmetic mean.

The descriptive statistics of the weights of the other three methods (IRWLS, MM, and LTS) are reported in the s.m..

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2024, Volume 19, Issue 1

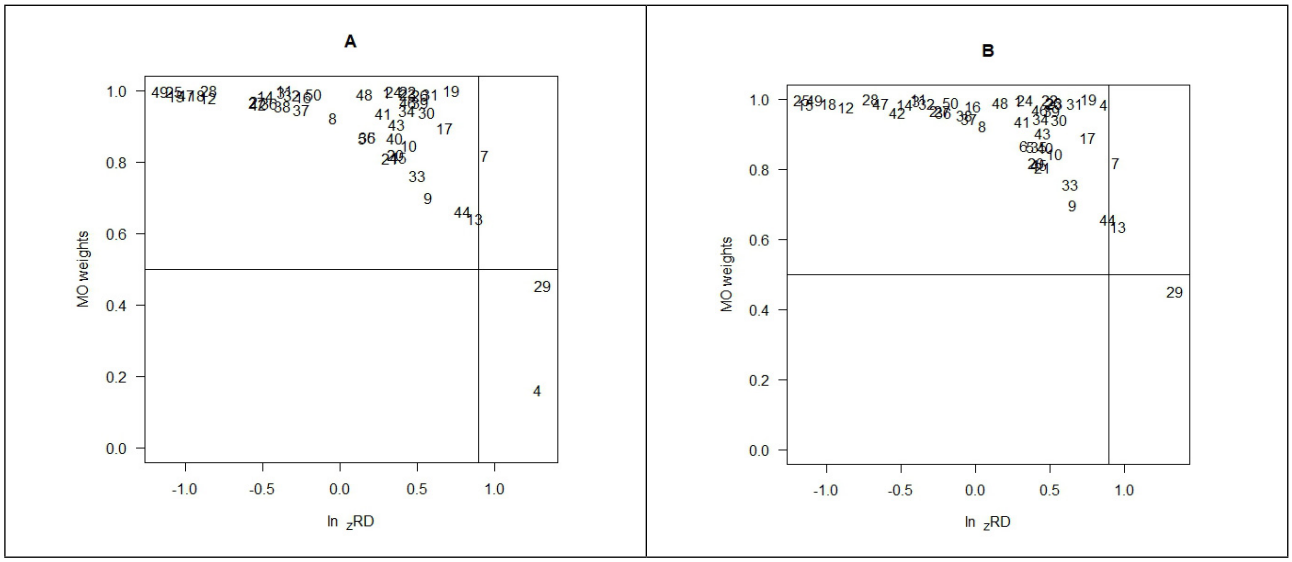


Figure 5. MO distribution of the weights given to observations in four quadrants. Data with the outlier (obs. n.4) (Panel A). Data without the outlier (Panel B)

CONCLUSIVE REMARKS

We have shown how the heteroskedastic pattern of the residuals leads to solutions that cannot be naively granted. Indeed, the Weighted Regression suggested as a not particularly sophisticated statistical method and as a standard solution to handle heteroskedasticity can be quite unsatisfactory with estimates very far from the expected ones, given the almost sure presence of outliers.

Our exemplification with data simulated according to a heteroskedastic model and with only one outlier created is not particularly illustrative of the advantages of using robust regression methods to handle potential or definite outliers, considering also that the handling of these observations is a very difficult task. However, this exemplification has shown how weighed regression even if iterated to removing the heteroskedasticity pattern can show an unsatisfactory behaviour compared to robust regression procedures.

In promoting the use of these robust procedures, special attention has been paid to the MO robust regression since its unique feature is to recover observations that are considered as outliers or at least not belonging to the bulk of observations after the first stage of its iterative process. In addition, its final figure with the residual classified as outlier or not and bad or good leverage points allows the researcher to make a sensible decision whether or not to exclude some observations from the dataset to be analysed.

A note of caution must be expressed regarding the interpretation of the determination coefficient (R^2) adjusted or not since in the context of the Weighted Regression or Robust regression it does not have the usual interpretation of the variance explained by the explanatory variables (see s.m.).

Finally, methods to deal with heteroskedasticity are not limited to weighted regression or to robust regression procedures. Indeed, since WLR estimates are as consistent and unbiased as those from OLS regression as long as the mean function in the regression model is correctly specified, it is possible to focus on the variability and to adopt methods that lead to bootstrapped standard errors computed nonparametrically by resampling from observed data [40] or to the Sandwich Standard Errors [40, 7.2.7 Sandwich Standard Errors for Least-Squares Estimates paragraph]. However, since it is not possible to rely with certainty on the fulfilment of their assumptions and since more technical statistical knowledge is required, it is strongly recommended to rely on the robust regression models and in particular on robust MO regression.

No funds have to be acknowledged.
No conflicts of interest have to be declared.
No acknowledgements have to be expressed.

REFERENCES

1. Atkinson A, Riani M. Robust Diagnostic Regression Analysis Springer 2000 New York
2. Rousseeuw PJ, Leroy AM. Robust Regression and Outlier Detection 1987 and 2003, John Wiley & Sons Inc.
3. Maronna RA, Martin RG, Yohai VJ. Robust Statistics: Theory and Methods 2006 John Wiley & Sons, Ltd. The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.
4. Maronna RA, Martin RD, Yohai VJ, Salibián-Barrera M. Eds Robust Statistics: Theory and Methods (with R) 2019 John Wiley & Sons Ltd.
5. Huber PJ. Robust Statistics 2005 John Wiley & Sons, Inc.
6. Huber PJ. Ronchetti EM. Robust Statistics. 2nd Ed. 2009 John Wiley & Sons, Inc.
7. Draper NR, Smith H. Applied Regression Analysis 3rd Ed. 1998 by John Wiley & Sons, Inc.) (Draper NR, Smith H. Applied Regression Analysis 3rd Ed. 1998 John Wiley & Sons, Inc. NY USA)
8. Kutner MH, Nachtsheim CJ., Neter J, Li W. Applied Linear Statistical Models 5th Edition-McGraw-Hill Irwin Companies, Inc., New York, NY, 2005, 1996, 1990, 1983, 1974 Pag 426-
9. Chatterjee S, Hadi AS. Regression analysis by example 2012 5th Ed. John Wiley & Hoboken, New Jersey. Chatterjee S, Price B. 1977, 2nd Ed. 1991, 3rd Ed. 1999, 4th Ed. 2006. Hadi AS, Chatterjee S. Regression Analysis by Example Using R. Sixth Ed. 2023
10. Hoaglin DC, Welsch RE. (1978). The hat matrix in regression and ANOVA. The American Statistician, 32(1), 17-22. doi: 10.2307/2683469
11. Orenti A, Marano G, Boracchi P, Marubini E. Pinpointing outliers in experimental data: the Hat matrix in Anova for fixed and mixed effects models JPH 2012, 9, 4 e 8663 – 1-13
12. Carroll RJ, Ruppert D. Transformation and Weighting in Regression Chapman and All NY 1988.
13. Rousseeuw PJ, Yohai, V. J. (1984). Robust Regression by Means of S-estimators. Robust and Nonlinear Time series, J. Franke, W. Härdle and R. D. Martin (eds.), Lectures Notes in Statistics 26, 256-272, New York: Springer.
14. Hoaglin DC, Welsch RE. (1978), The hat matrix in regression and ANOVA, Am. Stat., 32, 17-22.
15. Henderson HV, Velleman PF. (1981), Building multiple regression models interactively, Biometrics, 37, 391-411.
16. Cook RD, Weisberg S. (1982), Residuals and Influence in Regression, Chapman & Hall, London
17. Hoaglin DC, Mosteller F, Tukey JW. (1983), Understanding Robust and Exploratory Data Analysis, John Wiley & Sons, New York.
18. Paul SR. (1983), Sequential detection of unusual points in regression, The Statistician, 32, 417-424.
19. Stevens JP. (1984), Outliers and influential data points in regression analysis, Psychol. Bull., 95, 334-344.
20. Belsley DA, Kuh E, Welsch RE. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. (1980), John Wiley & Sons, New York.)
21. Cook RD, 1977 Detection of influential observation in linear regression, Technometrics 19, 15-18.
22. Draper NR, John JA. 1981. Influential observations and outliers in regression. Technometrics 23, 21-26
23. Atkinson AC, 1982 Regression diagnostics, transformations and constructed variables. J. R. Stat. Soc. Ser. B, 44, 1-36
24. Hocking RR. (1983) Development in linear regression methodology: 1959-1982, Technometrics 25, 219-249
25. Hocking RR, Pendleton OJ. (1983) The regression dilemma. Commun Stat (theory and Methods) 12, 497-527
26. SAS Institute Inc. 2016. SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc
27. Atkinson AC. (1983), Diagnostic regression for shifted power transformations, Technometrics, 25, 23-33
28. Velleman PF, Welsch RE. (1981), Efficient computing of regression diagnostics, Am. Stat., 35, 234-242.
29. Montgomery D C, Peck AE. Introduction to Linear Regression Analysis (1982), John Wiley & Sons, New York.
30. Nagelkerke, N. J. D. (1991) A note on a general definition of the coefficient of determination. Biometrika 78, 691–692.
31. Singer JD, Willett JB. (2003) Applied Longitudinal Data Analysis - Modeling Change and Event Occurrence. University Press Scholarship Online Oxford Scholarship Online.
32. Cohen J, Cohen P, West SG, Aiken LS. (2002) Applied multiple correlation/regression analysis for the social sciences third Ed. Lawrence Erlbaum Associates, Publishers 2003 Mahwah, New Jersey London)
33. Mosteller F, Tukey JW. (1977) Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley Publishing Company Reading, MA USA.
34. https://documentation.sas.com/doc/en/statcdc/14.2/statug/statug_nlin_examples02.htm.
35. Holland PW, Welsch RE. (1977) Robust regression using iteratively reweighted least-squares Communications in Statistics - Theory and Methods Communications in Statistics - Theory and Methods 6 (9), 813-827.
36. Street JO, Carroll RJ, Ruppert D. (1988). A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares. The American Statistician, 42(2), 152–154.
37. Heiberger RM, Becker RA. (1992). Design of a Function for Robust Regression Using Iteratively Reweighted Least Squares. Journal of Computational and Graphical Statistics, 1(3), 181–196.
38. Green PJ. (1984) Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives Journal of the Royal Statistical Society. Series B (Methodological) 46 (2), 149-192.
39. Orenti A, Marubini E. Robust regression analysis: a useful two stage procedure, Communications in Statistics - Simulation and Computation, 2021;50:16-37, doi: 10.1080/03610918.2018.1547400
40. Fox J, Weisberg S. An R Companion to Applied Regression 3rd Ed. 2019 Sage Publications, Inc 2455 Teller Road Thousand Oaks, California.

DATA SET: 50 observations with one created outlier (type: outlier = 0 / bulk = 1 observation n.4).
The data have been simulated, according to a heteroscedastic pattern, starting from the values (means and standard deviations rounded) of the Troponin (X, mg/L) and the weight (Y, g) of the infarcted myocardial area of beagle dogs obtained by heart dissection after being sacrificed shown in the book from: Cesana Bruno Mario, Antonelli Paolo and Pea Giuseppe. La Statistica per le Scienze Biomediche 2012 Libreria Universitaria.
Particularly, the 50 data (rounded to the second decimal figure) have been simulated according to a Gaussian distribution with mean equal to 14 (μ_X) and standard deviation equal ($\sigma_X = 6$).
In addition, the regression parameters are: slope (β_1) = 0.9 and the intercept (β_0) = 0.9.
Furthermore, a heteroskedastic error has been added by increasing the measurement error of the Y variable (created as the dependent variable of a straight linear regression) by an increasing quantity depending on the increase of the X variable.

R code for the simulation process of X and Y variables with heteroskedastic errors

```
# simulation of the X values;
set.seed(seed = 135791)
X_Trop=round(rnorm(n,muX,sigmaX),2)

# fixing the standard error of the heteroskedastic residuals;
sigma_err = 6.84

# simulation of errors according to a Gaussian distribution;
set.seed(seed = 24682)
err_sim=rnorm(n,mean = 0,sd =  sigma_err)

# calculation of heteroskedastic errors proportional to X variable;
hmin=1; hmax=4
# hmax and hmin are multiplicative factors of the error_Xmax and of the error_Xmin
# multiplicative (linear) factor equal to 1 for Xmin and equal to 4 for Xmax
# the following equation is the equation of a straight line passing through 1 and 4;
fm=sqrt(((hmax-hmin)*(X_Trop-min(X_Trop))/(max(X_Trop)-min(X_Trop)))+hmin)

# calculation of the heteroskedastic errors;
err_etsc= err_sim*fm

# calculation of the values of Y_Weight as the dependent variable of a linear regression
Y_Weight = round(beta0 + beta1*X_Trop + err_etsc,2)
dt_TropWeight = as.data.frame(cbind(X_Trop,Y_Weight))

# creation of only one outlier: observation n.4
out_indx= 4
outls= c( 1.59, 15.09)

# substitution of the observation n.4° in the data set dt_TropWeight with the outlier (outls);
dt_TropWeight_with_1outls= dt_TropWeigh
dt_TropWeight_with_1outls[out_indx,1:2] = outls
```

According to a SAS ® code for reading the data:

```
DATA Trop_weigth_outlier;
INPUT NUM X_Trop Y_Weight TYPE @@; *TYPE = OUTLIER = 0 / BULK = 1;
CARDS;
1   21.09  20.25    1  2   13.61  14.63    1
3   18.03  15.60    1  4    1.59  15.09    0
```

```
5   12.78  16.35    1  6   19.14  12.83    1
7   23.64  27.59    1  8   18.27  13.17    1
9   14.05   5.33    1  10  17.17  20.92    1
11  17.74  16.52    1  12  15.49  12.24    1
13  18.46  25.17    1  14  17.54  14.57    1
15  14.56  11.61    1  16   9.80  10.46    1
17  25.07  18.98    1  18  13.06  10.44    1
19   2.81   1.92    1  20  12.70   5.82    1
21  18.53  11.21    1  22   5.46   4.87    1
23   5.26   5.63    1  24  21.45  20.49    1
25  16.06  14.16    1  26   5.55   3.23    1
27  12.48  13.57    1  28  11.71  10.48    1
29  25.00  33.71    1  30  21.42  22.98    1
31   3.87   4.38    1  32  11.00   8.35    1
33  20.35  12.21    1  34  20.15  21.69    1
35  10.42  14.07    1  36  13.97  15.16    1
37  14.28  15.98    1  38  11.96  13.46    1
39   6.21   2.89    1  40  20.92  14.46    1
41  18.52  20.33    1  42  14.14  10.26    1
43  21.71  16.02    1  44  17.22  23.71    1
45  12.72   5.73    1  46   6.94   3.63    1
47  13.54  13.71    1  48   8.34   5.96    1
49  12.55  10.86    1  50  10.14   7.76    1
; RUN;
```

It has to be noted that the observation 4 (highlighted in bold) has been created as an outlier in the Y-direction by a huge increase of its ordinate (from 2.33 to 15.09) and keeping the original abscissa value from the simulation process. So, the dataset without the outlier is obtained by inserting the Y-value of 2.33 instead of 15.9 in observation n.4.

Linear Regression - Statistical Theory

The simple regression model: $Y_i = \alpha + \beta X_i + \varepsilon_i$ can be more conveniently written in matrix terms allowing for immediate extension to multiple regression. Consequently, the two parameter estimates in the case of a simple regression (a: intercept and b: regression coefficient/slope) are incorporated into a vector **b** (2 rows and 1 column) and, accordingly, they will be defined as b_0 and b_1, respectively, or as b_0 to b_k parameters estimates in the case of a multiple linear regression with k independent variables. So, in a sample of size n, the model pertinent to the i-th observation (case) for a linear regression and also for a general linear model is:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i \text{ with } i = 1, ..., n \qquad (1)$$

where: y_i is the random dependent variable (response); **x'**$_i$ is the i-th row of the matrix **X** of size n x (k+1). In observational studies aimed to evaluate the role of the independent variables in explaining the response we can think of $(\mathbf{x}_i', y_i)$ as a point in a (k+1) dimensional space. On the contrary, in experimental studies **X** is a non-random matrix defined by the structure of the experimental design and, consequently it is predetermined by the experimenter [10-11]. In both settings, the rank of the matrix **X** is the number of its independent column equal to k+1 with the first column consisting of all 1 for obtaining the intercept of the regression analysis or the grand mean of the experimental design. Of course, in the WLS regression with observation weights other than 1, the first column will consist of the actual values of the weights.
β is the (k+1) x 1 row vector of the parameters to be estimated with b_0 and b_1 usually used to indicate the intercept and the slope of the regression model, respectively; ε_i is a random error assumed to be identically, independently normally distributed (i.i.d.) with mean vector 0 and constant variance σ_ε^2;

obviously, the vector $\boldsymbol{\varepsilon}$ will be multivariate normally distributed with mean vector 0 and a diagonal variance-covariance matrix nxn $\mathbf{I}\,\sigma_{\varepsilon}^2$.

Furthermore, the residual (r_i) is defined as the difference between the observed value of the dependent variable (y_i) and the calculated/fitted y value defined $\hat{y}_i$ (read as y-hat); in particular: $r_i = y_i - \hat{y}_i$. So r_i corresponds to the estimate of ε_i when the equation is written with the estimates b_0 and b_1 instead of their parameters. In fact, $Y_i = \alpha + \beta X_i + \varepsilon_i \rightarrow \varepsilon_i = Y_i - \alpha + \beta X_i$ which corresponds in the sample as $r_i = Y_i - \left(b_0 + b_1 X_i\right)$ and $\hat{Y}_i = b_0 + b_1 X_i$. Of course, a corresponds to b_0 and b to b_1.

Furthermore, it should be noted that the variance of the error term ε_i, defined $\sigma_{\varepsilon_i}^2$, is equal to: $\sigma_{\varepsilon_i}^2 = E\left\{\varepsilon_i^2\right\} - \left(E\left\{\varepsilon_i\right\}\right)^2$. Since the expected value of ε_i ($E\{(\varepsilon_i\}$) is equal to 0, according to the assumptions of the regression model, the expected value of $\sigma_{\varepsilon_i}^2$ ($E\left\{\varepsilon_i^2\right\}$) is right $\sigma_{\varepsilon_i}^2$ with the conclusion that the squared residual (r_i^2) is an estimator of the error variance and its absolute value ($|r_i|$) is an estimator of its standard deviation obtained by the squared root of the variance. Of course, in the case of homoskedasticity, since the residual variances are all equal, $\sigma_{\varepsilon_i}^2$ can be replaced by σ^2 as the parameter of the error variance estimated by: $s^2 = \frac{1}{n-p}\sum_{i=1}^{n} r_i^2$.

Thus, it is very easy to conclude that the residuals (squared or absolute value) can be used to assess the relationship between the error variance (standard deviation) function with the pertinent independent or dependent or fitted variables in order to assess the presence of a heteroskedasticity pattern and ultimately model it.

For this purpose, it is also necessary to emphasize that in the presence of potential outliers in the dataset, regressing the standard deviation function should be preferred since it is less affected by the presence of outliers than regressing the squared residual function regression which is affected by the higher squared residuals. This very important point should be kept in mind in order to understand the role of the residuals in creating the weights for the WLS-R.

For easy reading, we report the formulas of the sample slope (b) and intercept (a) estimates from the ordinary least squares (OLS):

$$b = \frac{\sum_{i=1}^{n} Y_i X_i - n\bar{Y}\bar{X}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2};\ a = \bar{Y} - b\bar{X}$$

Regression analysis with the outlier

Table S1 shows the ANOVA table of the OLS regression analysis. The formulas are shown along with the actual values obtained from the OLS regression with the dataset shown above.

Table S1: ANOVA table of the simple linear OLS regression.

Source of variability	Degrees of freedom	SS	MS	Statistics F P	E(MS)
Regression Due to $b_1 \mid b_0$	1	$\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2$ 1363.18469	$b_1^2\sum_{i=1}^{n}(X_i-\bar{X})^2$ 1363.18469	69.63 <0.0001	$\beta_1^2\sum_{i=1}^{n}(X_i-\bar{X})^2+\sigma^2$
Residual (error variance) MSE	n – 2 50-2=48	$\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$ 939.89224	$s^2=\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{(n-2)}$ 19.58109		σ^2
Total (corrected)	n – 1 50-1=49	$\sum_{i=1}^{n}y_i^2-n\bar{y}^2$ 2303.07693	-		-
Correction term Due to b_0 (a – intercept))	1	$n\bar{y}^2$ 8989.79587	$n\bar{y}^2$ 8989.79587		$n\left(\sigma^2/n+\left(\alpha+\beta\bar{X}\right)^2\right)=$ $\sigma^2+n\left(\alpha+\beta\bar{X}\right)^2$
Total	N 50	$\sum_{i=1}^{n}y_i^2$ 11292.87280			

In addition, the estimated values of the intercept and slope along with their standard errors, t statistics, and significance level are reported in the Table S1.1.

Table S1.1

Estimates	Value	Standard error	Statistics t	p-value
Intercept (b_0)	0.42970	1.67672	0.26	0.7988
Slope (b_1)	0.90325	0.10825	8.34	<0.0001

Furthermore, the Square Root of the MSE (Mean Square Error) is equal to 4.42505 from $\sqrt{19.58109}$. Finally, the $R^2 = 0.5919$ and the adjusted R^2 (Adj R^2) = 0.5834;

Outlier diagnostics: theory

Of course, outliers can be in the Y-direction or in the X-direction; the latter are usually called "leverage points" to be further defined as "good" or "bad leverage points". Good leverage points are observations that are far from the bulk of the observations (observations close to the regression line determined by most of the data), but with no or almost irrelevant influence on the estimates. Otherwise, "bad leverage points" are far from the regression line determined by most of the data and can even have a dramatic impact on the estimates by pulling the regression line towards themselves. Often these "bad leverage points" are also outliers.

First, we need to consider the diagnostics based on the residuals obtained from OLS, but it should be emphasized that the OLS estimator has a very low performance (robustness) since only one observation (outlier) is sufficient to obtain parameter estimates that are much biased compared to those obtained only on the "bulk" of the observations. Furthermore, since the "breakdown point" in the one-dimensional estimation of location defined as the "smallest fraction of contamination (outlier observations) that can cause the estimator to take values arbitrarily far from the values estimated without any contamination", the OLS estimator has a breakdown point equal to 1/n, which tends to

zero when the sample size n is getting large [13]. So, it is possible to say that the estimator "breaks down" leading to the "breakdown point" expression.
First of all, it has to be remembered the "leverage h_{ii}" as the value on the diagonal of the so-called "hat matrix" **H** is defined. The name is due to the fact that this matrix transforms the observed vector **y** into its OLS estimates and then it is like putting a "hat" on the **y** vector: $\hat{\mathbf{y}} = \mathbf{Hy}$ and the "hat" is the symbol for "estimate". The hat matrix **H** is idempotent ($\mathbf{HH} = \mathbf{H}$) and symmetric ($\mathbf{H'=H}$) and is given by: $\mathbf{H} = \mathbf{X(X'X)^{-1}X'}$.
Most authors such as Hoaglin and Welsch [14], Henderson and Velleman [15], Cook and Weisberg [16], Hocking et al. [17], Paul [18], Stevens [19], and Belsley et al. [20] determine potentially influential points by looking at the "h", and paying particular attention to points for which h_{ii},>2p/n even if some people recommend a more conservative cut-off value of 3p/n, being "n" the number of observations used to fit the model and "p" the number of the parameters.
However, the hat matrix **H** completely neglects outliers in the Y-direction since **H** is based only on the X variables. Furthermore, the h_{ii} diagnostics are completely vulnerable to the "masking effect" consisting in the fact that one outlier masks another outlier. Masking occurs when a cluster of outlying observations skews the mean and the covariance estimates toward it, and the resulting distance of the outlying point from the mean is small; thus, real outliers in the X-direction can be masked by the effect of "good leverage points" that attract the regression line.
Another measure of leverage is the Mahalanobis distance or better Mahalanobis distance squared (MD_i^2) that should point to observations for which the explanatory part is far from the bulk of the data and which has a one-to-one relationship with the diagonal elements of the **H**-hat matrix. The MD_i^2 values can be compared with the 95% quantiles of the χ^2 distribution with k+1 degrees of freedom.
Moreover, other types of residual diagnostic can be computer, namely the "standardized residuals" (r_i/s, where s^2 is an unbiased estimator of s^2 when the measurement errors are independent and normally distributed with zero mean and standard deviation s), the "studentized residuals" or the "internally studentized residual" [$t_i = r_i/(s\sqrt{(1-h_{ii})}$] recommended by Hoaglin and Welsch [14], Cook and Weisberg [16], Paul [18], Stevens [19] Cook [21], Draper and John [22], Atkinson [23], Hocking [24], and Hocking and Pendleton [25].
It has to be stressed that it is possible to find a confusing denomination in the literature since the "studentized residuals" are sometimes called "standardized residuals". Finally, the term "studentized residual" is mostly applied to the studentized residuals obtained by: [$t_{(i)} = r_i/(s_{(i)}\sqrt{(1-h_{ii})}$] where $s_{(i)}$ is the estimate of σ from the regression carried out without the i-th case that are also called "*studentized deleted residuals*" or "*externally studentized residuals*" or also, according to Rousseeuw and Leroy (1983) [2] "*jakknifed residuals*" from the jackknife estimator technique of a parameter in which one systematically excludes one observation at a time from a data set, calculating the parameter estimate on the remaining observations, and then aggregating these calculated estimates. SAS® [26] calls these residuals "*RSTUDENT*" according to Belsey et al. [20] and reports that "The "RSTUDENT" residual differs slightly from "internally studentized residual" (called "STUDENT") since the error variance is estimated from $s_{(i)}^2$ without the i-th observation, not from s^2(calculated on all the observations)". Atkinson [27] referred to t(i) as *"cross-validatory residual*". Finally, Cook and Weisberg [16], and Velleman and Welsch [28] call t_i, an *"internally studentized residual"* and $t_{(i)}$ an *"externally studentized residual"* definitions that besides "jakknifed residuals" are, in our opinion, the more shareable and used in statistical jargon.
Observations with "RSTUDENT" larger than 2 in absolute value may require some attention and observations with an "internally studentized residual" greater than 3 (in absolute value) are generally considered outliers.
Assuming that the residuals are normally distributed and considering that the "studentized residuals" are practically equivalent to the "standardized residuals" when the sample size is more than 30, it is possible to say that they follow a standardized normal distribution with mean zero and variance equal to 1. Then, the probability of having a residual outside ±2 (or 2 in absolute value) is about 0.05, a value that can be considered too high and, consequently, it has to conservatively increased to ±2.5 for

decreasing the probability to about 0.01 or even to ±3 for having a probability value of about 0.0026. In fact, it has to be noted that there is a conservative approach in declaring an observation as an outlier to be perhaps deleted from the original dataset.
Taking into account that the influence of the i-th observation can be asserted by considering the results of the regression carried out both with and without that observation, some so-called "single-case diagnostics" have been proposed. Particularly, the Cook's squared distance [21] measures the change in the regression coefficients that would occur if a case was omitted.
Cook and Weisberg [16] and Montgomery and Peck [29] suggested that a value around 1.0 deserve attention since it is generally considered large.
A rule of thumb is that any observation with a Cook's distance greater than 4/n (where *n* is the number of the observations) is considered highly influential on the regression estimates and, consequently, should be considered as a potential outlier capable of biasing them.
Furthermore, Belsey et al. [20] proposed the DFFITS(i) statistics (very similar to Cook's distance) as a measure of influence on the prediction with a potential alarming threshold over $2\sqrt{(p/n)}$ and the DFBETASj(i) statistics, that are a scaled measure of the change in each parameter estimate (the j-th regression coefficient) calculated by deleting the i-th observation from the dataset with a cut-off value of $2/\sqrt{n}$ (better than just 2 without considering the sample size).
Another statistic to be considered is the "COVRATIO" that measures the change in the determinant of the covariance matrix estimated by deleting the i-th observation. Belsey et al. [20] suggest that an absolute value of (COVRATIO - 1) ≥ 3p/n deserves to be investigated as a potential outlier. Actually, the COVRATIO is the ratio between the determinant of the variance covariance matrix without the i-th observation and the determinant of the variance covariance matrix with the i-th observation included.
However, according to Rousseeuw and Leroy [2] as well, all these diagnostics have an interpretation no longer reliable when the data contain more than one outlier, and are susceptible to the masking effect with the unfortunate conclusion that they often fail to identify outliers.
Furthermore, we must remember the extension of most single-case diagnostics to multiple-case diagnostics with the Cook distance generalized by Cook and Weiberg [16] being the most relevant.
Finally, the "Resistant Diagnostic" has been proposed by Rousseeuw and Leroy [2, pages. 238-245) to which the interested readers are referred.

Outlier diagnostics: *data with the outlier*

The OLS residual statistics allow us to consider as potential outliers seven observations.
Particularly the residual statistics of the observation n.4 passed seven thresholds (leverage, externally studentized residuals, internally studentized residuals, Cook's distance, DFFITS, DBETAS intercept, and COVRATIO) but the residual statistics of the observation n. 29 also passed six thresholds (leverage, externally studentized residuals, Cook's distance, DFFITS, DBETAS slope, and COVRATIO). Then the residual statistics of the observations n. 19 and 31 passed two thresholds (leverage and COVRATIO); finally, the residual statistics of the observations n. 7, n. 17, and n. 23 passed only one threshold: DBETAS slope, leverage and COVRATIO, respectively.
Thus, it is almost obvious that the observation n. 4, modified to be an outlier in the Y-distance, has OLS residual statistics greater than too many thresholds causing it to be confirmed as a "true outlier".
Figure S1.1 and S1.2 show some plots of the OLS regression obtained from SAS® Proc REG with the keyword "influence".
On the left (Figure S1.1), there is the distribution of the studentized residuals (internally studentized residuals) with two vertical lines drawn at ±3 as the thresholds for considering the corresponding observation as an outlier and the Cook's distance with its threshold (see comments of the Figure 5) for considering the corresponding observation as a high leverage point. On the right, there is the diagram plot of the Y and X variables together with the regression line and the pointwise 95%CI of the fitted values (the two concave and convex lines delimit the black space near the line; as expected the point with the Y and X means as coordinates has the narrowest interval). In addition, there are the

95% prediction limits (95% CI of a generic Y value at the corresponding abscissa) shown as dotted lines that are apparently parallel owing to a minimum expected concave-convex pattern with the narrowest interval at the point with the means of Y and X as coordinates. It should be emphasized that the confidence probability of 95% has to be referred to each of these 95%CIs. To refer the confidence probability of 95% to all the intervals, the simultaneous 95%CIs must be calculated.

Figure S1.1

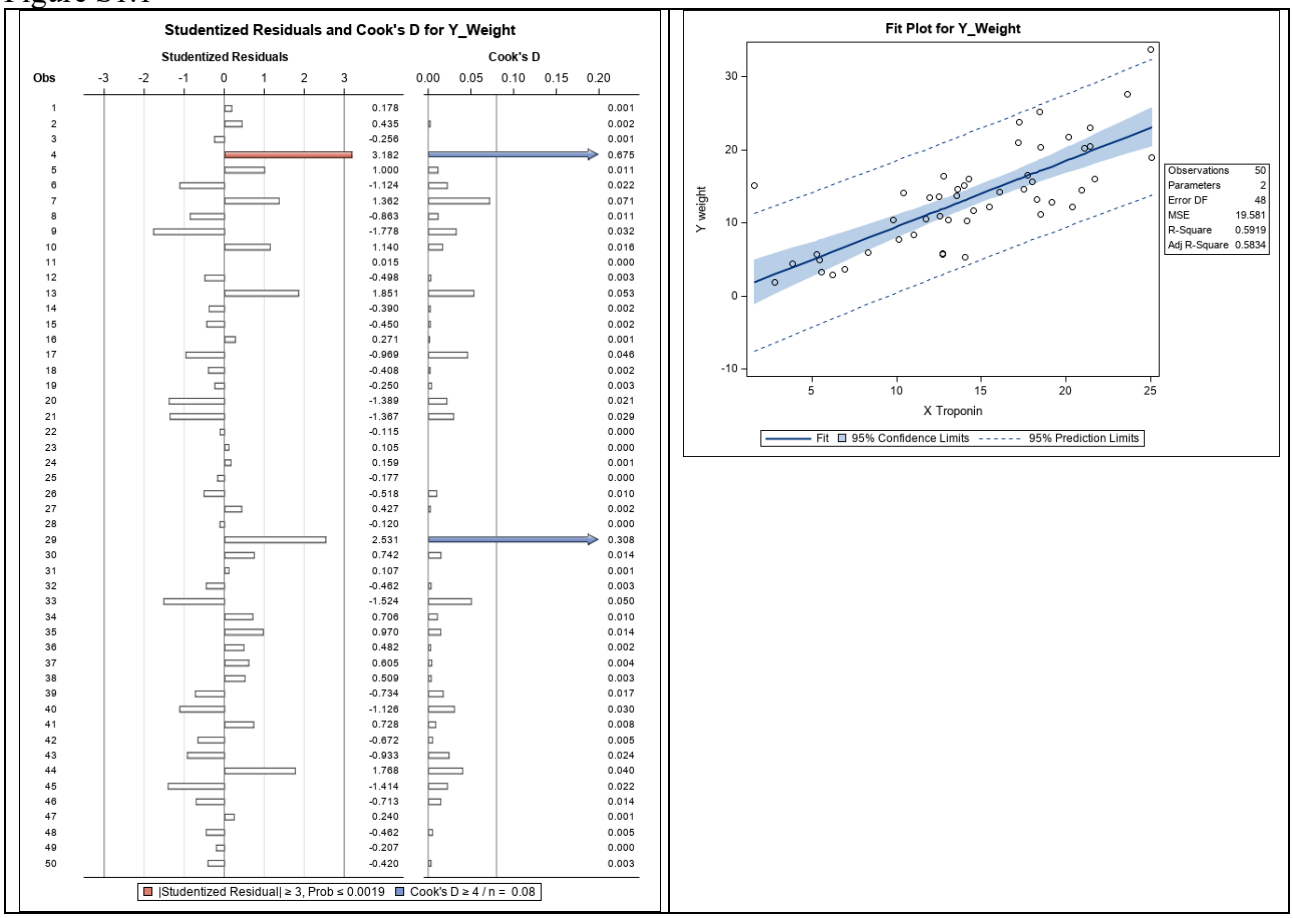


Figure S1.2.

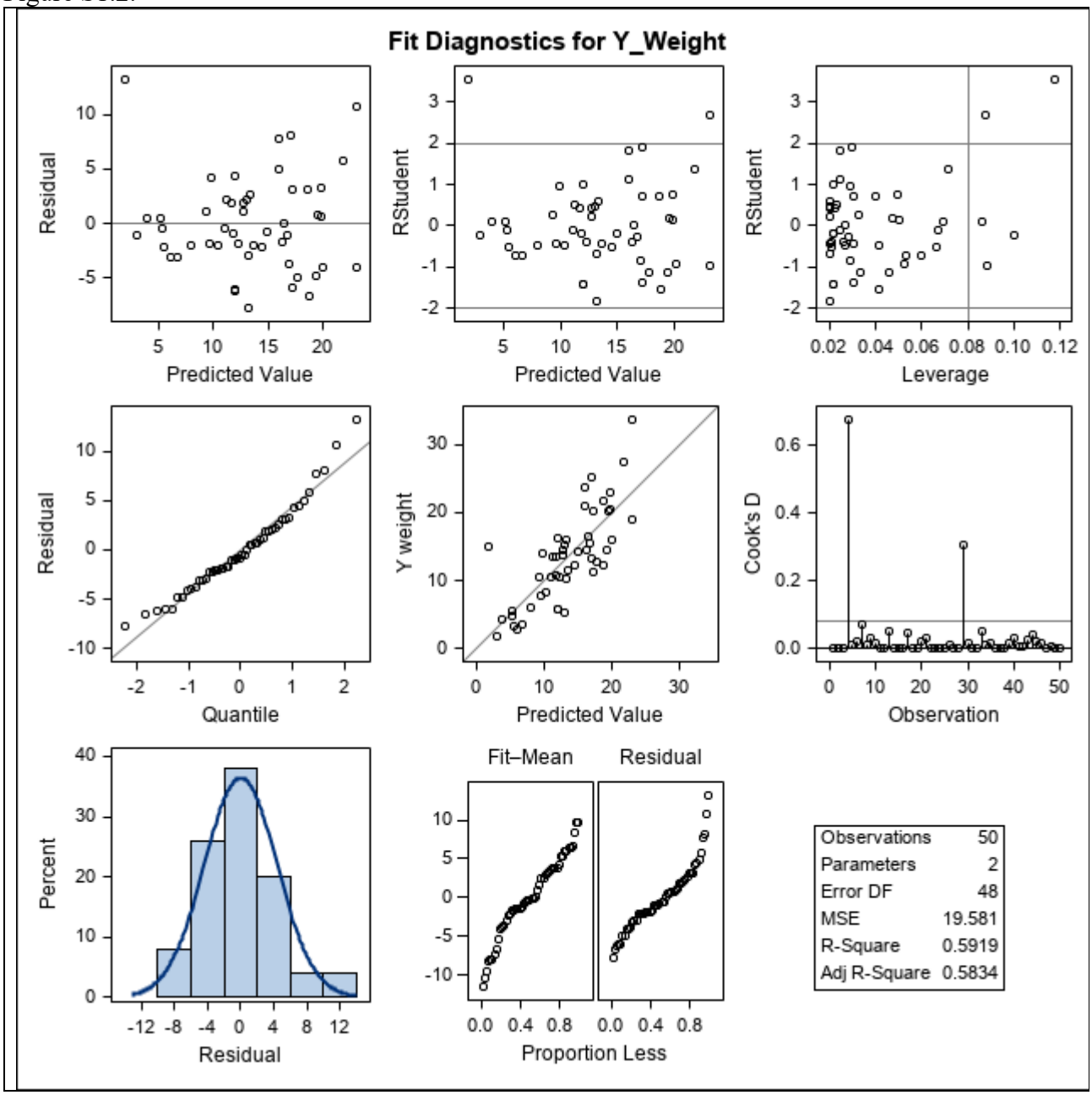


Figure S1.2 shows:

A.1)-Top row:

A.1.1)-left: plot of the "residuals" vs. the "predicted values"; the heteroskedasticity pattern is evident but it is better to consider the following plot.

A.1.2)-centre: plot of the "externally studentized residuals" vs. the "predicted values"; this plot should be considered primarily to judge the presence of a heteroskedasticity pattern. This plot corresponds to Figure 1-Panel A of the paper. The two horizontal lines at ±2 refer to the accepted thresholds to consider the "externally studentized residuals" that are inside as non-outliers;

A.1.3)-right: plot of the "externally studentized residuals" vs. the "leverage values". The two horizontal lines at ±2 refer to accepted thresholds for "externally studentized residuals"; the vertical line is drawn at the threshold of the leverage given by 2p/n equal to 4/50= 0.08.

In fact, there are two observations considered "outliers" with an externally studentized residual greater than 2 and leverage more than 0.8 (observations n.4 and n.29); then, there are three observations with leverage value greater than 0.8 (observations n.19, n.23 and n.31) and within the interval ±2.

A.2)-Centre row:

A.2.1)-left: QQ plot of the "residuals". It seems that the residuals are Gaussian distributed since they are well superimposed on the straight line obtained according to the Gaussian distribution. Indeed, the Shapiro-Wilk test does not reject the null hypothesis of a sample randomly drawn from a Gaussian distribution (P = 0.0976; Kolmogorov-Smirnov P >0.1500; Cramer-von Mises: P >0.2500; Anderson-Darling: P = 0.2217).
A.2.2)-centre: plot of the observed Y (Y_Weight) vs. the "predicted values" with the bisector equality line of the first Cartesian quadrant that is the place where the ordinates and the abscissas are equal and of the complete agreement between two measurement methods. It is obvious that in case of a perfect regression the observed and the fitted values are equal and lie on the equality line. The poorer the relationship, the further the points will move away from the bisector line.
A.2.3)-right: plot of Cook's distance D: it corresponds to the previous plot seen in Figure S1.1 rotated counterclockwise with a horizontal line drawn at 0.08 (threshold for this statistic to mark observations suspected to be outliers).
A.3)-Bottom row:
A.3.1)-left: histogram of the residuals overlaid with the Gaussian curve with mean and standard deviation of the sample "residuals".
A.3.2)-centre: plot of the Fit-Mean (fitted Y predicted values) on the left and the residuals (on the right) vs. their cumulative proportion. A straight-line pattern would suggest a uniform distribution. Since a Gaussian distribution is expected, we should see an italic "S" shaped line.
A.3.3)-right: some information about the regression analysis such as the number of the observations (n), the number of the parameters (2 in the case of the simple linear regression) the degrees of freedom (n – 2 = 48, in the case of the simple linear regression) of the error variance, Mean Square Error (MSE) equal to 19.581 for the OLS regression and the values of R^2 and adjusted R^2 (0.5919 and 0.5834, respectively).

The following Figure S3 shows the plot of the Rstudent (externally studentized) residuals vs. the fitted values. In the panel A there are the Rstudent residuals after the OLS regression corresponding to the iteration 0 ("Iter 0" on the top) of the iterative process given by the IRWLS estimator.
Panel B, Panel C, and Panel D show the Rstudent residuals obtained after the first, the second and the third iteration of the weighted regression performed with the IRWLS estimator.

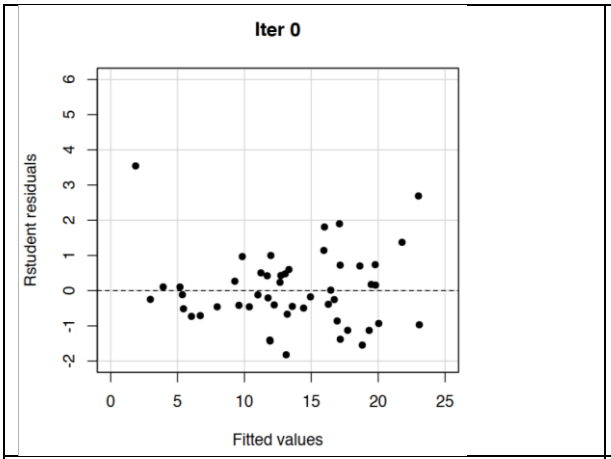


Figure S2-Panel A: residuals from OLS regression

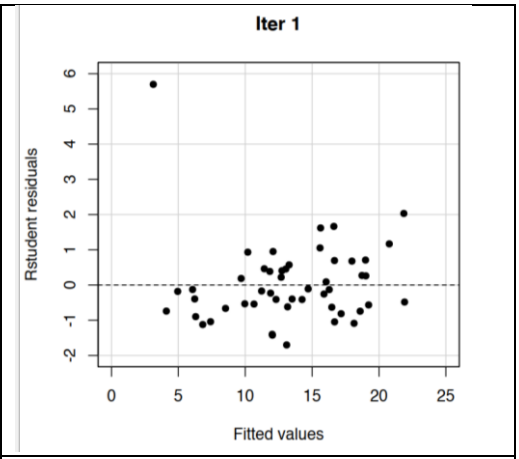


Figure S2-Panel B: residuals from the first weighted regression

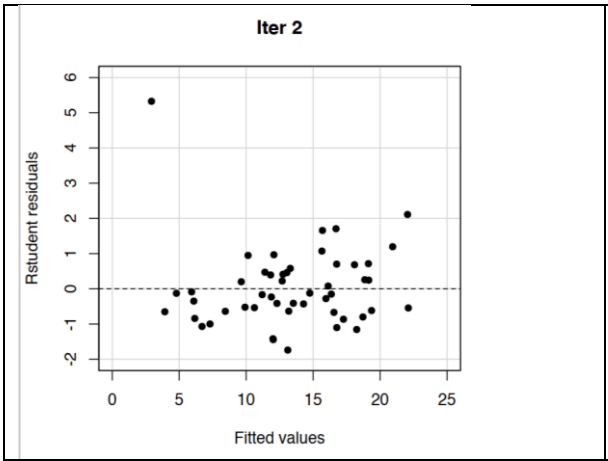


Figure S2-Panel C: residuals after the second weighted regression

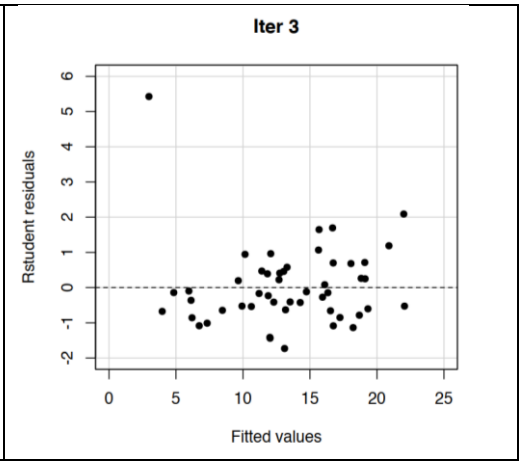


Figure S2-Panel D: residuals after the third weighted regression

It can be seen that the heteroskedastic pattern of the residuals, well evident in Figure 1-Panel A (after OLS regression), is somewhat reduced in Figure S2-Panel B with some points pushed towards the X axis. Hence, the heteroskedastic pattern practically does not change in the first and second iteration (second WS -Figure S2-Panel C- and third WS – Figure S2 – Panel D, respectively). Nevertheless, the parameter estimates of the successive iterations are somewhat different as shown in Table S2, as expected from the theoretical point of view. It should be noted that the first step shown in Table S2 corresponds to the OLS linear regression with weights equal to 1.

Table S2 – Dataset with the outlier.

Estimates	First OLS-R	Second (1WR)	Third (2WR)	Fourth (3WR)	Fifth (4WR)	Sixth (5WR)	Seventh (6WR)	Eight (7WR)
Intercept (b_0) (s.e.) Statistics t P	0.4297 1.6767 0.256 0.799	1.8534 1.4397 1.287 0.204	1.6301 1.4641 1.113 0.273	1.6912 1.4571 1.161 0.252	1.6741 1.4890 1.147 0.257	1.6789 1.4585 1.151 0.255	1.6776 1.4586 1.150 0.256	1.6779 1.4586 1.150 0.256
Slope (b_1) (s.e.) Statistics t P	0.9032 0.1083 8.344 <0.0001	0.8002 0.1058 7.562 <0.0001	0.8169 0.1055 7.740 <0.0001	0.8123 0.1056 7.693 <0.0001	0.8136 0.1056 7.707 <0.0001	0.8132 0.1056 7.703 <0.0001	0.8133 0.1056 7.704 <0.0001	0.8133 0.1056 7.704 <0.0001
MSE	4.425	4.524	4.488	4.497	4.494	4.495	4.495	4.495
R^2	0.5919	0.5346	0.5552	0.5522	0.5530	0.5528	0.5529	0.5528
Adj-R^2	0.5834	0.5431	0.5459	0.5428	0.5437	0.5435	0.5435	0.5435

From Table S2, it is possible to see that the intercept of the first WR is much greater than that of the OLS; hence, it decreases with a tendency to stabilize around values of 1.67. Otherwise, the slope of the WRs decrease to a plateau around 0.81. The standard errors of the estimates decrease during the iterative process.

It should also be noted the practically stable behavior of the MSE until a plateau around 4.495.

Considerations on the coefficient of determination

It is recalled that the R^2 and the adjusted R^2 ($R^2_{adjusted}$) are obtained respectively: $R^2 = 1 - \frac{(\mathbf{Y}-\mathbf{Xb})'(\mathbf{Y}-\mathbf{Xb})}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2}$; $R^2_{adjusted} = 1 - \frac{(1-R^2)(n-1)}{n-k-1}$ where k is the number of predictors and n is the total sample size. The adjusted R^2 should be preferred for judging the goodness of fitting since it increases only when an independent variable added to the model is statistically significant and

affects the dependent variable as opposed to the R^2 that increases when an independent variable is added to the model. Obviously, the adjusted R^2 value is always less than or equal to the R^2 value.

Note that the R^2 obtained by the WLS regression with the Y and X variables multiplied by their pertinent weights ($\mathbf{Y}_*$ **and** $\mathbf{X}_*$) is:

$$\mathbf{Y}_* = \mathbf{W}^{-1/2}\mathbf{Y}, \text{ and } \mathbf{X}_* = \mathbf{W}^{-1/2}\mathbf{X}; \quad R^2_{WLS} = 1 - \frac{\left(\mathbf{Y}_* - \mathbf{X}_*\mathbf{b}_*\right)'\left(\mathbf{Y}_* - \mathbf{X}_*\mathbf{b}_*\right)}{\mathbf{Y}_*'\mathbf{Y}_* - n\bar{Y}_*^2};$$

Where $\mathbf{b}_*$ is the WLS estimate of $\boldsymbol{\beta}$. In fact, it is informative to transform the equation to create a model that can be fitted with the OLS, even though the WLS estimates are usually calculated directly. Then, multiplying throughout the usual regression equation model by the squared root of the inverse of the weight matrix ($\mathbf{W}^{-1/2}$), one can obtain the above formula of the coefficient of determination (R^2_{WLS}) in the case of weighted regression.

The following formula of the coefficient of determination is not in matrix language:

$$R^2_{WLS} = 1 - \frac{\sum_{i=1}^{n} w_i \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} w_i \left[Y_i - \left(\sum_{i=1}^{n} w_i Y_i \Big/ \sum_{i=1}^{n} w_i\right)\right]^2}$$

Note that the above formula is essentially equivalent to the corresponding OLS formula except that instead of a weight always equal to 1, the weights have a specific value for each observation, resulting in the WLS estimator.

The weighted least squares output of some regression software packages includes R^2, the coefficient of determination (multiple in the case of more than one dependent variable). Users of these packages should treat this statistic with caution, because R^2 (adjusted R^2) does not have in weighted regression analysis the usual interpretation as in OLS regression. Indeed, the weighed R^2 is a measure of the proportion of the variation in the weighted Y than can be accounted by the weighted X. Interested readers are referred to Nagelkerke [30].

Furthermore, we agree with Willett and Singer [3l] "that it is not good to rely on any R^2 (even the pseudo R^2_{wls}) as a sole measure of goodness of fit".

Analysis without the outlier

The OLS regression results can be seen in Table 1 and Table 2 of the paper.

Outlier diagnostics: *data without the outlier*

One threshold was exceeded by the residual statistics of observations n. 7 (DBETAS slope), observation n. 13 and observation n. 44 (Rstudent or externally studentized residuals, for both).

Two thresholds were exceeded by the residual statistics of observations n. 4 (leverage and COVRATIO), observation n. 17 (leverage and Cook's D), observation n. 19 (leverage and COVRATIO), and observation n. 31 (leverage and COVRATIO).

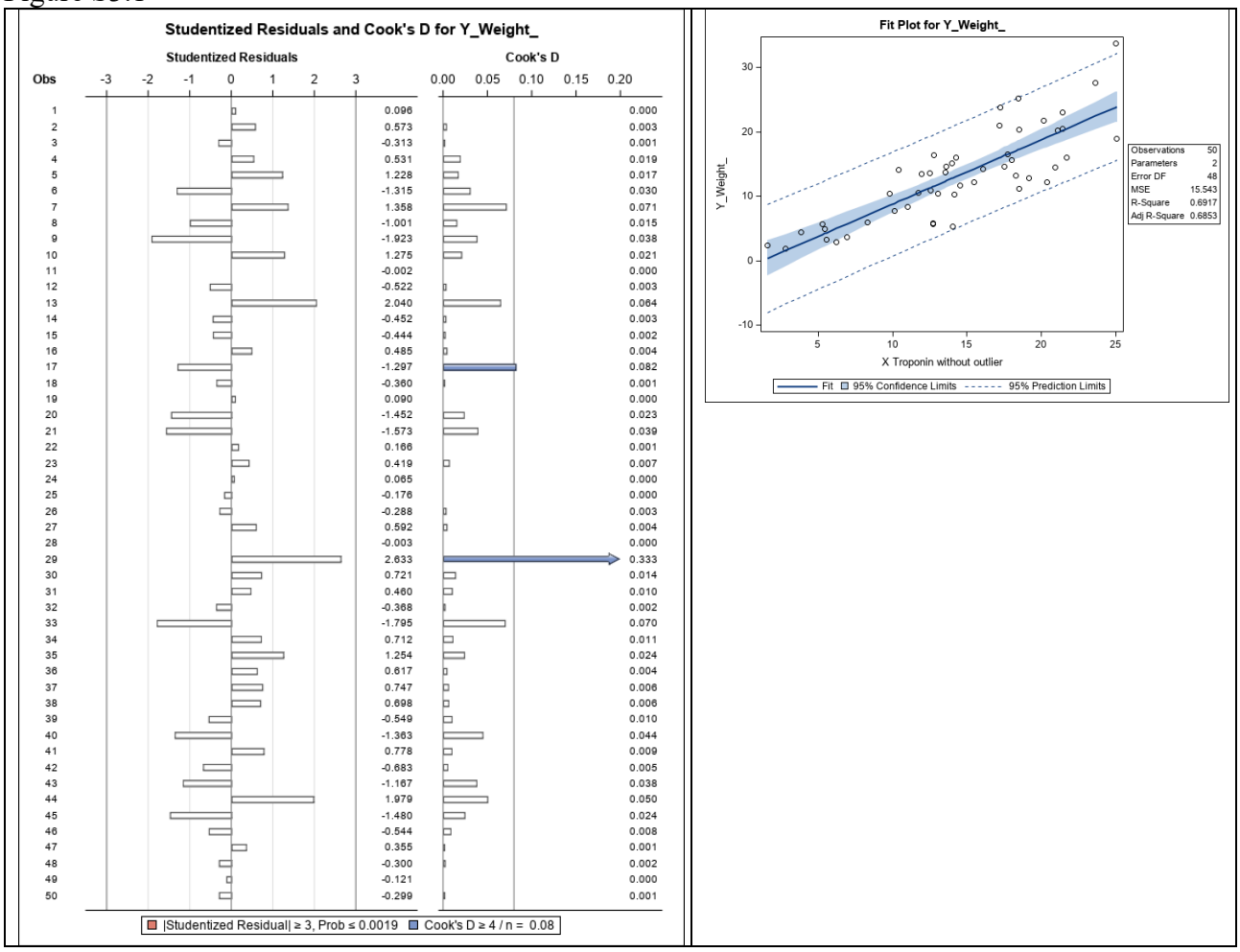
Finally, the residual statistics of observation n. 29 passed six thresholds (leverage, Rstudent or externally studentized residuals, Cook's D, DFFITS, DBETAS slope, COVRATIO).

Regarding these data, observation n. 29 with six thresholds exceeded can be considered an outlier. It is difficult to judge the case of observations n. 4, n. 17, n. 19, and n. 31 with two thresholds exceeded.

Finally, it is possible to conclude that the observations n. 7, n. 13, and n. 44 with only one threshold exceeded can be considered as such without further action.

However, all the observations described above from the dataset with and without the created outlier, have to be checked at least for imputation errors. Finally, it is difficult and almost impossible to decide what to do with the observations with more than two exceeded thresholds since deleting these observations can be considered an arbitrary decision.

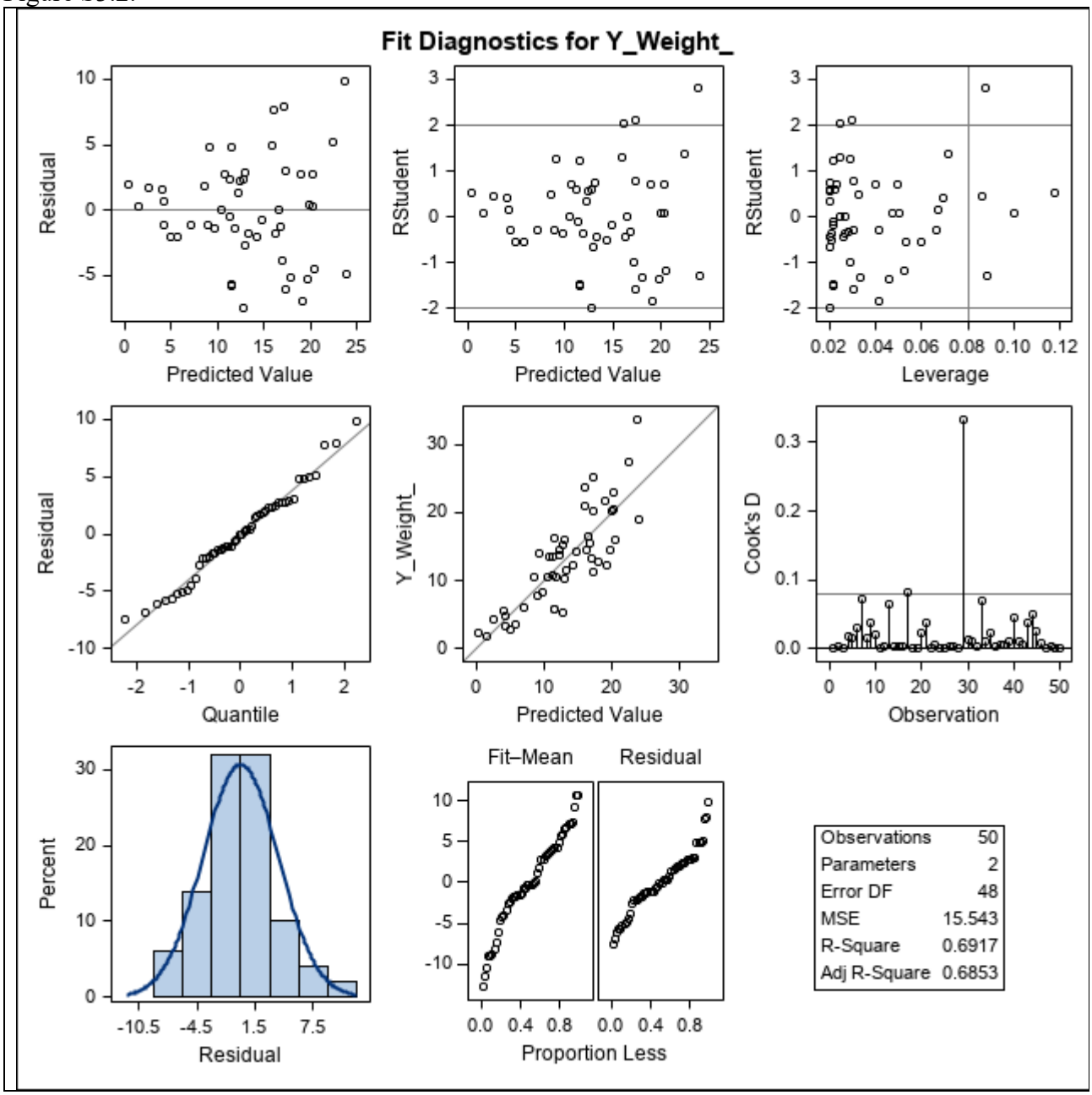
Figure S3.1



See Figure S1.1 for the comments. In this case the observation n.4 is not created as an outlier.

Figure S3.2 shows:
A.1)-Top row:
A.1.1)-left: plot of the "residuals" vs. the "predicted values"; the heteroskedasticity pattern is evident but it is better to consider the following plot.
A.1.2)-centre: plot of the "externally studentized residuals" vs. the "predicted values"; this plot is equal to the plot of Figure 1-Panel B in the paper.
A.1.3)-right: plot of the "externally studentized residuals" vs. the "leverage values". The two horizontal lines at ±2 refer to accepted thresholds for the "externally studentized residuals"; the vertical line is drawn at the threshold of the leverage given by 2p/n equal to 4/50= 0.08.
Indeed, there is only one observation considered as "outlier" with an externally studentized residual greater than 2 and leverage more than 0.8 (observation n.29); then there are four more observations with leverage value greater than 0.8 (observations n.4, n.17, n.19, and n.31).

Figure S3.2.



A.2)-Centre middle row:

A.2.1)-left: QQ plot of the "residuals". It seems that the residuals are Gaussian distributed since they are well superimposed on the straight line obtained according to the Gaussian distribution. Indeed, the Shapiro-Wilk test does not reject the null hypothesis of a sample randomly drawn from a Gaussian distribution (P = 0.5742; Kolmogorov-Smirnov P >0.1500; Cramer-von Mises: P >0.2500; Anderson-Darling: P > 0.2500).

A.2.2)-centre: plot of the observed Y (Y_Weight) vs. the "predicted values" with the bisector equality line of the first Cartesian quadrant, i.e. the place where the ordinates and the abscissas are equal. It is obvious that in the case of a perfect regression the observed and the fitted values are equal and lie on the equality line. The poorer the relationship, the further the points will move away from the aforementioned bisector line.

A.2.3)-right: plot of Cook's distance D: it corresponds to the previous plot in a counterclockwise direction with a horizontal line drawn at 0.08 (threshold for this statistic to mark observations suspected of being outliers).

A.3)-Bottom row:
A.3.1)-left: histogram of the residuals with the Gaussian curve with mean and standard deviation of the sample "residuals" superimposed.
A.3.2)-centre: plot of the Fit-Mean (fitted Y predicted values) on the left and the residuals (on the right) vs. their cumulative proportion "predicted values".
A.3.3)-right: some information about the regression analysis such as the number of observations (n), the number of parameters (2 in the case of simple linear regression) the degrees of freedom of the error variance or residual variance (n – 2 = 48, in the case of simple linear regression), the Mean Square Error (MSE) equal to 15.543 for the OLS regression and the values of R^2 and of adjusted R^2 (0.6917 and 0.6853, respectively).

Figure S4 Data without the outlier. OLS "RStudent" residuals (Iter0, Panel A) and "Rstudent" residuals after the first (Panel B), second (panel C) and third (Panel D) iterated weighted regression by means of the Iterated Re-Weighted Least Squares Estimator.

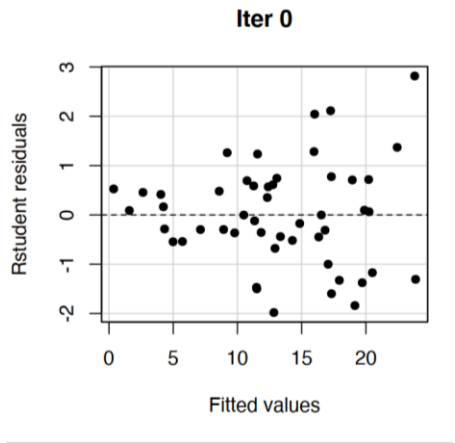


Figure S4.A: After OLS regression

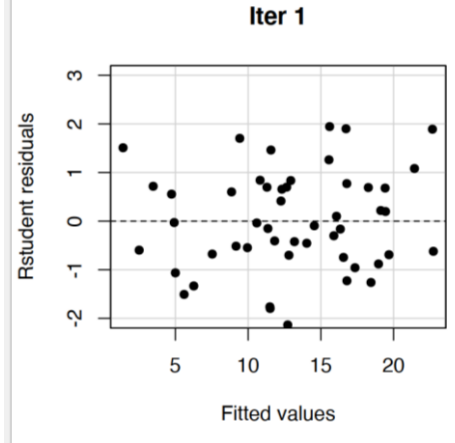


Figure S4.B: After the first weighted regression

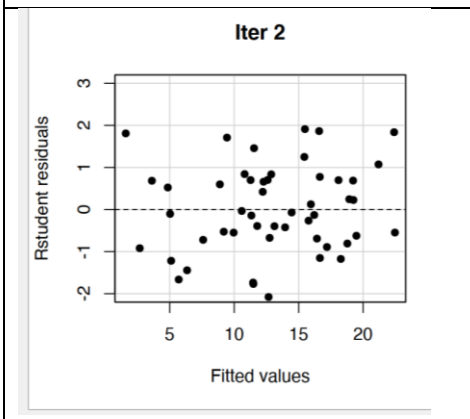


Figure S4.C: After the second weighted regression

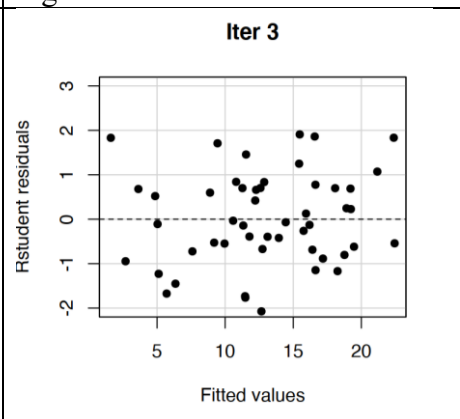


Figure S4.D: After the third weighted regression

It is possible to see that the heteroskedastic pattern is practically absent since the first iteration, even if some little adjustments (the RStudent residuals tend to be more uniformly distributed) have been made at the second iteration. Finally, the "Rstudent residuals" at the third iteration are practically equal to those of the second iteration. Nevertheless, in the successive iterations the parameter estimates are somewhat different as Table S3 shows.

Table S3 – Dataset without the outlier.

Estimates	First OLS-R	Second (1WR)	Third (2WR)	Fourth (3WR)	Fifth (4WR)	Sixth (5WR)	Seventh (6WR)
Intercept (b_0)	-1.2279	-0.0423	0.1746	0.1904	0.1918	0.1919	0.1919
(s.e.)	1.4939	0.3067	0.5009	0.4937	0.4931	0.4930	0.4930
Statistics t	-0.8220	-0.0700	0.3490	0.385	0.389	0.389	0.389
P	0.4150	0.9450	0.7290	0.701	0.699	0.699	0.699
MSE	3.942	2.596	2.309	2.287	2.285	2.285	2.285
R^2	0.6917	0.8185	0.8311	0.8319	0.832	0.832	0.832
Adj-R^2	0.6853	0.8148	0.8276	0.8284	0.8285	0.8285	0.8285

From Table S3, it can be seen that the intercept of the first WR shows a relevant increase in comparison to the OLS; then, the intercept value decreases with a tendency to stabilize around values of 0.1918. It has to recall that the fifth iteration corresponding to the fourth weighted regression. Otherwise, the slope of the WRs decreases to a plateau around 0.88 from the third iteration (second weighted regression). The standard errors of the estimates decrease during the iterative process. It should also be noted the decreasing behavior of the MSE to a plateau around 2.285. Of course, due to the absence of a relevant outlier, the iterative process is expected to converge with only a few iterations. Finally, it has to be noted that in this case the MSE decreases instead of the little increase shown for the dataset with the outlier.

Descriptive statistics of the weights given by the IRWLS, MM and LTS

Dataset without the outlier

IRWLS Method

The weights have a mean of 0.2886 (±0.06634, s.d.), median of 0.1096, first (Q_1) and third (Q_3) quartiles of 0.0651 and 0.17484, respectively, and a minimum and maximum of 0.0363 (obs. n.17) and 4.3291 (obs. n.4), respectively. No observations have a weight of 1. In addition, in this case the positive skewness of the distribution is evident since the median is much lower than the arithmetic mean. When these weights are "normalized" (divided by the sum of the weights multiplied by the number of observations) they sum equals the number of observations.

MM Method

The weights have mean of 0.9146 (±0.1159, s.d.), median of 0.9668, first (Q_1) and third (Q_3) quartiles of 0.8622 and 0.9891, respectively, a minimum and maximum of 0.4421 (obs. n.29) and 1.0000 (obs. n.28), respectively. In addition, in this case there is a distribution a little negatively skewed since the median is greater than the arithmetic mean.

LTS Method

There are 47 observations with weight equal to 1 and 3 observations equal to 0 (n.13, n.29, and n.44).

Dataset with the outlier

IRWLS Method

The weights have mean of 0.0912 (±0.03206, s.d.), median of 0.08574, first (Q_1) and third (Q_3) quartiles of 0.0688 and 0.1016, respectively, a minimum and maximum of 0.0517 (obs. n.17) and 0.1901 (obs. n.4), respectively. No observation has weight 1. In addition, in this case the distribution is almost symmetric since the median is very similar to the arithmetic mean.

MM Method

The weights have mean of 0.9069 (±0.1468, s.d.), median of 0.9675, first (Q_1) and third (Q_3) quartiles of 0.8779 and 0.9897, respectively, a minimum and maximum of 0.2000 (obs. n.4) and 0.9999 (obs.

n.28), respectively. In addition, in this case there is a distribution a little negatively skewed since the median is greater than the arithmetic mean.

LTS Method
There are 47 observations with weight equal to 1 and 3 observations equal to 0 (n.4, n.13, and n.29).

It is interesting to consider the degrees of freedom of the regression analyses with the considered 5 methods. Indeed, it has to stress that OLS, WLS with IRWLS, MO and MM have all 48 degrees of freedom for the error variance equal to n – 2 for the simple regression.
Otherwise, LTS analysis with 3 observations weighted as 0 has 45 degrees of freedom equal to 47 (the observations with weight equal to 1) minus 2.

Table S4. Basic notation for OLS and WLS method

Model	$\mathbf{y=X\beta+\varepsilon}$ **or** $y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}^{*}\beta_j + \varepsilon_i$	
	OLS	**WLS**
Assumptions		
In observational studies only	$\mathbf{x_i^*} \sim \mathbf{G_p}(\boldsymbol{\mu}_{x^*}, \boldsymbol{\Sigma}_{x^*})$	
In observational and experimental studies	$\boldsymbol{\varepsilon} \sim \mathbf{G_n}(\mathbf{0}, \sigma^2\mathbf{I_{(n)}}) \rightarrow \mathbf{y} \sim \mathbf{G_n}(\mathbf{X\beta}, \sigma^2\mathbf{I_{(n)}})$ $\varepsilon_i \sim G(0, \sigma^2) \quad \forall i = 1, 2, \ldots, n$	$\boldsymbol{\varepsilon} \sim \mathbf{G_n}(\mathbf{0}, {}_{\mathbf{W}}\sigma^2\mathbf{W^{-1}}) \rightarrow \mathbf{y} \sim \mathbf{G_n}(\mathbf{X\beta}, {}_{\mathbf{W}}\sigma^2\mathbf{W^{-1}})$ $\varepsilon_i \sim G(0, \sigma_i^2) \quad \forall i = 1, 2, \ldots, n$
Hat matrix	$\mathbf{H} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$ $h_{ii} = x_i'(\mathbf{X'X})^{-1}x_i \quad (3)$	${}_{\mathbf{W}}\mathbf{H} = \sqrt{\mathbf{W}}\mathbf{X}(\mathbf{X'WX})^{-1}\mathbf{X'}\sqrt{\mathbf{W}}$ $h_{ii} = \sqrt{w_i}x_i'(\mathbf{X'WX})^{-1}x_i\sqrt{w_i}$
Estimates		
Regression parameters	$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y} \quad (4) \qquad \mathrm{Cov}(\mathbf{b}) = \sigma^2(\mathbf{X'X})^{-1}$	${}_{\mathbf{W}}\mathbf{b} = (\mathbf{X'WX})^{-1}\mathbf{X'y} \quad (4) \qquad \mathrm{Cov}({}_{\mathbf{W}}\mathbf{b}) = {}_{\mathbf{W}}\sigma^2(\mathbf{X'WX})^{-1}$
Predicted values	$\hat{\mathbf{y}} = \mathbf{Xb} = \mathbf{Hy} \qquad \mathrm{Cov}(\hat{\mathbf{y}}) = \sigma^2\mathbf{H}$	${}_{\mathbf{W}}\hat{\mathbf{y}} = \mathbf{X}\,{}_{\mathbf{W}}\mathbf{b} = {}_{\mathbf{W}}\mathbf{Hy} \qquad \mathrm{Cov}({}_{\mathbf{W}}\hat{\mathbf{y}}) = {}_{\mathbf{W}}\sigma^2\mathbf{X}(\mathbf{X'WX})^{-1}\mathbf{X'}$
Residuals	$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I_{(n)}} - \mathbf{H})\mathbf{y} \qquad \mathrm{Cov}(\mathbf{e}) = \sigma^2(\mathbf{I_{(n)}} - \mathbf{H})$ $e_i = y_i - \hat{y}_i \qquad \mathrm{Var}(e_i) = \sigma^2(1 - h_{ii})$	${}_{\mathbf{W}}\mathbf{e} = \mathbf{y} - {}_{\mathbf{W}}\hat{\mathbf{y}} == (\mathbf{I_{(n)}} - {}_{\mathbf{W}}\mathbf{H})\mathbf{y} \qquad \mathrm{Cov}({}_{\mathbf{W}}\mathbf{e}) = {}_{\mathbf{W}}\sigma^2\mathbf{W}^{-1}(\mathbf{I_{(n)}} - {}_{\mathbf{W}}\mathbf{H})$ ${}_{\mathbf{W}}e_i = y_i - {}_{\mathbf{W}}\hat{y}_i \quad (5) \qquad \mathrm{Var}({}_{\mathbf{W}}e_i) = {}_{\mathbf{W}}\sigma^2 w_i^{-1}(1 - {}_{\mathbf{W}}h_{ii})$ ${}_{\mathbf{W}}\mathbf{r} = \sqrt{\mathbf{W}}\,{}_{\mathbf{W}}\mathbf{e} \qquad \mathrm{Cov}({}_{\mathbf{W}}\mathbf{r}) = {}_{\mathbf{W}}\sigma^2(\mathbf{I_{(n)}} - {}_{\mathbf{W}}\mathbf{H})$ ${}_{\mathbf{W}}r_i = \sqrt{w_i}\,{}_{\mathbf{W}}e_i \qquad \mathrm{Var}({}_{\mathbf{W}}r_i) = {}_{\mathbf{W}}\sigma^2(1 - {}_{\mathbf{W}}h_{ii})$
Residual sum of squares	$\mathrm{RSS} = \mathbf{e'e}$	$\mathrm{WRSS} = {}_{\mathbf{W}}\mathbf{r'}\,{}_{\mathbf{W}}\mathbf{r} = {}_{\mathbf{W}}\mathbf{e'W}\,{}_{\mathbf{W}}\mathbf{e}$
Mean Square	$\hat{\sigma}^2 = \frac{\mathrm{RSS}}{n - (p+1)}$	${}_{\mathbf{W}}\hat{\sigma}^2 = \frac{\mathrm{WRSS}}{n - p} \quad (7)$
Diagnostics		
Scaled residual	$\frac{e_i}{\hat{\sigma}}$	$\frac{{}_{\mathbf{W}}e_i}{{}_{\mathbf{W}}\hat{\sigma}} \qquad \frac{{}_{\mathbf{W}}r_i}{{}_{\mathbf{W}}\hat{\sigma}}$
Standardized residual	$\frac{e_i}{\sqrt{(1 - h_{ii})}}$	$\frac{{}_{\mathbf{W}}e_i}{\sqrt{w_i^{-1}(1 - {}_{\mathbf{W}}h_{ii})}} = \frac{{}_{\mathbf{W}}r_i}{\sqrt{1 - {}_{\mathbf{W}}h_{ii}}}$
Studentized residual	$\frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}$	$\frac{{}_{\mathbf{W}}e_i}{\sqrt{{}_{\mathbf{W}}\hat{\sigma}^2 w_i^{-1}(1 - {}_{\mathbf{W}}h_{ii})}} = \frac{{}_{\mathbf{W}}r_i}{\sqrt{{}_{\mathbf{W}}\hat{\sigma}^2(1 - {}_{\mathbf{W}}h_{ii})}}$
Studentized deletion residual*	$t_i = \frac{e_i}{\sqrt{\hat{\sigma}^2_{(i)}(1 - h_{ii})}}$ [1]	

* $\hat{\sigma}^2_{(i)}$ is the estimate of σ^2 when the entire regression is run again on the n sample without the i-th case.

Appendix A. Statistical tests for testing heteroscedasticity.

For completeness, heteroscedasticity can be formally tested by some statistical tests such as the Breusch-Pagan-Godfrey test [1A] and Godfrey [2A,3A] which has been extended by Cook and Weisberg [4A]. The test statistic is asymptotically distributed as a χ^2 distribution with k degrees of freedom (where k is the number of the predictors) under the null hypothesis of homoskedasticity and of normal distribution of the residual.
Since the Breusch-Pagan-Godfrey test statistic may not be accurate for non-normal data Bickel [5A] and Koenker [6A], among others, have proposed some variants. Furthermore, it has to be noted that this test is also not robust to multicollinearity.
In the open source R language, Breusch-Pagan-Godfrey test is performed by the 'ncvTest' function available in the "car" package (https://r-forge.r-project.org/projects/car/), by the 'bptest' function available in the "lmtest" package (https://CRAN.R-project.org/package=IMTest), by the 'plmtest' function available in the "plm" package (https://cran.r-project.org/package=plm) or by the 'breusch_pagan' function available in the "skedastic" package (https://CRAN.R-project.org/package=skedastic). The Koenker's variant is implemented in the package "lmtest" (https://CRAN.R-project.org/package=IMTest) of the open-source R language.
In SAS®, Breusch–Pagan can be obtained using the Proc Model. [7A]
White's test [8A] and Cook and Weisberg's test [9A] are another statistical test for heteroscedasticity that follows a chi-squared distribution, with degrees of freedom equal to number of estimated parameters minus 1 (a constant must be included). Furthermore, Waldman [10A] showed that White's test is equivalent to the algebraically modified Godfrey and Breusch-Pagan with choice of regressors as in White's test. The White's test can be performed by the "bptest" function from the "lmtest" package of the open-source R language.
Finally, we would like to mention the Goldfeld-Quandt test in its parametric and non-parametric version [11A] even if it is usually referred in its parametric form [12A,13A].
The Goldfeld–Quandt test checks for homoscedasticity in regression analyses by dividing the dataset into two groups (for this reason the test is sometimes called a two-group test) not necessarily of equal size nor contain all the observations. The groups are specified so that the observations for which the pre-identified explanatory variable takes the lowest values are in one subset, with higher values in the other. The test statistic uses the ratio of the mean square residual errors for the regressions on the two subsets and corresponds to an F-test of equality of variances. It should be remembered that the parametric test assumes that the errors have a normal distribution and that the design matrices for the two subsets of data are both of full rank.
Unfortunately, the Goldfeld–Quandt test is not very robust to specification errors. Thursby [14A] proposed a modification of the Goldfeld–Quandt test by using a variation of the Ramsey's "RESET" test [15A] in order to obtain some measure of its robustness.
In the open source R language, the Goldfeld-Quandt Test can be implemented using the 'gqtest' function of the "lmtest" package ((https://CRAN.R-project.org/package=IMTest) (parametric test only) or using the 'goldfeld_quandt' function of the "skedastic" package (https://CRAN.R-project.org/package=skedastic). (both parametric and nonparametric test).

References for Appendix A.
1A. Breusch TS, Pagan AR, A Simple Test for Heteroskedasticity and Random Coefficient Variation, Econometrica, 1979: 47 (5);1287–94, DOI:10.2307/1911963
2A. Godfrey L. Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables Econometrica, 1978: 46 (6); 1293-301.

3A. Godfrey L. Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables Econometrica 1978: 46 (6); 1303-10.
4A. Cook and Weisberg (Cook RD, Weisberg S. Diagnostics for Heteroskedasticity in Regression, Biometrika 1983:70(1);1–10, DOI:10.1093/biomet/70.1.1.1.
5A. Bickel PJ. 1978, Using residuals robustly I: Test for heteroscedasticity and nonlinearity, Annals of Statistics 1978; 6, 266-291.
6A. Koenker R. A note on studentizing a test for heteroscedasticity, Journal of Econometrics 1981: 17;107-l12.)
7A. SAS/ETS® 13.2 User's Guide. Cary, NC: SAS Institute Inc. Copyright © 2014, SAS Institute Inc., Cary, NC, USA)
8A. White HA. Heteroscedasticity - Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. Econometrica, 1980 48; 817-838.
9A. Cook RD, Weisberg S. Diagnostics for Heteroscedasticity in Regression. Biometrika 1983:70; 1-10)
10A. Waldman DM. A note on algebraic equivalence of White's test and a variation of the Godfrey/Breusch-Pagan test for heteroscedasticity. Economics Letters 1983: 13 (2–3); 197-200.
11A. Goldfeld SM, Quandt RE. Some Tests for Homoscedasticity". Journal of the American Statistical Association. 1965: 60 (310): 539–547),
12A. https://www.geeksforgeeks.org/goldfeld-quandt-test/ and
13A. https://en.wikipedia.org/wiki/Goldfeld%E2%80%93Quandt_test
14A. Thursby J. "Misspecification, Heteroscedasticity, and the Chow and Goldfeld-Quandt Tests". The Review of Economics and Statistics 1982: 64 (2); 314–321. doi:10.2307/1924311)
15A. Ramsey, J. B. (1969). "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis". Journal of the Royal Statistical Society, Series B. *31 (2): 350–371.*

Appendix B. General overview of some robust regression procedures

Robust regression procedures: a general overview.

It should be noted that interested readers should refer to at least one of the truly excellent books on this topic (Atkinson and Riani [1], Rousseeuw and Leroy [2] Maronna et al. [3,4], Huber [5], Huber and Ronchetti [6]) referred in this text.

In addition, from the preface of the first edition of the book by Maronna et al. [3] following sentence should be considered: "Robust methods have a long history that can be traced back at least to the end of the 19th century with Simon Newcomb [1B]. But its first great steps forward occurred in the 60s and the early 70s with the fundamental work of John Tukey [2B,3B] Peter Huber [4B,5B,6B] and Frank Hampel [7B,8B]. The applicability of the new robust methods proposed by these researchers was made possible by the increased speed and accessibility of computers."

It is also worth considering, always from the preface of the book from Maronna et al. [3], this other sentence: "robust methods remain largely unused and even unknown by most of the communities of applied statisticians, data analysts, and scientists that might benefit from their use. It is our hope that this book will help to rectify this unfortunate situation".

The well-known methods of robust estimation are: M estimation, S estimation, LST estimation and MM estimation. [9B].

1)-The M estimation, introduced by Huber [4B,10B] is the simplest approach both from a computational and theoretical point of view. Although it is not robust to leverage points, it is still widely used in data analysis when it can be assumed that contamination is mainly in the direction of the response. In fact, when there are outliers in the explanatory variables (leverage points), the method has no advantage over least squares.

The “M” in M-estimation stands for "maximum likelihood type".
There are several methods available to compute location and/or scale M-estimators. In principle one could use any of the general equation solving methods for such as the Newton–Raphson algorithm, but the computational method, called iterative reweighting, shows some advantages according to Maronna et al. [4, page 40].

2)-In the 1980s, several alternatives to M-estimation were proposed, such as the S-estimation. This method finds a line (plane or hyperplane) that minimizes a robust estimate of the scale (hence the method gets the S in its name) of the residuals [Rousseeuw and Leroy (1987, p. 263) [12B]. This method is highly resistant to leverage points and is robust to outliers in the response. However, this method was also found to be inefficient with some computational concerns according to Huber and Ronchetti (2009, p. 197) [10B]. Finally, Rocke [11B] showed that S-estimators can be sensitive to outliers even if the breakdown point is close to 0.5.

3)-A viable alternative is the Least Trimmed Squares (LTS) that was introduced by Rousseeuw [12B] and that is the preferred choice of Rousseeuw and Leroy [13B] and Ryan [14B] books, very useful for a pragmatic review of this topic. Least Trimmed Squares (LTS) estimation is a high breakdown value method, taking into account that the breakdown value is a measure of the proportion of contamination that an estimation method can withstand and still maintain its robustness. The performance of this method has been improved by the FAST-LTS algorithm proposed by Rousseeuw and Van Driessen [15B,16B]

4)-MM estimation, introduced by Yohai [17B,18B,19B] combines the high breakdown value estimation and the M estimation. It has the same high breakdown property as the S estimation but higher statistical efficiency.
MM-estimates have become increasingly popular and are one of the most commonly employed robust regression techniques.
The method proceeds by finding a highly robust and resistant S-estimate that minimizes an M-estimate of the scale of the residuals (the first M in the method's name). The estimated scale is then held constant while a close M-estimate of the parameters (the second M) is located.
Specifically, the MM estimation is based on the following three steps: A)-In the first step, it computes an initial consistent estimate $\hat{\beta}_0$ with a high breakdown point but possibly low normal efficiency; B)-In the second step, a robust M-estimate of the scale $\hat{\sigma}$ of the residuals is obtained based on the initial estimate; C)-Finally, in the third step, an M-estimate $\hat{\beta}$ starting at $\hat{\beta}_0$ is calculated [4].
Furthermore, the MM-estimator has the highest possible breakdown point of 0.5, and high efficiency under normality.
For the details of the algorithms, readers are referred to the manual of the statistical software used (package “RobStatTM” - https://CRAN.R-project.org/package=RobStatTM or package “Robustbase” https://CRAN.R-project.org/package=robustbase of the open source R language for the procedures shown in Maronna et al.’s books [3,4], even if they are at a high mathematical/statistical level, and to the manual of the of SAS®“PROC ROBUSTBASE” [20B].
As a final comment, the MM-estimates and the robust and efficient weighted least-square estimator (REWLSE) proposed by Gervini and Yohai [21B] have both a high breakdown point and a high efficiency and, according to a simulation study by Yu and Yao [22B], have the best overall performance among all the robust methods compared.

References for Appendix B

1B. Stigler S, Simon Newcomb, Percy Daniell and the history of robust estimation 1885–1920, Journal of the American Statistics Association, 1973, 68, 872–879.
2B. Tukey JW. A survey of sampling from contaminated distributions, in I. Olkin (ed.) Contributions to Probability and Statistics. 1960, Stanford University Press.
3B. Tukey JW. The future of data analysis, The Annals of Mathematical Statistics 1962, 33, 1–67.
4B. Huber PJ. Robust estimation of a location parameter, The Annals of Mathematical Statistics, 1964 35, 73–101.
5B. Huber PJ. A robust version of the probability ratio test, The Annals of Mathematical Statistics, 1965 36, 1753–1758.
6B. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions, in Proceedings of Fifth Berkeley Symposium of Mathematical Statistics and Probability, 1967 vol. 1, pp. 221–233. University of California Press.
7B. Hampel FR. A general definition of qualitative robustness, The Annals of Mathematical Statistics, 1971; 42, 1887–1896.
8B. Hampel FR. The influence curve and its role in robust estimation., The Annals of Statistics, 1974 69, 383–393.
9B. https://en.wikipedia.org/wiki/Robust_regression.
10B. Huber PJ, Ronchetti EM. Robust Statistics 2nd ed. 2009 Hoboken, NJ: John Wiley & Sons Inc.
11B. Rocke DM. Robustness Properties of S-Estimators of Multivariate Location and Shape in High Dimension. The Annals of Statistics 1996, 24,3,1327–1345.
12B. Rousseeuw PJ. Least median of squares regression. Journal of the American Statistical Association 1984 79, 871–880.
13B. Rousseeuw PJ, Leroy AM. (2003) [1986]. Robust Regression and Outlier Detection. John Wiley & Sons Inc.
14B. Ryan, TP. (2008) (1997). Modern Regression Methods. John Wiley & Sons Inc.
15B. Rousseeuw PJ, van Driessen K. (1999), A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41, 212–223.
16B. Rousseeuw PJ, van Driessen K. (2000), An algorithm for positive-breakdown regression based on concentration steps, in W. Gaul, O. Opitz and M. Schader (eds), Data Analysis: Modeling and Practical Applications, pp. 335–346. Springer Verlag.
17B. Yohai VJ. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression. Annals of Statistics 15:642–656.
18B. Yohai VJ, Stahel WA, Zamar RH. (1991). A Procedure for Robust Estimation and Inference in Linear Regression. In Directions in Robust Statistics and Diagnostics, Part 2, edited by Stahel WA, Weisberg SW. 365–374. New York: Springer-Verlag.
19B. Yohai VJ, Zamar RH. (1997). Optimal Locally Robust M-Estimates of Regression. Journal of Statistical Planning and Inference 64:309–323.
20B. SAS Institute Inc. 2016. SAS/STAT® 14.3 User's Guide. Cary, NC: SAS Institute Inc.
21B. Gervini D, Yohai VJ. A class of robust and fully efficient regression estimators. The Annals of Statistics 2002: 30:583–616.
22B. Yu C, Yao W. (2017) Robust Linear Regression: A Review and Comparison Communications in Statistics—Simulation and Computation. 46:8, 6261-82.

Appendix C. General overview of the MO robust procedure

The MO procedure is based on two stages to obtain both robust estimates of the model parameters and a valid classification of the outlying observations.

Particularly: the first stage jointly processes the response variable (Y) and the explanatory variable (X), which form together the so-called **Z** = (**Y, X**) matrix. Of course, it is also possible

to have several independent variables to form a matrix **X** with a matrix **Z** = (**Y, X**) with more than two columns. Then, the original dataset is split into two preliminary subsets (bulk and outliers) by using an approach based on the robust Minimum Covariance Determinant (MCD) according to Rousseeuw and Van Driessen [1C on page 213] calculated by minimizing the determinant of subsets of size equal to (n+p+1)/2 where "p" is the number of the parameters. Data with the MCD make up the preliminary bulk subset (pb) on which are calculated the preliminary OLS estimates of the model parameters: ${}_{pb}a_{OLS}$, ${}_{pb}b_{OLS}$, and ${}_{pb}\hat{\sigma}_{OLS}$.

These preliminary OLS estimates are used to compute the externally predicted regression diagnostics of the observations that form the preliminary outlier subset; particularly, externally predicted scaled residuals in Y-space and leverage in X-space ($\tilde{r}_i$, and $\tilde{h}_i$, respectively).

Observations with externally predicted scaled residual between the specified cut-offs (2.576 and 2.576 giving a 0.99 probability of inclusion in the interval) are either false outliers identified by MCD on **Z** matrix or good leverage points. These are all added back to the 'preliminary bulk subset' to obtain the 'confirmed bulk subset'.

These observations are labelled as "typical data" if $\tilde{h}_i \leq 2p/(n-m+1-2p)$ (where m is the number of observations in the preliminary outlier subset, p is the number of parameters and n is the sample size) or "good leverage points" if $\tilde{h}_i > 2p/(n-m+1-2p)$. It has to be noted that the leverage threshold of 2p/(n-m+1-2p) has been suggested by Marubini and Orenti [2C,3C].

Consistently, the remaining outliers are now forming the "confirmed outlier subset". The "confirmed bulk subset (cb)" is now used to compute the "confirmed" OLS estimates: ${}_{cb}a_{OLS}$, ${}_{cb}b_{OLS}$, and ${}_{cb}\hat{\sigma}_{OLS}$. It is noteworthy that $\tilde{r}_i$ are approximately distributed as a standard Normal variable (Salini et al. 2016) [4C] leading to justify the adopted cut-offs of -2.576 and 2.576 corresponding respectively to the 0.005 and 0.995 quantiles of the standard Normal distribution.

In the second stage, the confirmed OLS estimates start the iterative regression process of the M estimator: the weights are computed by using the biweight Tukey redescending function [5C] with constant c = 4.685 and keeping the scale parameter ${}_{cb}\hat{\sigma}_{OLS}$ fixed at each iteration. The final MO estimates are so attained: a_{MO}, b_{MO}, and $\sqrt{V}_{MO}$.

Where V_{MO} is the variance covariance matric corrected according to Maronna et al. [6C, pp. 100–101).

To label the different types of observations according to the final MO estimates, a graph is provided that plots the weights (ranging from 0 to 1) of the final iteration of the MO procedure, against the natural logarithm of the square root of the robust distance (ln ${}_{Z}$RD):

${}_{\bar{z}}RD_i = \sqrt{(\mathbf{z}_i - {}_{pb}\bar{\mathbf{z}})' {}_{pb}\mathbf{S}^{-1}(\mathbf{z}_i - {}_{pb}\bar{\mathbf{z}})}$, where $\mathbf{z}_i = (y_i, x_i)$ and ${}_{pb}\bar{\mathbf{z}}$, ${}_{pb}\mathbf{S}$ are the corresponding estimates of the mean vector and the covariance matrix computed on the preliminary bulk subset identified in the first stage. The Robust Distance indicates how far any observation is from the center of the ellipsoid forming the bulk, and the cut-off of such distance is shown in the plot by a vertical line drawn at the natural logarithm of the squared root of the 0.95 quantile of the $\chi 2$ distribution with degrees of freedom equal to the number of parameters. The Robust Distances together with the final MO weights allow for definitive labelling of the observations (see Figure 5A and 5B together with the pertinent comments). Note that the observations forming the "good leverage" set of points were in the first step labelled as outliers by the MCD estimator, but are subsequently not considered as outliers and thus are not down-weighted in the final MO iteration.

To decide whether it makes sense to eliminate all or some of the identified outliers from the original dataset it is of fundamental importance to carefully examine the observations forming the final outlier subset and consider: (i) their robust distance; (ii) the data generation process;

and (iii) their scientific/biological plausibility on the ground of subject matter knowledge. As a result, a reduced dataset can be obtained.
As a further comment, it is useful to say that a widely used measure of remoteness of observations from the centroid of this space is the Mahalanobis Square Distance [7C] but it is substantially influenced by the presence of outliers [8C]. Therefore, the MO procedure used the Robust Distances proposed by Rousseeuw and van Zomeren [9C] as an alternative.
Finally, MO procedure can be performed by an R code available on request from the corresponding author.

References for Appendix C.

1C. Rousseeuw PJ, van Driessen K. (2000), An algorithm for positive-breakdown regression based on concentration steps, in Gaul W, Opitz O, Schader M. (eds), Data Analysis: Modeling and Practical Applications, pp. 335–346. Springer Verlag.
2C. Orenti A, Marano G, Boracchi P, Marubini E. Pinpointing outliers in experimental data: the Hat matrix in Anova for fixed and mixed effects models IJPH – 2012;9:e86631-13.
3C. Orenti A, Marubini E. Robust regression analysis: a useful two stage procedure, Communications in Statistics - Simulation and Computation, 2021;50:16-37, doi: 10.1080/03610918.2018.1547400
4C. Salini S, Cerioli A, Laurini F, Riani M. (2016). Reliable Robust Regression Diagnostics. International Statistical Review, 84(1), 99–127.
5C. Beaton AE. Tukey JW. (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16:147–85.
6C. Maronna, R. A., D. R. Martin, and V. J. Yohai. 2006. Robust statistics: Theory and methods. Chichester, UK: John Wiley and Sons.
7C. Mahalanobis PC. (1936) On the generalised distance in statistics, in Proceedings of the National Institute of Sciences of India, 2:49–55.
8C. Li X, Deng S, Li L, Jiang Y. (2019) Outlier Detection Based on Robust Mahalanobis Distance and Its Application Open Journal of Statistics 9:15-26.
9C. Rousseeuw PJ, van Zomeren BC. (1990) Unmasking multivariate outliers and leverage points. J Am Stat Ass 85:633–639.